Factor misallocation in India

Gilles Duranton[*‡]
University of Pennsylvania

Ejaz Ghani[*¶]
World Bank

Arti Grover Goswami[*†]
World Bank

William Kerr[*§]
Harvard Business School

Work in progress, March 2016

Abstract: We propose a new approach to quantify the misallocation of manufacturing output and factors of production between establishments across Indian districts during the 1989-2010 period. We first examine the theoretical underpinnings of two classes of misallocation measures. Then, we distil a number of stylised facts about misallocation in India. We also demonstrate the validity of our misallocation metrics by connecting them to each other and to policy changes. Finally, we quantify the implications misallocation on output per worker. 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 and buildings plays a particularly important role in these challenges.

Key words: misallocation, India, productivity
JEL classification: D24, E24, O47

[*]Financial support from the World Bank, CEPR Private Enterprise Development in Low-Income Countries programme, and the Zell Lurie Center for Real Estate at the Wharton School is gratefully acknowledged. We appreciate an outstanding NBER discussion by Adam Guren and the comments and guidance from Ed Glaeser, Sandile Hlatshwayo, Colin Hottman, 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 UCLRA and sharing his data with us. Yeayeun Park and Yu Wang provided us with excellent research assistance. The views expressed here are those of the authors and not of any institution they may be associated with.

[‡]Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA. E-mail: duranton@wharton.upenn.edu; website: https://real-estate.wharton.upenn.edu/profile/21470/.

[¶]e-mail: eghani@worldbank.org.

[†]E-mail: agrover1@worldbank.org.

[§]E-mail: wkerr@hbs.edu; website: http://www.hbs.edu/faculty/Pages/profile.aspx?facId=337265.

1. Introduction

We study the misallocation of land and buildings inputs and other factors of production for the Indian manufacturing sector. Our paper makes five contributions. We first examine the theoretical underpinning of variance- and covariance-based measures of misallocation. Second, we 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. Third, we validate our misallocation 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 fourth contribution is to demonstrate how land and building misallocation fosters the misallocation of output in local areas. Finally, also show a strong connection of covariance-based measures of 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.

To conduct our study, we need 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, in a first step we contrast two classes of models of misallocation. Inspired by the seminal contributions of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), models of distortions with adjustment imply that misallocation can be measured by the variance of a simple productivity residual. Closer to the recent contribution of (Asker, Collard-Wexler, and De Loecker, 2014), we show that in models of shocks with lack of adjustment misallocation can be measured by the covariance between factor or

output shares and the same productivity residual as in the decomposition originally proposed by Olley and Pakes (1996). Hence, misallocation can be measured using either a variance of productivity residuals or one of the covariances of these residuals with factor or output shares. We also show how output misallocation relates to factor misallocation. Finally, our modelling is also useful to show how misallocation indices can be computed for districts of a country using productivity residuals that are estimated for industries that differ in factor intensity.

We note that our use of multiple measures of misallocation and computing them for subnational units represent two significant departures relative to extant literature Hopenhayn (2014*a*). 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 various measures of misallocation and output per worker, using subnational regions in cross-section and in panel.

Before doing this, we distill some stylised facts about misallocation. Interestingly, there is a lot of variation in misallocation across the different misallocation indices and for each index across districts. We also perform two important validation checks on our misallocation measures. The first one shows a connection of local policy reforms that affect property markets to the misallocation of land and output. This assessment is of independent interest as it allows us to measure how policies can affect misallocation. The second validation shows that, empirically, output misallocation relates to factor misallocation just as suggested by models of shocks without adjustment. This exercise is also of independent interest because it 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.

Finally, our 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. On the other hand, we find weak evidence regarding variance-based measures of misallocation. We acknowledge that the strong correlations we find between output and factor misallocation could be spurious. We use three strategies to strengthen these findings. The first is to show that our results are hold within and between districts and are robust to the

inclusion of control variables for factors and productivity. Second, we propose two instruments misallocation that use either time lags or the industry composition of districts. The last one simulates an economy with the same number of establishments, distribution of factors across firms, and covariances among them to derive similar effects of factor misallocation as in our regressions.

———————-

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. In addition we also show how factor misallocation relates to output misallocation.

In the fourth step,The

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.

India is well suited for our purpose. The government of India 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, many of which contain enough establishments for us to compute informative measures of misallocation.

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.

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. Misallocation in theory

To motivate the empirical approach we develop below, we consider the following simple production structure. There are n production units indexed by $i = 1,...,n$. We refer to production units as establishments to be consistent with the data we use. Each establishment produces a homogenous output according to:

$$Y_i = A_i F_i^a , \tag{1}$$

where A_i measures establishment i's total factor productivity. In line with past literature (Foster, Haltiwanger, and Syverson, 2008, Hsieh and Klenow, 2009), we refer to this quantity as an establishment's physical productivity (or TFPQ). There is a unique factor of production F_i available at the endogenously determined price P_F per unit.[1] We assume decreasing returns to scale in production: $a < 1$. The total endowment of factor is fixed to F in the economy.[2] Establishments are competitive. With no further loss of generality, we can normalise the price of output to unity.

As a first benchmark, we first consider that establishments operate without friction or distortion. After solving for the the first-order condition for profit maximisation, the demand for factor from establishment i is:

$$F_i = \left(\frac{a A_i}{P_F} \right)^{\frac{1}{1-a}} . \tag{2}$$

Market clearing for the factor of production implies:

$$F \equiv \sum_i F_i = \left(\frac{a}{P_F} \right)^{\frac{1}{1-a}} \sum_i A_i^{\frac{1}{1-a}} . \tag{3}$$

Next, we can use equation (3) to eliminate P_F from equation (2) and then substitute for F_i in the production function (1). Summing across establishments, we can finally write mean output per unit of factor in the economy as

$$Z^* = \frac{\sum_i Y_i}{F} = \left(\frac{\sum_i A_i^{\frac{1}{1-a}}}{F} \right)^{1-a} . \tag{4}$$

The competitive equilibrium, as described by equations (2) and (3) exists, is unique, and is efficient since equilibrium output per unit of factor as described by equation (4) is also the outcome of the planner's problem.

[1] We consider multiple factors of production.

[2] This is a simplified version of the models originally proposed by Lucas (1978) and Hopenhayn (1992). Decreasing returns can be justified by the limited span of control of the entrepreneur that runs each establishment. The heterogeneity of productivity among establishments reflects differences in ability among entrepreneurs. This model could be enriched to consider differentiated goods (and constant returns to scale) as in Melitz (2003). A demand system with constant elasticity of substitution would lead to results that are isomorphic to those obtained here for our purpose. See Hopenhayn (2014*a*) for further discussion.

Using equation (2), we can rewrite the establishment production function (1) in equilibrium as:

$$Y_i = \frac{P_F}{a} F_i \,. \tag{5}$$

Much of the literature regresses the log of establishment revenue on the log of factor using ordinary least squares (OLS). The residual of this regression is often referred to as its revenue productivity (or TFPR).[3] In the context of equation (5), this residual is the same for all establishments since profit maximisation implies that each establishment equates the marginal return to the factor of production with the (common) price of this factor. This residual is thus uninformative of the physical productivity of establishments.[4]

In the benchmark considered so far, there is no misallocation. To allow for misallocation, Restuccia and Rogerson (2008) , Hsieh and Klenow (2009), and much of the following literature impose a first class of restrictions to the setting above.[5] Namely, the output of each establishment i (still described by equation 1) is now taxed at the idiosyncratic rate τ_i. We can think of this class of models as models of "distortions with adjustments" (as establishments still choose factors without friction). The profit of an establishment is now given by:

$$\pi_i = (1 - \tau_i) A_i F_i^a - P_F F_i. \tag{6}$$

Profit maximisation then implies:

$$F_i = \left[\frac{a\,(1 - \tau_i) A_i}{P_F} \right]^{\frac{1}{1-a}} . \tag{7}$$

Using the market clearing condition $F = \sum_i F_i$ to eliminate P_F from equation (7) and re-inserting into the production function, we can write the following expression for mean output per unit of

[3]Like Bartelsman, Haltiwanger, and Scarpetta (2013), we call revenue productivity the residual of an OLS regression of log establishment revenue on log factors. Hsieh and Klenow (2009) define revenue productivity slightly differently as $P_i A_i$ in a model where the price of the output of each establishment, P_i, is endogenous and establishments operate with constant returns. Although this definition is not immediately applicable here, the model we use here is isomorphic to a one-factor version of their model.

[4]To make this residual differ across establishments (as is the case empirically), we could easily allow for some measurement error. This residual would nonetheless remain uninformative of physical productivity. In addition, in the regression of log revenue on log factor, the estimated coefficient is often interpreted as the factor share. While this is true in equation (1) when factors are exogenous, this is not case in equation (5) where the coefficient on log F is unity not a. This estimate of factor share is upward-biased because the (endogenous) amount of factor is correlated with unobserved physical productivity.

[5]See Restuccia and Rogerson (2013) for a review. We note in particular the work of Adamopoulos and Restuccia (2014) and Restuccia and Santaeulalia-Llopis (2014). Like us, they focus on land, but they are concerned with agricultural land in rural areas whereas we deal with land used by manufacturing establishments, which are often found in urban settings in developing countries.

factor:

$$Z^{\tau} = \frac{\sum_i Y_i}{F} = \frac{\sum_i A_i^{\frac{1}{1-a}} (1-\tau_i)^{\frac{a}{1-a}}}{F^{1-a} \left[\sum_i A_i^{\frac{1}{1-a}} (1-\tau_i)^{\frac{1}{1-a}} \right]^a}, \tag{8}$$

To gain insight into this expression, it is convenient to assume that the number of establishments, n, is large and that the distributions of physical productivity and tax rates are independent: $\tilde{\tau} \perp \tilde{A}$. Then, the law of large numbers and expression (4) imply that we can rewrite mean output per unit of factor as:

$$Z^{\tau} = Z^* \frac{\mathbb{E}\left[(1-\tilde{\tau})^{\frac{a}{1-a}} \right]}{\left\{ \mathbb{E}\left[(1-\tilde{\tau})^{\frac{1}{1-a}} \right] \right\}^a}, \tag{9}$$

where the operator $\mathbb{E}[.]$ denotes expectations. Because $a < 1$, Jensen's inequality implies that $Z^{\tau} < Z^*$ provided the distribution of tax rates is non-degenerate. The equilibrium is no longer efficient because the marginal product of the factor is not equalised across establishments as a consequence of idiosyncratic taxation. More highly taxed establishments are suboptimally small and the magnitude of the distortion increases with the tax. Hsieh and Klenow (2009) also show that, if $\tilde{A}$ and $1/(1-\tilde{\tau})$ are jointly normally distributed, output per unit of factor declines with the variance of $\log(1-\tilde{\tau})$.

This is an important and useful result. To understand its usefulness, we can use equation (7) and the production function, and show that

$$Y_i = \frac{P_F}{a(1-\tau_i)} F_i \,. \tag{10}$$

As with equation (5), regressing the log of Y_i on the log of F_i estimates again a coefficient on F_i of one but the residual is now $-\log(1-\tilde{\tau}_i)$ (up to a constant). Although this residual is still uninformative regarding the physical productivity of an establishment, its distribution can be used to compute the variance of $\log(1-\tilde{\tau})$, which, in turn, affects output per unit of factor negatively. Put more directly, the variance of revenue productivity across establishments is a measure of misallocation that affects output per worker.

Misallocation may also occur because establishments operate under imperfect information (David, Hopenhayn, and Venkateswaran, 2016) or because price wedges prevent establishments from adjusting factors after a shock (Asker *et al.*, 2014).[6] To model this type of constraint in a highly simplified manner, we return to our initial setting with no taxation but now consider that physical

[6] Another strand of literature (e.g. Banerjee and Moll, 2010, Midrigan and Xu, 2014) considers frictions in financial markets that prevent establishments from growing in size as their productivity grows.

productivity is given by $A_i = A_i^0 \times A_i^1$ where A_0 is initial productivity and A^1 is a subsequent productivity shock. Without loss of generality when establishments know the distribution of productivity shocks, we assume that the expected shock is equal to unity, $\mathbb{E}(\tilde{A^1}) = 1$, and that initial productivity and the subsequent shock are orthogonal, $\tilde{A^0} \perp \tilde{A^1}$.

Importantly, establishments make their factor decision based only on their knowledge of A_0, either because A_1 is unknown or because establishments face price wedges that prevent them from adjusting their quantity of factor after a productivity shock. We can think of this class of model as generating "misallocation through a lack of factor adjustment".

The maximisation of expected profit by competitive establishments implies:

$$F_i = \left(\frac{a\, A_i^0}{P_F} \right)^{\frac{1}{1-a}} . \tag{11}$$

Using equation (11) and recalling that $A_i = A_i^0 \times A_i^1$, mean output per unit of factor is given by:

$$\underline{Z} = \frac{\sum_i Y_i}{F} = \frac{\sum_i A_i^0 A_i^1 F_i^a}{F} = \frac{P_F}{a} \sum_i A_i^1 \frac{F_i}{F} . \tag{12}$$

Because $\sum_i F_i = F$, using again equation (11) to substitute for $\frac{P_F}{a}$ in equation (12), output per worker can be usefully rewritten as:

$$\underline{Z} = \left(\frac{\sum_i (A_i^0)^{\frac{1}{1-a}}}{F} \right)^{1-a} n\, \mathbb{E}\left[\frac{F_i}{F} A_i^1 \right] . \tag{13}$$

When the number of establishments is large, we can apply the law of large numbers. Then, orthogonality between $\tilde{A^0}$ and $\tilde{A^1}$ implies that $n\, \mathbb{E}\left[\frac{F_i}{F} A_i^1 \right] = n\, \mathbb{E}\left[\frac{F_i}{F} \right] \mathbb{E}\left[A_i^1 \right] = n \times \frac{1}{n} \times 1 = 1$. Hence, output per worker simplifies into:

$$\underline{Z} = \left(\frac{\sum_i (A_i^0)^{\frac{1}{1-a}}}{F} \right)^{1-a} . \tag{14}$$

Because $A_i = A_i^0 \times A_i^1$ with $\mathbb{E}\left[A_i^1 \right] = 1$, Jensen's inequality implies that $\sum_i (A_i^0)^{\frac{1}{1-a}} < \sum_i (A_i)^{\frac{1}{1-a}}$. Then by equation (4), it is immediate that $\underline{Z} < Z^*$. Again, misallocation occurs because the marginal product of the factor is not equalised across establishments. The difference with the previous class of models is that this lack of equalisation occurs because of a lack of factor adjustment following a productivity shock. In models of distortions with adjustments, the lack of equalisation of the marginal product of factor occurs instead because of a tax wedge between the marginal return to the factor and its price.

Empirically, it will be useful to consider groups of establishments potentially subject to common shocks. Within those groups, the productivity shocks A^1 may be correlated. This could be the case with establishments located in the same district g if, for instance, $A_i^1 = A_i^{1'} + A_{g(i)}^1$ where $A_i^{1'}$ is an idiosyncratic productivity shock and $A_{g(i)}^1$ is a productivity (or demand) shock affecting the entire district. In this case, the expectation term for the product of establishment factor share and the productivity shock in equation (13) is no longer equal to the product of the expected value of the productivity shock and the expected factor share since the covariance between those two terms is in general different from zero. For a group g of n_g establishments, the expression analogous to equation (15) for output per unit of factor is now:

$$Z_g = \left(\frac{\sum_{i \in g} (A_i^0)^{\frac{1}{1-a}}}{F_g} \right)^{1-a} n_g \, \mathbb{E}\left[\frac{F_i}{F_g} A_i^1 \right] = \left(\frac{\sum_{i \in g} (A_i^0)^{\frac{1}{1-a}}}{F_g} \right)^{1-a} \left(n_g \text{cov}(\frac{F_i}{F_g}, A_i^1) + 1 \right) , \qquad (15)$$

where $F_g = \sum_{i \in g} F_i$ and the last simplification makes use of the law of large numbers. In turn, equation (15) suggests that, empirically, we can measure the effects of misallocation on output per unit of factor using the covariance between establishments' shares of factor and their productivity shock A^1. This measure of misallocation is of course akin to that originally proposed by Olley and Pakes (1996) and used by much of the subsequent literature.[7]

The covariance measure of misallocation in equation (15) obviously requires knowing about the productivity shock A^1 and the factor shares of establishments F_i/F_gwithin their group. Factor shares can be read directly from the data. For the productivity shocks, it is easy to see from equation (11) and the production function that,

$$Y_i = \frac{P_F}{a} A_i^1 \, F_i \, . \qquad (16)$$

Like with equations (5) and (16), regressing the log of establishment revenue on the log of factors estimates yet again a coefficient of one for the factor (and not the factor share a). Crucially, the productivity residual will now be the log of A_i^1, the productivity shock. This is easy to understand. Factor choices are made on the basis of initial productivity and are thus orthogonal to subsequent productivity shocks. While A^1 is only one component of physical productivity A, it is the component that drives misallocation. Put differently, by regressing log establishment revenue

[7]We note that the OP measure of misallocation — sometimes referred to as the OP decomposition — has been used to explore issues such as changes in factor allocation in the telecom industry after deregulation in the US (Olley and Pakes, 1996), the effects of structural reforms in developing countries (Eslava, Haltiwanger, Kugler, and Kugler, 2006), productivity differences across countries (Bartelsman *et al.*, 2013), and the role financial institutions in factor allocation (Midrigan and Xu, 2014). It has been extended to a dynamic setting by Melitz and Polanec (2013).

on log factors the residual we recover is now the productivity shock. We still mismeasure physical productivity but the estimated residual is what we need to measure misallocation and to assess its effects on output per unit of factor.

To sum up, misallocation can be measured with the variance of revenue productivity in models of distortions with adjustment while it can be be measured with the covariance between factor shares and revenue productivity in models with a lack of factor adjustment. In addition, in models of distortions with adjustment, the covariance between measured productivity and factor shares is only indicative of the correlation between physical productivity and the tax (if any). This can be seen from equation (16), where (estimated) revenue productivity is $1/(1-\tau)$ while factor use depends on $[(1-\tau)A]^{1/1-a)}$ by equation (7). In this class of model, this covariance is thus only an indirect and, perhaps at best, second-order determinant of output per unit of factor (Hopenhayn, 2014*b*). In models with a lack of factor adjustment, the dispersion of revenue productivity has no predictive power regarding output per worker once the covariance between establishment factor share and the productivity shock is taken into account as can be seen from equation (15). The same equation indicates that if we could manage to recover initial productivity A_0 (perhaps through a more sophisticated productivity estimation), the dispersion of estimated productivity should affect output per unit of factor positively instead of negatively in models of distortions with adjustment.

As different models suggest using different measures of misallocation, a first important objective of our work below is to uncover empirically robust measures of misallocation. We use the models above as broad guides but will refrain from giving our results an overly strong structural interpretation. Our limited empirical knowledge about misallocation provides few restrictions regarding the structural models that might be considered. Instead of proposing yet another detailed structure, we want to provide some robust empirical features to guide future theorising.

Our second objective is then to assess empirically the effect of misallocation on output per unit of factor as suggested by equations (9) or (15). Again, we want to be cautious and avoid having to rely strongly on theory at this stage. Existing estimates of the output cost of misallocation are highly model-dependent and sensitive to parameters which we know little about (Hopenhayn, 2014*a*). This is the case for instance of the factor share a in models of distortions with adjustment, where the effects of idiosyncratic taxes are greatly magnified as a approaches one. Our choice below will be to rely on the data more than on potentially arbitrary parameter values imposed from theory.

Our third objective is to relate measures of misallocation to specific potential sources of misallocation. This is important for at least three reasons. First, this investigation can provide further validation of the measures of misallocation we use. Second, a better understanding of the sources of misallocation will also be useful to guide theory. Third, knowing about the sources of misallocation is also a crucial first step towards better policies and less misallocation.

3. Measuring misallocation in practice

The approach we develop below consists essentially of a set of regressions where we use various measures of misallocation either as the explanatory variables of interest or as dependent variables. Before turning to these regressions, we first need to discuss our measures of misallocation and the productivity estimates used to construct them.

Models of distortions with adjustments in the line of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) suggest measuring misallocation with the variance of revenue productivity. For a group of establishments g, this measure can be formally written as:

$$M_g^{\text{TFP}} \equiv \sum_{i \in g} (\varphi_i - \overline{\varphi}_g)^2 \,, \tag{17}$$

where $\overline{\varphi}_g \equiv \frac{1}{n_g} \sum_{i=1}^{n_g} \varphi_i$ is mean productivity across establishments in group g with n_g establishments.

Models of lack of adjustments suggest measuring misallocation with the covariance between establishment factor share and revenue productivity:

$$M_g^F \equiv -n_g \operatorname{cov}_g(s_{ig}, \varphi_i) = -(\Phi_g - \overline{\varphi}_g) \,, \tag{18}$$

where $s_{ig} \equiv F_i / F_g$ is the factor share of establishment i in group g and $\Phi_g \equiv \sum_{i=1}^{n_g} s_{ig} \varphi_i$ is the share-weighted productivity of establishments in that group. This last quantity can be interpreted the average productivity with which output is produced. Hence, the covariance measure of allocation M_g defined by equation (18) is also equal to the difference between un-weighted and weighted productivity, where the weights are the factor shares of each establishment. Misallocation is less when this difference between weighted and unweighted productivity is larger or, equivalently, when the covariance between shares and productivity is higher. Note that, relative to usual practice in past work (e.g. Bartelsman *et al.*, 2013) and to ease the reading of our results, equation (18) multiplies the covariance term by minus one so that an increase in the index corresponds to an increase

in misallocation. $M^F = 0$ corresponds to factor intensity being uncorrelated with establishment productivity while more negative values corresponds to positive covariances between the use of factor and productivity leading to less misallocation. It is possible to imagine positive values for our misallocation index M^F. Loosely speaking, such situations correspond to a 'worse than random' match between factor usage and productivity.

While our modelling has so far considered a unique factor of production. In our empirical analysis, we consider a three-factor extension to the production function proposed in equation (1):

$$Y_i = A_i \, T_i^{\alpha} \, K_i^{\beta} \, L_i^{\gamma} \,, \tag{19}$$

where T is land and buildings, K is other fixed assets, and L is labour. We explicitly separate between land and buildings and other fixed assets such as machines and equipment because this distinction is important in our context. Empirically, we estimate:

$$\log Y_i = \alpha \log T_i + \beta \log K_i + \gamma \log L_i + \varphi_i \,, \tag{20}$$

where φ_i is the residual.

With several factors of production, a different misallocation index can be computed for each factor or group factor depending on the weights used for the factor shares in equation (18). Hence, we can define measures of misallocation for employment, M^L, land and buildings, M^T, other fixed assets, M^K, or any combination of these. Note that we can also use output shares as weights in equation (18) and define a misallocation index for output, M^Y.

While this multiplicity is noted in past literature, it is mostly viewed as a minor annoyance forcing a choice upon the analyst, typically between output or employment, for the computation of the shares entering a single misallocation index. We note instead that these different measures of misallocation are conceptually distinct. For instance, a given factor of production might be more misallocated than another depending on the relative frictions affecting input markets. We also note that a random allocation of factors for all inputs will lead to a non random allocation of output since, in expectation, more productive establishments will produce more output out of the same expected amount of inputs.

Next, recall that we are interested in measuring misallocation at the district level. At the same time, the misallocation index defined in equation (18) requires the productivity measure φ_i and the factor shares to be comparable across establishments in group g. Summing directly across establishments that belong to different industries is potentially problematic. To understand this

point, consider two industries with different capital intensity. 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 actually be using more capital even though its productivity is lower. Hence, for comparability purpose, establishments need to belong to the same industry (e.g., textiles or basic metals).

To obtain consistent measures of misallocation by district, we proceed as follow. Consider district g where establishments can be partitioned into m industry subgroups denoted $j = 1, 2, ..., m$. After dropping the district index g to ease notations, we can rewrite equation (18) as:

$$M^F = -n \operatorname{cov}(s_i, \varphi_i) = -\left(\sum_{i=1}^{n} s_i \varphi_i - \frac{1}{n} \sum_{i=1}^{n} \varphi_i \right), \tag{21}$$

where n is the total number of establishments in the district. Using the fact that the share of an establishment is equal to the share of that establishment in its industry times the share of that industry in the district, $s_i = s_{ij} \times s_j$, we can rewrite equation (21) as:

$$\begin{aligned} M^F &= -\left(\sum_j \sum_{i \in j} s_j \, s_{ij} \varphi_i - \frac{1}{n} \sum_j \sum_{i \in j} \varphi_i \right) \\ &= -\left[\sum_j s_j \left(-M_j + \frac{1}{n_j} \sum_{i \in j} \varphi_i \right) - \frac{1}{n} \sum_j \sum_{i \in j} \varphi_i \right], \end{aligned} \tag{22}$$

where the last simplification arises directly from the definition of M^F in equation (18). After defining mean industry productivity in the district, $\Phi_j \equiv \sum_{i=1}^{n_j} s_{ij} \varphi_i$, and further simplifications, we can rewrite expression (22) as:

$$M^F = \sum_j s_j \, M_j - \sum_j \left(s_j - \frac{n_j}{n} \right) \overline{\varphi}_j \,. \tag{23}$$

Because $-\sum_j \left(s_j - \frac{n_j}{n} \right) = 0$ as the shares of both factors and establishments sum to unity, equation (23) can be finally written as:

$$M^F = \tilde{M} + M_\Delta \tag{24}$$

with $\quad \tilde{M} \equiv \sum_j s_j \, M_j^F \quad$ and $\quad M_\Delta \equiv -\sum_j \left(s_j - \frac{n_j}{n} \right) \left(\overline{\varphi}_j - \overline{\varphi} \right).$

Hence, misallocation for establishments in a district can be decomposed into the sum of two simple terms. The first, $\tilde{M}$, is a within-industry (and district) misallocation term. It sums misallocation for each district and industry across industries, weighting them by their share s_j in the

district. The second term, M_Δ, is a measure of between-industry misallocation. It is actually the between-group analog of the covariance term used in equation (18). This makes strong intuitive sense: total misallocation in a district is the sum of the misallocation of each of its industries where each industry is weighted by its factor share and misallocation across industries.

We use the above decomposition to compute consistent measures of misallocation at the level of each district using measures of misallocation computed within each district and industry and across industries within each district.[8] Beyond offering a consistent way to compute district misallocation, this decomposition is useful because it clarifies what we do when we measure misallocation at the district level relative to aggregate misallocation for the entire economy. A straightforward application of equation (24) shows that aggregate misallocation is the weighted sum of district misallocation and a between-district misallocation term.

The variance misallocation index described by equation (17) and the covariance misallocation indices (24) all rely on a productivity residual φ_i estimated from the regression (20) of log establishment revenue on log capital, labour, and land and buildings. As cautioned by the models above,we do not expect this productivity residual to be equal to the natural log of physical productivity A_i. Instead, model of distortions with adjustment interpret it as a measure of the idiosyncratic tax faced by establishments $\log(1/(1-\tilde{\tau}_i))$. Models of lack of factor adjustment interpret the productivity residual as the productivity shock, which is only one component of physical productivity. In both cases, this productivity residual is nonetheless an appropriate measure of the distortions faced by establishments.

We nonetheless worry about relying heavily on a mis-estimated productivity residual, albeit appropriately so. To deal with that worry, we estimate equation (20) using a variety of alternative approaches including considering two factors of production instead of three (L and $T+K$), constraining total factor share to unity, or using total gross revenue instead of value added as a measure of establishment revenue.

We also experiment with approaches that attempt to recover log physical productivity, $\log A$ by correcting for factor endogeneity. Unfortunately, the approaches developed by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP) require a panel of firms, which is not available to

[8] For instance, it is possible to imagine situations where more productive establishments employ more workers within each industry, but that industries that host more productive establishments on average have access to substantially fewer workers than less productive industries (thinking for instance about gender specialisation across industries). This would be a case of low misallocation within industries but high misallocation across industries.

us from the ASI and NSSO surveys. Following an approach initially developed by Sivadasan (2009), we can nonetheless implement an approach in the same spirit as LP that only requires information aggregated by districts. More specifically, 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. . We appropriately tailored it to our needs. Note that this approach is broadly consistent with our modeling above when we assume that productivity shocks that affect establishments have both an idiosyncratic and a district component.

There are also more specific worries regarding the covariance indices of equation (20). While the approach we use to compute misallocation at the district level by aggregating misallocation computed for each district and industry is consistent, other aggregations are possible. We also worry about possible sampling problems when computing measures of misallocation for industries within districts, where, in some cases, the number of establishments is small. Hence, we compute simpler indices based on the covariance between 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).[9] We can also measure misallocation using only a sum of within-industry misallocation without considering the between-industry term.

Following the approach of Bartelsman *et al.* (2013) in their cross-country comparison, we also use national shares to aggregate industries. Another alternative is, before summing across sector, to subtract the average tendency of each industry to be misallocated in the country. This 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 by equation (24), we compare our results below to those obtained with these alternative measures.

[9]To avoid the results being driven by a small number of extremely large establishments with particularly high or low productivity, 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 use these corrected factor measures to compute equation (18) 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.

4. Data and descriptive facts about misallocation in India

Before we develop our empirical framework further, some knowledge of the data at hand is required.

4.1 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 Aghion, Burgess, Redding, and Zilibotti (2008), Allcott, Collard-Wexler, and O'Connell (2016) or Ghani, Goswami, and Kerr (2014) from which the first part of this data description builds on. The 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, Goswami, and Kerr (2012), 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.

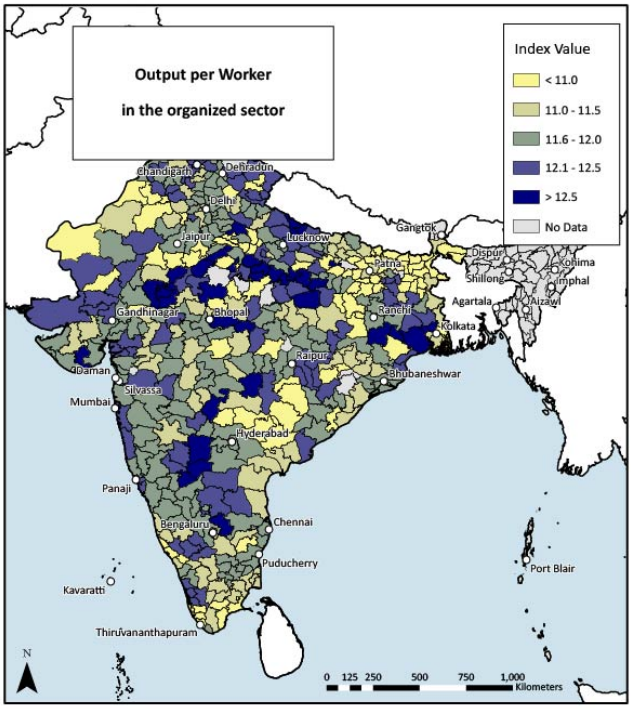
In both sectors, we can measure output through value added, count total employment, and obtain a measure of machinery, equipment, and other fixed assets. For land and buildings matters are more complicated. Establishments in both the organised and unorganised sectors report information on the value of the land and buildings that they own.[10]

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 to 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 A. In robustness checks, we distinguish between owners and renters and duplicate our main regressions.

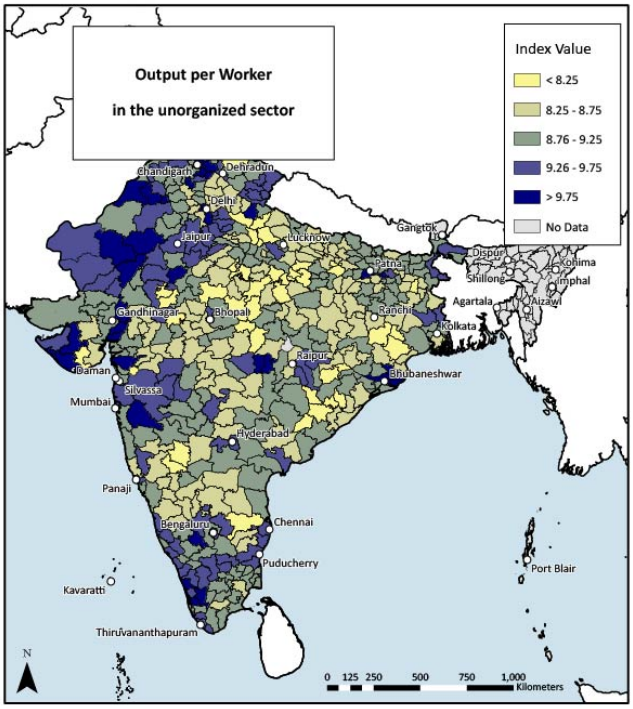
The raw data from the five survey years contain 262,911 observations in the organised sector and 955,234 observations in the unorganised sector. Following the clean-up process described in Appendix A, the final sample that we work with for productivity and misallocation metrics consists of 145,829 observations 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. 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.[11]

[10]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. We also note that the organised sector surveys offer the distinction across land and buildings consistently for all the survey years under consideration. For the unorganised sector, this separation between land and building stops in 2000. To remain consistent, we always consider land and buildings together.

[11]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.



Panel A



Panel B

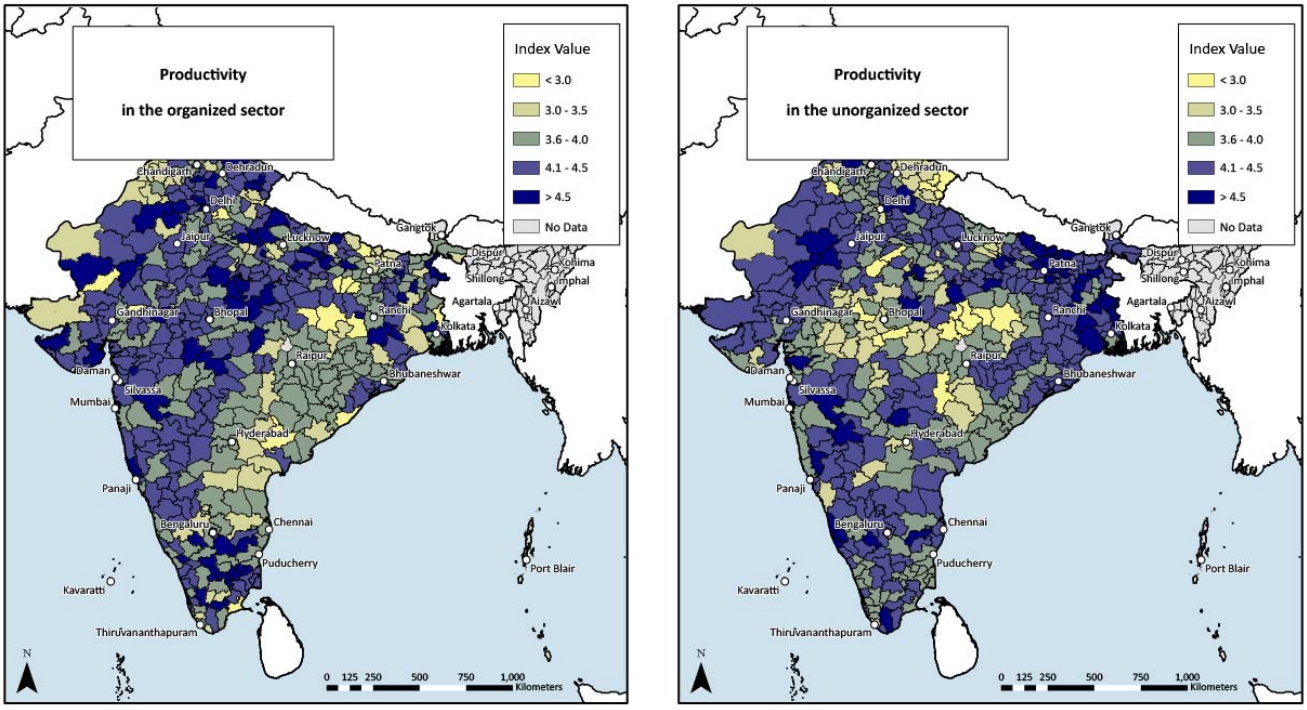
Figure 1: Map of output per worker in Indian districts, by quintile, 2010

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

Figures 1 shows a map of output per worker. This map is 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.[12]

Turning to productivity estimates, our preferred approach estimates a separate coefficient for land and buildings, other fixed assets, and labour for each industry in each sector. We treat the same 'industry' in the organised and unorganised sectors as separate because they tend to differ widely in their factor intensity. For instance, the mean share of other fixed assets across industries is 0.27 in the organised sector versus 0.20 in the unorganised sector whereas for land and build-

[12]This map combines the organised and unorganised sectors. As might be expected, output per worker is nearly everywhere much higher in the organised sector. The correlation between output per worker in the organised and unorganised sector is somewhat modest, around 0.1. For instance, output per worker is high in the unorganised sector relative to the organised sector in Kerala, while the opposite hold true in West Bengal.



Panel A Panel B

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

ings we estimate a mean coefficient of 0.11 for the organised sector and 0.22 in the unorganised sector. For labour, the coefficients are relatively similar at 0.76 and 0.78 for the organised and unorganised sectors, respectively. These differences may reflect differences in access to factors 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.

We note that for a majority of sectors, the sum of the coefficients on the different factors is above one. As made clear in section 2, we do not expect these coefficients to reflect true factor shares. This is nonetheless more than what is typically estimated (Syverson, 2011) and is robust to the exact estimation technique. We believe two phenomena are potentially at work. First, manufacturing establishments that are constrained by frictions on factor markets may not be able reach their optimal size. It could also be the case that frictions 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 use relatively less factors after conditioning out their productivity. This implies that in equilibrium factors could

increase less than proportionately with output.

Figure2B represents our TFP estimates averaged by district. Like the map of output per worker, this map reveals a strong contrast between the more productive western parts of the country and the less productive eastern areas.It also shows again areas of higher productivity around IndiaŠs 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 IndiaŠs spatial disparities in output per worker are to some extent productivity disparities.

4.3 Some descriptive facts about misallocation in India

Figures 3A and 3B provide maps of two misallocation indices: the variance of revenue productivity as described in equation 17 and the covariance between shares of land and buildings and revenue productivity computed as per equation (24).

XXXXXXXXXXXXXXXXXXXXXXXX

REST OF THIS SECTION TO BE REWRITTEN

XXXXXXXXXXXXXXXXXXXXXX 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.

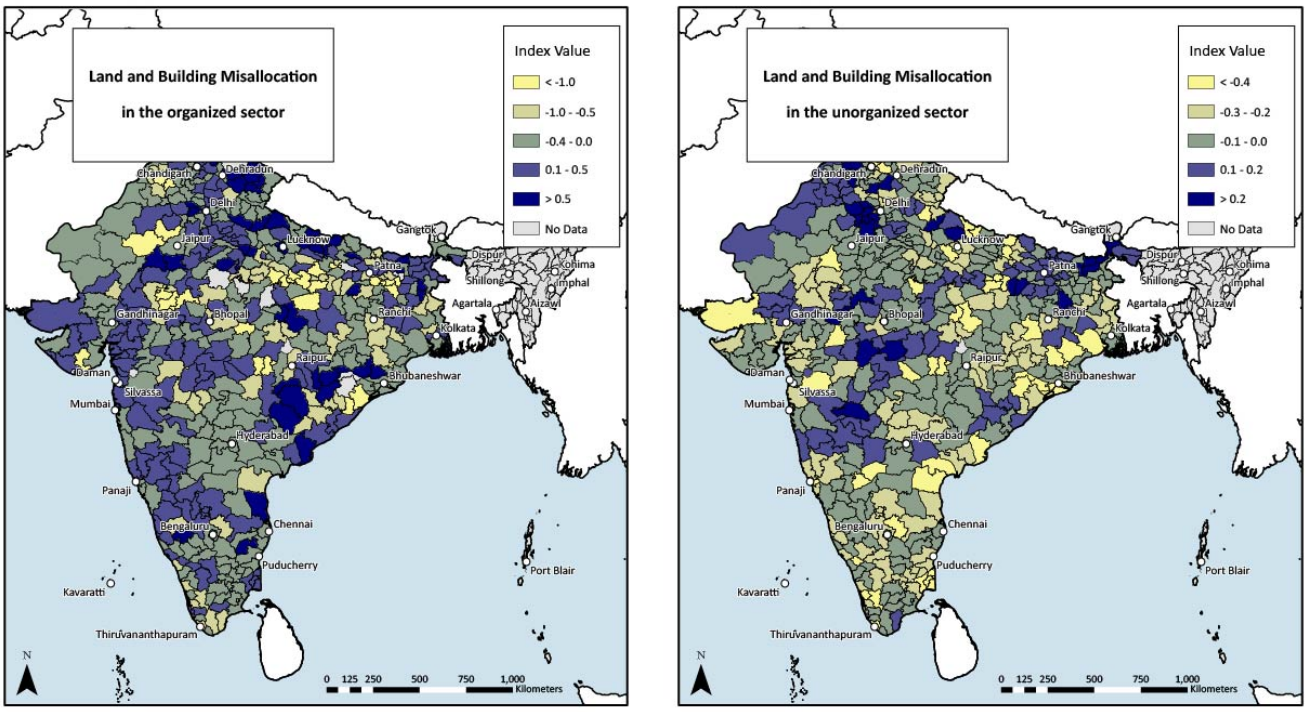
Table 1 reports descriptive statistics for indices of misallocation for Indian districts. We report statistics for the full manufacturing sector, followed by breakouts of the organised and unorganised sectors. A number of features are worth highlighting. Recalling that higher index values indicate greater 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, Haltiwanger, and Scarpetta (2009) compute overall input misallocation indices for 18 countries at different levels of economic development. For the sake

Table 1: Misallocation indices across districts

	Output	Labour	Land & Buildings	Other fixed assets
Panel A. Full Indian manufacturing sector				
District mean 1989	-0.97	-0.44	-0.52	-0.60
District mean 1994	-0.75	-0.14	-0.28	-0.26
District mean 2000	-0.91	-0.28	-0.38	-0.54
District mean 2005	-0.79	-0.30	-0.30	-0.46
District mean 2010	-0.70	-0.20	-0.19	-0.31
District mean 1989-2010	-0.82	-0.27	-0.33	-0.43
District s.d. 1989	0.60	0.44	0.52	0.63
District s.d. 1994	0.54	0.27	0.41	1.40
District s.d. 2000	0.58	0.35	0.47	0.59
District s.d. 2005	0.48	0.42	0.52	0.52
District s.d. 2010	0.46	0.43	0.43	0.61
District s.d. 1989-2010	0.53	0.38	0.47	0.75
Panel B. Organised sector				
District mean 1989-2010	-0.46	-0.07	-0.11 // -0.03	-0.10
District s.d. 1989-2010	0.43	0.26	0.37//0.41	0.46
Panel C. Unorganised sector				
District mean 1989-2010	-0.60	-0.06	-0.09	-0.19
District s.d. 1989-2010	0.29	0.12	0.20	0.49

Notes: 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.



Panel A Panel B

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

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 withBartelsman *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).

The second key feature of table 1 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 1 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 1. 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.[13]

5. The policy determinants of misallocation: A first validation of our metrics

While a number of studies calculate and compare metrics of misallocation, there has been limited scope to validate them given the aggregate nature of these studies. An advantage of our focus on a cross-section of districts is that we can demonstrate their general validity and usefulness by studying changes in district-level misallocation associated with policy changes. We expect economic policies to affect the idiosyncratic taxes that establishments face, wedges in factor prices, and the uncertainty that establishments face when they make factor choices. Because our results emphasise the importance of land and buildings as a key source of misallocation, we focus on two important policy reforms in that area: the repeal of the Urban Land (Ceiling and Regulation) Act (ULCRA) and changes in state-level stamp duties (taxes on land sales).[14] 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 abolition of the Urban Land (Ceiling and Regulation) Act has the advantage of being unantici-

[13]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.

[14]In work in progress, we explore a broader range of policy changes associated with labour reforms, foreign direct investment, industry licencing, etc.

pated, and in our analyses we control for trends in many other local traits that could potentially be correlated with this policy change. 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.

5.1 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.[15] The regulation and market constraints reduced the incentives of

[15]Exemptions to ULCRA were certainly pervasive (Srinivas, 1991). There is nonetheless a broad consensus in the literature that ULCRA was imposing binding constraints. Again, our objective is to show that removing the the constraints created by ULCRA affected misallocation. Trying to estimate the elasticity of misallocation with respect to the stringency of land use regulations is beyond our scope here.

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.[16] The ULCRA reform was mostly unanticipated. 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.

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 for 252 Indian districts as the dependent variable. Our key explanatory variable is an indicator variable for states that repealed ULCRA early, as listed above.[17] We estimate:

$$\Delta_{2010-2000} M_g = a_0 + a_1 \text{EarlyRepeal}_i^{\text{ULCRA}} + b_1 X_g + \epsilon_g \,, \tag{25}$$

where $M = M^Y$, $M = M^T$, or $M = M^{\text{TFP}}$ depending on the regression.

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. We also control for past misallocation to control flexibly for trends.

XXXXXXXXXXXXXXXXXXXXXXXXXXX

WILL HAVE TO REWRITE FROM HERE

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

[16] 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).

[17] 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.

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 ULCRAŠs repeal.

Column 2 shows this result is robust to adjustments in the covariates modelled, although the basiASIC controls are important for the statistical precision with which we can measure ULCRAŠs 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 ULCRAŠs 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.

5.2 Stamp duty

High stamp duties are another trait of Indian property markets.[18] These taxes are collected whenever a real property is transacted. While these taxes are between zero and 5% in most North American jurisdictions, they tend to be much higher in India. There is also a lot of variation across states and over time. For instance, the lowest rate in 1989 was found in Tripura at 5%, while the highest rate of 21.2% was in West Bengal. 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, Duranton, and Turner, 2012).

High stamp duties impose high compliance costs on taxpayers and lead to widespread avoidance through under-reporting (Alm, Annez, and Modi, 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.

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

[18]Stamp duties in India are imposed under the Indian Stamp Act, 1899, as amended several times over the years at the central government level. 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 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.

to 2003 as the core explanatory variable. We again control for initial misallocation and the traits of districts.

XXXXXXXXXXXXXXXXXXXXXXXXXXX

WILL HAVE TO REWRITE FROM HERE

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

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.

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.

6. How different measures of misallocation interact: A second validation of our metrics

We now turn to how different forms of misallocation interact at the district level. This second empirical exercise is interesting in its own rights. It also provides a second validation of our metrics.

Consistent with equation (18) and after dropping group indices for simplicity, output misallocation can be written as:

$$M^Y = -n\,\mathrm{cov}(s_i^Y, \varphi_i) = -\sum_{i=1}^{n} s_i^Y (\varphi_i - \overline{\varphi})\,, \tag{26}$$

where $s_i^Y = Y_i/Y = Y_i/\sum_j Y_j$ is the share of output of establishment i in total output.

Consider a straightforward three-factor extension of the model of lack of factor adjustment where the production function is given by (19). Recall that establishments decide on their factors knowing initial productivity A^0 but not the productivity shock $\tilde{A}^1$. The maximisation of expected profit implies $R\,T_i = \alpha\,\mathbb{E}\,[Y_i]$, $r\,K_i = \beta\,\mathbb{E}\,[Y_i]$, and $w\,L_i = \gamma\,\mathbb{E}\,[Y_i]$ where R is the rental rate of land and buildings, r is the interest rate on other fixed assets, and w is the wage.

Aggregating across all establishments using the law of large numbers, it is easy to obtain that $R\,T = \alpha\,Y$, $r\,K = \beta\,Y$, and $w\,L = \gamma\,Y$ where $T \equiv \sum_i T_i$, $K \equiv \sum_i K_i$, and $L \equiv \sum_i L_i$ are the district aggregate values for land and buildings, other fixed assets, and labour respectively. Then it obviously follows from these simple results that

$$\begin{aligned}
\mathbb{E}\left[s_i^Y\right] &= \frac{R\,T_i}{(\alpha+\beta+\gamma)Y} + \frac{r\,K_i}{(\alpha+\beta+\gamma)Y} + \frac{w\,L_i}{(\alpha+\beta+\gamma)Y} \\
&= \frac{R\,T_i}{(\alpha+\beta+\gamma)\frac{R\,T}{\gamma}} + \frac{r\,K_i}{(\alpha+\beta+\gamma)\frac{r\,K}{\alpha}} + \frac{w\,L_i}{(\alpha+\beta+\gamma)\frac{w\,L}{\beta}} \\
&= \frac{\alpha s_i^T}{\alpha+\beta+\gamma} + \frac{\beta s_i^K}{\alpha+\beta+\gamma} + \frac{\gamma s_i^L}{\alpha+\beta+\gamma}\,,
\end{aligned} \tag{27}$$

where s_i^T, s_i^K, and s_i^L are the shares of establishment i of land and buildings, other fixed assets, and labour respectively. Using equation (27), it is easy to rewrite equation (26) as:

$$M^Y = \frac{\alpha}{\alpha+\beta+\gamma} M^T + \frac{\beta}{\alpha+\beta+\gamma} M^K + \frac{\gamma}{\alpha+\beta+\gamma} M^L\,. \tag{28}$$

Hence by regressing output misallocation on factor misallocations, we should recover the factor shares. XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXx

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:

$$M_{gt}^Y = a_0 + a_1 M_{gt}^T + a_2 M_g^K + a_3 M_{gt}^L + a_4 M_{gt}^{\mathrm{TFP}} + b_t + b_g + b_1 X_{gt} + \epsilon_{gt} \tag{29}$$

Table 2: Output misallocation and factor misallocation, 1989-2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: district misallocation for value added								
Explanatory variable: misallocation for,								
Land/buildings	0.42^a		0.41^a	0.37^a			0.42^a	0.36^a
	(0.052)		(0.050)	(0.057)			(0.053)	(0.056)
Other assets	0.16^a		0.17^a	0.19^a			0.16^a	0.19^a
	(0.036)		(0.035)	(0.043)			(0.036)	(0.042)
Employment	0.41^a	0.48^a	0.42^a	0.41^a			0.41^a	0.40^a
	(0.046)	(0.043)	(0.047)	(0.061)			(0.046)	(0.061)
All assets		0.47^a						
		(0.033)						
TFP					0.73^a	0.71^a	0.046	0.17^b
					(0.083)	(0.11)	(0.047)	(0.070)
Fixed effects	N	N	state	district	N	district	N	district
Observations	1,576	1,576	1,576	1,576	1,576	1,576	1,576	1,576
R-squared	0.69	0.68	0.70	0.79	0.11	0.48	0.69	0.79

Notes: OLS regressions with year effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

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 factorŠs 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

Table 3: Output misallocation and factor misallocation, 1989-2010, for richer specifications

	(1) TFP LP	(2) TFP LP +var(TFP)	(3) w.factors	(4) w. factors +var(TFP)	(5) org.	(6) unorg.	(7) owners	(8) w. weights
Dependent variable: district misallocation for value added								
Explanatory variable: misallocation for,								
Land/buildings	0.46^a	0.45^a	0.38^a	0.37^a	0.28^a	0.35^a	0.40^a	0.38^a
	(0.051)	(0.061)	(0.055)	(0.055)	(0.064)	(0.064)	(0.057)	(0.050)
Other assets	0.13^a	0.11^a	0.19^a	0.18^a	0.19^a	0.24^a	0.18^a	0.19^a
	(0.024)	(0.025)	(0.043)	(0.042)	(0.054)	(0.042)	(0.043)	(0.035)
Employment	0.31^a	0.35^a	0.39^a	0.39^a	0.23^a	0.67^a	0.40^a	0.36^a
	(0.067)	(0.077)	(0.061)	(0.061)	(0.068)	(0.11)	(0.062)	(0.044)
TFP	-0.17		0.20^c	0.056	-0.027	0.18^b	0.12^c	
	(0.11)		(0.12)	(0.046)	(0.026)	(0.075)	(0.068)	
Land/b. (owners)							-0.064^b	
							(0.027)	
Other explanatory variables:								
Land/buildings			-0.079^c	-0.074^c				
			(0.042)	(0.042)				
Other assets			0.062^a	0.061^a				
			(0.023)	(0.023)				
Employment			0.013	0.011				
			(0.022)	(0.022)				
TFP				-0.051				
				(0.11)				
Observations	1,576	1,229	1,575	1,575	1,525	1,522	1,526	-
R-squared	0.76	0.78	0.79	0.79	0.62	0.66	0.79	0.81

Notes: OLS regressions with year and district fixed effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

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.

Table 4: Output misallocation and factor misallocation, 1989-2010 for alternative measures of misallocation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: district misallocation for value added in the combined sample								
Explanatory variable: misallocation for,								
Land/buildings	0.38^a	0.37^a	0.80^a	0.51^a	0.36^a	0.32^a	0.37^a	0.42^a
	(0.050)	(0.057)	(0.062)	(0.041)	(0.054)	(0.056)	(0.057)	(0.057)
Other assets	0.18^a	0.15^a	0.15^a	0.087^a	0.17^a	0.15^a	0.19^a	0.20^a
	(0.037)	(0.045)	(0.048)	(0.023)	(0.044)	(0.043)	(0.043)	(0.044)
Employment	0.35^a	0.38^a	0.52^a	0.33^a	0.41^a	0.40^a	0.41^a	0.40^a
	(0.046)	(0.063)	(0.069)	(0.034)	(0.059)	(0.068)	(0.061)	(0.058)
Observations	1,576	1,576	1,576	1,576	1,576	1,576	1,576	1,576
R-squared	0.69	0.79	0.89	0.88	0.80	0.78	0.79	0.82

Notes: OLS regressions with district and year effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Misallocation indices are computed as follow. Direct district aggregation in columns 1. Aggregation of industries ignoring the between-industry misallocation term in column 2. Direct district aggregation taking out national industry misallocation weighted by local shares in column 3. National industry shares in column 4. Direct district aggregation with factors re-weighted by their production function coefficients in column 5. Misallocation computed from OLS productivity estimates with a constant return constraint in column 6, from OLS productivity estimates with three factors in column 7, and from OLS productivity estimates with two factors in column 8.

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.[19]

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.

********** Interestingly, we note that in our results, the sum of the coefficients on the factor misallocation terms are always close to unity. This is obviously consistent with firms operating not far from constant returns. We also note that the factor shares that we obtain when estimating the empirical counterpart to equation (28) differ from the factor shares we obtain when we estimate establishment productivity. There are two possible reasons for that. The first that the price of land and buildings is at the margin much higher than what establishments pay on average. This would be consistent with the spirit of our model. A second interpretation is that factor choices are not made simultaneously but sequentially with land and buildings being chosen first. Then labour and capital become functions of land and buildings and unsurprisingly the estimation of the empirical counterpart to equation (28) indicates a much greater importance of land and buildings than we observe in terms of expenditure. **********

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

[19]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.

land and employment today may worsen the misallocation of output or value added in the future.

7. Productivity consequences of misallocation

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 ŚdeeperŠ 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:

$$\log Y_{gt} = a_0 + a_1 M_{gt}^T + a_2 M_g^K + a_3 M_{gt}^L + a_4 M_{gt}^{\text{TFP}} + b_t + b_g + b_1 X_{gt} + \epsilon_{gt} \tag{30}$$

or simpler

$$\log Y_{gt} = a_0 + a_1 M_{gt}^Y + b_t + b_g + b_1 X_{gt} + \epsilon_{gt} \tag{31}$$

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 ŚpureŠ 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.

Table 5: Output per worker and output misallocation, 1989-2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Base	w. district fe	w. weights	revenue as dv	w. controls	w. controls + fe	organised sector	unorg. sector
Dependent variable: log value added per worker								
Output misallocation	-0.30^a	-0.35^a	-0.20^a	-0.33^a	-0.44^a	-0.46^a	-0.26^a	0.037
	(0.045)	(0.042)	(0.043)	(0.043)	(0.024)	(0.029)	(0.036)	(0.027)
TFP					0.43^a	0.47^a		
					(0.049)	(0.063)		
Land/buildings					0.17^a	0.22^a		
					(0.030)	(0.044)		
Other assets					0.43^a	0.35^a		
					(0.019)	(0.026)		
Employment					0.12^a	0.17^a		
					(0.013)	(0.022)		
District effects	N	Y	Y	Y	N	Y	Y	Y
Observations	1,576	1,576	24,418,322	1,576	1,575	1,575	1,526	1,573
R-squared	0.11	0.69	0.84	0.69	0.80	0.90	0.68	0.75

Notes: OLS regressions with year effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

The results of table 7 suggest fairly large effects associated with factor misallocation. For the full

Table 6: Output per worker and factor misallocation, 1989-2010 for alternative measures of misallocation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: Dependent variable: log value added per worker in the combined sample								
Explanatory variable: TFP or misallocation for,								
Output	-0.64[a]	-0.63[a]	-0.039[a]	-1.32[a]	-0.44[a]	-0.69[a]	-0.35[a]	-0.26[a]
	(0.063)	(0.060)	(0.013)	(0.13)	(0.058)	(0.052)	(0.042)	(0.039)
District effects	N	Y	Y	Y	Y	Y	Y	Y
Observations	1,576	1,576	1,576	1,576	1,576	1,576	1,576	1,576
R-squared	0.24	0.74	0.66	0.71	0.71	0.75	0.69	0.68

Notes: OLS regressions with year effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Misallocation indices are computed as follow. Direct district aggregation in columns 1. Aggregation of industries ignoring the between-industry misallocation term in column 2. Direct district aggregation taking out national industry allocation weight by local shares in column 3. National industry shares in column 4. Direct district aggregation with factors re-weighted by their production function coefficients in column 5. Misallocation computed from OLS productivity estimates with a constant return constraint in column 6, from OLS productivity estimates with three factors in column 7, and from OLS productivity estimates with two factors in column 8.

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

Table 7: Output per worker and misallocation, 1989-2010 for alternative specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: log value added per worker								
Explanatory variable: misallocation for,								
Land/buildings	-0.37^a	-0.39^a		-0.40^a	-0.22^a	-0.26^a		-0.27^a
	(0.089)	(0.075)		(0.072)	(0.049)	(0.045)		(0.041)
Other assets	-0.18^b	-0.11^c		-0.13^b	-0.070^c	-0.14^a		-0.15^a
	(0.071)	(0.059)		(0.056)	(0.037)	(0.034)		(0.032)
Employment	0.21^b	0.091		0.065	-0.079	-0.12^b		-0.14^a
	(0.089)	(0.072)		(0.073)	(0.065)	(0.046)		(0.046)
TFP			0.31^b	0.60^a	0.19^a		-0.098	0.60^a
			(0.13)	(0.12)	(0.061)		(0.20)	(0.12)
Other explanatory variables:								
Output					0.45^a			
					(0.016)			
Land/buildings						0.27^a		0.27^a
						(0.040)		(0.039)
Other assets						0.32^a		0.32^a
						(0.026)		(0.025)
Employment						0.21^a		0.19^a
						(0.024)		(0.023)
TFP							0.46^b	-0.0092
							(0.19)	(0.12)
Fixed effects	N	district	district	district	district	district	district	district
Observations	1,576	1,576	1,576	1,576	1,576	1,575	1,576	1,575
R-squared	0.14	0.71	0.66	0.72	0.90	0.89	0.66	0.90

Notes: OLS regressions with year effects in all columns. Clustered standard errors (by district) in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

better factor allocation are perhaps even larger.[20] [21]

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 al-

[20] 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.

[21] 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.

located, 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 decileŮarguably still well below perfect efficiency and most likely well below the U.S. averageŮis 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 basiASIC 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 policy changes like the 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 efficiency of factor input measured by Bartelsman *et al.* (2009) across 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 canŠt 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* 104(6):1667–1697.

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.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell. 2016. How do electricity shortages affect industry? Evidence from India. *American Economic Review* 106(3):587–624.

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

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. Dynamic inputs and resource (mis) allocation. *Journal of Political Economy* 122(5):1013–1063.

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

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2009. Measuring and analyzing 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. 2002. India - Urban Land Reform. Processed, available at www.alain-bertaud.com.

Bertaud, Alain. 2011. Mumbai FAR/FSI conundrum. Processed, available at www.alain-bertaud.com.

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.

David, Joel, Hugo Hopenhayn, and Venky Venkateswaran. 2016. Information, misallocation, and aggregate productivity,. *Quartely Journal of Economics* 131:forthcoming.

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

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.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr. 2012. Is India's manufacturing sector moving away from cities? Working Paper 17992, National Bureau of Economic Research.

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

Hopenhayn, Hugo. 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5):1127–1150.

Hopenhayn, Hugo. 2014*a*. Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics* 6(0):735–770.

Hopenhayn, Hugo. 2014*b*. On the measure of distortions. Processed, UCLA.

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

Joshi, Vijay and Ian M. D. Little. 1991. *IndiaŠs Economic Reforms: 1991Ű2001*. New York (NY): Oxford University Press.

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.

Morris, Sebastian and Ajay Pandey. 2009. Land markets in India. Distortions and issues. In Nirmal Mohanty, Runa Sarkar, and Ajay Pandey (eds.) *India Infrastructure Report. Land - A Critical Resource for Infrastructure*. New Delhi: Oxford University Press, 13–19.

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. 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11(4):707–720.

Restuccia, Diego and Richard Rogerson. 2013. 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. Thesis, Massachusetts Institute of Technology, Department of Urban Studies and Planning.

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. 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):2482–2484.

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

Appendix A. Data issues

A. 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:

$$ror_{it} = \rho_i + \eta_t + \epsilon_{it}\,,$$

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

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

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:

$$T_{jit} = ror_{it} \times rent_{jit}\,,$$

where T_{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:

$$T_{jit} = ror_{Ti} \times rent_{Tjit},$$

where K_{jit} is the value of asset, $K \in \{\text{land,buildings}\}$ hired by establishment j in district i at time t and $rent_{T_{jit}}$ is the corresponding rent for asset K, ror_{T_i} is the rate of return on asset T in district i. The value of asset K used in the production function is then simply the sum of the asset T 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.

B. Data clean-up

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 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 8 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. Two features 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 consists of 145,829 observations from the organised sector

Table 8: Descriptive statistics

	1989	1994	2000	2005	2010
Panel 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
Panel 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
Panel 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
Panel 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
Panel 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
Panel 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

Notes: Descriptive statistics taken from ASI and NSSO.

and 575,989 observations from the unorganised sector.

Appendix B. 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 ŞpushŤ 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 IndiaŠs 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)Ůa form of land readjustment system similar in its principle to the ones that had been used in Korea, Taiwan, and GermanyŮhave 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.