The misallocation of land and other factors of production in India

Gilles Duranton
Wharton School, University of Pennsylvania

Ejaz Ghani
World Bank

Arti Grover Goswami
World Bank

William Kerr
Harvard Business School

World Bank Working Paper – January 2015

Abstract: We quantify the misallocation of manufacturing output and factors of production between establishments across Indian districts during the 1989-2010 period. We first distil a number of stylised facts about misallocation in India, and we demonstrate the validity of our misallocation metrics by connecting them to regulatory changes in India that affected real property. With this background, our study next quantifies the implications and determinants of factor and output misallocation. Although more-productive establishments in India tend to produce more output, factors of production are grossly misallocated. A better allocation of output and factors of production is associated with greater output per worker. Misallocation of land plays a particularly important role in these challenges.

Acknowledgements: Financial support from the World Bank and CEPR Private Enterprise Development in Low-Income Countries programme is gratefully acknowledged. We appreciate the comments and guidance from Ed Glaeser, Vijay Jagannathan, Somik Lall, Jevgenijs Steinbuks, and conference and seminar participants. We are also grateful to Yoichiro Kimura for helping us to understand the repeal of ULCRA and sharing his data with us. The views expressed here are those of the authors and not of any institution they may be associated with.

1. Introduction

We study the misallocation of land and buildings inputs and other factors of production for the Indian manufacturing sector. Our paper makes four contributions. We first document that factor misallocation among establishments is substantially worse than that of output, and that districts vary substantially in the efficiency with which factors are allocated to the most productive establishments. Second, we validate these indices for studying land and building misallocation by showing their close connection to local policy and tax reforms in property markets. Building from these baselines, our third contribution is to demonstrate how land and building misallocation fosters the misallocation of output in local areas, in both the current context and also looking forward. We also show a strong connection of land and building misallocation to reduced labour productivity for districts.

The economic magnitudes of our results are substantial and provide an important lens for considering how to best foster economic development. Economists and policy makers have traditionally sought to encourage development and growth through increasing factor productivity and fostering factor accumulation. We show that a better factor allocation across establishments would also generate very large economic gains. We quantify that a one standard deviation decrease in the misallocation of land and buildings is associated with about a 25% increase in output per worker. This is equivalent to a six-fold increase in the land supply for manufacturing establishments in these districts. We also observe that the misallocation of other inputs to firms hampers performance, but land and building access appears especially important in India. This parallels accounts in the press of the exceptional and stifling nature of Indian land markets. These insights provide an important input to the achievement of economic growth and the World Bank's twin goals of reducing extreme poverty and promoting shared prosperity.

Our paper provides three distinct contributions to the academic literature. First, as we discuss below, there have been many studies arguing the importance of misallocation. The consequences of misallocation are usually inferred indirectly by asking a model what would be the aggregate consequences of reduced misallocation. That is, extant claims about the importance of misallocation rely on measures of aggregate misallocation and computations of counterfactuals from particular models. In contrast, we provide direct evidence about the importance of misallocation by investigating empirically the link between factor misallocation across establishments and output per worker, using subnational regions in cross-section and in panel.

Second, increasing the productivity of factors of production, fostering their accumulation, and reducing their misallocation can only be viewed as proximate causes of economic growth and development. Importantly, as part of our metric validation, we begin to make connections between policies and factor misallocation. Our findings suggest that factor misallocation is not exogenously determined but is instead affected by policies. More generally, we think of our results as emphasising the importance of "frictions" as key determinant of the efficiency of factor allocation and, in turn, prosperity. Policies can be a source of friction. Better policies can also reduce frictions and improve allocation. Our emphasis on frictions differs from many models in the literature that rely on idiosyncratic distortions as the root cause of misallocation.

Third, we highlight the uniquely important role played by land and buildings in misallocation. We attribute this to the fact that choosing a location is a decision that conditions many others and cannot easily be changed, especially in an environment with poorly functioning land markets. More productive establishments will have a tough time buying more machines or employing more workers if they have no room to accommodate them. Land may also be a uniquely important asset for establishments that seek to expand since it can be used as collateral for external financing. While we do not take a stance regarding the exact mechanism through which land and buildings affect factor misallocation, our results highlight that land used for non-agricultural production may play a more important role than hitherto thought. Better land policies can make land more broadly available and reduce the frictions associated with land transactions. We are nonetheless aware that the elimination of frictions on the land market would require more than better land use regulations and a more efficient taxation of properties. Better-functioning land markets also require clearly defined property rights, a reliable land registry, and a number of other institutional improvements. This is a considerable challenge in a country like India where property rights for land and buildings are poorly defined and often conflict with tenancy rights.

To conduct our study, two main challenges need to be overcome. The first is to develop a methodology that allows us to explore the effects of misallocation and its determinants in a cross-section of districts. To do this, we develop indices of misallocation in the spirit of the decomposition originally proposed by Olley and Pakes (1996). In essence, misallocation can be measured by the difference between un-weighted mean establishment productivity and mean establishment productivity weighted, typically, by output. This difference is proportional to the covariance between establishment productivity and output. A higher covariance indicates a more efficient allocation as more productive establishments produce more output.

Our analysis starts with two simple observations. First, Olley-Pakes misallocation indices can be computed not only for output but also for each factor of production separately. Second, these misallocation indices need not be computed exclusively at the country level. They can also be produced for subnational units such as districts. Interestingly, there is a lot of variation in misallocation across the different misallocation indices and for each index across districts. We then think about misallocation as an intermediate outcome which we can relate to final outcomes of interests such as output per worker (or establishment productivity). We can also relate these intermediate misallocation outcomes to deeper causes such as policies or the local characteristics of districts.

More specifically, our analysis proceeds in five steps. The first is to estimate the productivity of all establishments in the data and factor shares for all industries. Establishment productivity is a necessary input to measure misallocation, and we consider several approaches to estimating productivity. Our second step is to compute misallocation indices for output and for each factor of production at the district level. The main difficulty here is that misallocation is most meaningfully computed at the industry level since industries differ in their factor intensity. Measures of misallocation at the district-industry level must thus be aggregated across industries to obtain a district-level measure. After distilling some stylised facts about misallocation across districts, our third step validates these metrics for studying land and building access by showing a connection of them to unanticipated local policy reforms that affect property markets.

In the fourth step, we quantify the relationship between various forms of misallocation. This step allows us to tease out the unique role played by the misallocation of land and buildings among Indian establishments. This misallocation of land and buildings is at the root of much of the misallocation of output. The last step examines the effects of factor misallocation on establishment productivity and output per worker. This analysis affords statements about which forms of misallocation matters more in Indian districts. Again, land and buildings appear especially important given their relatively small cost share.

The second main challenge is one of data. To compute measures of misallocation at the district level, establishment/plant data are needed as large firms may be present in many districts. Although it is always possible to apportion consolidated firm-level accounts across their various establishments, it is easier and more accurate to exploit establishment-level information. Our analysis is also best conducted using detailed information about the balance sheets of establishments to distinguish between different types of fixed factors. Although our approach can be applied to more aggregated factors (e.g., all fixed inputs), some of our more interesting findings come from our ability to distinguish more finely across different forms of capital, such as land and building vs. other fixed assets. Our analysis also requires a large country with enough subnational units to exploit sufficient cross-sectional information. Each subnational unit also need to be large enough for our measures of misallocation to carry some information and not be driven only by sample variation. Sampling issues are all the more important since our indices of misallocation must be first calculated at the level of each sector and district and then aggregated across sectors within each district. Finally, our country of study should be heterogeneous enough for there to be sufficient variation in misallocation.

India is well suited for our purpose. The government conducts regular censuses of production at the establishment level and requires establishments to report their assets with an unusual level of detail. India is also an extremely large country composed of several hundred districts, most of which contain enough establishments for us to compute informative measures of misallocation.

Although we discuss our methodology and findings in relation to the literature in greater detail below, we note that our work mainly contributes to the misallocation literature recently pushed forward by the seminal contributions of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). As already mentioned we explicitly measure how factor misallocation affects output per worker using a cross-section of subnational units rather than infer it indirectly from a model. We also show how factor misallocation affects output misallocation. Finally, we provide evidence regarding the effects of some economic policies on factor allocation.

2. Background and data

Our analysis considers five surveys from two major sources of data: the Annual Survey of Industries (ASI) and National Sample Survey Organization (NSSO). These data have been used extensively in prior research on the Indian economy, including Ghani et al. (2014) from which the first part of this data description builds on. ASI is a survey of the organised sector undertaken annually by the Central Statistical Organization, a department in the Ministry of Statistics and Program Implementation of the Government of India. Under the Indian Factory Act of 1948, all establishments employing more than 20 workers without using power or 10 employees using power are required to be registered with the Chief Inspector of Factories in each state. This register is used as the sampling frame for the ASI. The ASI extends to the entire country, except the states of Arunachal Pradesh, Mizoram, and Sikkim and Union Territory (UT) of Lakshadweep.

The ASI is the principal source of industrial statistics for the organised manufacturing sector in India. It provides statistical information to assess changes in the growth, composition, and structure of organised manufacturing sector comprising activities related to manufacturing processes, repair services, gas and water supply, and cold storage. As noted in Ghani et al. (2012b), organised manufacturing contributes a substantial majority of India's manufacturing output, while the unorganised sector accounts for a large majority of employment for Indian manufacturing workers. Manufacturing activity undertaken in the unorganised sector, such as households (own-account manufacturing enterprise) and unregistered workshops, is covered by the NSSO. The distinction between organised and unorganised sectors broadly captures the difference between formal and informal sectors.

Our study considers data taken from 1989-90, 1994-95, 2000-01, 2005-06, and 2010-11. In the first four instances, we use contemporaneous ASI and NSSO surveys. In the last period, we use 2009-10 ASI data and 2010-11 NSSO data. For simplicity, we normally refer to a sample survey by its starting year—for example, the sample survey in the year 2005-06 is referred to as the 2005 survey. We use 2010 in the tables and text to refer to the years 2009-10 for ASI data and years 2010-11 for NSSO data.

Establishments in both the organised and unorganised sectors provide information on the value of the land and buildings that they own. Although both sectors provide this information, only the organised sector surveys offer the distinction across land and buildings consistently for all the survey years under consideration.[1] For the unorganised sector, this separation between land and building stops in 2000. Thus, in this sector, we always consider land and buildings together.

Some businesses rent the land that they operate on. The questionnaires for both sectors ask renting establishments to provide information on the rent paid for land and buildings. The NSSO goes a step further and requires establishments report the value of the land and buildings that they rent. The organised sector survey also provides information on the rent paid and/or received for land and buildings for the years 2000, 2005, and 2010. To impute the value of land and buildings rented by establishments, we use the reported values for the land and buildings when available. Otherwise, our imputations use rents and estimated local capitalisation rates as described in appendix box 1. Appendix tables 1a and 1b provide details and exact wording of the questions pertaining to land and building values that appear in the respective survey questionnaires. In robustness checks, we duplicate our main regressions for the samples of only owners and only renters and find similar results.

The raw data from the five survey years contain 262,911 observations in the organised sector and 955,234 observations in the unorganised sector. In the organised sector, we retain establishments whose status is declared to be open, thereby dropping establishments that were closed during the

[1] The survey provides data on opening value, closing value, gross value, depreciation and so on for land and buildings. Closing net value of land and buildings is taken as the value of land and buildings owned. To obtain the value of the total land and buildings owned by establishments in the organised sector, we simply sum the separate values of land and buildings.

reference year, non-operational, deleted due to deregistration, or out of coverage (that is belonging to defence, oil storage, etc.). This step leaves 216,898 observations in the organised sector. Next, we drop establishments that do not belong to the manufacturing sector. Such industries may relate to mining and quarrying, fishing and aquaculture, or services sectors (e.g., electricity; water supply; transportation; wholesale/retail trade). This pruning yields a sample count of 203,031 for the organised sector and 655,571 for the unorganised sector. Finally, we delete observations with missing, null, negative, or extremely large values of output (being greater than 1 billion rupees), output per worker (being null or greater than 1 million rupees), or employment (being over 500,000). At this stage, we also delete from the NSSO surveys the states that are not surveyed by ASI. We also delete observations that have blank state codes. These exclusions result in a sample of 169,814 observations for the organised sector and 651,808 for the unorganised sector.

Table 1 reports a number of descriptive statistics about the data we use and its coverage. Panel A provides a raw count of observations after basic pruning for each type of establishment and year of data. Panel B reports the corresponding number of establishments in Indian manufacturing. Panel C contains the number of establishments that report the value of land and building *owned*, while panel D reports average values of land and building owned by establishments in Indian manufacturing after winsorization at the 1% level. Panel E contains some information about post imputation values and the transformation of rental values into asset values.

Finally, panel F reports the revenue shares of land and building for various subsets of Indian manufacturing establishments and years of data. As these trends are of interest in their own right, Ghani et al. (2014) provides additional tabulations and notes on how these trends compare to that of energy usage, the primary subject of Ghani et al. (2014). Two key results are worth noting here. First, the early stability of land and building usage per output unit and its rise in 2010 are consistently observed when preparing the data with alternative procedures. Second, a decomposition exercise finds that almost all of the increase in land and building intensity is occurring within districts, rather than being due to reallocation of activity over districts (e.g., establishments moving toward high-priced areas). This within-district feature motivates the emphasis in this paper on district-level analyses.

To estimate productivity and misallocation, we need to further clean up the data by dropping establishments with negative value added, missing raw materials value, unknown district names, and/or locations in small and conflict states. These include Andaman and Nicobar Islands, Dadra & Nagar Haveli, Daman & Diu, Jammu & Kashmir, Tripura, Manipur, Meghalaya, Mizoram, Nagaland and Assam. The final sample that we work with for productivity and misallocation metrics consists of 145,829 observations from the organised sector and 575,989 observations from the unorganised sector. This leaves us with over 320 districts each year for the two sectors, and appendix table 2 provides exact counts for each sector and year. These numbers roughly correspond to the number of districts for which land and building values are available. Although we end up considering only about half of the 630 Indian districts in our study, the included districts represent over 95% of establishments, employment, and output in the manufacturing sector throughout the study period.[2]

3. Measuring misallocation

3.1 Output and factor misallocation

Firms differ enormously in what they produce and how they produce it. Even within narrowly defined market segments, firms are very heterogeneous in how productive they are (Fox and Smeets, 2011; Syverson, 2011). For example, a firm at the top decile of productivity is twice as productive as a firm in the bottom decile for a typical manufacturing industry in the United States (Syverson, 2011). For India and China, a firm at the top productivity decile may be five times as productive as a firm at the bottom decile (Hsieh and Klenow, 2009).

A distribution of firm productivity within a given industry, however narrowly defined, is not necessarily a sign of inefficiency. Managers of the most productive firms may not be able to supervise all the workers in the industry (Lucas, 1978). Some poorly productive firms may also produce some highly specific varieties that some consumers value (Melitz, 2003). High transport

[2] Our analyses also use a number of social, economic and geographic characteristics of districts as well as information about specific policies. We describe these data as they are introduced.

costs may also lead to less productive firms serving customers in remote areas. At the same time, a number of features may prevent firms from operating efficiently. So far the literature has focused mainly on idiosyncratic taxes on inputs and outputs faced by firms (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) experiencing financial market frictions (Banerjee and Moll, 2010; Midrigan and Xu, 2014) and poor management (Bloom et al., 2013).

This said, most models of heterogeneous firms predict that more productive firms should be producing more output and using more factors of production. More specifically, these models imply that the most productive firms should be the ones with the highest revenue and should also be using the largest amounts of factors of production. It is easy to understand that if a more productive firm uses a smaller amount of land, less capital, and fewer workers than a less productive firm, total output would increase by reallocating some amount of these factors of production from the less to the more productive firm. Then, from this, it follows that the ranking of firms by productivity should be the same as the ranking of firms by factor usage. Productivity and factor usage should be perfectly (rank-)correlated under a perfectly efficient allocation of factors across firms.

Thus, a less than perfect correlation between productivity and factor usage indicates a misallocation of factors across firms. The worse the correlation between productivity and factor usage, the more misallocated are factors of production and the less output is produced relative to an efficient allocation. Using this insight, Olley and Pakes (1996) propose to measure misallocation using the covariance between firm productivity and output. Box 1 provides some technical details. This measure of allocation is also equal to the difference between un-weighted and weighted productivity, where the weights are the firm shares of output. The Olley-Pakes (OP) measure of misallocation—sometimes referred to as the OP decomposition—has been widely used to explore issues such as changes in factor allocation in the telecom industry after deregulation (Olley and Pakes, 1996), the effects of structural reforms in developing countries (Eslava et al., 2006), productivity differences across countries (Bartelsman et al., 2013), and the role financial institutions in factor allocation (Midrigan and Xu, 2014).

Box 1. Measuring output and factor misallocation

Define the share-weighted aggregate productivity of a group g of n establishments as:

$$\Phi_g = \sum_{i=1}^{n} s_i \varphi_i \,, \tag{1}$$

where s_i the share of establishment i in the group (with $\sum_{i=1}^{n} s_i = 1$) and φ_i a measure of total factor productivity. For now, we can think of the 'share' as a weight. It will receive a precise definition below.

As discussed by Melitz and Polanec (2013), there are many ways to decompose aggregate productivity. Following Olley and Pakes (1996), it is particularly insightful to note that:

$$\Phi_g = \overline{\varphi}_g + \sum_{i=1}^{n} (s_i - \bar{s}_g)(\varphi_i - \overline{\varphi}_g) = \overline{\varphi}_g + n \operatorname{cov}(s_i, \varphi_i) \,, \tag{2}$$

where $\overline{\varphi}_g \equiv \frac{1}{n} \sum_{i=1}^{n} \varphi_i$ is the unweighted productivity mean across establishments in group g and $\bar{s}_g (= 1/n)$ is the mean establishment share.

Then, we can define misallocation in group g as

$$M_g = -(\Phi_g - \overline{\varphi}_g) = -n \operatorname{cov}_g(s_i, \varphi_i) \,. \tag{3}$$

This misallocation index is a negative function of the covariance between the share and productivity of establishments so that misallocation is minimised when the covariance between shares and productivity is highest. Relative to usual practice in past work (e.g. Bartelsman *et al.*, 2013) and to ease the reading of our results, equation (3) multiplies the 'misallocation' covariance term by minus one so that an increase in the index corresponds to an increase in misallocation.

There are many ways to measure the 'share' of an establishment in a group. In our work, we use establishment shares for output, value added, and each factor of production. These various shares allow us to capture various dimensions of misallocation. After noting labour by L, capital by K and breaking it into land T, buildings B, and other fixed assets OK, output by Y, and value added by VA, we can define the measures of misallocation M^L, M^K, M^T, M^B, $M^{T\&B}$, M^{OA}, M^Y, and M^{VA} for labour, capital, land, buildings, land and buildings, other fixed assets, output, and value added.

[...]

Box 1 (continued). Measuring output and factor misallocation

Note also that computing M_g in equations (2) and (3) requires the productivity measure φ_i and factor usage used to compute the share to be comparable across establishments in group g. For this, establishments need to belong to the same industry (e.g., textiles or basic metals). We compute productivity as residual. In our baseline productivity measure, we subtract measures of factor usage weighted by their output coefficient from value added for each establishment in a standard manner. Such productivity measures are not comparable across industries or within the same industry across the organised and unorganised sector as we estimate different coefficients for factors.

For our baseline, we define each group as an industry in a sector and a district. We can then define misallocation in district d as

$$M_d = \sum_{g \in d} S_g M_g \,. \tag{4}$$

where S_g is the share of group g in district d computed in the same fashion as the establishment share s within group g above. For instance, to compute the misallocation of value-added in district d in the organised sector, we sum the misallocation indices (3) over all industries in the organised sector in this district weighting each industry by its share in the aggregate value-added of district d. To compute the misallocation of employment in district d for the combined organised and unorganised sectors we perform the same operation weighting by the employment share of each industry in each sector (organised and unorganised) separately. We discuss alternative measures of district misallocation in the text.

Relative to extant literature, we expand the OP approach in two directions. First, as already noted, while the literature computes misallocation indices using output or value added, analogue misallocation indices can be computed through weighting establishments by their use of any given factor of production instead of output. More formally, depending on what is used to measure establishment share, we can build measures of allocation efficiency for output, value added, employment, land, land and buildings, etc. This variety of measures allows us to explore how various measures of allocation efficiency are related to each other and which ones matter to determine aggregate outcomes.

Second, we compute misallocation indices for subnational units. At this point it is important to clarify what a measure of misallocation within a district captures. Computing the misallocation of, say, employment in each district separately yields a measure of the misallocation of workers within each district, taking as given the distribution of employment across districts. Hence, the misallocation of employment within districts is only one component of the misallocation of employment in the entire country. For instance, it is possible to imagine situations where more productive establishments employ more workers within each district, but that districts that host more productive establishments on average have access to substantially fewer workers than less productive districts. This would be a case of a low level of misallocation within districts but a high level of misallocation across districts. While we leave the analysis of misallocation across districts for ongoing work as described in the conclusions, we note that our focus on within-district misallocation is warranted by the fact that within- and between-district misallocations probably have very different root causes and most likely call for different policies.

3.2 Issues with measuring misallocation at the district level

Our primary metrics of misallocation focus on the combined organised and unorganised manufacturing sectors, and we also directly compute the misallocation within each sector. We compute indices of misallocation for output, value added, and factors of production taken individually, such as employment, or in subsets, such as land and buildings or all fixed assets. We then use these measures of misallocation across Indian districts in various regressions that seek to validate their usefulness and describe the implications of factor misallocation.

Leaving aside for now the issues associated with the specifics of our regressions, our approach raises two broad concerns. The first is that the measures of misallocation that we use, while intuitive and informative, are not structural. Below, we compare the estimates obtained from our regressions to counterfactuals from the models used in the literature so far, typically finding quite consistent results.

The second main issue is that our approach requires measures of misallocation at the district level, while the OP approach generates measures of misallocation within industries. To reconcile this requirement of a district-level misallocation measure with industry-level measures, the most obvious solution is to compute misallocation at the industry level within each district and then aggregate across industries within each district.

Aggregating industry-level indices of misallocation within each district to obtain a district index of misallocation can be performed in several ways. Our baseline computation sums industry misallocation across all industries and weights each industry by its corresponding local share of manufacturing activity. For instance, to compute the misallocation of employment within a district, employment misallocation is summed across industries using the employment share of industries in the district as weights.

While this is a natural and straightforward way to aggregate across industries, other aggregations are possible. In their cross-country comparison, Bartelsman et al. (2013) use constant sector shares. Indices with constant sector shares provide a partial measure of district misallocation that does not account for the possibility of more misallocated sectors having a tendency to predominantly locate in more misallocated districts. Another drawback of constant-share aggregation is that misallocation is less accurately measured in smaller industries with fewer establishments. From a measurement perspective, it is better to give a small weight to an industry that is locally small even though this industry might be much larger nationally. Another alternative is to sum misallocation across industries weighting each by its local share, as in our main design, but also subtract the average tendency of each sector to be misallocated in the country. This type of index measures excess misallocation relative to what would be expected given the local composition of industrial activity. Although we focus on our preferred measure of misallocation described in box 1, we compare our results to those obtained with constant-share aggregations and with aggregations that isolate excess misallocation.

Among possible alternative measures of misallocation, we also compute simpler indices based on the covariance between output or a given factor of production and productivity across all establishments regardless of their industry. The main advantage of measuring misallocation in this manner is that the covariance is measured over a much larger sample (all manufacturing establishments instead of all establishments in an industry). The main problem with this alternative is the following. When, say, capital intensity differs between two industries, it is not immediately obvious that the lower productivity establishment in the more capital-intensive industry should be using less capital than the more productive establishment in the less capital-intensive industry. If the differences in productivity are small enough, the establishment in the more capital-intensive industry should be using more capital even though its productivity is lower. Because we find factor coefficients to be relatively constant across industries in our productivity estimations, this problem may not be as important as it first appears. It remains nonetheless that this simpler metric will overstate the true level of misallocation.

A possible refinement is to correct for the use of factors by each establishment by the intensity with which the industry of this establishment uses factors. Then we can use these corrected factor measures to compute equation (3) across all establishments irrespective of their industry. For instance, in an industry that uses twice as much capital as the average industry, we can multiply the amount of capital used by establishments in this industry by 0.5 to normalize its capital use and then compute misallocation directly across all establishments.

The OP measure of misallocation described by equations (3) and (4) in box 1 is not the only available measure of the efficiency of factor allocation across establishments. In a celebrated article, Hsieh and Klenow (2009) propose a model of heterogeneous firms facing idiosyncratic distortions. They show that under some conditions, the efficiency of factor allocation across firms can be measured by the dispersion of observed firm productivity.[3] This alternative measure of misallocation has been widely used, and Restuccia and Rogerson (2013) provide a discussion. We note in particular the work of Adamopoulos and Restuccia (2014) and Restuccia and Santaeulalia-Llopis (2014). They focus on land as well, but they are concerned with agricultural

[3] Following Foster et al (2008), they call this measure of productivity revenue productivity (TFPR). This is the concept of productivity usually estimated by economists and it embeds the price at which output is sold. TFPR is thus a measure of the ability of firms to generate revenue. It stands in contrast with physical productivity (TFPQ), which measures the ability of firms to produce output. We return to this distinction below.

land in rural areas whereas we deal with land used by manufacturing establishments, which are often found in urban settings in developing countries.

The OP misallocation measure for output/value added, the OP misallocation measure for factors, and the Hsieh-Klenow misallocation measure have interesting and subtle differences. Our measure of output misallocation essentially captures the covariance between output and productivity. Recall that output is given by: $Y = TFP.L^{a}.(T\&B)^{b}.(OA)^{c}$ where *TFP* is an estimated productivity residual, *L* is employment, *T&B* is land and buildings and *OA* is other fixed assets. In situations where more productive establishments use less of factors relative to less productive establishments, the covariance between productivity and output decreases and our measure of output misallocation increases. However, the misallocation of factors will also allow poorly productive establishments to remain in business. Then, a worse misallocation of factors will perhaps increase the variance of productivity for active establishments. In turn, this increases the covariance between output and productivity. Hence, our measure of output misallocation will embed both the "negative" direct effect of factor misallocation and a "positive" indirect effect of market selection. Our measures of misallocation for individual factors are not directly affected by this second effect. They only consider whether more productive firms utilize more of a given factor. This is why looking at factor misallocation may be more informative in some contexts than looking at output misallocation. Finally, HK misallocation only considers the dispersion of productivity. That is, relative to the OP measure of output misallocation, it only accounts for the indirect market selection effect and does so in the opposite direction relative to the OP metric for output or value added.

Given these differences, we also follow Hsieh and Klenow (2009) and use the dispersion of establishment productivity to measure misallocation and check the robustness of some of our key results. Unfortunately, the HK approach offers only one general measure of misallocation. Unlike our extension of the OP approach, it does not allow us to distinguish between output and factor misallocation or identify differences in misallocation among factors.

As made clear by this discussion, there is no perfect measure of misallocation at the district level. As described below, the productivity measure that we use in equation (3) of box 1 is estimated as a residual from a regression using imperfect data. Then, computing industry misallocation in each district raises some sampling issues as the number of surveyed establishments is numerically large—larger, for example, than the survey size of the Annual Survey of Manufacturers in the United States—but not a universal census for each district. Finally, aggregating across industries raises further challenges, as just highlighted. Systematic measurement error is a worry only to the extent that it is correlated with misallocation. That we may over- or understate misallocation in all districts is not an issue as our main sources of variation are either differences across districts or changes over time within districts. It is nonetheless possible to imagine that measurement error varies systematically with misallocation. For instance, establishment productivity in industries subject to greater misallocation may be measured more accurately as the frictions that drive misallocation make the use of factors of production less responsive to demand shocks, or vice versa. This would lead us to understate misallocation in more misallocated industries and perhaps in more misallocated districts. Adding to this, classical measurement error will lead us to underestimate the true effect of factor misallocation when used as an explanatory variable in a regression. Although we estimate large economic effects of factor misallocation, our results may understate the true impact.

3.3 Estimating total factor productivity

To compute the indices of misallocation defined above, we need a measure of establishment productivity. Since productivity is not directly observed, it is usually estimated as a residual measuring the ability of an establishment to produce conditional on the inputs that it uses. There are two main issues associated with this standard approach. First, as already noted, in a large majority of cases, we measure the revenue that an establishment receives, not the physical quantity of output it produces. Even when the number of units produced is observed, it is unclear what this measure means in most industries since product quality is highly heterogeneous. So while we are able to observe the ability of firms to generate revenue, we are not able to observe their ability to produce a quantity and quality of output.

The second important estimation issue with production functions relates to the endogeneity of inputs. Any firm-specific demand or productivity shock will affect both the residual of the production function estimation and the demand for factors of production. This endogeneity of input choices has received much attention since Olley and Pakes' (1996) seminal work. In our work below, we follow the approach subsequently developed by Levinsohn and Petrin (2003) (LP) which relies on the use of intermediate inputs, in particular energy consumption, to detect demand and productivity shocks. The key idea that underlies this approach is that demand and productivity shocks will affect energy consumption but not capital stocks, which are decided before those shocks are known.

The LP approach for productivity estimation requires a panel of firms, which is unfortunately not available to us from the ASI and NSSO surveys that contain district identifiers. Using information for establishments in the same location and industry, we can nonetheless implement an approach in the same spirit as LP that only requires information from the panel of districts. The main idea is to use energy consumption at the district level to detect temporal local demand shocks in a manner parallel to that done at the plant level with the LP methodology. This approach was initially developed by Sivadasan (2009). We appropriately tailored it to our needs as explained formally in appendix box A1.

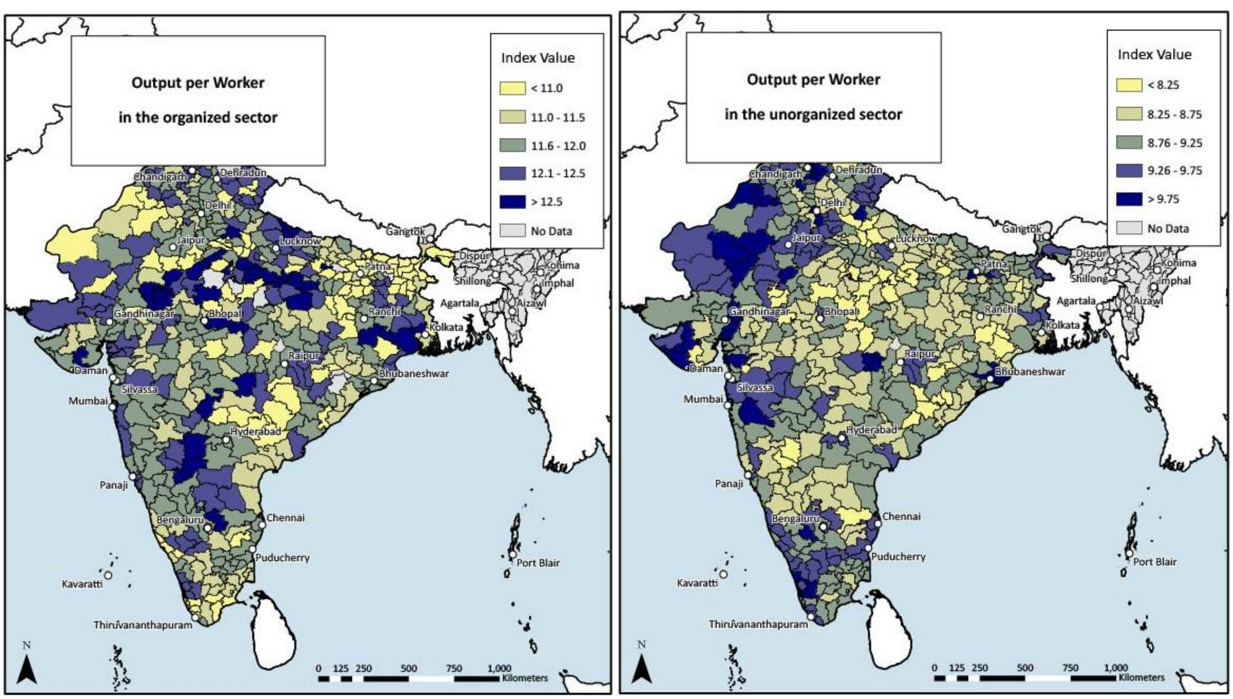
We show below that we obtain similar results when using OLS to estimate productivity levels. That OLS and our LP-Sivadasan measures of productivity should deliver very similar results is consistent with our findings that extremely large frictions generate significant misallocation. It is then only natural to expect very little factor adjustments for firms following demand shocks. A more subtle worry is whether capital is actually the long-run factor for Indian firms. We find very similar outcomes in estimations that treat labour as the long-run factor.

In summary, to alleviate concerns focused on specific approaches to estimating productivity, we replicate our main regressions for measures of misallocation computed from a variety of productivity estimators. To alleviate concerns coming from sampling issues in small industries, we replicate our results imposing a higher selection threshold for local industries and with various weighting strategies. Finally, to alleviate concerns regarding our aggregation of industry misallocation indices at the district level, we replicate our results for a variety of aggregation approaches.

3.4 Some descriptive facts about output per worker and productivity in India

Figures 1a and 1b show maps of output per worker in the organised and unorganised sector, respectively. Both maps are consistent with well-known aggregate statistics. Districts with high output per worker are found in the north western part of the country (e.g., Gujarat, Rajasthan, Haryana, Punjab, Maharashtra). These are also states with a high GDP per capita relative to the rest of the country. On the other hand, low output per worker is more prevalent in the eastern part (e.g., Uttar Pradesh, Bihar, Orissa). High output per worker is also evident around major population centres, including Bangalore and Chennai as examples outside of the states already mentioned. Finally, output per worker is typically higher in coastal areas relative to the interior.

Figure 1: Map of output per worker in Indian districts, by quintile, 2010



(a) Organised sector (b) Unorganised sector

Comparing the scales of figures 1a and 1b, it is immediate that output per worker is nearly everywhere much higher in the organised sector. There is also an apparent correlation between output per worker in the organised and unorganised sectors across districts, but it is modest at

0.09. Output per worker is high in the unorganised sector relative to the organised sector in Kerala, while the opposite hold true in West Bengal.

Turning to productivity estimates, table 2 reports the estimated coefficients for 22 two-digit manufacturing industries. We start with the combined estimations for the organised and unorganised sectors and then treat each sector separately. A number of important features emerge from this table.

First, the organised and unorganised sectors differ in their capital and employment intensity. The mean share of capital (i.e., total fixed assets) across industries is 0.41 in the organised sector versus 0.31 in the unorganised sector. These differences may reflect differences in access to capital or other conditions that determine the operation of establishments in both sectors. They may also capture specialisation in different segments of these industries. Regardless of their origin, these differences are large enough that we want to estimate productivity allowing for factor coefficients to differ across sectors for the same industry.

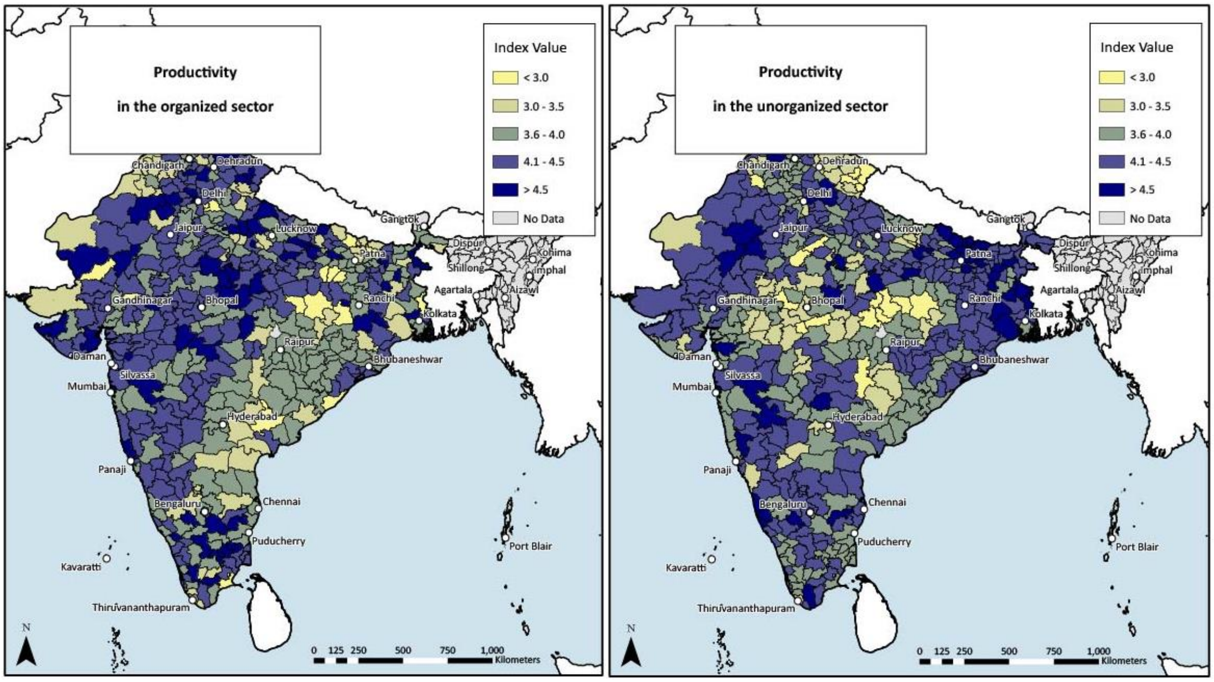
A large majority of industries appear in Table 2 to operate under increasing returns, since the sum of the employment and capital coefficients often exceeds one. This is not an artefact of our estimation technique. Similar results are obtained when productivity is estimated with OLS techniques and modelling these two factors of production or separating the main components of capital: land, buildings, and other fixed assets. Measuring increasing returns in the ability of establishments to generate revenue is puzzling as we expect most establishments to face only imperfectly elastic demand. We believe two phenomena are at work. First, they could partly reflect true increasing returns faced by all establishments that are constrained by frictions and cannot reach their optimal size. Measured increasing returns could also reflect frictions that systematically affect larger and more productive establishments. For instance, strict labour regulations may affect large establishments more than small ones (e.g., Banerjee and Moll, 2010), so that large and more productive plants employ relatively fewer workers. This type of phenomena could then lead to coefficients on factors of production that are biased upwards in productivity estimations. That is, while plants operate under constant or decreasing returns in reality, it may look like they operate under increasing returns as more productive establishments use relatively fewer factors. Regressions could attribute differences across establishments to differences in factor usage, making factors look more productive than they really are.

We thus face the empirical challenge that we need to know establishment productivity in order to measure frictions in factor allocation across establishments. At the same time, these same frictions can distort measures of establishment productivity. To address with this problem, we duplicate our results using productivity estimates that impose constant returns to scale.

More generally, our main need here is to obtain the best possible estimate for productivity at the establishment level. For this, we focus on the LP-Sivadasan approach. We also experiment satisfactorily with OLS approaches to ensure robustness and to model more flexible functional forms that can consider more factors of production (e.g., break down total fixed assets into land, buildings, and other fixed assets). When we estimate OLS TFP with three factors of production (land and buildings, capital, and employment), the average coefficient on land and building is 0.13 and the average coefficient on other fixed assets is 0.28 in the organised sector. These two coefficients sum to 0.41 which is the same as the average coefficient on TFP with our main LP estimation. When we distinguish between land and buildings, the average coefficients are 0.04 on land, 0.11 on buildings, and 0.27 on other fixed assets. These coefficients sum to just above 0.41. These elasticities of establishment value added with respect to land and buildings are used below to assess the effects of increased factor availability.

Figures 2a and 2b represent our TFP estimates averaged by district for the organised sector and unorganised sector, respectively. Like the earlier maps of output per worker, these two maps reveal a strong contrast between the more productive western parts of the country and the less productive eastern areas. These two maps also again show areas of higher productivity around India's largest metropolitan areas, most notably Kolkata, Chennai, Bangalore and Delhi. Although a full development or growth accounting exercise is beyond the scope of our study, the comparison these figures is suggestive that India's spatial disparities in output per worker are to some extent productivity disparities. There is no correlation between the organised and unorganised sectors with respect to district-level TFP.

Figure 2: Map of TFP in Indian districts, by quintile, 2010



(a) Organised sector (b) Unorganised sector

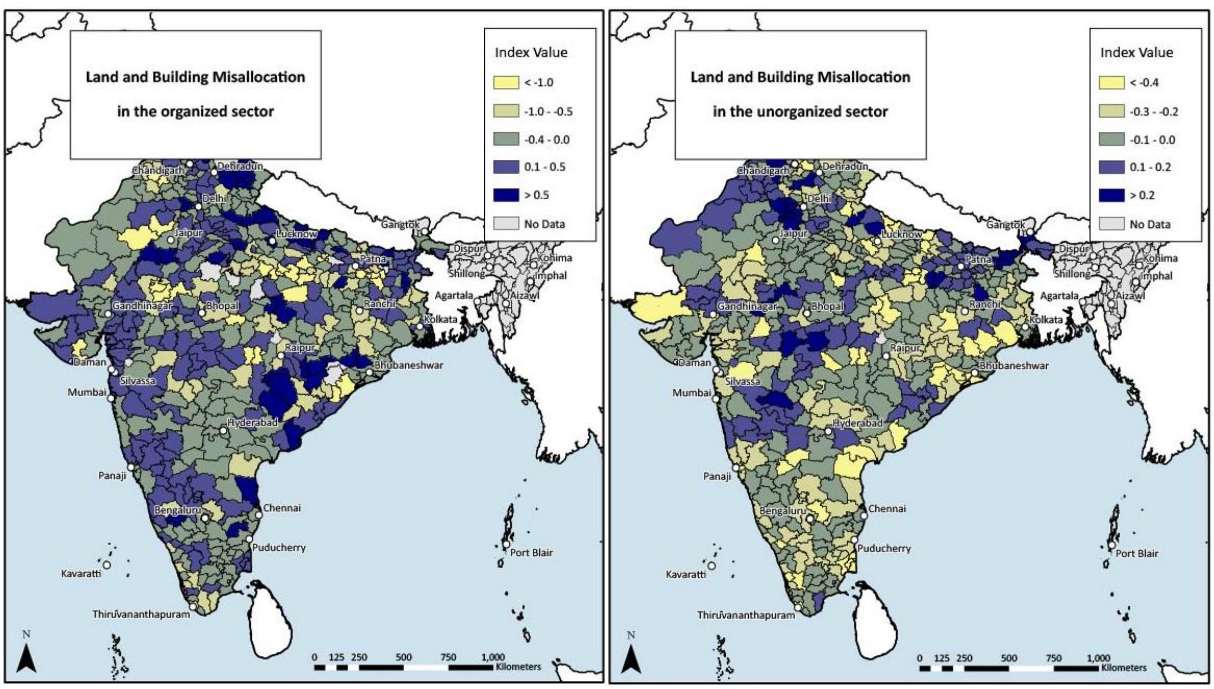
3.5 Some descriptive facts about misallocation in India

Figures 3a and 3b provide maps of our preferred misallocation index for land and buildings in the organised and unorganised sector, respectively. Darker colours indicate greater misallocation. Both sectors display a negative correlation between output per worker and misallocation. The more misallocated districts in the northeast, south, and interior are all districts with low output per worker. The next section documents and quantifies these relationships more precisely.

Table 3 reports descriptive statistics for indices of misallocation for Indian districts. We report statistics for the full manufacturing sector, followed by breakouts of the organised and unorganised sectors. A number of features are worth highlighting. Recalling that higher index values indicate greater misallocation, we estimate less value-added misallocation in panel A when considering the full manufacturing sector than when considering each sector independently. This is not surprising given the high concentration of output in the organised sector and its more productive establishments. Combining the organised and unorganised distributions increases the correlation of these output shares and productivity.

These district-level indices for misallocation of output and value added are reasonable compared to prior literature. Bartelsman et al. (2009) compute overall input misallocation indices for 18 countries at different levels of economic development. For the sake of comparison we compute an average input factor misallocation index for the full manufacturing sector, weighting indices of factor input misallocation by their average share in production. Despite some computational differences with Bartelsman et al. (2009), these indices are roughly comparable. The mean over 1989-2010 for Indian districts is about -0.33. This is marginally worse than the United States (as a whole) calculated by Bartelsman et al. (2009). More generally, Indian districts are less misallocated than the majority of the 18 countries considered by Bartelsman et al. (2009).

Figure 3: Map of output misallocation in Indian districts, by quintile, 2010



(a) Organised sector (b) Unorganised sector

The second key feature of table 3 is the considerable variation of misallocation across districts. The standard deviations for the indices of misallocation for output and value added are two-thirds of their levels for the full manufacturing sector in panel A. For the organised sector, the mean and standard deviation are comparable. This suggests considerable differences in misallocation within India, a fact not previously documented by the literature. The differences in misallocation within India are even larger than the differences across countries estimated by Bartelsman et al. (2009). The difference between the country with the lowest misallocation and that with the highest in their data is about 0.7. This corresponds to about 1.3 standard deviations for misallocation of value added in the full manufacturing sector of India, and 1.6 standard deviations in the organised sector.

A third key feature in table 3 is the extreme misallocation of individual factors of production, which contrasts with the lower levels of misallocation of output and value added. While the average 1989-2010 misallocation of output is -0.71 for the organised sector, it is only -0.27 for employment and -0.33 for buildings and land. The levels are even lower for each sector individually at approximately -0.1 for land and building factors.

Recall that zero values correspond to a situation where factor allocation is orthogonal to productivity. Hence, while more productive establishments manage to produce more than less productive establishments, allocations of some factors of production are barely better than random. Given the large variation in factor misallocation across districts, this actually indicates that there are many districts in India where factor allocation is worse than random. This troubling feature is vividly illustrated by figures 3a and 3b, where land and building misallocation indices are positive in about 40% of the districts.

Three candidate explanations exist for this gap between the indices of misallocation for output and value added and similar indices for individual factors of production. The first is that the data measure output much better than factors. While it is perhaps relatively easy for an establishment to know what its revenue is, it may be much harder to know what its capital is. This argument, however, could work in the opposite direction as it is not immediately obvious whether it is output or labour that is better measured. There is also an incentive for firms to hide output more than factors of production.

The second explanation is that even if factors are allocated equally across establishments (or at random), we still expect more productive establishments to produce more output. As argued above, vast differences in establishment productivity (e.g., due to managerial skill) may then explain why a reasonably low level of misallocation of output can co-exist with extreme factor misallocation. The important counterpart to this statement is that improvements in factor allocation may yield large output gains. We return to this issue below. Finally, we also need to keep in mind that a highly productive establishment with little land may be able to offset this by being particularly intensive in employment or in other forms of capital. For now, we only consider the covariance between factors and productivity but not how factors co-vary.

Finally, two more subtle patterns also emerge from table 3. First, the misallocation of employment and land is worse than the misallocation of buildings (when it can be separated from land) and other fixed assets. Second, there appears to be a mild trend towards a worsening output and factor misallocation over time.

As we turn to our regression analyses, our focus will be on the total misallocation in the manufacturing sector. At times, we will describe results that consider variation just within the organised or unorganised sectors. Similar to Table 3, it is important to highlight that results for the combined sample will not be an average of these sector-level results. This is because the full sample also accounts for the fact that the frictions at the root of factor misallocation will affect whether an establishment belongs to the organised or the unorganised sector. It further captures the size of productivity differences between the sectors and their relationship to establishment scale.[4]

4. Validation of misallocation metrics

While a number of studies calculate and compare metrics of misallocation, there has been limited scope to validate them. An advantage of our focus on factor misallocation—and in particular

[4] The following extreme example helps form intuition. Consider a situation where more productive establishments are either unaffected by frictions or completely crippled by them (i.e., frictions prevent them from hiring more than 10 workers). Whether a more productive establishment is affected is random. In this extreme case, there is no misallocation in the upper part of the employment distribution (i.e., in the organised sector) since only productive establishments unaffected by frictions are represented. On the other hand, establishments with low employment (i.e., in the unorganised sector) will either be constrained productive establishments and unconstrained poorly productive establishments. Measured misallocation will be extremely low in the organised sector and high in the unorganised sector. Combining the two sectors together will also yield a misallocation worse than any sector individually.

land misallocation—is that we can demonstrate their general validity/usefulness by studying changes in district-level misallocation around two important policy reforms in India: the repeal of the Urban Land (Ceiling and Regulation) Act and changes in state-level stamp duties (taxes on land sales). This section shows that reforms that reduced frictions in the functioning of land markets are associated with a reduction in the misallocation of land (and also output).

Our empirical strategies are differences-in-differences estimations around policy changes. While our emphasis is mostly on the validation of our metrics, the empirical connection of land misallocation to policy determinants is interesting in its own right and worthy of close study. The ULCRA reform has the advantage of being unanticipated, and in our analyses we control for trends in many other local traits that could correlated with policy adoption. We are cautious to note that other factors or policies might be adjusting alongside those that we study, and these unobserved factors could impact our results in terms of the likely impact of a policy reform. For our core focus, observing these strong linkages of our misallocation indices to frictions thought to reduce allocative efficiency provides greater confidence in our measurement design and their useful for quantifying the role of factor misallocation.

4.1. The repeal of Urban Land (Ceiling and Regulation) Act

In 1976, the parliament of India enacted the Urban Land (Ceiling and Regulation) Act (ULCRA) with the main objective of limiting the concentration of urban land. ULCRA imposed ceiling limits for holdings of vacant land, prohibited transfers of land and buildings, and restricted building construction in 64 of the largest urban agglomerations (central cities and their suburbs). More specifically, this regulation distinguished between four groups of cities. It applied to all of the largest cities and other cities with population larger than 200,000 in 1971 in 17 states and three UTs. Because the regulation applied to both ownership and 'possession' of land, it constrained both owners and renters (lessees). ULCRA further imposed potentially draconian penalties to offenders, including the destruction of newly built properties or the forced purchase of properties by the government at a symbolic price. Kimura (2013) describes how these regulations severely constrained the operations of the land and property markets in areas where ULCRA applied.

Despite the intentions of parliament, there is little empirical evidence that the equity objectives of ULCRA were fulfilled (Sridhar, 2010). In fact, the law artificially restricted the supply of urban land (e.g., by freezing large areas of land in legal dispute), bid up land prices, and encouraged corruption (Joshi and Little, 1991; Bertaud, 2002). Importantly, ULCRA also prevented private developers from assembling land for subsequent development. For almost a quarter century, ULCRA practically halted legal development of land by the private sector in urban areas unless exemptions were obtained (Srinivas, 1991). The regulation and market constraints reduced the incentives of landholders to invest in building construction. Thus, a large proportion of firms were both land and building constrained by way of ULCRA.

In 1999, the Repeal Act gave rights to state governments of India to repeal ULCRA. The ULCRA reform was mostly anticipated. A number of states and UTs repealed ULCRA by 2003, including Delhi, Gujarat, Haryana, Karnataka, Madhya Pradesh, Orissa, Punjab, Rajasthan, and Uttar Pradesh. By contrast, Andhra Pradesh, Assam, Bihar, Maharashtra, and West Bengal kept ULCRA effective until 2008.[5]

To assess the effects of the repeal of ULCRA on misallocation, we estimate a series of regressions where we consider the district-level change in misallocation from 2000 to 2010 as the dependent variable. Our key explanatory variable is an indicator variable for states that repealed ULCRA early, as listed above. We consider the set of late adopters to be the control group as repeals in 2008 are unlikely to have substantial consequences by 2010, especially for ASI establishments surveyed in 2009-10. Estimations are unweighted cross-sectional regressions that include 252 Indian districts. We cluster standard errors by state to represent the state-level choices being made to repeal ULCRA.

$$M_{i,2010}^{L} - M_{i,2000}^{L} = a_0 + a_1 EarlyRepeal_i^{ULCRA} + b_1 X_i + \epsilon_i$$

[5] The negative effect of ULCRA is still evident in the land use patterns of Mumbai, many years after ULCRA was abrogated by the State of Maharashtra in 2007 (Bertaud, 2011). Siddiqi (2013) provides an in-depth analysis of the political economy of ULCRA adoption and its repeal in Mumbai (Maharashtra). With the repeal of ULCRA in 1999, about 25,000 acres of land have been freed. However, only 10,000 of these are in developable zones, while the remaining 15,000 fall in areas with restrictions—such as coastal zones and forest lands (Sridhar, 2010).

Two main estimation worries are that the repeal of ULCRA may have coincided with other district features that affected misallocation and that the initial sample of ULCRA districts may have been highly selected in a way that affects the results. To minimise these worries, we include a number of control variables *X*. The most essential is the initial level of misallocation. This is important because districts where ULCRA applied may have experienced greater misallocation in 2000. We also have a battery of additional controls for the initial traits of districts (e.g., population density, local demographics, local infrastructure traits) as listed in the notes to table 4.

Table 4 reports our main results for the effect of the repeal of ULCRA. Panel A considers misallocation of land and buildings as the outcome variable. We focus on estimations that consider the full manufacturing sector, combining organised and unorganised sectors together. The negative coefficient indicates that the early state-level repeal of ULCRA is associated with a stronger decline in land and building misallocation during 2000 to 2010 compared to late adopters. The 0.057 coefficient is quite substantial and represents about one-tenth of a standard deviation of land and building misallocation. Using results later reported in Table 7, the 0.057 decrease in misallocation for land and buildings of associated with the repeal of ULCRA corresponds to an increase in output per worker of about 3.7%.

Panel b considers the change in misallocation for value added for districts. ULCRA is associated with reduced misallocation on this dimension as well. In fact, it appears the declines may have been greater here than on land and buildings in economic terms. In terms of standard deviations, the ULCRA repeal is associated with a decline of about one-fifth of a standard deviation in value added. This larger impact may seem puzzling to start, but it is quite possible under the scenarios highlighted above where improved factor allocation is magnified by inherent differences in establishment productivity (e.g., better land access is amplified by a capable manager). Current research is also considering whether the improved functioning of land and property markets aids in other factor acquisition through improved property rights and borrowing conditions. As suggestive evidence, unreported analyses find the allocation of other fixed assets also improved with ULCRA's repeal.

Column 2 shows this result is robust to adjustments in the covariates modelled, although the basic controls are important for the statistical precision with which we can measure ULCRA's effect due to many other changes underway in India during this period. Columns 3-6 show that these results are quite robust to variations in methodology. We show similar findings when misallocation is computed without aggregating first by industry, when using OLS estimates of productivity, when weighting districts by initial employment, and when following the HK approach. The results are actually stronger statistically than our primary metrics in column 1. The point estimates of the coefficients are not directly comparable due to the different scales and variances of the metrics. The results are also robust to different approaches towards extreme values like winsorization.

Unreported estimations consider misallocation in distributions specific to the organised and unorganised sectors. The repeal of ULCRA is more closely associated with reduced misallocation for factors in the unorganised sector than the organised sector: the unorganised sector coefficient is -0.035 (se=0.021), which represents again a tenth of a standard deviation for the sector. Yet, it is quite clear overall that the major impact of ULCRA came less from changes in misallocation within each sector vs. changes in the relative sizes of sectors and their joint distribution for misallocation. As one example, the log change in the value of land and buildings in the organised sector with ULCRA's repeal is 0.722 (se=0.261) for the organised sector, compared to 0.420 (se=0.197) for the unorganised sector. These relative changes in factor allocations across sectors appear to have been more central for the reductions in misallocation compared to changes within sectors.

4.2 Stamp duty

High stamp duties are another trait of Indian property markets.[6] These taxes are collected whenever a real property is transacted. While these taxes are between zero and 5% in most North

[6] Stamp duties in India are imposed under the Indian Stamp Act, 1899, as amended several times over the years at the central government level. Under these central acts, each state has the authority to enact its own stamp duties, so that the specific features of the stamp duties, while broadly similar across the states, also take on state-specific characteristics. Within the broad definition of stamp duties imposed on sale and purchase of business transactions, including property, there are two sub-classifications: (i) Judicial stamp duty is usually a small fee collected by the

American jurisdictions, they tend to be much higher in India. There is also a lot of variation across states and time. For instance, the lowest rate is found in Tripura at 5%, while the highest rate of 21.2% was in West Bengal early in our period of study. West Bengal dramatically lowered its stamp duty, reaching 5% in 2003. These taxes represent an important friction and have received some academic attention outside of India, affecting for example the Canadian housing market by lowering prices and the number of transactions (e.g., Dachis et al., 2012).

High stamp duties impose high compliance costs on taxpayers and lead to widespread avoidance through under-reporting (Alm et al, 2004; Morris and Pandey, 2009). This in turn adversely affects the possibility of using land as collateral for construction financing. High stamp duties also discourage land transactions, and as a consequence reduce the supply of land on the market. High stamp duties are thought to be at the root of a $3.4 billion scam on the use of fraudulent stamp papers by Abdul Karim Telgi that was reported in India in 2002.

Table 5 uses a long-differenced estimation similar to Table 4 to consider the relationship between changes in stamp duties at the state level and local misallocation. We consider changes in misallocation from 1989 to 2010 as the outcome variable, with changes in stamp duties from 1989 to 2003 as the core explanatory variable. We again control for initial misallocation and the traits of districts.

Panel A of column 1 shows that an increase in stamp duties at the state level is associated with rising misallocation for land and buildings during the period of study. To provide a sense of magnitude, a one standard deviation increase in the change in stamp duties is associated with a one-tenth of a standard deviation increase in misallocation of land and buildings. Panel B shows a very similar result for changes in value-added misallocation, and these results are generally robust across columns 2-6. Inclusion or exclusion of district covariates does not impact these results nor their statistical precision. The one exception is that we do not observe a decline in the HK metric of misallocation with this reform.

---

court for litigation purposes, and (ii) Non-judicial stamp duty is a onetime charge on the transfer of immovable property based on the value of the transaction.

Similar to our ULCRA analysis, we have explored variations within the organised and unorganised sectors in isolation. Higher stamp duties are associated with more misallocation of value added and land and buildings for establishments in the organised sector. The economic magnitudes for the organised sector are similar to that for the manufacturing sector overall at about one-tenth of standard deviation. Higher stamp duties are not associated with significant change in misallocation among establishments in the unorganised sector. Similar to ULCRA, much of the action appears to come through changes in the relative sizes of the two sectors and their joint distributions (e.g., lower stamp duties allow more productive establishments in the organised sector to differentially expand). While these numbers should be taken with proper caution given the caveats noted earlier, they suggest potentially large frictions caused by nearly punitive taxes on the transfer of a fundamentally important asset.

5.3 Local characteristics

As a related exercise, although perhaps more mundane than studying policy adjustments, we ascertain that the traits of districts that correlate to our misallocation indices make intuitive sense. Unreported analyses evaluate which of the control variables used in tables 4-5 most closely relate to factor and output misallocation. Rather than wade through an exhausting set of tables, the following discussion summarises a large number of regressions.

We explore the urban dimension by regressing misallocation on various measures of urban scale, mostly population density and total population. We find some interesting patterns, with district population and district density yielding similar results. There is a negative association of population with misallocation of value added in the organised sector, no effect in the unorganised sector, and a positive effect when both sectors are combined. This is consistent with urban scale reducing misallocation in the organised sector but also leading to a larger unorganised sector where misallocation is greater. For the misallocation of land and buildings, the coefficients of the urban variables are generally positive, perhaps consistent with the notion that frictions in the land and property markets are more important in larger cities.[7]

[7] Interestingly for firms in the organised sector where we can distinguish between land and buildings, the coefficients on our measures of urban scale are negative for buildings and positive for land.

A limitation to these results is that their statistical significance often disappears when many more control variables are added. At the same time, it is well-known that productivity and output per worker are higher in larger and denser cities. The literature usually attributes the productivity advantage of cities to agglomeration economies. Our results suggest that agglomeration economies are not directly caused by reduced misallocation in larger cities, a result consistent with the findings of Combes et al. (2012) for France.

We also consider the effects of relative location, accessibility, and roads. We only find weak indications that the misallocations of value added and land and building decrease with a higher composite index of infrastructure (e.g., paved roads, telecom access, sanitation/water). We also find some effects associated with the distance to the major cities, but they go in opposite direction. Interestingly, state-level fixed effects can have important explanatory power. India is a federation, and institutions and policies differ across states. Consistent with our policy results above, these differences affect misallocation levels.

Finally, misallocation is negatively correlated with many proxies for development and local wealth such as access to banking, the fraction of scheduled castes and tribes in the local population, the male-female sex ratio, or access to power. It is difficult with our data to assess whether these are causal effects or merely correlations. For instance, one can easily imagine why a more limited access to banking and the financial system could worsen misallocation through credit constraints. At the same time, it is also possible that the same institutional features that drive misallocation also limit the presence of the financial sector. Either way, similar to the two policy reforms that we studied, we gain confidence in our metrics when observing these reasonable connections to other local district traits.

5. Consequences of misallocation

Having considered the validity/usefulness of the factor misallocation metrics, we now turn to how different forms of misallocation interact at the district level. We also quantify the relationship of local misallocation to productivity.

5.1 How different forms of misallocation interact

We study how forms of misallocation interact in the cross-section and also dynamically. We start by measuring the extent to which the misallocations of factors of production explain the contemporaneous misallocation of output and value added. These exercises consider both cross-sectional relationships and panel co-movements. We then ask whether the misallocation of factors of production in one period has a dynamic effect on the future misallocation of output or value added (e.g., high land misallocation leading to worsening output misallocation).

To answer the first set of questions, we regress the index of misallocation for output or value added on the corresponding factor misallocation indices. These regressions take the form:

$$M_{i,t}^{Y} = a_0 + a_1 M_{i,t}^{L} + a_2 M_{i,t}^{T\&B} + a_3 M_{i,t}^{OA} + b_t + \epsilon_{i,t},$$

where the misallocation of output in districts i and year t is regressed on the misallocation of employment, the misallocation of land and buildings, the misallocation of other fixed assets, a vector of year fixed effects, and an error term. Estimations cluster standard errors by district.

Table 6a reports the results. Estimations in panel A are unweighted district-year observations, while estimations in panel B weight districts by their initial 1989 employment levels. Column 1 considers our benchmark estimation that combines the organised and unorganised sectors, includes three factors of production, and considers misallocation of value added as the outcome variable. The relationships are quite powerful, with all forms of factor misallocation contributing to the misallocation of value added. Land and building misallocation is particularly important, with a one standard deviation increase in that factor's misallocation corresponding to a 0.6 standard deviation increase in value-added misallocation. In economic terms, land and building misallocation is the most important explanatory factor, with the misallocation of other fixed assets being the weakest (its coefficient is 0.00039 (se=0.00005)). A one standard deviation increase in employment misallocation corresponds to a 0.4 standard deviation increase in value-added misallocation.

Going into more detail, the larger coefficient on land and buildings compared to employment is striking for two reasons. First, there is greater cross-district variation in the misallocation of land and buildings than for employment or fixed assets, as reported in Table 3. We are thus finding a particularly strong coefficient on the factor with the greatest variability, per the economic interpretations given above. Second, recall that the elasticity of value added with respect to land and buildings is only about 0.13, whereas the elasticity of value added with respect to employment is 0.66 on average across industries in the organised sector. Hence, even though the land and buildings factor accounts for a small fraction of final output and value added, it plays a disproportionate role in explaining the misallocation of final output. On the other hand, the misallocation of other capital appears to account for very little of the misallocation of final output. Column 2 shows similar results when combing the two forms of fixed assets together.

To assess the robustness of these results, we perform a number of checks. Column 3 finds similar results when further adding district fixed effects to focus on panel variation. Unreported analyses also find very similar outcomes when modelling state fixed effects. In both cases, the relationships are usually statistically undistinguishable from our base estimation.

Columns 4-6 show variations on the calculation of misallocation. Column 4 yields slightly smaller magnitudes when calculating misallocation indices directly across establishments from all sectors, without allowing for industry production differences. This is consistent with the notion that ignoring the industry dimension will lead to greater mis-measurement of factor misallocation, but as important the gaps are quite small. Unreported analyses also find very similar results in variations that use national industry weights instead of local shares, ignore sectors but normalize factor usage by the inverse of the estimated factor shares, and so forth. Column 5 also shows similar results using OLS TFP metrics, and we also obtain similar results with other productivity estimation approaches. Column 6 shows robust results when using misallocation of output as the dependent variable rather than misallocation of value added.

While we should be cautious and refrain from any structural interpretation, these results shed an interesting light on misallocation in Indian manufacturing. The fairly high R-squared values for the regressions of table 6a are also interesting and suggestive of the importance of factor misallocation for accounting for the frequently discussed misallocation of output. Finally, we should keep in mind that these results are all the more striking since these indices are likely to be mis-measured. Overall, these findings are consistent with the notion that land (and to a lesser extent employment) is the least flexible factor of production and that its misallocation likely breeds the misallocations of other factors and output.[8]

Unreported analyses separate the organised and unorganised sectors. While our focus is on the combined misallocation that includes effects within both sectors and their relative sizes, it is again intriguing to isolate variation within each group. Results in both sectors are statistically significant with magnitudes for the organised and unorganised sectors of about 0.30 and 0.55, respectively. For both sectors, the standard deviation of the value-added misallocation index is comparable to the standard deviation of the land and building misallocation index, so these figures are comparable and indicate a somewhat higher importance for the informal sector of land misallocation.

To assess whether factor misallocation has dynamic implications for the misallocation of output and value added, we regress changes in output or value added misallocation from 1989 to 2010 on the initial levels of misallocation for factors in 1989. These estimations take the form:

$$M_{i,2010}^{Y} - M_{i,1989}^{Y} = a_0 + a_1 M_{i,1989}^{Y} + a_2 M_{i,1989}^{L} + a_3 M_{i,1989}^{T\&B} + a_4 M_{i,1989}^{OA} + \epsilon_i$$

Because changes in the misallocation of output or value added are likely to be poorly measured, we also control for the initial value of misallocation of output or value added.

Table 6b reports the results. There is considerable evidence for mean reversion in misallocation over time, but it is unfortunately not possible to tell how much of this is caused by true mean reversion versus measurement error. Panel A of table 6b also shows that initial factor

[8] Recall that the share of employment in production is more than four times as high as the share of land and buildings and employment is about as poorly misallocated as land and buildings. Hence, should the misallocation of land and buildings be the same as that of employment with the two being independent, we would expect a much larger coefficient for employment misallocation in the regressions of table 6a.

misallocation has only small and relatively fragile effects on future changes in value added misallocation in unweighted estimations. By contrast, panel B finds a much stronger connection when weighting districts by their initial size. This difference may indicate economic consequences from land misallocation that are particularly concentrated in larger initial centres of manufacturing activity, or the stronger results may reflect reduced measurement error when focusing on places with more plants in 1989 to calculate misallocation. These connections strengthen in our variations reported in Column 2-6.

To summarize, we find that factor misallocation explains output misallocation. Land and buildings appear to play a particularly important role in this respect. Our findings regarding the dynamic implication of factor misallocation are more mixed but suggestive that misallocation of land and employment today may worsen the misallocation of output or value added in the future.

5.2 Effects of misallocation on output per worker and productivity

We now turn to the effects of factor misallocation on output per worker. We think of factor allocation as a proximate cause that explains output per worker in Indian districts. In turn, factor allocation is determined by a number of 'deeper' determinants, similar to our earlier policy analyses. By analogy with the empirical growth literature, this exercise should be thought more as growth accounting (Caselli, 2005) or in the spirit of the early cross-country growth regressions looking at proximate factors of growth (Barro, 1991).

More specifically, we estimate regressions of the following form:

$$Y_{i,t} = a_0 + a_1 M_{i,t}^{L} + a_2 M_{i,t}^{T\&B} + a_3 M_{i,t}^{OA} + b_t + \epsilon_{i,t},$$

where the dependent variable is now output per worker in district *i* in year *t*. The explanatory variables are again various measures of factor misallocation and year fixed effects. In some robustness checks, we also introduce state or district fixed effects. However, we do not consider further determinants of output per worker. For instance, we suspect that a larger population or a better access to infrastructure boost output per worker in India, as has been identified elsewhere.

These determinants of output per worker are also likely to influence factor misallocation. This said, attempting to estimate the ‘pure’ effect of factor misallocation may not make much sense. Deeper determinants of output per worker will affect output per worker through a variety of channels, including factor misallocation. When assessing proximate factors of output per worker, it is not appropriate to include deeper determinants.

Table 7 reports our main results in the same format as the earlier work. For the combined sample in column 1, the coefficient on the misallocation of land and buildings is large in absolute value at -0.645. By contrast, there is no connection to the misallocation of employment, and the impact of the misallocation of other fixed assets is negative and statistically significant, but extremely small. Similar conclusions are reached in other specifications, including the introduction of district fixed effects in Column 3 that focus on panel variation in the data.

Land and buildings misallocation is associated with a reduced labour productivity in both sectors of manufacturing (organised sector coefficient is -0.428 (se=0.097), unorganised sector coefficient is -0.279 (se=0.109)), with employment misallocation being extremely important for the unorganised sector, far exceeding the role of land and buildings. We also examined the HK misallocation index. This measure relies on the variance of estimated productivity within each district and is not tied to a particular factor. Unreported analyses regress output per worker on this summary measure of misallocation and yield a coefficient of -0.757 for the organised sector, -1.535 for the unorganised sector, and -1.685 for the combined sample. These three coefficients are highly statistically significant, and the R-squared values are similar to those obtained with our core approach. On the other hand, we do not observe a systematic relationship between labour productivity growth and initial land and building misallocation.

The results of table 7 suggest fairly large effects associated with factor misallocation. For the full manufacturing sector, a standard deviation in the misallocation of land and buildings represents about a 25% reduction in output per worker. These effects are sizeable, and they grow modestly as other forms of misallocation are also considered. We also find very similar magnitudes when using the HK metric. Factor misallocation appears to affect the unorganised sector more, perhaps because it affects smaller firms more or because it forces more productive establishments into the unorganised sector. More important, the overall effect of factor misallocation on the two sectors is larger than its effects on either sub-sector in isolation due to its impact on their relative sizes and joint distribution.

We can compare the effects of a better factor allocation with the effects of an increase in the supply of factors. In the organised sector, recall that the coefficient on land and buildings in the production function is 0.13. Hence, a doubling of land and buildings made available for manufacturing activities would imply a 9% increase in output. While a doubling in manufacturing land and buildings seems like a large increase, recall that the supply of land in India is extremely constrained, and particularly so for manufacturing activity. By contrast, we estimate a one standard deviation reduction in land and building misallocation is associated with a 25% increase in output per worker. To match this effect, the supply of land and buildings would need to increase more than six-fold. While there are gains from expanding the supply of land, the gains from a better factor allocation are perhaps even larger.[9,10]

Although our approach is novel, we can also roughly compare our results with those of Hsieh and Klenow (2009). While this comparison is somewhat of a stretch, it is instructive. When Hsieh and Klenow (2009) use their model to compare output for actual Indian manufacturing during 1998-2005 period in a hypothetical Indian economy where factors would be perfectly efficiently allocated, they find a difference of about 100%. That is, they calculate that eliminating misallocation completely from Indian manufacturing would lead to a doubling of output per worker. If India was brought the same level of efficiency as the United States, they compute a gain of still about 50%. Relative to our findings, the calculations of Hsieh and Klenow (2009) appear to understate the potential gains from better factor allocation. Ballpark estimates for the organised sector in 2000 suggest that moving from a median level of misallocation to the top decile—arguably still well below perfect efficiency and most likely well below the U.S.

[9] As a second point of comparison, moving from the 10th to the 90th percentile of districts in terms of efficiency in the use of land and buildings is associated with more than a 50% increase in output per worker. For a district in India, a 50% increase in output per worker corresponds to moving from the 20th percentile of the distribution of output per worker to about the median.

[10] While a reduction in the misallocation of inputs and an increase in their supply should be conceptually distinguished, in practice, many policies appear to affect both. For instance, we show that ULCRA worsened the misallocation of land and buildings where it applied. It is also widely argued that ULCRA restricted the supply of land and buildings.

average—is associated with a 50% increase in output per worker. These very large numbers do not even consider the potential gains from reallocating factors from the less productive unorganised sector to the more productive organised sector nor do they consider the gains from reallocating factors from less productive districts to more productive districts.

6. Conclusions

Our exploration of the impact of land and building misallocation across Indian districts yields a rich set of findings, which we summarize in four basic groups.

- First, and very important from a methodological perspective, we validate the usefulness of misallocation indices for land and building inputs at the district level through evaluations of local policy changes like the unanticipated ULCRA repeal and stamp duty changes.
- Second, we provide novel statistics about the importance and severity of Indian factor misallocation. The misallocation of factor inputs is substantially worse than the misallocation of outputs. While more productive establishments produce more in India, they are often doing so in environments with factor allocations that are as good as random and sometimes even worse. Indeed, the variation in the misallocation of factor inputs across India districts is of the same magnitude as the variation in the misallocation of factor input measured by Bartelsman et al. (2009) for 18 countries of the world.
- Third, a higher misallocation of inputs is closely connected to a greater misallocation of output. These relationships are evident in cross-section levels for districts and panel adjustments over time. Land and buildings, which are arguably the most fixed of assets, appear to play a particularly important role in driving the misallocation of output. Initial land and building misallocation for India in 1989 can forward predict future increases in the misallocation of output.
- Finally, misallocation has important implications in terms of productive efficiency. A standard deviation in the misallocation of land and buildings is associated with a 25% difference in output per worker. These effects are seemingly even larger than those predicted by extant models. We estimate a one standard deviation improvement in land

and building misallocation would have an output boost equal to a six-fold increase in land supply.

While these results are provocative, more remains to be explored. Continuing first with our district-level focus, we see four key areas for future work.

- One is to further refine the connections of the misallocation of each input factor with the overall misallocation of output. This extension of the Olley-Pakes methodology will provide a sharper sense of relative input frictions and the areas upon which policy should focus. Likewise, we have provided in this paper accounts of changes within the organised and unorganised sector, relative to the whole. We are seeking to build a formal decomposition of the relative changes in these components and their joint positions.
- Second, this paper demonstrates the reasonable connection of land and building misallocation indices with the reforms impacting these inputs, and we hope to extend this to a similar consideration for labour inputs, which is of potentially overwhelming importance given that it accounts for a very large share of output. To assess the effects of policies on the misallocation of labour, we plan to use changes in labour regulation that took place in the late 1980s and early 1990s compiled by Besley and Burgess (2008) and the dismantling of the Raj Licence system during the early 1990s (Aghion et al., 2008) as our main sources of variation.
- A third vein investigates whether land misallocation is amplified and entrenched by causing misallocation in other inputs to firms. This paper finds that land misallocation is the greater among equals in terms of input misallocation, but it does not examine the propagation channels or aggregated effects that it creates. We seek to specifically investigate this feedback loop in terms of financial misallocation. The development of well-functioning banking and capital markets is an important step for economic growth. Most bank loans require some form of collateral to be provided to guarantee the loan. Land is simply the best form of collateral possible due to its immobility (the debtor can't run-off with land) and its high rate of re-usability in other contexts besides its current form. Thus, an important hypothesis to be investigated is that land misallocation breeds local misallocation of bank loans.
- A final, clear extension is to understand misallocation in other sectors like services.

In addition, it is also very important to connect these within-district findings to large differences in misallocation across districts and how they aggregate to national misallocation. Our analysis treats each district in isolation, which has several implications. One consequence is that we naturally miss the role of cross-district variation in factor allocations for explaining the aggregate misallocation of output that studies frequently consider. This may be of substantial importance for India given the historical constraints placed on business location choices and industrial policies that sought broad-based regional participation. We need to better understand these between- vs. within-district components to enable richer accounting exercises and to document more clearly the statistical and economic inference made with panel data from districts. We are seeking to extend the Olley-Pakes decomposition in this direction, and we will need to consider its properties relative to other widely used forms like Hsieh and Klenow (2009). Knowing the within/between breakdown of misallocation is important since the policy tools to target both types of misallocation are likely to differ. For instance, further integration of financial markets within India or a reduction in the cost of migration between areas will likely affect misallocation (of capital and labour) across places first and foremost. Other policies such as a relaxation of the strict zoning policies in some Indian cities such as Mumbai are, on the other hand, expected to improve the efficiency of factor allocation in those areas where these regulations are the most binding.

References

Adamopoulos, Tasso and Diego Restuccia. 2014. The size distribution of farms and international productivity differences. *American Economic Review*, forthcoming.

Aghion, Philippe, Robin Burgess, Stephen J. Redding, and Fabrizio Zilibotti. 2008. The unequal effects of liberalization: Evidence from dismantling the license Raj in India. *American Economic Review* 98(4):1397–1412.

Alm, James, Patricia Annez, and Arbind Modi. 2004. Stamp Duties in Indian States: A Case for Reform. World Bank Policy Research Working Paper #3413.

Asif, Mohammed. 1999. Land Acquisition Act: Need for an Alternative Paradigm. *Economic and Political Weekly* 24(25): 1564–66.

Baily, M., C. Hulten and D. Campbell. 1992. Productivity Dynamics in Manufacturing Plants. *Brookings Papers on Economic Activity: Microeconomics* 1992: 187-249.

Banerjee, Abhijit V. and Benjamin Moll. 2010. Why does misallocation persist? *American Economic Journal: Macroeconomics* 2(1):189–206.

Barro, Robert J. 1991. Economic growth in a cross section of countries. *Quarterly Journal of Economics* 106(2):407–443.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2009. Measuring and analysing cross-country differences in firm dynamics. In Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts (eds.) *Producer dynamics: New evidence from micro data*. Chicago: University of Chicago Press, 15–76.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review* 103(1):305–334.

Bertaud. Alain. 2011. Mumbai FAR/FSI conundrum. Available at: alainbertaud.com/AB-Mumbai-FSI-Conundrun-Revised_June-2013_k.pdf

Bertaud, Alain. 2002. India – Urban Land Reform. Available at: www.alain-bertaud.com/AB_%20India_%20Urban_Land_Reform.doc

Bertaud, Alain and Jan K. Brueckner. 2005. Analyzing building-height restrictions: Predicted impacts and welfare costs. *Regional Science and Urban Economics* 35(2):109–125.

Bertaud, Alain and Stephen Malpezzi. 2003. The spatial distribution of population in 48 world cities: implications for economies in transition. Draft manuscript, University of Wisconsin

Bertaud, Alain and Stephen Malpezzi. 2001. Measuring the Costs and Benefits of Urban Land Use Regulation: A Simple Model with an Application to Malaysia. *Journal of Housing Economics*, 10 (3): 393-418.

Besley, Timothy and Robin Burgess. 2000. Land reform, poverty reduction, and growth: Evidence from India. *Quarterly Journal of Economics* 115(2):389–430.

Besley, Timothy and Robin Burgess. 2004. Can labor regulation hinder economic performance? Evidence from India. *Quarterly Journal of Economics* 119(1):91–134.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. Does management matter? Evidence from India. *Quarterly Journal of Economics* 128(1): 1-51.

Business Standard. 2009. 'The Land Hurdle Again. 27 March 2009.

Caselli, Francesco. 2005. Accounting for cross-country income differences. In Philippe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth*, volume 1A. Amsterdam: North-Holland, 680–741.

Choudhary, Tarun. 2009. Use of Eminent Domain Process and Its Critique. Published in Mohanty, N. R. Sarkar and A. Pandey (eds). (2009). *India Infrastructure Report—Governance Issues for Commercialisation*, 3inetwork, Oxford University Press, New Delhi.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80(6):2543–2594.

Dachis, Ben, Gilles Duranton, and Matthew A. Turner. 2012. The effects of land transfer taxes on real estate markets: evidence from a natural experiment in Toronto. *Journal of Economic Geography* 12(2):327–354.

Duranton, Gilles and Diego Puga. 2014. Urban land use. Chapter forthcoming in: Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds), *Handbook of Regional and Urban Economics*, vol. 5, Amsterdam: North-Holland.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2006. Plant turnover and structural reforms in Colombia. *IMF Staff Papers* 53(0):58–75.

Fernandes, Walter. 1998. Land Acquisition (Amendment) Bill, 1998: Rights of Project-Affected People Ignored. *Economic and Political Weekly* 33(42–3): 2703–06.

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. Aggregate Productivity Growth: Lessons from Microeconomic Evidence, in Dean, E., M. Harper and C. Hulten (eds.) *New Developments in Productivity Analysis* (Chicago: University of Chicago Press): 303-372.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98(1):394–425.

Fox, Jeremy T. and Valérie Smeets. 2011. Does input quality drive measured differences in firm productivity? *International Economic Review* 52(4):961–989.

Ghani, Ejaz, Goswami, Arti Grover, and William R. Kerr. 2012b. Is India's Manufacturing Sector Moving Away from Cities? NBER working paper 17992.

Ghani, Ejaz, Goswami, Arti Grover, and William R. Kerr. 2014. Spatial dynamics and energy usage in India. Processed, Harvard Business School.

Government of India (GOI). 1985. *The Land Acquisition Act, 1894.* Government of India.

Griliches, Zvi and Jacques Mairesse. 1998. Production functions: The search for identification. In Steinar Strom (ed.) Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium. Cambridge (MA): Cambridge University Press, 169–203.

Griliches, Z. and H. Regev. 1995. Productivity and Firm Turnover in Israeli Industry: 1979-1988. *Journal of Econometrics*, 65: 175-203.

Hsieh, Chang-Tai and Peter J. Klenow. 2009. Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics* 124(4):1403–1448.

Hsieh, Chang-Tai and Peter J. Klenow. 2014. The Life Cycle of Plants in India and Mexico. *Quarterly Journal of Economics*, forthcoming.

Jain, M. K. 1993. Emerging trend of suburbanisation in India, 1971–81: an analysis of data on standard urban areas. Occasional Paper No. 2, Census of India, Registrar General of India, New Delhi.

Joshi, Vijay and Ian M. D. Little. 1991. *India's Economic Reforms: 1991–2001*. New York: Oxford University Press.

Kimura, Yoichiro. 2013. Property rights improvement and credit reallocation: Evidence from urban India. Processed, Boston University.

Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2):317–342.

Lucas, Robert E., Jr. 1978. On the size distribution of business firms. *Bell Journal of Economics* 9(2):508–523.

Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6):1695–1725.

Melitz, Marc J. and Saso Polanec. 2013. Dynamic Olley-Pakes productivity decomposition with entry and exit. Processed, Harvard University.

Midrigan, Virgiliu and Daniel Yi Xu. 2014. Finance and misallocation: Evidence from plant-level data. *American Economic Review*, forthcoming

Mohanty, Nirmal. 2009. Eminent Domain Powers: Rationale, Abuse, and Way Forward. Published in Mohanty, N. R. Sarkar and A. Pandey (eds), India Infrastructure Report—Governance Issues for Commercialisation, 3inetwork, Oxford University Press, New Delhi.

Morris, Sebastian and Ajay Pandey. 2009. Land Markets in India Distortions and Issues. Published in Mohanty, N. R. Sarkar and A. Pandey (eds). (2009). *India Infrastructure Report—Governance Issues for Commercialisation*, 3inetwork, Oxford University Press, New Delhi.

Olley, G. Steven and Ariel Pakes. 1996. The dynamics of productivity in the telecommunication equipment industry. *Econometrica* 64(6):1263–1297.

Restuccia, Diego and Richard Rogerson. 2013. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11(4):707–720.

Restuccia, Diego and Richard Rogerson. 2008. Misallocation and productivity. *Review of Economic Dynamics* 16(1):1–10.

Restuccia, Diego and Raul Santaeulalia-Llopis. 2014. Land misallocation and productivity. Processed, University of Toronto.

Siddiqi, Faizan Jawed. (2013). Governing Urban Land: Political Economy of the ULCRA in Mumbai. A thesis submitted to the department of Urban Studies and Planning. Massachusetts Institute of Technology.

Sivadasan, Jagadeesh. 2009. Barriers to competition and productivity: Evidence from India. *BE Journal of Economic Analysis & Policy* 9(1.42):1–64.

Sridhar, Kala. S. 2004. Cities with suburbs: evidence from India. Working Paper No. 23/2004, National Institute of Public Finance and Policy (http://www.nipfp.org.in/working paper/wp23.pdf).

Sridhar, Kala. S. 2007. Density gradients and their determinants: evidence from India. *Regional Science and Urban Economics* 37(3): 314–344.

Sridhar, Kala. S. 2010. Impact of Land Use Regulations: Evidence from India's Cities. *Urban Studies* 47(7): 1541-71.

Srinivas, Lakshmi. 1991. Land and Politics in India: Working of Urban Land Ceiling Act, 1976. *Economic and Political Weekly*, 26(5): 2482–4.

Syverson, Chad. 2011. What determines productivity? *Journal of Economic Literature* 49(2):326–365.

Thorner, Daniel. 1976. *The Agrarian Prospect in India.* Allied Publishers, New Delhi

Turner, Bengt and Stephen Malpezzi. 2003. A Review of Empirical Evidence on the Costs and Benefits of Rent Control. *Swedish Economic Policy Review* 10(0): 11-56.

Appendix box 1. Owners and renters harmonization and imputations

To impute the value of land and buildings used by establishments, we obtain the rate of return on land and buildings for all the years under consideration from establishment-level data for the unorganised sector by dividing rental values by the declared asset values. We repeat the same step for land and buildings separately for the years 1989 and 1994. Next, these rates of returns are winsorised by state and year and the median rates of return by state, district and year are computed. Examining the raw values of rates of return on land and buildings at the establishment level, we note that for the year 2005, the valuation of rented land and buildings often differs considerably the valuations in previous or subsequent years for the same district. Thus, for imputation purposes, we follow the following procedure. First, for all years, except 2005, we estimate:

$$ror_{it} = \alpha_i + \beta_t + \epsilon_{it}\,,$$

where ror_{it} is the median rate of return for district i in year t, α_i are the district fixed effects while β_t are the year fixed effects. Then, the rate of return for the year 2005 is computed as:

$$ror_{i2005} = \alpha_i + (\beta_{2000} + \beta_{2010})/2\,.$$

Since, these predictions are based on regressions, a number of negative or extremely high values for rates of return are obtained. In such cases, we replace ror_{i2005} as missing and address it in the same way as cells with missing values on rates of return for district-and-year combinations. These missing district-year cells are, in turn, replaced by the mean rates of return on land and building by state and year.

Once the rate of return on land and buildings is derived, we obtain the value of land and buildings hired by the establishments in the unorganised sector as:

$$T\&B_{jit} = ror_{it} \times rent_{jit}\,,$$

where $T\&B_{jit}$ is the value of land and building hired by establishment j in district i at time t and $rent_{jit}$ is the corresponding rent. ror_{it} is the rate of return on land and buildings in district i at time t. The value of land and buildings used in the production process is then simply the sum of the land and building hired and owned.

For the formal sector, we need the rates of return on land and buildings separately. However, for these returns, we have data from for the unorganised sector for the years 1989 and 1994 only. We obtain the median rates of return separately for land and buildings by state, district and year and then take the average rates of return over the two years. For the missing district-year cells, we use the mean rates of return by state.

Once this rate of return is obtained separately for land and buildings, we derive the separate values of land and buildings hired by the establishments in the formal sector as:

$$K_{jit} = ror_{Ki} \times rent_{Kjit},$$

where K_{jit} is the value of asset, $K \in \{land, buildings\}$ hired by establishment j in district i at time t and $rent_{Kjit}$ is the corresponding rent for asset K, ror_{Ki} is the rate of return on asset K in district i. The value of asset K used in the production function is then simply the sum of the asset K hired and owned. The value of land and buildings used in the production process is then the sum of above obtained land and buildings values.

Appendix box 2. Sivadasan's modified Levinsohn and Petrin methodology

We follow Appendix B of Sivadasan (2009) where he reports his methodology. We appropriately adjusted it for our panel analysis over the 1989-2009 period.

We assume that our value-added production function $v = f(l,k,\omega)$ is part of a more general production function separable in all intermediate inputs $Y = g(f(l,k,\omega),h(\Gamma,\omega))$, where Γ is a vector of intermediate inputs.

Let ι be one intermediate input, which Levinsohn and Petrin (2003) (LP) assume has a demand function of the form $\iota_{it} = \iota_t(\omega_{it},k_{it})$. Other possible state variables not explicitly included in the above input demand function include prices of inputs and output(s). We assume input and output prices are fixed across firms within the same industry, but allow for the common prices to change over time by indexing the input demand function by t. Assuming monotonicity, i.e., input choice is strictly increasing in productivity for all relevant capital levels, the input demand function can be inverted to obtain an implicit value for the unobserved productivity: $\omega_{it} = \omega_t(\iota_{it},k_{it})$.

Then, assuming the monotonicity condition holds, we can estimate the coefficients on the labor inputs by estimating the following regression:

$$v_{it} = \beta_l l_{it} + \phi_t(\iota_{it},k_{it}) + \eta_{it} \qquad \text{(Step 1)}$$

where

$$\phi_t(\iota_{it},k_{it}) = \beta_k k_{it} + \omega_t(\iota_{it},k_{it}).$$

We use quantity of electricity consumed ς_t as the input proxy ι_t. We specify $\omega_t(\iota_{it},k_{it})$ as a polynomial function in its arguments (including the absorbed intercept term and dropping the firm index i for expositional convenience) as follows:

$$\begin{aligned}
\omega_t(\varsigma_t,k_t) \quad = \quad & \alpha_{11}\varsigma_t + \alpha_{12}\varsigma_t^2 + \alpha_{13}\varsigma_t^3 + \alpha_{14}k_t + \alpha_{15}k_t\varsigma_t + \alpha_{16}k_t\varsigma_t^2 \\
& + \alpha_{17}k_t^2 + \alpha_{18}k_t^2\varsigma_t + \alpha_{19}k_t^3 \\
& + \alpha_{21}t_2\varsigma_t + \alpha_{22}t_2\varsigma_t^2 + \alpha_{23}t_2\varsigma_t^3 + \alpha_{24}t_2k_t + \alpha_{25}t_2k_t\varsigma_t + \alpha_{26}t_2k_t\varsigma_t^2 \\
& + \alpha_{27}t_2k_t^2 + \alpha_{28}t_2k_t^2\varsigma_t + \alpha_{29}t_2k_t^3 \\
& + \alpha_{31}t_3\varsigma_t + \alpha_{32}t_3\varsigma_t^2 + \alpha_{33}t_3\varsigma_t^3 + \alpha_{34}t_3k_t + \alpha_{35}t_3k_t\varsigma_t + \alpha_{12}\varsigma_t^2 + \alpha_{36}t_3k_t\varsigma_t^2 \\
& + \alpha_{37}t_3k_t^2 + \alpha_{38}t_3k_t^2\varsigma_t + \alpha_{39}t_3k_t^3
\end{aligned}$$

where $t_2 = 1$ for years 2000 or 2005, and $t3 = 1$ for years 2009.

Identifying the coefficient on the capital variable requires additional assumptions and a second stage estimation procedure. The moment condition that LP propose uses panel information to identify the capital coefficient. LP assume that:

$$E[k_{i,t}.\{\omega_{i,t} - E[\omega_{i,t}|\omega_{i,t-1}]\}] = 0. \tag{1}$$

This follows from a behavioral assumption that capital does not respond to "surprises" in productivity, or equivalently from assuming that $\{\omega_{i,t}\}_{i=1}^{\infty}$ follows a stochastic first order Markov process.

The LP methodology could be adapted to a repeated cross-section context by making the broader assumption that $\omega_{i,t}$is uncorrelated with the choice of capital $k_{i,t}$, (which is arguably fixed in the short run). This moment condition is discussed by Griliches and Mairesse (1998), but they suggest this assumption may be too restrictive, as capital is likely to respond to any persistent component of ω_{it}. Instead we propose a less-restrictive moment condition, which can be used in the repeated cross-section context. Instead of using last period's productivity for each firm (unobservable in our data), we use the average productivity in the previous period for a closely matched industry-location-size cell (observable in our data) as the predictor for this period firm productivity. This attempts to approximate the moment condition in equation 1 as closely as possible, given the limitations of our data.

To implement this approach, we sub-divide the data into industry-location-size cells and estimate the average productivity for each cell in every period. Then our modified moment condition replacing equation 1 is given by:

$$k_{i,t}.\{\omega_{i,t} - E[\omega_{i,t}|\bar{\omega}_{i,t-1}]\} = 0 \tag{2}$$

where

$$\bar{\omega}_{i,t-1} = \frac{1}{m_{j_i}} \sum_{s=1}^{m_{j_i}} \omega_{s,t-1} \tag{3}$$

where j_i indexes the industry-location-size cell to which firm i belongs, and m_{j_i} is the number of observations in cell j.

As in the LP methodology, we then identify the coefficient on the capital variable (β_k) by considering the second step regression:

$$v^*_{i,t} = (\beta_k k_{i,t} + E[\omega_{i,t}|\bar{\omega}_{i,t-1}] + \eta^*_{i,t} \tag{Step 2}$$

where $v^*_{i,t} = v_{i,t} - (\beta_l l_{i,t} + \beta_n n_{i,t})$ and $\eta^*_{i,t} = \{\omega_{i,t} - E[\omega_{i,t}|\bar{\omega}_{i,t-1}]\} + \eta_{i,t}$.

The specific estimation algorithm to obtain the capital coefficient is as follows:

i. Start with a candidate estimate of the capital coefficient β_{k*}.

ii. From the results of the first stage regression, obtain:

$$\hat{\phi}_t = \hat{v}_t - \hat{\beta}_l l_t.$$

iii. Then obtain:

$$\widehat{\omega_t} = \hat{\phi}_t - \hat{\beta}_{k*} k_t.$$

iv. Estimate the mean productivity for each industry-size-location cell using:

$$\widehat{\bar{\omega}_{t-1}} = \frac{1}{m_j} \sum_{s=1}^{m_j} \widehat{\omega_{s,t-1}}$$

where m_j is the number of observations in cell j.

v. Regress $\hat{\omega}_t$ on $\widehat{\bar{\omega}_{t-1}}$ and $\widehat{\bar{\omega}^2_{t-1}}$ and use the predicted values to form $E[\widehat{\omega_t|\bar{\omega}_{t-1}}]$.

vi. Obtain $\hat{v}^*_t = v_t - \hat{\beta}_l l_t$.

vii. Form $\hat{\eta}^*_t = \hat{v}^*_t - \beta_{k*} k_t - E[\widehat{\omega_t|\bar{\omega}}_{t-1}]$.

viii. Estimate β_k by minimizing the sum (over all the firm-year observations) of the squared residuals in Step 2:

$$\min_{\beta_{k*}} \left\{ \sum_i \left(\hat{v}^*_{it} - \beta_{k*} k_{it} - E[\omega_{it}|\bar{\omega}_{it-1}]\right)^2 \right\}$$

As discussed in Levinsohn and Petrin (2003), a bootstrapping procedure is used to estimate the standard errors.

Appendix A. Land issues in India

Efficient use of land is critically dependent on necessary institutions, such as a transparent system to convert land use, a clear definition of property rights, a robust system of land and property valuation, and a strong judicial system for addressing public concerns to facilitate land markets, land transactions, and land use changes. India, however, faces serious shortcomings on all fronts.

Land use and the intensity of land development are severely restricted. Everywhere, there are severe restrictions on the transition of land from agricultural to non-agricultural use. This limits urban expansion. Zoning also imposes constraints within cities where obsolete land uses are often maintained. For instance, old cotton mills still occupy large areas of land in central Mumbai and Ahmedabad (Bertaud, 2002).

There are also extremely restrictive regulations in urban areas imposing ceilings on building heights and maximum floor space index (FSI) levels. FSI in India is seldom above 1.6 which is very low when compared to values ranging from 5 to 15 in other cities of Asia. Low FSIs increase the demand for land as more land is required to build a given area of floor space. In turn, this leads to higher property prices. Low FSIs also have a negative impact on the spatial structure of cities. By unreasonably reducing the amount of floor space that can be built in centrally located areas, and by making land recycling difficult, some regulations tend to “push” urban development toward the periphery. As a result, commuting trips become longer, public transport becomes difficult to operate, and urban infrastructure has to be extended further than what would have been the case if land supply had been unconstrained (Bertaud and Brueckner, 2005). Low FSIs may also affect the spatial distribution of jobs due to their impact on residential suburbanisation (Sridhar, 2010). Furthermore, blanket FSIs are implemented in Indian cities, covering large areas. This obstructs the ability to use land use regulations as a policy instrument to strategically increase densities around infrastructure networks.

The process of evaluation for land that is publicly purchased offers wide opportunities for corruption. Valuation is done by the public agent acquiring the property without much specific guideline except that the value should be determined by current use. While landowners cannot develop agricultural land, public land can easily be converted to residential or commercial use. This provides strong incentives to entrepreneurs to lobby public official and have them use eminent domain powers to purchase agricultural land and convert it. For instance, the land acquired for Yamuna Expressway connecting Noida and Agra in Uttar Pradesh was purchased by the state government at Rs 50 per square meter. However, a decade later, parcels in the same location is sold at Rs 15,000 per square meter.

Rental properties in India are also subject to rent control. The negative impacts effects of rent controls are well documented both in India as well as internationally. It is believed that the longer a property is under rent control the higher is the difference between controlled rent and market rent. This motivates the tenants to never leave the premises. To make matters worse, there is also a provision to transmit the rented property to their relatives with the right to occupy their apartment after their death. De facto, a large part of the property right is being progressively transferred from the landlord to the tenant, except for the right to sell which remains with the landlord. Therefore, many rent controlled buildings are old and badly maintained. However, no redevelopment can occur until the tenants move voluntarily out the building. Since owners cannot vacate their premises, rent control creates the perverse incentive for landlords to see their property deteriorate or even collapse. Rent control, thus, contributes to a decrease in land supply because buildings which are under rent control cannot be redeveloped or even renovated. It also creates additional frictions that prevent land from being used efficiently.

The combined effect of multiple layers of poorly conceived central, state, and municipal regulations contribute to an artificial urban land shortage in India (Bertaud and Malpezzi, 2001). As a result urban land prices are abnormally high in relation to India’s household income, and households and businesses consume less floor space than they could afford if the regulatory environment were reformed.

Recognizing the negative impact and social tensions created by the rent control laws, the Government of India came out with a Model Rent Legislation (MRL) in 1992. Following the formulation of the model rent legislation, many states have repealed their old acts and formulated new acts. These states include Karnataka (new act in 1999), Rajasthan (2001), Maharashtra (1999) and West Bengal (1997). Several other states are in the process of reforming their acts. The rent control reforms are already witnessing some positive results. For instance, in Maharashtra an increase in rent of space held by the General Insurance Corporation from Rs. 51,000 per month to Rs. 680,000 per month led the corporation to look for rental space in other areas to store its old records and release the prime property for more productive uses (Business Standard, 2001).

In sum, over the years, India has used a number of restrictive regulations to curb the development of urban land. These regulations were pursued with a dual objective: equity in distribution of land (e.g. ULCRA, rent control act) and curtailing urban congestion and densification (e.g. by progressively reducing FSI). Despite their noble intentions, these regulations achieved neither their egalitarian nor low-density objectives. Indian cities have the largest and densest slums in the world. Furthermore, these regulations account for much of the land price increase and extremely low consumption of floor space and land. By contrast, Town Planning Schemes (TPS)—a form of land readjustment system similar in its principle to the ones that had been used in Korea, Taiwan, and Germany—have achieved much success in cities such as Ahmedabad. The average area developed each year in Ahmedabad is roughly equivalent to about 3.2% of the current built-up area of the Municipality of Ahmedabad. Besides the gradual reforms to ULCRA and rent control, two other initiatives are being tried that might open up land markets in a few states to overcome the existing constraints. These are: (i) township development and (ii) land pooling and readjustment.

Table 1: Land and building inputs in manufacturing plants

	1989	1994	2000	2005	2010
A. Raw count of data observations after basic pruning					
Organised sector, urban areas	22,899	29,024	15,296	20,379	19,323
Organised sector, rural areas	10,457	13,625	10,275	14,722	13,778
Unorganised sector, urban areas	40,584	53,262	131,652	40,062	49,818
Unorganised sector, rural areas	53,846	53,051	131,125	39,562	49,719
Total	127,786	148,962	288,348	114,725	132,638
B. Estimated number of plants in India using sample weights on raw counts in panel A					
Organised sector, urban areas	54,675	64,921	62,835	63,044	72,750
Organised sector, rural areas	20,767	28,114	36,266	41,537	45,400
Unorganised sector, urban areas	2,823,783	2,967,126	5,048,512	4,901,061	7,396,214
Unorganised sector, rural areas	10,451,079	9,059,888	11,830,554	11,932,791	10,283,594
Total	13,350,303	12,120,050	16,978,167	16,938,433	17,797,958
C. Raw count of data observations from panel A that report land and building values					
Organised sector, urban areas	17,790	22,724	12,605	16,597	16,255
Organised sector, rural areas	9,358	12,211	9,474	13,263	12,498
Unorganised sector, urban areas	23,515	34,763	79,605	23,834	27,775
Unorganised sector, rural areas	46,406	91,102	72,247	33,446	40,261
Total	97,069	160,800	173,931	87,140	96,789
D. Raw average land and building values for reporting plants (mean weighted by multiplier)					
Organised sector, urban areas	193,290	259,760	400,218	461,964	754,869
Organised sector, rural areas	303,117	404,557	635,906	698,194	1,183,264
Unorganised sector, urban areas	3,540	5,498	6,998	8,535	15,457
Unorganised sector, rural areas	978	1,011	1,390	1,725	3,607
E. Post imputation and weighting, the total land and building values for plants (in millions)					
Organised sector, urban areas	14,204	21,820	36,188	39,426	77,397
Organised sector, rural areas	7,311	12,915	28,063	33,632	62,956
Unorganised sector, urban areas	10,889	21,213	38,638	33,134	136,622
Unorganised sector, rural areas	9,795	9,585	18,018	19,212	44,452
Total	42,200	65,533	120,908	125,404	321,427
F. Aggregate land and building usage/output levels					
Organised sector, urban areas	0.073	0.089	0.129	0.106	0.135
Organised sector, rural areas	0.080	0.095	0.129	0.102	0.129
Unorganised sector, urban areas	0.533	0.463	0.399	0.306	0.881
Unorganised sector, rural areas	0.414	0.345	0.359	0.341	0.552
Total	0.128	0.144	0.188	0.145	0.248

Notes: Descriptive statistics taken from Annual Survey of Industries (ASI) and National Sample Statistics (NSSO).

Table 2: Estimated coefficients with LP-Sivadasan methodology for labour and capital

		Full sector		Organised sector		Unorganised sector	
	Industry Description	Labour	Capital	Labour	Capital	Labour	Capital
15	Food products and beverages	0.85	0.31	0.60	0.46	0.90	0.25
16	Tobacco	0.70	0.04	0.64	0.44	0.69	0.02
17	Textiles	0.85	0.34	0.62	0.39	0.82	0.26
18	Wearing apparel; dressing and dyeing of fur	0.80	0.47	0.24	0.55	0.74	0.44
19	Tanning and dressing of leather	0.77	0.24	0.58	0.43	0.72	0.21
20	Wood and some products of wood	0.75	0.27	0.70	0.34	0.75	0.28
21	Paper and paper products	0.73	0.31	0.69	0.41	0.76	0.25
22	Publishing, printing and reproduction of recorded media	0.68	0.50	0.57	0.44	0.67	0.42
23	Coke, refined petroleum products and nuclear fuel	0.71	0.35	0.69	0.39	0.77	0.35
24	Chemicals and chemical products	0.79	0.31	0.66	0.43	0.73	0.32
25	Rubber and plastic products	0.83	0.46	0.73	0.41	0.81	0.41
26	Other non-metallic mineral products	0.72	0.34	0.78	0.43	0.70	0.31
27	Basic metals	0.76	0.30	0.67	0.38	0.75	0.41
28	Fabricated metal products, except machinery and equipments	0.87	0.39	0.69	0.39	0.87	0.36
29	Machinery and equipment n.e.c.	0.95	0.26	0.77	0.37	0.92	0.26
30	Office, accounting and computing machinery	0.41	0.38	0.77	0.41	0.23	0.21
31	Electrical machinery and apparatus n.e.c.	0.78	0.47	0.72	0.42	0.75	0.32
32	Radio, television and communication equipment	0.61	0.53	0.81	0.34	0.59	0.42
33	Medical, precision and optical instruments	0.77	0.44	0.62	0.42	0.81	0.37
34	Motor vehicles, trailers and semi-trailers	0.96	0.27	0.62	0.41	0.87	0.34
35	Other transport equipment	0.83	0.24	0.66	0.38	0.84	0.30
36	Furniture; manufacturing n.e.c.	0.72	0.30	0.75	0.34	0.71	0.28
	Unweighted industry average	0.77	0.34	0.66	0.41	0.74	0.31

Notes: Table reports the industry-level coefficient values for labour and capital using the Sivadasan (2009) version of Levinsohn-Petrin (2003) productivity calculations for repeated cross-sections.

Table 3: Misallocation indices across districts

	Output	Value-added	Labour	Buildings	Land	Other K
A. Full Indian manufacturing sector						
District mean 1989	-0.90	-0.97	-0.44	-0.52		-0.60
District mean 1994	-0.60	-0.75	-0.14	-0.28		-0.26
District mean 2000	-0.82	-0.91	-0.28	-0.38		-0.54
District mean 2005	-0.68	-0.79	-0.30	-0.30		-0.46
District mean 2010	-0.57	-0.70	-0.20	-0.19		-0.31
Average over surveys	-0.71	-0.82	-0.27	-0.33		-0.43
District SD 1989	0.56	0.60	0.44	0.52		0.63
District SD 1994	0.48	0.54	0.27	0.41		1.40
District SD 2000	0.52	0.58	0.35	0.47		0.59
District SD 2005	0.44	0.48	0.42	0.52		0.52
District SD 2010	0.47	0.46	0.43	0.43		0.61
Average over surveys	0.50	0.53	0.38	0.47		0.75
B. Organised sector						
District mean 1989	-0.40	-0.50	-0.10	-0.15	-0.05	-0.16
District mean 1994	-0.34	-0.47	-0.08	-0.12	-0.01	-0.07
District mean 2000	-0.33	-0.49	-0.08	-0.09	-0.02	-0.11
District mean 2005	-0.32	-0.46	-0.09	-0.13	-0.08	-0.11
District mean 2010	-0.24	-0.40	-0.02	-0.05	0.02	-0.03
Average over surveys	-0.32	-0.46	-0.07	-0.11	-0.03	-0.10
District SD 1989	0.38	0.44	0.29	0.38	0.37	0.55
District SD 1994	0.36	0.43	0.25	0.35	0.43	0.46
District SD 2000	0.39	0.42	0.27	0.39	0.44	0.45
District SD 2005	0.34	0.43	0.23	0.35	0.39	0.38
District SD 2010	0.37	0.43	0.28	0.38	0.41	0.46
Average over surveys	0.37	0.43	0.26	0.37	0.41	0.46
C. Unorganised sector						
District mean 1989	-0.60	-0.60	-0.01	-0.05		-0.02
District mean 1994	-0.53	-0.58	0.01	-0.06		-0.28
District mean 2000	-0.65	-0.60	-0.10	-0.13		-0.19
District mean 2005	-0.76	-0.71	-0.15	-0.13		-0.30
District mean 2010	-0.51	-0.49	-0.05	-0.08		-0.16
Average over surveys	-0.61	-0.60	-0.06	-0.09		-0.19
District SD 1989	0.30	0.27	0.09	0.18		0.26
District SD 1994	0.33	0.30	0.11	0.20		1.16
District SD 2000	0.34	0.30	0.12	0.18		0.32
District SD 2005	0.41	0.36	0.16	0.28		0.37
District SD 2010	0.26	0.25	0.11	0.16		0.32
Average over surveys	0.33	0.29	0.12	0.20		0.49

Notes: Table presents average values and standard deviations calculated across Indian districts. Plant-level survey weights are used to create district-level values. Values presented in this tables are unweighted statistics for district-level values. A more negative value indicates less misallocation.

Table 4: Changes in misallocation following the repeal of ULCRA, 2000-2010

	Baseline estimation with extended controls	Using basic set of control variables only	Without industry aggregation step	Using OLS TFP metrics	Weighting by log initial district employment	Using Hsieh-Klenow metric
	(1)	(2)	(3)	(4)	(5)	(6)
A. Change in misallocation for land and buildings						
ULCRA repeal	-0.057+	-0.059+	-0.171+++	-0.194++	-0.057+	n.a.
	(0.028)	(0.030)	(0.056)	(0.074)	(0.029)	
Adjusted R-squared	0.378	0.382	0.346	0.394	0.380	
B. Change in misallocation for value added						
ULCRA repeal	-0.136++	-0.127+	-0.293+++	-0.284++	-0.132+	-0.066+++
	(0.059)	(0.059)	(0.093)	(0.120)	(0.062)	(0.018)
Adjusted R-squared	0.477	0.481	0.477	0.418	0.471	0.433

Notes: Estimations quantify the change in misallocation levels surrounding the repeal of the Urban Land (Ceiling and Regulation) Act. Estimations are cross-sectional regressions that include 252 Indian districts. The outcome variable in Panel A is the change in land and building misallocation from 2000 to 2010. The outcome variable in Panel B considers misallocation of value added. The primary explanatory variable is a (0,1) indicator variable for a state that repeals ULCRA by 2003. Estimations control for the initial value of the studied misallocation. Basic controls further include 12 initial traits of districts: log population density, log population, log share of urban population, log built-up area, the log share of built-up area, percent graduates, an infrastructure composite index, log minimum travel time to the 10 largest cities, a measure of local age profiles, the share of population with access to banking, the male-female sex ratio, and the share of district population in scheduled casts and tribes. The infrastructure composite index considers the population share with telecom access, the share with power access, the share of villages with paved roads, and the percent share with safe water. Extended controls further include log distance to national highway, log distance to state highway, log distance to railroads, and log distance to the closest metropolitan area. Regressions are unweighted and report standard errors clustered by state.

Table 5: Changes in misallocation associated with stamp duty changes, 1989-2010

	Baseline estimation with extended controls	Using basic set of control variables only	Without industry aggregation step	Using OLS TFP metrics	Weighting by log initial district employment	Using Hsieh-Klenow metric
	(1)	(2)	(3)	(4)	(5)	(6)
A. Misallocation for land and buildings						
Changes in stamp duty 1989-2003	0.019++	0.019++	0.019++	0.014	0.020++	n.a.
	(0.007)	(0.007)	(0.009)	(0.010)	(0.007)	
Observations	306	306	306	306	306	
Adjusted R-squared	0.453	0.456	0.348	0.367	0.449	
B. Misallocation for value added						
Changes in stamp duty 1989-2003	0.017++	0.018+++	0.021+	0.016	0.017+++	-0.001
	(0.006)	(0.006)	(0.010)	(0.011)	(0.006)	(0.002)
Observations	306	306	306	306	306	302
Adjusted R-squared	0.525	0.528	0.478	0.470	0.520	0.497

Notes: Estimations quantify the change in misallocation levels surrounding adjustments in state-level stamp duties. The outcome variable in Panel A is land and building misallocation from 1989 to 2010. The outcome variable in Panel B considers misallocation of value added. The primary explanatory variable is the change in state-level stamp duty imposed on land transactions from 1989 to 2003. Basic and extended controls are the same as defined in Table 4. Regressions are unweighted and report standard errors clustered by state.

Table 6a: Value added misallocation as a function of factor misallocation

	Baseline estimation	Considering aggregate fixed assets	Including district fixed effects	Without industry aggregation step	Using OLS TFP metrics	Considering output misallocation as DV
	(1)	(2)	(3)	(4)	(5)	(6)
A. Unweighted estimations						
Land and building	0.623+++		0.582+++	0.472+++	0.365+++	0.641+++
misallocation	(0.041)		(0.043)	(0.039)	(0.059)	(0.040)
Other fixed assets	0.000+++		0.000+++	0.007+++	0.162+++	0.000+++
misallocation	(0.000)		(0.000)	(0.001)	(0.043)	(0.000)
Employment	0.398+++	0.518+++	0.361+++	0.451+++	0.428+++	0.264+++
misallocation	(0.053)	(0.043)	(0.062)	(0.050)	(0.047)	(0.053)
Total fixed assets		0.493+++				
misallocation		(0.031)				
Observations	1816	1816	1816	1816	1824	1816
Adjusted R-squared	0.626	0.638	0.671	0.663	0.652	0.616
B. Weighting districts by initial employment levels						
Land and building	0.622+++		0.544+++	0.591+++	0.618+++	0.587+++
misallocation	(0.041)		(0.110)	(0.058)	(0.073)	(0.064)
Other fixed assets	0.000++		0.000++	0.006+++	-0.025	0.001++
misallocation	(0.000)		(0.000)	(0.001)	(0.062)	(0.000)
Employment	0.323+++	0.512+++	0.232+++	0.335+++	0.418+++	0.178+++
misallocation	(0.047)	(0.047)	(0.072)	(0.067)	(0.049)	(0.052)
Total fixed assets		0.425+++				
misallocation		(0.045)				
Observations	1804	1804	1804	1804	1812	1804
Adjusted R-squared	0.638	0.614	0.678	0.690	0.710	0.612

Notes: Estimations quantify the relationship between value added misallocation levels and that of factor inputs. Observations are district-year values. Regressions include year fixed effects, are unweighted, and report standard errors clustered by district.

Table 6b: Change in value added misallocation as a function of initial factor misallocation

	Baseline estimation	Considering aggregate fixed assets	Without industry aggregation step	Using OLS TFP metrics	Considering output misallocation as DV	Excluding district covariates
	(1)	(2)	(3)	(4)	(5)	(6)
A. Unweighted estimations						
Value added	-0.956+++	-0.967+++	-1.070+++	-0.978+++	-1.024+++	-0.994+++
misallocation, 1989	(0.076)	(0.069)	(0.095)	(0.091)	(0.089)	(0.080)
Land and building	0.176		0.138	0.139	0.251+	0.217
misallocation, 1989	(0.129)		(0.174)	(0.180)	(0.134)	(0.134)
Other fixed assets	0.121		0.015	0.002	0.110+	0.107
misallocation, 1989	(0.077)		(0.120)	(0.124)	(0.066)	(0.079)
Employment	0.018	0.045	0.182	0.107	-0.037	0.016
misallocation, 1989	(0.087)	(0.071)	(0.118)	(0.117)	(0.086)	(0.088)
Total fixed assets		0.269+++				
misallocation, 1989		(0.057)				
Observations	309	309	309	309	309	309
Adjusted R-squared	0.525	0.536	0.479	0.470	0.474	0.520
B. Weighting districts by initial employment levels						
Value added	-0.942+++	-0.903+++	-1.254+++	-1.266+++	-1.009+++	-0.962+++
misallocation, 1989	(0.133)	(0.132)	(0.107)	(0.131)	(0.127)	(0.130)
Land and building	0.441++		0.669+++	0.671+++	0.405++	0.594+
misallocation, 1989	(0.224)		(0.213)	(0.257)	(0.199)	(0.302)
Other fixed assets	-0.016		-0.145	-0.161	0.005	-0.101
misallocation, 1989	(0.118)		(0.123)	(0.141)	(0.091)	(0.182)
Employment	-0.035	0.106	0.079	0.176	-0.049	-0.193
misallocation, 1989	(0.121)	(0.088)	(0.127)	(0.135)	(0.114)	(0.156)
Total fixed assets		0.273+++				
misallocation, 1989		(0.069)				
Observations	309	309	309	309	309	309
Adjusted R-squared	0.572	0.575	0.611	0.628	0.534	0.492

Notes: Estimations quantify the relationship between the change in value added misallocation levels from 1989 to 2010 and initial factor misallocation in 1989. Observations are district-level values. Additional controls include the extended set of district-level traits in Table 4. Regressions report standard errors clustered by district.

Table 7: Labor productivity as a function of factor misallocation

	Baseline estimation	Considering aggregate fixed assets	Including district fixed effects	Without industry aggregation step	Using OLS TFP metrics
	(1)	(2)	(3)	(4)	(5)
A. Unweighted estimations					
Land and building	-0.645+++		-0.490+++	-0.547+++	-0.170++
misallocation	(0.088)		(0.080)	(0.076)	(0.086)
Other fixed assets	-0.001+++		-0.001+++	0.002++	-0.257+++
misallocation	(0.000)		(0.000)	(0.001)	(0.069)
Employment	-0.103	-0.165	-0.091	-0.042	0.194++
misallocation	(0.147)	(0.136)	(0.099)	(0.122)	(0.082)
Total fixed assets		-0.584+++			
misallocation		(0.064)			
Observations	1816	1816	1816	1816	1824
Adjusted R-squared	0.154	0.169	0.577	0.201	0.108
B. Weighting districts by initial employment levels					
Land and building	-0.396++		-0.486+++	-0.346++	-0.012
misallocation	(0.163)		(0.146)	(0.175)	(0.130)
Other fixed assets	-0.001++		-0.001+++	-0.001	-0.250+++
misallocation	(0.000)		(0.000)	(0.004)	(0.083)
Employment	0.299	0.239	0.025	0.198	0.226
misallocation	(0.265)	(0.226)	(0.152)	(0.255)	(0.183)
Total fixed assets		-0.353+++			
misallocation		(0.096)			
Observations	1804	1804	1804	1804	1812
Adjusted R-squared	0.118	0.127	0.575	0.134	0.122

Notes: See Table 6a.

Appendix Table 1a: ASI data fields

Items	Years
Section: Fixed Assets	
Gross value	
- Opening value	1989, 1994, 2000, 2005, 2009
- Additions during the year	1989, 2000, 2005, 2009
- Revaluations	1989, 2000, 2005, 2009
- Actual additions	1989, 2000, 2005, 2009
- Deductions and adjustments	1989, 1994, 2000, 2005, 2009
- Closing value	1989, 1994, 2000, 2005, 2009
Depreciation:	
- Up to year beginning	1989, 2000, 2005, 2009
- Provided during the year	1989, 2000, 2005, 2009
- Total depreciation	1989, 2000, 2005, 2009
Net value	
- Opening value	1989, 2000, 2005, 2009
- Closing value	1989, 2000, 2005, 2009
Section: Other Expenses	
Rent paid	1989, 1994, 2000, 2005, 2009
Section: Other Incomes	
Rent received	2000, 2005, 2009

Note: Assets are separated into land and building components. Values are in Rupees. The rent figures on land includes royalties on mines, quarries and similar assets.

Appendix Table 1b: NSSO data fields

Items	Land	Buildings	Land and Buildings
Section: Fixed assets owned and hired			
Market value of assets (Rs.) as on last date of reference period			
- Owned	1989, 1994	1989, 1994	2000, 2005, 2010
- Hired	1989, 1994	1989, 1994	2000, 2005, 2010
Net opening balance	1989, 1994	1989, 1994	
Net additions to *owned* assets during reference year			
- new	1989, 1994	1989, 1994	
- used	1989, 1994	1989, 1994	
- own construction	1989, 1994	1989, 1994	
- total	1989, 1994	1989, 1994	2000, 2005, 2010
Depletion of assets during reference year			
- sold	1989, 1994	1989, 1994	
- discarded	1989, 1994	1989, 1994	
Depreciation	1989, 1994	1989, 1994	
Net closing balance	1989, 1994	1989, 1994	
Rent payable on *hired* assets during reference period	1989, 1994	1989, 1994	2000, 2005, 2010

Note: Values are in Rupees. Year entry in the land, building and land and buildings columns indicate the survey years for which this information was collected.

Appendix Table 2: District count for descriptive sample

Year	District count	Where at least one plant reports land and building values
ASI sample:		
1989	416	414
1994	429	424
2000	359	358
2005	395	395
2010	401	398
NSSO sample:		
1989	396	396
1994	414	413
2000	419	419
2005	420	420
2009	428	428