Efficiency and Equity of Land Policy in Developing Country Cities: Evidence from the Mumbai Mills Redevelopment*

Michael Gechter† Nick Tsivanidis‡

November 14, 2017

Please find the latest version HERE

Abstract

Land policy in developing country cities places significant restrictions on formal sector developers but often fails to reign in informal development. To what extent does this pattern reduce city efficiency, and how are the effects shared between the rich and poor? We address these questions in three steps. First, we exploit a unique natural experiment in Mumbai that led 15% of central city land occupied by the city's defunct textile mills to come onto the market for redevelopment in the 2000s. Second, we use a "deep learning" approach to measure slums from satellite images, and combine this with administrative sources to construct a uniquely spatially disaggregated dataset spanning the period. Third, we develop a quantitative general equilibrium model of a city featuring formal and informal housing supply to guide our empirical analysis. We document a large increase in the supply of formal construction on mill sites, and find substantial spillovers on nearby locations that led slums to redevelop into formal housing. Our findings suggest that land policy can reduce the efficiency of developing country cities by misallocating land away from its optimal use, but policies that promote formal housing supply may have unintended consequences for equity by reducing the stock of relatively affordable housing in slums.

*Preliminary. We wish to thank Russell Cooper, Jonathan Eaton, Pablo Fajgelbaum, Jon Hersh, Chang-Tai Hsieh, Dan Keniston, Kala Krishna, Vidhadyar Phatak, Debraj Ray, Steve Redding, Jim Tybout and participants at several workshops for helpful comments. We could not have collected the data needed for our analysis without invaluable assistance from Vikas Dimble, Sahil Gandhi, Dilip Mookherjee, and Pronab Sen. We are grateful to Jordan Graesser for sharing computer code to extract spatial features from satellite images. Han Chen, Inhyuk Choi, JD Costantini, Suraj Jacob, Peiquan Lin, Alejandra Lopez, Rucha Pandit, Pankti Sanghvi, Christine Suhr, Sneha Thube, and Nishali Withanage provided excellent research assistance.

†Department of Economics, The Pennsylvania State University. Email: mdg5396@psu.edu.

‡Booth School of Business, University of Chicago. Email: ntsivanidis@uchicago.edu.

1 Introduction

In developed country cities, businesses concentrate in central business districts (CBDs) and residential housing is built at higher densities in central areas. This efficient structure increases the productivity of firms (who benefit from agglomeration) and reduces commute costs (by allowing more individuals to live close to jobs). Cities in developing countries look very different: they are often characterized by low-density sprawl, with slums prominent in central areas often interspersed between office blocks (Henderson et. al. 2016). While this structure is likely inefficient, it may benefit the poor who live in slums by providing access to jobs in the CBD. Observers suggest this spatial organization occurs in part because de jure land policy combined with limited enforcement prevents the allocation of land to its most efficient use.[1] However, establishing this link is challenging since it is rare to observe large-scale changes in land policy and hard to measure outcomes at small spatial scale in developing country cities where such data are rarely available.

This paper assesses the effects of land policy on efficiency and equity in developing country cities. In doing so, we make three contributions to the understanding of land market regulation and city structure in developing countries. First, we document the effects on land value, land use, and employment of a uniquely large-scale policy change in Mumbai that substantially increased the ability of the owners of the city's defunct textile mills (once occupying 15% of central city land) to redevelop. Second, to our knoweldge we provide the first use of a "Deep Learning" approach to identifying slums from satellite images. We demonstrate this provides a dramatic improvement in quality of slum measurement over frontier methods. Third, we develop a novel quantitative general equilibrium model of a city featuring formal and informal housing and use it to guide our empirical analysis of the effects of the change in policy on city structure.

We have two main results. First, we find that the policy change led to a large direct effect on the supply of tall, formal buildings on mill sites. This supports the idea that land policy can drastically misallocate land away from its efficient use. Second, we find substantial spillovers on surrounding locations: the price of formal properties rose, and slums were redeveloped into formal buildings. This provides some of the first evidence of housing market spillovers within developing country cities, and highlights that they can have unintended consequences for low-skill residents. We also show that employment increased near mill sites, suggesting the policy increased productivity in the presence of spillovers by increasing employment densities in central locations. Our findings suggest that restrictive land use policies can promote an inefficient city structure, but governments should be aware that efforts to promote high densities of formal housing may have unintended consquences for equity by reducing the stock of nearby informal housing.[2] The results from structurally estimating our model results (forthcoming) will allow us to assess the equity-efficiency trade-off of slums in general, and the impact of policies such as public housing construction on previous slum sites on the equity implications of pro-formalization policies.

[1] e.g. Bertaud (2011), Ellis & Roberts (2016).

[2] This statement holds under the assumption that poor are more likely to live in slums and are less than perfectly compensated to move out when slums are converted. We provide empirical results in support of the former and institutional context in support of the latter.

Observers have consistently criticized urban land policy in India as inefficient, failing to capture the potential benefits of agglomeration and promoting low-density cities characterized by slums, unused land, congestion and sprawl (Bertaud, 2011; Bertaud & Brueckner, 2005; Brueckner & Lall, 2015; Harari, 2016). While India remains of independent interest as the country with the world's second highest urban population (UN, 2015), examples of restrictive land policy and similar city structure can be found in cities across Asia and Sub-Saharan Africa (Ellis & Roberts, 2016; Henderson et. al. 2016). However, due to a lack of data and natural experiments it is challenging to identify the contribution of distortionary land use regulations on these spatial configurations.

In this paper, we exploit a unique change in land policy in the center of one of the world's largest megacities. Mumbai's 58 textile mills, covering 15% of land in the central city (602 acres), hark back to a time of industrial prominence during the late 18th and early 19th centuries. However, as the industry declined after World War II and the city turned increasingly to employment in services, the city's land use regulations effectively prohibited redevelopment of the mills. As we describe in detail in section 3, this changed in the 2000s when these laws were amended to greatly improve the ability of owners to sell and develop these sites.

Evaluating the effects of policies on outcomes such as land use, property prices, residential populations and employment within developing country cities is typically complicated by a lack of data at small-scale geographies. We confront this by constructing a number of new datasets. To our knowledge, we provide the first application of deep learning via a convolutional neural network (CNN) to identify slums using satellite images from 2001 and 2016. We provide the CNN with a training dataset consisting of a satellite image of suburban Mumbai and a shapefile of identified slums.[3] The network learns the features of the image that best predict the status (slum or non-slum) of pixels; we use the trained network to predict the locations of slums in Mumbai city in 2016 and 2001. The predicted slums line up extremely well with identified slums in our validation sample (R^2 of 0.94), about 2.5 times better performance than frontier methods used in the literature (R^2 of 0.36).

We supplement this with a number of additional administrative data sources. First, we digitized from records from the Maharashtra Department of Registrations of Stamps that provides values of formal residential and commercial floorspace across more than 900 subzones in the city between 1994 and 2012. Second, we geolocate all formal establishments with more than 10 workers using addresses provided in the 2005 and 2013 Economic Census. In ongoing work, we obtain employment by the universe of all establishments in these years across approximately 8,000 enumeration blocks by digitizing hand drawn maps for each block obtained from the Census Office and the Ministry of Statistics. Third, we construct population totals by scheduled caste and scheduled tribe (SCST) and non-SCST (which we interpret as proxies for low- and high-skilled workers) in 2001 and 2011 across 220 election wards obtained from the Bombay Municipal Corporation Election Office. Taken together, these datasets allow us to track changes in slums, formal property prices, employment and residence across the city over time.

To guide our empirical analysis, we develop a new quantitative general equilibrium model of a city. Low-

[3] Slums are identified based on built structure quality by the Slum Rehabilitation Authority of Mumbai in 2016.

and high-skill workers decide where to live and work. Formal and informal firms produce across the city using floorspace and labor. Some locations are more productive than others, and we allow for productivity spillovers so that a location's productivity can depend on the density of nearby employment. A key feature of out model is that we allow for both formal and informal housing. Land owners in each location choose between building formal and informal units based on their relative price. Formal buildings are attractive both because they can command a higher price per unit of floorspace (through higher amenities provided to residents) and because their construction technology allows for taller units to be built on a unit of land. However, formal development is subject to distortions which summarize the many de jure and de facto restrictions faced by developers. We allow for spillovers in residential amenities so that utility from living in a neighborhood depends on the density of formal housing in surrouding locations.

We model the law change in Mumbai as the lifting of prohibitively high barriers to redeveloping mill sites. Due to their central location, the model predicts that owners will develop tall, formal buildings. In the presence of spillovers in amenities, this increases the relative value of formal floorspace nearby, causing redevelopment of slums. Some of the new floorspace on mill sites is used for commercial purposes, increasing the employment density in nearby locations. In the presence of spillovers in productivity, employment rises nearby as a result of increased productivity. Log-linearizing the model's equilibrium conditions, we derive regression equations that decompose these effects into a portion due to the direct effects of mill sites within a location as well as an indirect effect due to spillovers from mill sites nearby. These specifications resemble others used in the literature not derived from general equilibrium theory (e.g. Autor, Palmer and Patak 2013). We take them to the data to provide evidence of the policy's effects along these channels as well as to identify the presence of spillovers.

Our identifying assumption is that trends in unobservable amenities and productivities were uncorrelated with a location's proximity to mill sites, within administrative wards (24 large neighborhoods of the city) and conditional on characteristics such as distance to the CBD. While we believe this is reasonable given the uncertainty around the legality of the regulatory change, we provide evidence in support of this hypothesis by examining the impact of proximity to mill sites on the price of formal properties between 1994 and 2012.[4] We find no differential trends in prices in the years leading up to the law change, supporting our assumption.

We have four main empirical findings. First, we find that the law change caused almost all of the mill sites to develop into tall, formal buildings by 2016. This suggests the land had been previously misallocated. Second, we show that the price of formal property increased in surrounding locations. Through the lens of the model, this is rationalized by the presence of spillovers. Third, the share of land occupied by slums falls in nearby locations, consistent with the increased return to formal development in these locations. Fourth, we find that employment rises near mill sites which the model rationalizes through the presence of spillovers in productivity. Results from our structural model are forthcoming.

Taken together, our results suggest that restrictive land use policies reduce the efficiency of cities by mis-

[4] This is the only variable available to us in years preceding the policy change.

allocating land away from its optimal use. However, the substantial spillovers we document, especially those specific to developing countries through the conversion of slums into formal buildings in nearby locations, suggest that policymakers should be aware of the potential effects on equity from changing regulations. More generally, our methodology to measure slums with high accuracy by combining CNNs with satellite imagery provides a way to improve the measurement of city structure in developing countries in future work.

In forthcoming results, we structurally estimate the model and use it to quantitatively evaluate the impact of slums and land policy. First, we simulate the effect from replacing all slums with formal housing in the center city. The change in city output as well as welfare across low- and high-skill workers map directly to the general impact of slums on efficiency and equity. Second, we evaluate how policies such as the construction of public housing on previous slum sites can impact the equity implications of pro-formalization policies.

The rest of the paper proceeds as follows. Section 2 discusses the paper's contribution to the literature. Second 3 presents the context of Mumbai and its textile mills. Section 4 describes the data and the methodology we use to measure slums from satellite images. Section 5 develops the model and section 6 outlines the reduced form framework it delivers and presents our results. Section 7 outlines the structural estimation and quantitative results, while section 8 concludes.

2 Related Literature

Our paper makes several contributions to the literatures on urban and development economics.

In development, we contribute to the literature on slums in developing countries (see Bruecker and Lall 2015 and Marx, Stoker and Suri 2013 for reviews). Most have focused on titling (Field 2007), slum upgrading (Field and Kremer 2008, Harari and Wong 2017) and relocation from slums to public housing (Barnhardt, Field and Pande 2017). More closely related is Henderson, Venables and Regan (2017) who study the development of Nairobi as slums transition to formal construction with the growth of city over time. Our paper differs along three key dimensions. First, we open up the "black box" of development frictions that might constrain formal developers using a large real-world change in land use policy. Second, we measure slums using CNNs which provides a general methodology to predict slums from satellite images alone compared to tracing the perimeters of each informal building. Third, we develop a quantitative model of city featuring informal and formal development that can be fit to any arbitrary two-dimensional geography, and use the regression equations implied by the model to guide our analysis of the change in policy.

Within a large body of work on the effects land use regulations in developed countries, only a small strand directly addresses the endogenity of land use regulation (e.g. Ihlanfeldt 2007; Zhou, McMillen, and McDonald 2008; Saiz 2010; Turner, Haughwout, and Van Der Klaauw 2014). Most highlight how regulations have the potential to distort spatial organization, with implications for aggregate performance (e.g. Hsieh and Moretti 2017). However, evidence for developing country cities is much sparser. While many highlight a potential role played by land policy in promoting sprawl and inefficient structure in Indian cities,[5] this literature has relied on cross-sectional comparisons across cities. In contrast, we combine a natural experiment in the form of a large policy change with new sources of spatially disaggregated data in Mumbai to provide evidence on the causal link between land policy and city structure in developing country cities. Similar to the literature in developed countries (e.g. Rossi-Hansberg, Sarte and Owens 2010), we show that housing market spillovers operate in developing countries and can have transformative effects through the redevelopment of informal housing absent in richer country settings.

In order to measure changes in the locations of slums, we turn to the use of satellite images. While a relatively large literature within economics has used nighttime satellite imagery to address missing data on economic activity at the region or city level,[6] a much smaller but growing strand has turned to daytime satellite images that provide high enough resolution to measure outcomes within cities. The downside to working with these images is that the volume of data they contain mean that they are less interpretable. Often, authors extract single features from these data (e.g. green space, luminosity) that are readily interpretable within the specific context.[7] However, such methods are not amenable to the measurement of land use since a single index is unlikely to capture sufficient information. Recently, papers in geography and economics have begun to incorporate potentially large sets of spatial features extracted from satellite images to identify land use, often using machine learning methods such as random forests to combine the features into a single predictive model (see Goldblatt, You, Hanson, and Khandelwal (2016) for an example at the intersection of the two fields, classifying land use as built or non-built). Kuffer, Pfeffer, and Sliuzas (2016) provides a review papers in geography using this approach to identify slum cover. However, we find these methods insufficiently accurate to conduct analysis at the city block level. Instead, we provide the first application of a convolutional neural network (CNN) (also known as "deep learning") for slum identification, making use of insights from Yuan (2017) to our problem.[8] The CNN can be thought of as generating spatial features tailored to the identification of slums (Goodfellow, Bengio, and Courville (2016)), and we see a large increase in predictive performance from its use.

Lastly, we contribute to the growing literature using quantitative models to study the internal structure of cities (Ahlfeldt et. al. 2015; Allen et. al. 2015; Owens et. al. 2017; Tsivanidis 2017). Our approach is most closely related to Redding and Sturm (2016) who combine a model with multiple worker groups with the wartime bombing of London to evaluate spillovers from the destruction of neighboring houses. We differ from the authors' approach by developing a model featuring both informal and formal housing supply to fit our developing country context, and by log-linearizing the model's equilibrium equations to provide regression

[5] See Bertaud and Brueckner, 2005, Brueckner and Sridhar, 2012, Duranton et. al. 2015, Harari 2015, Glaeser, 2011, Sridhar, 2010.

[6] See Donaldson and Storeygard (2016) for a review.

[7] For example, Marx, Stoker, and Suri (2014) measure the luminosity of roofs in slums in Nairobi, Burgess et. al. (2012) use the color bands to measure deforestation in Indonesia, and Henderson, Regan, and Venables (2016) use LiDAR images to measure building height in addition to manually tracing building footprints.

[8] There are relatively few applications of CNNs in economics to date. Jean, Burke, Xie, Davis, Lobell, and Ermon (2016) and Engstrom, Hersh, and Newhouse (2016) are two exceptions.

equations that we take to the data to shed light on the channels predicted by the model and inform us about the presence of spillovers.

3 Background: The Redevelopment of Mumbai's Textile Mills[9]

Mumbai's 58 textile mills, covering 15% of land in the central city (602 acres), hark back to a time of industrial prominence during the late 18th and early 19th centuries. However, as the industry declined after World War II and Mumbaikers turned increasingly to employment in services. The coup de grace for most of the mills came during an 18-month strike spanning 1982 and 1983. Many closed. The government of the state of Maharashtra did not allow sale of the mill lands, however, considering them an asset securing payment of workers' wages lost during the strike.

During the 1991 economic reforms, the state government was pushed to allow development on unused mill lands subject to government approval. In response, the state government drafted Regulation 58 of the city's Development Control Rules (DCR 58). DCR 58 stipulated that any mill requesting permission for redevelopment turn over one-third of the land in question to the Maharashtra Housing and Development Authority (MHADA) for construction of public housing and one-third to the Bombay Municipal Corporation (BMC) for development of public open space. Mumbai's low level of open space per capita relative to cities like Delhi and Madras explains some of the enthusiasm for using mill land for this purpose. There was also hope that some of the space could be used to widen roads and alleviate congestion.

What amounted to a 66% tax on land did not prove attractive to mill owners and relatively little redevelopment took place during the 1990s. In 2001 the state government ammended DCR 58 so that mill owners could deduct the floorspace represented by existing structures from the area subject to the 66% tax. In many cases this had the effect of reducing the amount of land to be given up to almost zero. The various authors articles in D'Monte (2006) offer different stories for the origins of the DCR 58, with different roles played by mill owners. It would be naive to believe that mill owners played no role in pushing for the change, but we show in Section 6.1 that there is no evidence from differential land prices that either the mill owners or the government made anticipatory investments in mill areas. Given the ammendment's subsequent tortured legal history, it may be that owners or potential buyers believed the ammendment had little chance of becoming law and being implemented.

Despite passing in 2001, the ammendment was not implemented until 2003 when a specific formula to compute the amount of land to be surrendered was issued, meaning that no development was allowed to take place from 2001 to 2003. Even once the formula was available, a group of NGOs was able to obtain a stay on the implemention from the Bombay High Court starting in 2005. The mill owners appealed the decision to the Indian Supreme Court, which finally ruled in their favor in 2006 and allowed large-scale redevelopment to proceed. Analysis of satellite images over the decade suggests that this was the key decision that drove very

[9] We draw heavily on D'Monte (2006) for the historical background described in this section.

large-scale redevelopment. Appendix Table A.1 provides details on redevelopments for which floor counts are available. Many of the former mills are now the sites of some of the tallest buildings in Mumbai, representing dramatic increases in the amount of floorspace available in the center of the city.

As a motivating example, Figure 1 shows satellite images for the centrally located Apollo, Simplex and Hindoostan Mills in 2000 and 2016. Panel (a) shows these sites, outlined in blue, within the broader neighborhood. The mills themselves are long industrial buildings with tiled, corrugated roofs. We also see a large number of informal structures (with small, clustered roofs) near Apollo Mills in the top left. Panel (b) shows how these sites transformed by the 2016. There are three key takeaways from comparing these images. First, all undergo transformative development with very tall, formal buildings constructed in each site.[10] Second, new formal construction has taken place near mill sites. This can be seen at the bottom of Apollo Mills (top left), as well as above Simplex Mills (top right). Third, part of this new construction came from converting slums into formal housing (below Apollo Mills). A large section of slums have also been cleared at the center. However, not all nearby slums have been cleared. This suggests that a "stickiness" in the adjustment of land use, which could be due (in part) to conversion costs.

[10]Not every part of each mill site is developed. In parts of our analysis, we measure the new floorspace by combining satellite images with additional sources on the number of floors of skyscrapers in Mumbai.

Figure 1: Apollo, Simplex and Hindoostan Mills

(a) 2000

(b) 2016

4 Data

We construct a number of new datasets measuring the evolution of slum cover, floorspace prices, scheduled caste and tribe population, and formal establishment location within an India megacity. A number of these administrative datasets are at a level of spatial granularity previously unavailable to researchers as we describe in detail below. One key methodological contribution of this paper is that we bring new methods to bear on the problem of slum identification from satellite images, with substantial improvements in performance compared to previous efforts.

4.1 Measuring Slums from Satellite Images

Data on land use in developing countries is often incomplete or misreported in official statistics. This is particularly true for slum areas. In Mumbai, maps of informal settlements were completely non-existent before 2016. This represents a challenge for us, since formalization is a key developing-country specific response to spatial spillovers that we wish to measure.

To address the lack of historical land use data we turn to satellite images, which are available as far back as 2001. Identifying slum areas from a satellite image is a classification or regression problem where an observation is a spatial unit (such as a pixel representing 50 x 50 cm on the ground or a city block) and the outcome variable can either be an indicator for slum status or a continuous variable measuring, for example, the slum area of the spatial unit. The explanatory variables are the red, green, and blue values of pixels within the spatial unit and the pixels surrounding it.[11] The researcher fits a model of the outcome as a function of the explanatory variables in a "training" area outside the area of interest. Slum identification involves plugging the explanatory variables of spatial units in the area of interest into the estimated model to generate the predicted slum status for each spatial unit in the area of interest. The problem is extremely high-dimensional. For instance, a vector of the red, green, and blue values forming a 128 x 128-pixel square around a pixel of interest has $3 \times 128^2 = 49,152$ elements.

To date, the literature in geography on using satellite images to identify slums[12] recommends reducing the dimension of the problem by performing specific, known calculations to extract a set of "spatial features"[13] from the image in a first stage, then regressing the slum measure of a spatial unit on the spatial features. Slum identification involves plugging the spatial features of the area of interest into the fitted second stage model to generate predicted slum status of each spatial unit in the area of interest.

In this paper, we take a different approach based on deep learning. Rather than use pre-specified procedures to generate spatial features, the deep learning approach fits a model made up of a series of parameterized linear

[11] Other values associated with surrounding pixels, such as near-infrared, may also be used.

[12] See Kuffer et al. (2016) for a review.

[13] These include standard features such as the Normalized Difference Vegetation Index (NDVI), which measures the difference between the amount of near-infrared and red light emitted, as well as "textural features" such as the Histrogram of Oriented Gradients (HoG), which measures the average change in image brightness vertically and horizontally at each pixel yielding higher values in more complex settled areas such as slums (Engstrom et al. (2016)). The dimension of the spatial features is usually in the double or low triple digits.

and non-linear transformations called a Convolutional Neural Network (CNN) directly to the high-dimensional explanatory variable vector and the outcome variable. The "deep" in deep learning refers to the potentially substantial number of linear and non-linear transformations employed. By directly using the explanatory variables, the CNN avoids the information loss resulting from aggregation to pre-specified spatial features.

We proceed as follows. First, we describe the inputs into our classification procedure. Second, we evaluate the "spatial features" methods and show their performance is insufficient for our purposes. Third, we describe our deep learning approach and show it performs remarkably well in measuring the location of slums in our verification sample.

Defining and measuring slum status in the training area To construct our outcome variable, we first need to define what constitutes a slum before mapping their location in the city. We adopt a building structure-based definition, which is most natural given our use of satellite imagery for measurement.[14] Our data come from two sources. First, the Slum Rehabilitation Authority of Mumbai conducted a survey of all slums in the city in 2016 and provide a map of the results which we digitized. The SRA designated slum status based on the type of built structures and available amenities, and thus their slum map captures both "notified" (i.e. recognized by the government) and "non-notified" slums.[15] Second, we augment this with enumeration block maps from the 2011 census we were able to purchase from the census office. These contain the location of individual buildings across the city, which are classified as "pucca" or "kutcha" structures, which group the materials of the wall and roof into two categories. Roughly speaking, these categories correspond to what one might think of as slum and non-slum.[16] While we find the SRA map to have the best coverage of slum-like buildings, perhaps unsurprisingly since the survey was conducted for this sole purpose, we cross-check the two sources with our satellite images to resolve any inclusion/exclusion errors using the census maps.

Results using existing approaches We first implement the spatial characteristics approach described above by following one of the frontier papers in the geography literature, Engstrom, Sandborn, Yu, Burgdorfer, Stow, Weeks, and Graesser (2015). We use 32 spatial features, comprising all outputs from the Fourier Transform, HoG, Lacunarity, Local Binary Patterns, NDVI, and PanTex extraction procedures[17] implemented using Jordan Graesser's `spfeas` Python package on 8 x 8 and 16 x 16-pixel grids around each image pixel. Engstrom et al. (2015) define the spatial unit of interest as a 50 x 50 cm pixel, but we acheived better performance by taking the average of each spatial feature within city blocks defined as the complement of the intersection of the city

[14] We recognize there are other ways to define slums, for example based on whether residents have tenure.

[15] SRA uses a topographical survey based on satellite images and LiDAR to first map out the location of slums across the city before visiting the clusters with enumerators. Their built-structure based approaches corresponds well with what we seek to measure in this paper. See Dhikle et. al. (2017) for a description.

[16] From the census enumerator guide: "A Pucca building may be treated as one which has its walls and roof made of the following materials. Wall materials: Stones (duly packed with lime or cement mortar), G.I/metal/asbestos sheets, Burnt bricks, Cement bricks, Concrete. Roof Material: Machine-made tiles, Cement tiles, Burnt bricks, Cement bricks, Stones, Slate, G.I./Metal/Asbestos sheets, Concrete. Buildings, the walls and/or roof of which are predominantly made of materials other than those mentioned above such as unburnt bricks, bamboos, mud, grass, reeds, thatch, plastic/ polythene, loosely packed stone, etc., may be treated as Kutcha buildings."

[17] See Engstrom et al. (2015) and Kuffer et al. (2016) for descriptions of each procedure.

footprint and the space taken up by primary, secondary, tertiary, and unclassified roads on OpenStreetMap. Like Engstrom et al. (2015), we use Breiman (2001)'s Random Forest algorithm for the second-stage model. We use the area of each block that is covered by slums according to the SRA map as our dependent variable.

The Mumbai metropolitan area is split up into two districts: Mumbai District and Mumbai Suburban. Almost all of the mill sites were in Mumbai District so this is our main area of interest. We use Mumbai Suburban as our training area. In 2016, we can assess the performance of a given image-based slum identification method by comparing the slum cover of Mumbai District blocks predicted by the method to the actual slum cover according to the SRA map. Figure B.2 plots the results of this exercise, with the points representing blocks and the fitted line the result of regressing the blocks' SRA slum area on the predicted slum area of the spatial features procedure (spfeas). The R^2 of the fit, a common evaluation metric in the literature, is given at the bottom of the graph.

The R^2 is comparable to the 0.45 value reported in Engstrom et al. (2015).[18] This level of measurement error, however, is inadequate for our purposes since we are interested in the relatively subtle task of determining how the change in slum cover between 2001 and 2016 was affected by proximity to former mill sites within neighborhoods.

Our Approach using Deep Learning To improve performance, we apply new methodology to the problem of identifying slums from satellite imagery. The approach is based on a CNN architecture. As described above, a CNN transforms potentially high-dimensional inputs into predictive outputs through a parametrized sequence of linear and nonlinear transformations. A standard logit model, which is linear in parameters after a nonlinear transformation, is analogous to the final two steps in the sequence when the dependent variable is binary. The difference is that the input to the logit model in a CNN would already be a transformation of the vector of red, green, and blue values of pixels surrounding the pixel of interest. The fact that many of the transformations in a CNN are linear provides computational tractability despite the thousands of parameters involved[19], and can easily be parallelized. Due to these advantages and the growing availability of multi-processor computing environments, CNNs took the world of image processing by storm in 2012 and have enabled many of the technologies associated with artificial intelligence such as self-driving cars (see Goodfellow et al. (2016) and Gershgorn (2017) for relatively non-technical histories).

We depart further from existing approaches by adapting Yuan (2017)'s building footprint detection method to define the dependent variable at the pixel level to be a signed distance from the boundary of a slum area. That is, Y_i for pixel i is an integer representing the number of pixels between i and the boundary of the nearest slum, with the integer being negative if i is on the interior of a slum. Figure B.3 demonstrates the approach. Panel a shows a part of the training area containing large slum (small, low structures) and non-slum (taller buildings)

[18]The difference in performance is perhaps unsurprising given that Engstrom et al. (2015) run their analysis only on built-up areas of Accra, while we consider all areas of Mumbai district. We note that predicting the share of a given block covered by slums results in much worse performance, acheiving an R^2 of only 0.02.

[19]Our own CNN includes about 600,000 parameters.

areas. Panel b shows signed distance from the boundary of an SRA slum. The lightest areas are deepest in the interior of a slum. Following Yuan (2017), we implement our CNN using the Theano framework in Python.

The results are remarkably good. Qualitatively, Figure 2 compares how the algorithm's predicted slum locations compare with the data. Panel (a) overlays the slum shapefile data over a satellite image of an area near the center of the city. The large red area is the well-known Dharavi slum (featured in Slumdog Millionaire). Panel (b) overlays the predicted slum areas according to our CNN over the same section of the city, slums are denoted in red (with blue outlines representing predicted boundaries). Notably, the CNN picks up small pockets of structures and is also able to distinguish the formal structures within Dharavi that are not slums. Quantitatively, our CNN offers a dramatic improvement over existing methods. Figure 3 reproduces Figure B.2, with CNN predictions in place of spfeas predictions. A linear regression of the SRA slum cover area per block on our CNN predictions explains 94% of the variation in block-level slum cover. We note that we overpredict for blocks with large slum areas, but see this as a virtue of our approach we believe the SRA may be biased against including non-notified slums in their maps.

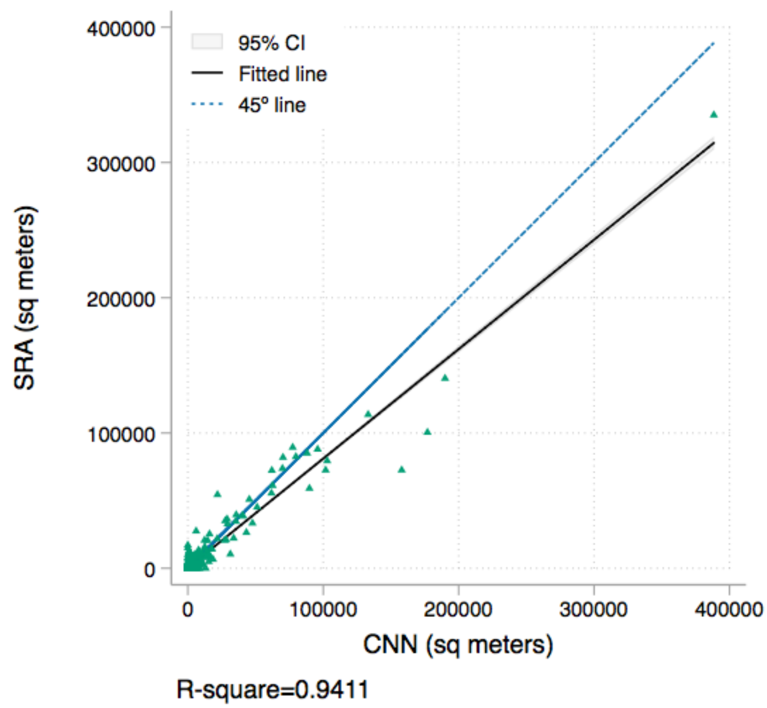
Figure 2: Out-of-Sample Prediction vs Actual Slums

(a) SRA Slums

(b) Predicted Slums

For earlier years, we lack primary data sources on the location of slums and therefore adjust our 2016 SRA map by overlaying it on top of images from 2001 and 2005 to identify structures in our training sample that

Figure 3: Validation Sample: SRA vs CNN Slums



have remained unchanged. We then extend the map to slum structures existing in early years but not 2016 by identifying structures in the images which looked identical to those present in both years. We produce validation data for 2001 in the same way. We do not yet acheive quite the same performance in the 2001 validation data (R^2 of 0.74), but still do far better than existing approaches.

4.2 Additional Adminstrative Data Sources

In addition to the satellite images, we have collected a number of new data administrative sources at a high level of spatial granularity within Mumbai.

Our employment data comes from the Fifth and (newly available) Sixth Economic Census (EC) of India. This covers the universe of establishments in 2005 and 2013 respectively. While the raw data provides the block in which each establishment is located (out of around 8,000), the Ministry of Statistics does not make their spatial location available to researchers. However, our team has been able to access enumeration block maps for both rounds covering Mumbai district. These are hand-drawn by the enumerator assigned to each block during the survey. Each map contains a sketch of the roads (labelled), landmarks and buildings within the block. Our research assistants are finishing geocoding each of these maps. Therefore, the results we present in this draft geolocate firms using the addresses of formal establishments with more than 10 workers provided in the EC. We clean and geolocate these addresses using the Google Maps API.[20] We experimented with

[20]The address formats are irregular. Therefore, we parse the addresses and extract their components (ie, street, building, postal code) using the natural language processing C library libpostal, which has been trained on OpenStreetMap data.

different combinations of address components and benchmarked each against a random sample of addresses we located manually. This gives us a distribution of error for coordinates found using each method, which our final longitude-latitude combinations minimize.

We construct data on residential population using the 2001 and 2011 censuses. As before, geographic coverage is typically only available by town. However, our team was able to access the block-level datasets which report the total number of residents as well as residents in scheduled castes and scheduled tribes (SCST). We use the latter as a proxy for skill. We obtained the enumerator maps for each block in 2011. Unfortunately, the maps for 2001 have been destroyed. Instead, we accessed population in both years from the Bombay Municipal Corporation Election Office for 221 election wards across the city. These report both total and SCST population and are created by the Election Office (EO) using the block-level data from the population census.[21]

Data on floorspace values come from annual official assessments produced by the city. These are published each year in the city's Ready Reckoner for around 900 "subzones" across Mumbai and Mumbai suburban districts. We digitized paper copies of the assessments going back to 1994.[22] While these official assessments are partially based on transaction records, we are currently collecting plot-level transaction records available from the government of Maharashtra between 2002 and 2015. We acknowledge the problematic nature of working with formal sector property price data in the Indian context, where transaction prices may be underreported for tax avoidance, but believe that city assessments should be less prone to this issue.[23] Moreover, this force would tend to exert a downward bias on house price appreciation near mill sites, implying our estimates should be more conservative than otherwise.

We recover the locations of mill sites primarily by georeferencing a map from the Correa Report, a 1996 government-commissioned plan for the mill lands. Finally, we use commuting microdata from the 2005 Mumbai Municipal Corporation transportation survey.

5 Model

To guide our analysis, this section presents a general equilibrium model of a city. The key difference with existing quantitative urban models (e.g. Ahlfeldt, Sturm, Redding, and Wolf (2015)) is that we incorporate informal housing and informal employment. The model serves two purposes. First, it disciplines our empirical analysis by providing regression specifications we take to the data. Second, it provides a framework to undertake quantitative analysis once we estimate its structural parameters in section 7.

The city consists of a discrete set of locations $i \in \{1, \ldots, I\}$. Individuals are mobile and choose where to live and work. In each location, there can be both formal and informal housing whose supply is decided by

[21]The EO ombine the block-level maps and data to district the city into electoral zones. Unfortunately, while the EO does keep a record of historical population totals they construct by election ward, they did not retain the 2001 enumeration blocks maps.

[22]Unfortunately, while the present day Ready Reckoner prices are available online, historical records had to be obtained in paper copy in Mumbai and only certain years were made available to our research team. We therefore have data covering 1990-2000, 2003 and 2013. We drop the three initial years due to data reliability issues.

[23]We are in process of validating Ready Reckoner valuations for 2016 with property price data collected online.

landowners. Firms in each location produce using commercial floorspace and labor. Locations differ in their attractiveness for individuals and firms, which is determined in equilibrium by productivity, amenities, land supply and commute costs. We consider the DCR 58 law change as a shock to the supply of land available for development.

5.1 Workers

There is a fixed number of high- and low-skilled residents in the city denoted by $\bar{L}_g$ for $g \in G = \{H, L\}$.[24] Each individual chooses a location to live i and a location to work j. In each location, workers can live in either informal or formal residential housing denoted by $k \in \{I, F\}$. Each individual ω has Cobb-Douglas preferences over a freely-traded numeraire good and housing. Indirect utility for a choice (i, j, k) of where to live, where to work and which type of housing to reside in is given by

$$U_{ijk\omega} = \frac{u_{ig}^k w_{jg} (r_{Ri}^k)^{\beta - 1}}{d_{ij}} \epsilon_{ijk\omega}.$$

Here u_{ig}^k are the amenities enjoyed by type-g workers living in i in type k housing, w_{jg} is the wage earned by group g workers in location j, r_{Ri}^k is the price of residential housing of type k in location i, and $d_{ij} = \exp(-\rho t_{ij})$ is the iceberg disutility cost of commuting. We allow amenities to differ by the type of housing (to reflect differences in quality between formal and informal units) and by group (to reflect differences in preferences across skill groups for neighborhoods and housing types). We take these amenities to be exogenous for now, later we allow them to depend on neighborhood characteristics.

Individuals have an idiosyncratic preference for each (i, j, k) tuple $\epsilon_{ijk\omega}$ which is drawn iid from a Frechet distribution with unit scale and shape θ.[25] Standard results imply that the mass of workers living and working in different locations is given by

$$L_{Rig}^k = \lambda_g^U (u_{ig}^k (r_{Ri}^k)^{\beta-1})^\theta \Phi_{Rig} \tag{1}$$

$$L_{Fjg} = \lambda_g^U w_{jg}^\theta \Phi_{Fig} \tag{2}$$

where $\Phi_{Rig} = \sum_j (w_{jg}/d_{ij})^\theta$ reflects the access to jobs from location i and $\Phi_{Fjg} = \sum_{i,k} (u_{ig}^k (r_{Ri}^k)^{\beta-1}/d_{ij})^\theta$ reflects the access to workers from location j. The constant λ_g^U is determined in equilibrium and is invariant across locations.[26] Overall worker welfare $\bar{U}_g$ is given by

$$\bar{U}_g = \gamma \left[\sum_{i,j,k} \left(u_{ig}^k w_{jg} (r_{Ri}^k)^{\beta-1} / d_{ij} \right)^\theta \right]^{1/\theta}. \tag{3}$$

[24] This is the closed city assumption with infinite mobility costs between the city and the rest of the country. In quantitative exercises we consider the alternative extreme assumption of zero mobility costs so that population can move freely in and out of Mumbai (the open city assumption).

[25] For simplicity we assume this is constant across groups, but we relax this later.

[26] In particular, $\lambda_g^U \equiv \bar{L}_g (\gamma / \bar{U}_g)^\theta$ where $\gamma = \Gamma\left(1 - \frac{1}{\theta}\right)$.

Average income of residents of i is determined by the probability of commuting to different employment destinations conditional on living in i

$$\bar{w}_{ig} = \sum_j \frac{(w_{jg}/d_{ij})^\theta}{\sum_g (w_{jg}/d_{ij})^\theta} w_{jg}.$$

Housing market clearing then requires that the supply of housing is equal to the demand. Given supplies of formal and residential floorspace H^F_{Ri} and H^I_{Ri} and Cobb-Douglas preferences, this requires that

$$r^k_{Ri} = (1-\beta)\frac{\sum_g \bar{w}_{ig} L^k_{Rig}}{H^k_{Ri}} \tag{4}$$

5.2 Firms

In each location, firms produce the freely traded good under perfect competition. Some of these firms produce in formal buildings, while others produce in informal sites using a Cobb-Douglas technology over commercial floorspace and labor

$$Y_{jk} = A_{jk} L^\alpha_{Fjk} (H^k_{Fj})^{1-\alpha}$$

where A_{jk} is productivity in location j and housing type k, L_{Fjk} is the total labor used in production and H_{Fj} is the amount of commercial floorspace. We assume that worker skill-groups are perfect substitutes in production, but allow for differences in the units of effective labor provided by each worker type. In particular, we normalize the effective units provided by low-skill workers to one and assume that each high skill worker provides $Z_H > 1$ units of effective labor. Thus, $w_{jH}/w_{jL} = Z_H$ in all locations.

Taking wages as given, demand for labor from firms is given by

$$L_{Fjk} = \left(\frac{\alpha A_{jk}}{w_j}\right)^{\frac{1}{1-\alpha}} H^k_{Fj} \tag{5}$$

Zero profits for firms pin down the price of commercial floorspace from

$$r^k_{Fj} = (1-\alpha) A_{jk}^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w_j}\right)^{\frac{\alpha}{1-\alpha}} \tag{6}$$

5.3 Floorspace Use

For each type of floorspace use, a share ϑ_{ki} is allocated to residential purposes while the remainder is allocated to commercial use. Floorspace use decisions are made by land owners which pin down these shares through a no arbitrage condition. A tax equivalent of zoning restriction lowers the return to commercial use relative to residential use by a factor $1-\xi_i$. No arbitrage therefore requires that

$$\begin{aligned} \vartheta_{ki} = 1 \quad & r^k_{Ri} > (1-\xi_i) r^k_{Fi} \\ \vartheta_{ki} \in (0,1) \quad & r^k_{Ri} = (1-\xi_i) r^k_{Fi} \\ \vartheta_{ki} = 0 \quad & r^k_{Ri} < (1-\xi_i) r^k_{Fi} \end{aligned} \tag{7}$$

On informal plots we assume there is minimal government enforcement so that (with an abuse of notation) $\xi_i = 0$ on those plots.

Let r^k_i denote the price of floorspace in housing of type k. No arbitrage ensures this is equalized acorss uses. In locations completely specialized in residential use, this is simply r^k_{Ri}. The same applies for those specialized in commercial use. For mixed use locations, we denote $r^k_i = r^k_{Ri}$ and let $r^k_{Fi} = \frac{1}{1-\tau^k_i} r^k_i$. The return earned by land owners is simply r^k_i.

5.4 Housing Supply

Setup In each location i, there are T_i total units of land available. Land on these plots can be allocated between formal, informal and vacant use. There are a continuum of plots. Each plot is owned by an atomistic land owner who decides how to develop their land, taking prices and neighborhood characteristics as given. Since each land owner is small, they themselves have no affect on aggregate outcomes and there is no coordination between them.

Plot owners choose between the three types of development (formal, informal and vacant). Under the formal technology, the land owner combines land and capital according to the Cobb-Douglas technology $H^F_i = T^{1-\eta}_i K^\eta_i$, so that $h^F_i = k^\eta_i$ units of housing is constructed per unit of land if k_i units of capital are used per unit of land. Capital is available at the same price across the city, at a price normalized to one. When the formal technology is used, therefore, the land owner solves $\max_k r^F_i k^\eta_i - k_i$. This implies that $k_i = (\eta r^F_i)^{\frac{1}{1-\eta}}$. Using that $h^F_i = k^\eta_i$, we see that the resultant density of formal development is given by $h^F_i = (\eta r^F_i)^{\frac{\eta}{1-\eta}}$. Land owners face a tax on profits at rate τ_i. Profits per unit of land from formal development are therefore $\pi^F_i = \tilde{\eta}(1-\tau_i)(r^F_i)^{\frac{1}{1-\eta}}$ where $\tilde{\eta} \equiv \eta^{\frac{\eta}{1-\eta}}$.

Under the informal technology, single-story structures can be built using land as the sole input so that one unit of housing that can be produced per unit of land. Profits per unit of land from informal development are therefore $\pi^I_i = r^I_i$.

Plot owners have the outside option of leaving their plots vacant in which case they get a constant profit which we normalize to $\pi^V_i = 1$.

Land Use Decisions We assume each plot owner has an idiosyncratic profitability from each use represented by the vector $\epsilon = (\epsilon^F, \epsilon^I, \epsilon^V)$. In particular, this means that the land use allocation problem for an owner with profit shock vector ϵ is given by

$$\max \left\{ \pi^F_i \epsilon^F, \pi^I_i \epsilon^I, \epsilon^V \right\}.$$

We assume each ϵ vector is drawn iid from a Frechet distribution with shape parameter $\kappa > 1$. This implies the land use shares in i are given by

$$\lambda_{Fi} = \frac{\tilde{\tau}_i (r_i^F)^{\frac{\kappa}{1-\eta}}}{1 + \tilde{\tau}_i (r_i^F)^{\frac{\kappa}{1-\eta}} + (r_i^I)^\kappa} \tag{8}$$

$$\lambda_{Ii} = \frac{r_{Ii}^\kappa}{1 + \tilde{\tau}_i (r_i^F)^{\frac{\kappa}{1-\eta}} + (r_i^I)^\kappa} \tag{9}$$

$$\lambda_{Vi} = 1 - \lambda_{Fi} - \lambda_{Ii}. \tag{10}$$

where we have defined $\tilde{\tau}_i \equiv (\tilde{\eta}(1 - \tau_i))^\kappa$.

Given the results over construction density above, the total supply of formal and residential floorspace are given by

$$H_i^F = \lambda_{Fi} T_i \tilde{\eta} (r_i^F)^{\frac{1}{1-\eta}} \tag{11}$$

$$H_i^I = \lambda_{Ii} T_i \tag{12}$$

5.5 Equilibrium

We now define general equilibrium in the city.

Definition. Given vectors of exogenous location characteristics $\{T_i, u_{ig}^k, A_{jk}, t_{ij}, \xi_i, \tau_i\}$, city group-wise populations $\{\bar{L}_g\}$ and model parameters $\{\beta, \alpha, \rho, \theta, \kappa, \eta\}$, an equilibrium is defined as a vector of endogenous objects $\{L_{Rig}, L_{Fjg}, w_j, r_i^k, \vartheta_{ki}, \lambda_{ki}, \bar{U}_g\}$ such that

1. **Labor Market Clearing** The supply of labor by individuals (2) is consistent with demand for labor by firms (5),
2. **Floorspace Market Clearing** The market for residential floorspace for each housing type clears (4) and its price is consistent with residential populations (1), firms earn zero profits (6) and floorspace shares are consistent with landowner optimality (7),
3. **Land Use and Floorspace Supply** The share of land allocation to formal and informal use in each location (8-9) and the supply of floorspace on each type of land use (11-12) is consistent with landowner optimality.
4. **Closed City** Populations add up to the city total, i.e. $\bar{L}_g = \sum_{i,a} L_{Riag} \, \forall g$.

5.6 Introducing Spillovers in Amenities and Productivity

A long literature points to the importance of spillovers in cities.[27] We therefore relax the assumption that amenities and productivities of locations are exogenous. Equilibrium in this extension is defined analagously to the previous section, which we omit for brevity.

Amenities We allow amenities to depend on an exogenous component, as well as an endogenous component that depends on the surrounding density of formal housing

$$u_{ig}^{k} = \bar{u}_{ig}^{k} \left[\sum_{j} d_{ij}^{-\delta_U} (H_{Fi}/T_i) \right]^{\mu_{U.g}} \tag{13}$$

where $d_{ij} = \exp(-t_{ij})$, $\bar{u}_{ig}^{k}$ is the exogenous component of amenities, δ_U controls the rate at which amenities decay with commute times and $\mu_{U.g}$ reflects the overall preferences of type-g indvidiuals to live near formal housing.

When $\mu_{U.g}$ is large, group-g's preferences for residential neighborhoods depend a lot on the composition of housing nearby. We think of this as a reduced form way of capturing the different features of neighborhoods with lots of formal vs informal housing (e.g. cleaner streets, wider roads, larger retail space). These spillovers create linkages between residential locations across space, and will drive the model's predictions from the response in neighborhoods when locations nearby experience large increases in formal housing supply due to the DCR 58 change.[28]

Productivities Similarly, we allow productivity to depend on the density of surrounding employment. To keep things simple, we assume the common component depends on overall surrounding employment density and is common to formal and informal establishments

$$A_{jk} = \bar{A}_{jk} \left[\sum_{s} d_{js}^{-\delta_A} (L_{Fs}/T_s) \right]^{\mu_A} \tag{14}$$

Here μ_A controls the strength of productivity spillovers and δ_A controls the rate at which they decay with commute times. These spillovers create similar linkages between employment locations across space.

[27] This idea dates back at least to Adam Smith (1776), and was articulated more fully in Marshal (1890). Two prominent examples establishing this relationship are Ciccone and Hall (1996) using regional data and Ahlfeldt et. al. (2016) using intra-city data. See Rosenthal and Strange (2004) for a review.

[28] We could also model these amenities as depending on neighborhood composition (e.g. the high-skill ratio) rather than the density of formal housing. Then, if the high-skill prefer to live in formal housing DCR 58 will (indirectly) lead to similar changes in endogenous amenities by increasing the number of high-skilled residents on mill sites. Since our current method is simpler and more direct, this is what we pursue in this paper.

6 Using the Model to Guide Our Empirical Approach

In this section, we use our model to derive reduced form regression specifications to take to the data. Since we model the DCR 58 change as a shock to the supply of land across the city, our aim is to derive log-linear relationships between changes in the data we observe and land supply.

This serves two purposes. First, the results are directly informative about the sign and magnitude of key structural parameters including the strength of spillovers.[29] Second, we believe the reduced form relationships between exposure to mill sites and outcomes such as formal property prices, slum share, employment and residential demographics are of independent interest in understanding the effects of land policy on city structure.

We proceed by considering each outcome seperately, first deriving a specification from the model then by evaluating its predictions in the data. We make a number of simplifying assumptions and derive partial equilibrium relationships holding employment outcomes constant when considering residential outcomes and vice versa. This greatly simplifies the algebra without losing much intuition. We postpone a full general equilibrium analysis when all outcomes are interlinked to the quantitative exercises.

6.1 Formal Residential Floorspace Prices

Deriving Our Specification In this section, we (i) hold outcomes in the labor market fixed (i.e. w_j and A_{jk} are constant), (ii) assume that all housing is used for residential purposes only (i.e. $\alpha = 1$) and (iii) suppose that there is only one group of workers. Our aim is to link the change in floorspace prices to changes in land supply. Under these simplifying assumptions, equilibrium in residential floorspace markets is determined by the floorspace market clearing condition (4), resident supply (1), housing supply (11) and amenities (13).

Letting $\hat{x} = x'/x$ denote the gross change in variable x across equilibria, and recalling that we hold wages fixed (so that $\hat{\bar{w}}_i = \hat{\Phi}_{Ri} = 1$), we can write these equations in relative changes as

$$\begin{aligned}
\hat{L}^F_{Ri} &= \hat{u}^{\theta}_{Fi} \hat{r}_{Fi}^{-\theta(1-\beta)} \\
\hat{r}_{Fi} &= \frac{\hat{L}_{Ri}}{\hat{H}_{Fi}} \\
\hat{H}_{Fi} &= \hat{\lambda}_{Fi} \hat{T}_i \hat{r}_{Fi}{}^{\frac{\eta}{1-\eta}} \\
\hat{u}_{Fi} &= \hat{\bar{u}}_{Fi} \left[\sum_j \gamma_{ij} \hat{H}_{Fj} \right]^{\mu_U}
\end{aligned}$$

where the shares $\gamma_{ij} = \frac{d_{ij}^{-\delta}(H^F_j/T_j)}{\sum_s d_{is}^{-\delta}(H^F_s/T_s)}$ reflect the share of location j in the overall endogenous portion of amenity in location i. This share will be high if j is either very close to i (i.e. large $d_{ij}^{-\delta}$) or contains a lot of formal housing density (i.e. large H_{Fj}/T_j).

[29] While we will not be able to interpret results as structural parameters due to the log-linear approximations used in deriving these relationships, identification will be more transparent than in the structural estimation of the next section.

To reduce this to a log-linear system, we log-linearize the change in amenities $\hat{u}_{Fi} \approx \hat{\bar{u}}_{Fi}\left[\prod_j \hat{H}_{Fj}^{\gamma_{ij}}\right]^{\mu^U}$. Intuitively, this looks very similar to the arithmetic weighted mean above, but now the change is approximated by a weighted geometric mean. Substituting this into the expression for floorspace prices and rearranging, we get a single relationship between floorspace prices and floorspace supply

$$(1+\theta(1-\beta))\Delta \ln r_{Fi} = \theta\mu^U \sum_j \gamma_{ij}\Delta \ln H_{Fj} - \Delta \ln H_{Fi} + \theta\Delta \ln \bar{u}_{Fi}$$

Stacking this system of equations and defining the matrix $\Gamma \equiv [\gamma_{ij}]$, we have that

$$(1+\theta(1-\beta))\Delta \ln r_F = (\theta\mu^U\Gamma - I)\Delta \ln H_F + \theta\Delta \ln \bar{u}_F$$

There are two effects of the change in formal floorspace supply on formal sector house prices. First, the increases in supply causes prices to fall (captured by $-I$). Second, the increased stock of formal housing causes formal sector prices to increase around the city through spillovers reflected through the term $\mu^U\Gamma$. μ^U controls the overall size of this externality, while Γ captures the spatial diffusion of the externality which is determined by the spatial decay δ contained in the weights γ_{ij}.

The last step is to link the change in formal housing supply to changes in land available for development. The change in formal floorspace supply is given by $\hat{H}_{Fi} = \hat{\lambda}_{Fi}\hat{T}_{Fi}\hat{r}_{Fi}{}^{\frac{\eta}{1-\eta}}$. Stacking these conditions and substituting out for formal housing supply, we find that

$$\Delta \ln r_F = (1-\eta)\left((1+\theta(1-\beta)(1-\eta))\, I - \mu_U\theta\eta\Gamma\right)^{-1}\left(\theta\mu_U\Gamma - I\right)\left(\Delta \ln \lambda_F \ln T_F\right) + \epsilon_F$$

where ϵ_F is a structural error term.[30] Define $B \equiv (1-\eta)\left((1+\theta(1-\beta)(1-\eta))\, I - \mu_U\theta\eta\Gamma\right)^{-1}\left(\theta\mu\Gamma - I\right)$ to be the matrix of reduced form coefficients. To simplify this relationship further, we consider the partial relationship holding land use shares constant (i.e. $\Delta \ln \lambda_F = \mathbf{0}$). Then we can write the change in formal sector residential floorspace prices across the city as

$$\Delta \ln r_{Fi} = B_{ii}\Delta \ln T_{Fi} + \sum_j B_{ij}\Delta \ln T_{Fj} + \epsilon_{Fi} \tag{15}$$

Intuition This regression equation captures two forces. The first term captures the direct effect of the policy change, i.e. whether location i had its formal land supply increased through the DCR 58 change. The second term captures the indirect of the policy change: by increasing the supply of formal housing in surrounding neighborhoods, the amenities in location i rise in turn increasing the price of formal housing. Note that both of these effects are heterogeneous: the entries of B can differ across locations due the spatial heterogeneity contained in the Γ weight matrix. For example, locations which contain a lot of formal housing relative to nearby will put more weight on the change in own location changes so that B_{ii} will be large.

[30]In particular, $\epsilon_{Fi} \equiv (1-\eta)\left((1+\theta(1-\beta)(1-\eta))\, I - \mu\theta\eta\Gamma\right)^{-1}(\theta\mu\Gamma - I)\theta\Delta \ln \bar{u}_F$

To build intuition for how the regression coefficients will be informative about our structural parameters, it is informative to consider two special cases. First, imagine $\mu^U = 0$ so that there are no spillovers. Then (15) reduces to

$$\Delta \ln r_{Fi} = -\frac{1-\eta}{1+\theta(1-\beta)(1-\eta)}\Delta \ln T_{Fi} + \epsilon_{Fi}$$

In this case, there are no spatial linkages across locations and the only change in land supply that matters is the change within the location under consideration. This only has a negative effect on property prices, by increasing the supply of floorspace.

In contrast, in the model with spillovers there are two supply shocks: one the supply of floorspace and another to the supply of amenities. To see this, consider the case where $\mu^U > 0$, $\delta \to \infty$ so that the only location that matters for spillovers in amenities in location i is the location itself. Then $\Gamma = I$ and equation (15) becomes

$$\Delta \ln r_{Fi} = \frac{(1-\eta)(\theta\mu_U - 1)}{1+\theta\left((1-\beta)(1-\eta) - \mu_U \eta\right)}\Delta \ln T_{Fi} + \epsilon_{Fi}$$

The net effect is now ambiguous. Floorspace prices fall due to the positive supply shock and rise due to the increase in amenities.

If we assume the change in land supply due to the policy is uncorrelated with unobserved factors affected property price appreciation (more on this below), this dicussion highlights how a regression similar to (15) will be informative about the presence of spillovers. First, if we observe that locations containing mill sites experience increases in floorspace prices, this suggests spillovers are strong since the price appreciation due to the increase in amenities outweighed the price depreciation due to the increase in the stock of formal housing. Second, if we observe that the presence of mill sites in nearby locations drive increases in property prices (conditional on mill area within the area itself) then the model predicts that this should only be driven by spatial spillovers.[31]

Empirical Specification In its exact form, estimating (15) would be non-trivial since the coefficient matrix B is a non-linear function of the underlying structural parameters. The fact that it was derived by log-linear approximation would make it hard to interpret the resulting estimates. Instead, we run an empirical analogue of the form

$$\Delta \ln r_{Fi} = \beta_1 \times \text{Mill Site Own}_i + \beta_2 \times \text{Mill Site Exc}_i + \gamma_w + \gamma' X_i + \epsilon_{Fi} \tag{16}$$

where we control for factors other than mill sites which may be affecting price appreciation through ward fixed effects γ_w and a set of initial characteristics X_i (such as distance to the CBD). Mill Site Own$_i$ is the share of land occupied by mills within a location (in this case, subzones) while Mill Site Exc$_i$ corresponds to the share of mills in a disk around the location centroid excluding the location itself. We typically run this in

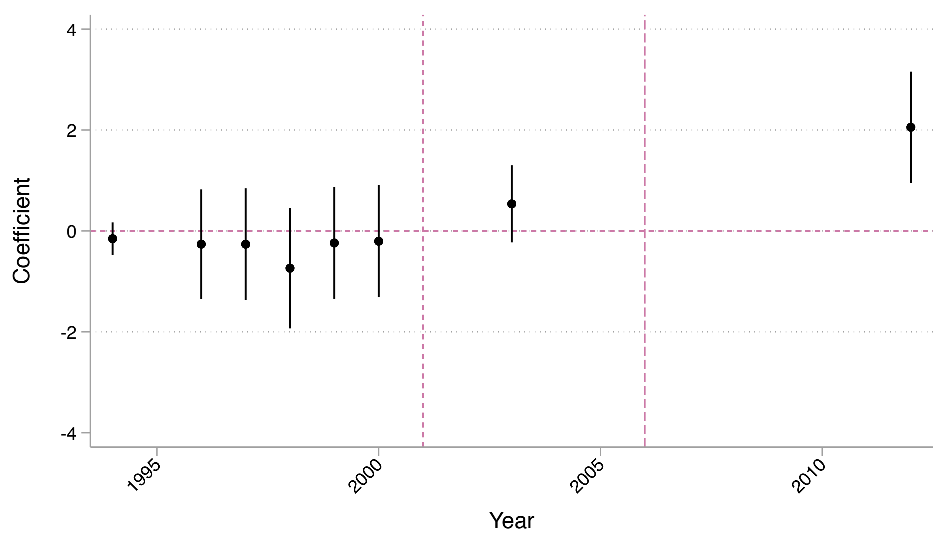
[31] We note that the specification (15) resembles those seen elsewhere in the literature. For example, Autor, Palmer and Patak (2013) examine the impact of changes in rent control regulation in Boston and break down the impact of the policy into an own-property effect as well as a spillover effect from surrouding lots.

long-differences between 2012 and 1993.

Our identification assumption is that conditional on observable characteristics, a location's relation to mill sites within wards is uncorrelated with unobserved factors driving house price growth. This might be violated if, for example, the mill development had been anticipated in years preceding the law change. Alternatively, the amendment itself might have been driven by lobbying from mill owners who wanted to take advantage of growing prices in the area, in which case appreciation near mills would be a cause rather than a consequence of the law change. The discussion in section 3 suggests that neither was the case. In order to provide evidence in support of this assumption, we exploit the annual data on formal floorspace prices we have back to 1993 and run an event study version of (16) where all coefficients and fixed effects are allowed to vary by year. We focus on the β_{1t} coefficients to examine how the house price appreciation associated with the share of mill sites in a subzone evolved over time relative to the base year of 1993.

Figure 4 plots the results. The first dashed line indicates the announcement of the DCR 58 change in 2001, while the second denotes the supreme court decision in 2006 which ultimately spurred development. There is no change in house price appreciation in subzones more or less exposed to mill sites in the years leading up to the law change. After the initial accouncement, we see that prices start to grow near mill sites: the differences in prices between a subzone completely covered by mills and one containing no mills is about double ($101\% = \exp(0.7) - 1) * 100$) its value in 1993. By 2012, 6 years after the policy uncertainty was resolved by the supreme court, this semi-elasticity had risen from 0.7 to 2. If the amendment had either been anticipated before 2001 or caused by growth in surrounding areas, we would expect this price differential to have been growing in the years leading up to the DCR 58 change. We therefore interpret these results as supporting our identification assumption.

Figure 4: Impact of Mill Share on Residential Floorspace Prices by Year



Plot show coefficients from regression of log residential floorspace price on the mill share measure (750m) interacted with year dummies, and plots the coefficient on the mill measure in each year. Additional controls include region fixed effects and polynomial in log distance to the CBD, both interacted with year dummies,and subzone fixed effects. Omitted category is coefficient in 1990. Solid line is 2001 when DCR 58 change occured, dashed line in 2006 when uphelp by the Supreme Court. Only subzones within 3km of mill site included. Standard errors clustered at subzone level.

In light of this evidence, Table 1 turns to the estimation of (16) in which we interpret the coefficients as causal estimates of the impact of exposure to mill sites on formal floorspace price growth over the period. In column 1, we run the specification including only the own-location mill site measure to start. A one standard deviation increase in the mill share leads to a 9% $(= (\exp(.062 * 1.438) - 1) * 100)$ rise in house prices, an economically large effect. In column 2 we expand this mill measure to reflect the share of mills in a 750m disk around each subzone centroid. The semi-elasticity rises from 1.438 to 2.354, suggesting the additional impact of spillovers from exposure to mills in distant locations. The next two columns attempt to formally distinguish these own-exposure and spillover effects using the full specification. Column 3 shows large effects from mill exposure only in surrounding locations. Column 4 shows this is robust to controlling for the share of mills within a location. Interpretted through the lens of the model, the final column provides the first piece of evidence of spillovers from exposure to mill sites.

Table 1: The effect of mill share on log floorspace prices

	Res	Res	Res	Res	Comm	Comm	Comm	Comm
Mill Share Own X Post	1.438***	2.354***		1.171**	0.936***	1.457***		0.739**
	(0.500)	(0.552)		(0.478)	(0.354)	(0.482)		(0.332)
Mill Share Exc X Post			1.759**	1.221*			1.241**	0.901
			(0.738)	(0.694)			(0.566)	(0.548)
N	368	368	368	368	368	368	368	368
Subzone FE	X	X	X	X	X	X	X	X
Post FE X Ward FE	X	X	X	X	X	X	X	X
Post FE X logDistCBD Poly	X	X	X	X	X	X	X	X
Own Measure	Own	Disk	Own	Own	Own	Disk	Own	Own

Note: Table shows regression of log floorspace prices on mill measure interacted with Post dummy, where Post equals 1 in 2012 and 0 in 1990. Observations are weighted by area. Columns (1)-(4) use residential prices as the outcome, columns (5)-(8) use commercial prices. Only subzones whose edge is closer than 3km from nearest mill site are included. logDistCBD Poly is a 4th-order polynomial in log distance to the CBD. Own measure is either the share of land within a sub zone containing mill sites (own), or the share of land within 750m disk from centroid (disk). Exc measure is the share of land in a 750m disk from the subzone centroid, excluding the subzone itself, that contains mill sites. Standard errors clustered at subzone level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

6.2 Slum Share

We now turn to measuring the reduced form effect of mill sites on the share of slums in a location. Totally differentiating the expression for the share of slums in a location in (9), and considering only the partial change holding informal prices constant we find that[32]

$$
\begin{aligned}
\Delta \ln \lambda_{Ii} &= -\frac{\kappa}{1-\eta}\lambda_{Fi}\Delta \ln r_{Fi} \\
&= -\frac{\kappa}{1-\eta}\lambda_{Fi}B_{ii}\Delta \ln T_{Fi} - \frac{\kappa}{1-\eta}\lambda_{Fi}\sum_j B_{ij}\Delta \ln T_{Fj} + \epsilon_{Ii}
\end{aligned}
$$

[32]While this is an unrealistic assumption since prices of informal housing will likely rise, so long as the rise in formal prices exceeds that of informal prices the model's predictions for the reduced form coefficient sign is unchanged.

Intuitively, the increase in formal house prices leads land owners to redevelop slums into formal housing. The rate at which they do so is determined by the formal housing supply elasticity $\frac{\kappa}{1-\eta}$. We think of this as a measure of redevelopment frictions. The second line uses (15) to relate the change in formal house prices to exposure to mill sites. Once again, the model predicts that both the amount of mills in and around a location should determine slum redevelopment.

Empirical Specification and Results Our deep learning procedure generates a set of polygons representing the locations of slums inferred from the 2001 and 2016 satellite images. Therefore, we can run our regression analysis at a fine geographic level: city blocks as defined in Section 4. Many city blocks have no slums in one of the two years, so there can be intensive (i.e. reduction in slum share) and extensive responses (i.e. elimination of slums) of the share of a block covered by slums to the mill redevelopment.

We therefore estimate the analog of (15) by Poisson pseudo maximum likelihood, following Santos Silva and Tenreyro (2006). We note, however, that in this case both β_1 and β_2 should be interpreted as measuring spillovers. No mill sites actually contained slums, so any change in slum share on blocks containing mills must be due to the price effects of mill redevelopment on nearby plots. Table 2 shows the results.

Table 2: The effect of mill share on formalization

	Slum share	Slum share	Slum share	Slum share
Mill Share Own X Post	-3.696***	-4.764**		-3.424***
	(1.021)	(1.865)		(1.076)
Mill Share Exc X Post			-2.194	-0.736
			(1.441)	(0.973)
N	664	664	664	664
Block FE	X	X	X	X
Post FE X Ward FE	X	X	X	X
Post FE X logDistCBD Poly	X	X	X	X
Own Measure	Own	Disc	Own	Own

Note: Table shows Poisson pseudo maximum likelihood (Santos Silva and Tenreyo 2006) estimates of the elasticity of the share of a city block taken up by slums with respect to mill measures interacted with Post dummy. Post equals 1 in 2016 and 0 in 2001. logDistCBD Poly is a 4th-order polynomial in log distance to the CBD. Own measure is either the share of land within a block containing mill sites (own), or the share of land within a 500m-radius disc around the block centroid (disc). Exc measure is the share of land in a 500m-radius disc around the block centroid, excluding the block itself, that contains mill sites. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The first column shows the semi-elasticity of formal land supply with respect to the share of city block taken up by mills. The effect is large, implying that a one-standard deviation increase in the share of mill sites is associated with a 23% $(= (\exp(.072 * -3.696) - 1) * 100)$ decrease in the share of the block taken up by slums. The second column shows that the semi-elasticity increases in magnitude if we consider the mill share of a 500m radius disc[33] around the block centroid. In the last two columns we see that higher mill share in the 500m disc around the centroid, excluding the block itself, results in fewer slums on the block although this

[33] We decrease the radius of the disc because blocks are smaller than subzones.

effect is imprecisely estimated and decreases in magnitude when controlling for the mill share of the block.

6.3 Employment

To derive regression specifications linking employment outcomes to changes in land supply, we (i) hold outcomes in the residential market fixed (i.e. u_{ik} and L_{Rik} are constant), (ii) assume that all floorspace is used for commercial purposes only (i.e. $\beta = 1$) and (iii) suppose there is only one group of workers and firms (who produce using only formal floorspace). Under these assumptions, equilibrium in the labor market is determined by the zero profits condition (6), labor demand (5), labor supply (2), housing supply (11) and productivities (14).[34]

As before, we write these equations in relative changes as

$$\begin{aligned}
\hat{L}_{Fj} &= \hat{w}_j^{\theta} \\
\hat{L}_{Fj} &= \left(\frac{\hat{A}_j}{\hat{w}_j}\right)^{\frac{1}{1-\alpha}} \hat{H}_{Fj} \\
\hat{r}_{Fj} &= \hat{A}_j^{\frac{1}{1-\alpha}} \hat{w}_j^{-\frac{\alpha}{1-\alpha}} \\
\hat{H}_{Fi} &= \hat{\lambda}_{Fi} \hat{T}_{Fi} \hat{r}_{Fi}^{\frac{\eta}{1-\eta}}
\end{aligned}$$

where $\hat{A}_i \approx \hat{\bar{A}}_i \left[\prod_j \hat{L}_{Fj}^{\xi_{ij}}\right]^{\mu_A}$ is the log-linearized change in productivity and $\xi_{ij} = \frac{d_{ij}^{-\delta_A}(L_{Fj}/T_j)}{\sum_s d_{is}^{-\delta_A}(L_{Fs}/T_s)}$ is the share of location j in i's overall endogenous productivity in the initial equilibrium. This share will high if j is either very close to i (large $d_{ij}^{-\delta_A}$) or contains a high employment density (large L_{Fj}/T_j).

As shown in the appendix, this system can be reduced to express the change in commercial floorspace prices and employment in a location as a function of its relation to the change in land supply across the whole of the city

$$\Delta \ln L_{Fj} = C_{jj} \Delta \ln T_j + \sum_s C_{js} \Delta \ln T_j + \nu_j \tag{17}$$

where the coefficient matrix is given by $C \equiv \left(\frac{1}{\theta(1-\alpha)}\left(1 + \theta(1-\alpha) + \frac{\eta\alpha}{1-\eta}\right) I + \frac{\mu_A}{1-\alpha}\frac{1-2\eta}{1-\eta}\Xi\right)^{-1}$ where $\Xi \equiv [\xi_{ij}]$ and ν is structural residuals containing changes in model unobservables. Once again, we have considered the partial response holding land use share constant (i.e. $\Delta \ln \lambda_{Fi} = 0$)

[34] We note that we are more agnostic in this section about how informative the reduced form results can be about the structural parameters. In deriving our labor market regression equations, we are holding outcomes in the residential market fixed. The model predicts that employment near mill sites should increase as a result from the rise in commercial floorspace, yet it is also possible that labor supply increases in these neighborhoods from an increase in residential densities, which would also increase employment. The reduced form employment specifications are similar in this case when we allow for both margins of adjustment (with the same expected sign). We highlight this caveat and leave the task of seperating these forces to the quantitative section.

Intuition Consider the case without spillovers $\mu_A = 0$. Then $C = \frac{\theta(1-\alpha)}{1+\theta(1-\alpha)+\frac{\eta}{1-\eta}\alpha} I$ and

$$\Delta \ln L_{Fj} = \frac{\theta(1-\alpha)}{1+\theta(1-\alpha)+\frac{\eta}{1-\eta}\alpha} \Delta \ln T_j + \nu_j$$

In this case, employment growth only depends on the change in land supply within a location and we should expect to see no dependency of employment on changes in land supply nearby.[35] It is only when both $\mu_A > 0, \delta_A \neq \infty$ that the second term in (17) is non-zero. Thus, the presence of productivity spillovers will be identified by the dependence of employment growth in a location on the change in land supply in nearby locations, conditional on the change within the location itself. The reason is that the increase in land supply increases employment within a location, which in turn increases the productivity of surrounding locations in the presence of spillovers.

Results As described in Section 4, our employment data are indicators for categories of employment (i.e., 15-49 workers) associated with the addresses of formal establishments listed in the 2005 and 2013 Economic Censuses. To examine an effect free without imputation, we currently investigate an extensive margin semi-elasticity: the effect of the mill share of 500 x 500 m grid cells and surrounding areas on the number of formal establishments operating in the cell. The regression specification is exactly analogous to 16 so we omit it in the interest of space. Table 3 provides the results.

Table 3: The effect of mill share on formal establishment counts

	log(Establishments)	log(Establishments)	log(Establishments)	log(Establishments)
Mill Share Own X Post	1.020*	2.267***		0.480
	(0.545)	(0.850)		(0.715)
Mill Share Exc X Post			2.207**	1.714
			(0.854)	(1.158)
N	774	774	774	774
Grid cell FE	X	X	X	X
Post FE X Ward FE	X	X	X	X
Post FE X logDistCBD Poly	X	X	X	X
Own Measure	Own	Disc	Own	Own

Note: Table shows regression of log cell establishments on mill measure interacted with Post dummy, where Post equals 1 in 2013 and 0 in 2005. Mill measure is the share of land in a 1000m disc from cell centroid which contains a mill site. Exc own subzone specifications construct the mill measure using only land in the disc which is outside of the cell itself. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

There is substantial measurement error in the dependent variable coming from issues in geolocating establishments and the artificial use of a rectangular grid to define observations, [36], making the estimates imprecise. Nevertheless, there appears to be a positive effect of mill share on growth in the number of establishments in

[35]The specification is similar when only own-location spillovers are present, i.e. $\mu_A \neq 0, \delta_A \to \infty$, in which case $C = \frac{\theta(1-\alpha)}{1+\theta(1-\alpha)+\frac{\eta}{1-\eta}\alpha+\frac{\mu_A}{\theta}\frac{1-2\eta}{1-\eta}} I$ and only the coefficient on the change in own-location land supply changes.

[36]Future drafts will use city blocks instead.

a grid cell between 2005 and 2013 and the effect appears concentrated in the mill share of a 1000m radius disc around the cell centroid omitting the cell itself.[37] As a benchmark, the estimates from the second column indicate that a one standard deviation increase in mill share of a 1000m disc around the grid cell centroid is associated with a 14% (= $(\exp(.059 * 2.267) - 1) * 100$) increase in the number of formal establishments operating in the cell over the period. Our findings are thus generally supportive of the idea that the mill redevelopment generated productivity spillovers for nearby firms.

We note, however, that an increase in the number of formal establishments may be associated with a decrease in the number of informal establishments, leaving the overall effect on employment ambiguous (in addition to the possibility of employment reductions by incumbents accompanying entry). We are therefore preparing data on block-level totals (including formal and informal employment) from the 2005 Economic Census to accompany the block-level totals we already have from the 2013 Economic Census. To further corroborate the finding of positive productivity spillers, in the following subsection we detail a positive the effect of mill share on commercial floorspace prices.

6.4 Commercial Floorspace Prices

Intuitively, the response of commercial floorspaces to changes in land supply is similar to employment growth and is given by the following equation (see appendix for derivation).

$$\Delta \ln r_{Fj} = \tilde{C}_{jj} \Delta \ln T_{Fj} + \sum_{s} \tilde{C}_{js} \Delta \ln T_{Fj} + \tilde{\nu}_j \tag{18}$$

where $\tilde{C} \equiv C \frac{1}{1-\alpha} \left(\mu_A \Xi - \frac{\alpha}{\theta} I \right)$ and $\tilde{\nu}$ is a structural residual. This specification is informative about the presence of spillovers in the same way that we saw for residential floorspace prices. When $\mu_A = 0$, then $\tilde{C} = -\frac{\alpha}{1+\theta(1-\alpha)+\frac{\eta}{1-\eta}\alpha} I$ so that the change in land supply reduces floorspace prices within a location (through increasing supply) and has no effect on surrounding locations. By contrast, when $\mu_A > 0$ the change in land supply can increase prices within a location (if the increase in productivity is strong enough to offset the positive supply shock). In addition, when $\delta_A \neq 0$ then the second term in (18) will be non-zero. Thus, both a positive coefficient on the change in land supply within a location as well as an effect from changes in nearby locations will be informative about the presence of productivity spillovers.

Results The last four columns of Table 1 provide further evidence of productivity spillovers. For instance, the point estimate on the mill share of a 750m disc around a subzone's centroid, but excluding the subzone itself, shows that a one standard deviation increase in mill share is associated with a 6% (= $(\exp(.062 * .901) - 1) * 100$) increase in commercial floorspace prices. The estimate just escapes significance at the 10% level.

[37] To the extent that some still-operating mills were converted to residential use, there may be a mechanical downward bias in the effect of the mill share of the grid square itself.

6.5 Share of High-Skilled Workers

The results from this section are forthcoming.

7 Structural Estimation and Quantification

7.1 Solving for the Model's Unobservables

In our data we observe residential populations by skill and housing type L^k_{Rig}, formal sector residential and commercial prices r^F_{Ri} and r^F_{Fi}, land use shares λ_{ki}, total land area T_i, commute times (which determine d_{ij}), and employment by formal and informal establishments L^F_{Fj} and L^I_{Fj}. The following proposition shows that, for any vector of model parameters, this data is sufficient to invert the model in order to recovers its unobservables.

Proposition 1. (Model Inversion) *Given data on residence L^k_{Rig}, employment L^k_{Fj}, formal sector floorspace prices r^F_{Ri}, r^F_{Fi}, land use shares λ_{ki}, total land are T_i and commute costs d_{ija}, there is a unique vector of unobservables $\{u^k_{ig}, A_{ik}, w_{jg}, H^k_{Fi}, H^k_{Ri}, \vartheta_i, \tau_i, \xi_i, r_{Ii}\}$ that rationalize the observed data as an equilibrium of the model.*

7.2 Structural Estimation

The results from this section are forthcoming.

7.3 Quantitative Exercises

The results from this section are forthcoming.

8 Conclusion

References

Ahlfeldt, G. M., D. M. Sturm, S. J. Redding, and N. Wolf (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica 83*(6), 2127–2189.

D'Monte, D. (Ed.) (2006). *Mills for sale: the way ahead*. Mumbai: Marg Publications.

Engstrom, R., J. Hersh, and D. Newhouse (2016). Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being and Geographic Targeting. pp. 1–48.

Engstrom, R., A. Sandborn, Q. Yu, J. Burgdorfer, D. Stow, J. Weeks, and J. Graesser (2015). Mapping slums using spatial features in Accra, Ghana. *2015 Joint Urban Remote Sensing Event, JURSE 2015*.

Gershgorn, D. (2017, jul). It's not about the algorithm: the data that transformed AI research - and possibly the world.

Goldblatt, R., W. You, G. Hanson, and A. K. Khandelwal (2016). Detecting the boundaries of urban areas in India: A dataset for pixel-based image classification in google earth engine. *Remote Sensing 8*(8).

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Henderson, V., T. Regan, and A. J. Venables (2016). Building the City: Sunk Capital, Sequencing and Institutional Frictions. *Working Paper*.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353*(6301), 790–4.

Kuffer, M., K. Pfeffer, and R. Sliuzas (2016). Slums from space-15 years of slum mapping using remote sensing. *Remote Sensing 8*(6).

Marx, B., T. Stoker, and T. Suri (2014). There Is No Free House : Ethnic Patronage and Diversity in a Kenyan Slum.

Santos Silva, J. M. C. and S. Tenreyro (2006). The Log of Gravity. *The Review of Economics and statistics 88*(November), 641–658.

Yuan, J. (2017). Automatic Building Extraction in Aerial Scenes Using Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

A Additional Tables

Table A.1: Selected Mill Redevelopments

Name of the Mill	Current Status	Name of the New Development	# of Floors
Apollo Mills (South)	Development completed	Lodha Bellissimo	50
Gokuldas Morarjee Mills no.1	Development completed	Ashok Towers	49
Hindoostan Spinning & Weaving Mills no.2	Development completed	Kalpataru Heights	39*
Mafatlal Mills no.3	Development completed	Marathon Futurex	38*
Modern Mills	Development completed	Belvedere Court	40*
Morarjee Textile Mills	Development completed	Ashok Towers	49*
Ruby Mills	Development completed	The Ruby	40*
Simplex Mills	Development completed	Planet Godrej	51*
Standard Mills no.1	Development completed	Sheth Beaumonde, Chaitanya Towers	26*
Ambika Mills	Development in progress	Namaste Tower	66
Bharat Mills	Development in progress	Indiabulls Blu	56
Bombay Dyeing (Spring Mills)	Development in progress	Island City Center, Spring MillsTower	61
Elphinstone Mills (South)	Development in progress	Indiabulls Sky Suites, Finance Center	75*
Hindoostan Spinning & Weaving Mills no.3 (Crown Mills)	Development in progress	DB Orchid Crown	68
Jupiter Mills (South)	Development in progress	Indiabulls Sky Forest	80*
Khatau Makanji Spinning & Weaving Mills	Development in progress	Monte South	60
Kohinoor Mills no.3 (North)	Development in progress	Kohinoor Square	48
Mumbai Textile Mills (Sakseria Mills)	Development in progress	Lodha The Park	78
New Islam Mills	Development in progress	One Avighna Park	64*
Poddar Processors (Edward Mills)	Development in progress	Indiabulls Blu	50
Prakash Cotton Mills	Development in progress	Raheja Imperia	60
Raghuvanshi Mills	Development in progress	Namaste Tower	66
Shree Ram Mills	Development in progress	Palais Royale	88*
Shriniwas Mills	Development in progress	World One	119
Mafatlal Mills no.1	Mill demolished with development planned	Piramal Aranya, Byculla Zoo	70
Mafatlal Mills no.2	Mill demolished with development planned	Piramal Aranya, Byculla Zoo	70

*Only includes # of floors above ground

B Additional Figures

Figure B.1: Location of Mills

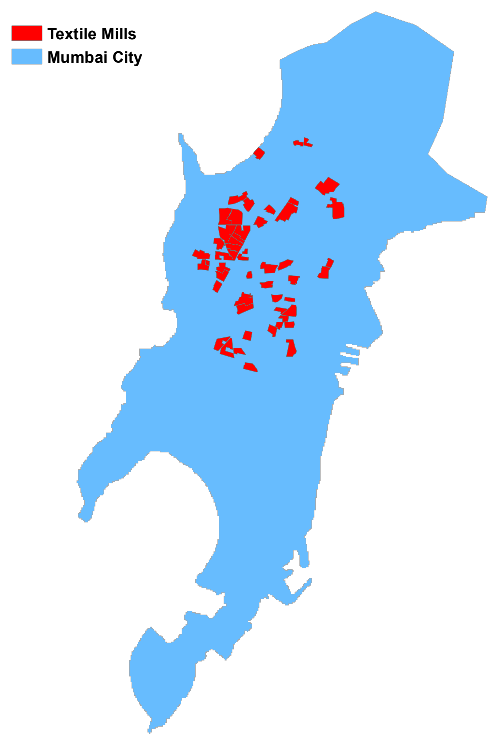


Figure B.2: Validation Sample: SRA vs SPFEAS Slums

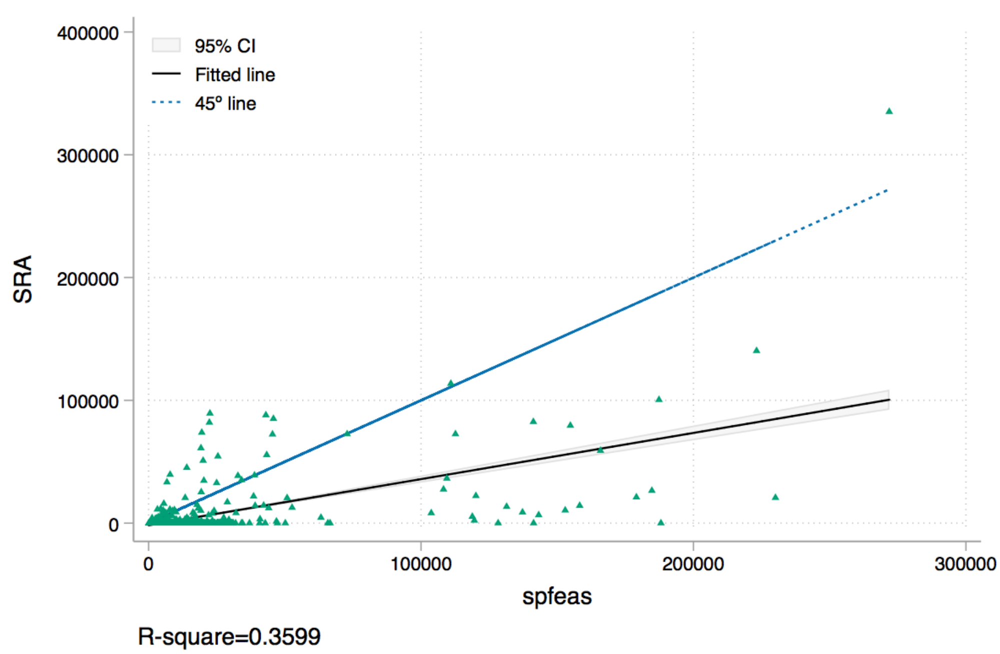


Figure B.3: Generating signed distance to a slum boundary

(a) Raw image containing slum and non-slum areas

(b) Signed distance from a slum boundary

C Proofs

Proof of Proposition 1. **Wages** From the commuting market clearing condition $L^k_{Fj} = \sum_{i,g} \pi_{j|ig} L_{Rig}$ we have

$$
\begin{aligned}
L_{Fj} &= \sum_i \frac{(w_j/d_{ij})^\theta}{\sum_s (w_s/d_{is})^\theta} L_{Ri} \\
\text{where } L_{Fj} &= \sum_k L^k_{Fj} \\
L_{Ri} &= \sum_{k,g} L^k_{Rig}
\end{aligned}
$$

since workers are perfect subsitutes in production and the differences in effective units of labor supplied (captured by Z_H) are constant over space. That this has a unique solution comes from Ahfelft et. al. (2016).

Floorspace Use Wedges These rationalize the different prices of residential and commercial floorspace we observe in the data, and are retrieved from $1 - \xi_i = \frac{r^F_{Ri}}{r^F_{Fi}}$.

Informal Sector Floorspace Prices and Distortions on Formal Development The land use equations are

$$
\begin{aligned}
\lambda_{Fi} &= \frac{\tilde{\tau}_i (r^F_i)^{\frac{\kappa}{1-\eta}}}{1 + \tilde{\tau}_i (r^F_i)^{\frac{\kappa}{1-\eta}} + (r^I_i)^\kappa} \\
\lambda_{Ii} &= \frac{r^\kappa_{Ii}}{1 + \tilde{\tau}_i (r^F_i)^{\frac{\kappa}{1-\eta}} + (r^I_i)^\kappa}
\end{aligned}
$$

$$\lambda_{Vi} = 1 - \lambda_{Fi} - \lambda_{Ii}.$$

The LHS of each is observed. On the RHS, only r_i^F is observed. From the vacant share we can obtain the denominator. We can then rewrite the first two equations as

$$\lambda_{Fi} = \lambda_{Vi}\tilde{\tau}_i^T (r_i^F)^{\frac{\kappa}{1-\eta}}$$
$$\lambda_{Ii} = \lambda_{Vi}(r_i^I)^{\kappa}$$

Since λ_{ki}, r_i^F is observed, this allows us to recover $\tilde{\tau}_i^T$ and r_i^I.

Amenities We now have all residential house prices, as well as wages, so amenities are retrieved (to scale) from $L_{Rig}^k = \lambda_g^U (u_{ig}^k (r_{Ri}^k)^{\beta-1})^{\theta} \Phi_{Rig}$.

Productivities These are obtained from $r_{Fj}^k = (1-\alpha) A_{jk}^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w_j}\right)^{\frac{\alpha}{1-\alpha}}$.

Floorspace Quantities and Use Shares Total residential floorspace is given by $H_{Ri}^k = \frac{(1-\beta)}{r_{Ri}^k} \sum_g \bar{w}_{ig} L_{Rig}^k$ which can now be computed. Similarly, total commercial floorspace is given by $H_{Fj}^k = \left(\frac{(1-\alpha)A_{jk}}{r_{Fj}^k}\right)^{\frac{1}{\alpha}} L_{Fj}^k$. Finally, use shares are simply $\vartheta_i^k = \frac{H_{Ri}^k}{H_{Ri}^k + H_{Fj}^k}$. ■

D Additional Derivations

Regression Specifications

Employment and Commercial Floorspace Prices Using the equations in the main text, we begin by equating labor supply and demand to get

$$\left(\frac{1+\theta(1-\alpha)}{\theta(1-\alpha)} I + \frac{\mu_A}{1-\alpha}\Xi\right) \Delta \ln L_F = \Delta \ln(\lambda_F \ln T) + \frac{\eta}{1-\eta} \Delta \ln r_F + \frac{1}{1-\alpha} \Delta \ln \bar{A}$$

From the zero profit condition for commercial floorspace, we get a second equation linking employment, floorspace prices and land:

$$\Delta \ln r_F = \frac{1}{1-\alpha}\left(\mu_A \Xi - \frac{\alpha}{\theta} I\right) \Delta \ln L_F + \frac{1}{1-\alpha} \Delta \ln \bar{A}$$

Substituing this inot the labor market clearing condition, we find that

$$\left(\frac{1}{\theta(1-\alpha)}\left(1+\theta(1-\alpha)+\frac{\eta\alpha}{1-\eta}\right) I + \frac{\mu_A}{1-\alpha}\frac{1-2\eta}{1-\eta}\Xi\right) \Delta \ln L_F = \Delta \ln(\lambda_F \ln T) + \frac{1}{(1-\alpha)(1-\eta)} \Delta \ln \bar{A}$$

Defining $C \equiv \left(\frac{1}{\theta(1-\alpha)}\left(1+\theta(1-\alpha)+\frac{\eta\alpha}{1-\eta}\right) I + \frac{\mu_A}{1-\alpha}\frac{1-2\eta}{1-\eta}\Xi\right)^{-1}$ and considering a partial change holding land use share fixed ($\Delta \ln \lambda_{Fi} = 0$) yields (17) in the text. Substituting this back into the expression for commercial floorspace prices above delivers (18).