Spatial Spillovers from Urban Renewal: Evidence from the Mumbai Mills Redevelopment*

Michael Gechter† Nick Tsivanidis‡

March 31, 2020

PRELIMINARY AND INCOMPLETE:

PLEASE DOWNLOAD UPDATED PAPER HERE

Abstract

Developing country cities are characterized by informal housing–slums–but as incomes grow their governments will pursue a host of urban renewal policies that promote the construction of modern, formal sector housing. This paper examines spatial spillovers from urban renewal using a unique policy experiment in Mumbai that led 15% of central city land occupied by the city's defunct textile mills to be redeveloped during the 2000s. We digitize a host of new spatially disaggregated datasets on population, employment and house prices, and provide the first application of a deep convolutional neural network to measure changing slum cover from daytime satellite imagery. We find reduced form evidence of sizable spatial spillovers that impact surrounding locations by (i) increasing formal sector house prices and reducing slum cover, (ii) reducing informal employment density with no increase from the formal sector and (iii) increasing the share of high-skill residents and reducing population density. We disentangle the source of these spillovers by developing a quantitative urban model with formal and informal land and labor markets, and use it to quantify the equity-efficiency trade-off associated with slums and urban renewal policies.

*We wish to thank Russell Cooper, Jonathan Eaton, Pablo Fajgelbaum, Jon Hersh, Chang-Tai Hsieh, Dan Keniston, Kala Krishna, Abhay Pethe, 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, Dongyang He, Isabel Hincapie, Suraj Jacob, Peiquan Lin, Alejandra Lopez, Guoxuan Ma, Rucha Pandit, Pankti Sanghvi, Andy Sheu, Christine Suhr, Nicholas Swanson, Sneha Thube, and Nishali Withanage provided excellent research assistance. Financial support from the DigitalGlobe Foundation and the Tata Centre for Development at University of Chicago is gratefully acknowledged.

First Version: December 2017.

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

‡Haas School of Business and Department of Economics, University of California at Berkeley. Email: ntsivanidis@berkeley.edu.

1 Introduction

Vast swathes of developing country cities will be built in the coming decades. In India, inflows of over 250m people into cities by 2030 mean that millions of square meters of commercial and residential floorspace will need to be constructed: the equivalent of a new Chicago every year (McKinsey 2010). While developing country cities today are characterized by low-quality, informal structures or "slums", many policies will be used to promote the construction of taller, "formal" sector buildings using high-quality modern technologies. These range from direct policies like slum clearance to indirect ones such as the relaxation of land use regulations, building height restrictions and rent control that incentivize developers to build new construction. However, the presence of externalities in cities means these "urban renewal" policies that increase the supply of high-quality housing can have potentially transformative effects on entire neighborhoods surrounding the directly affected areas.

A number of questions arise that are crucial to understand the dynamics of cities in the developing world and inform government policy. How will these investments spill over into house prices nearby? Will appreciation drive gentrification of affected neighborhoods and displacement of existing slums? How will the distribution of formal and informal economic activity be affected? Finally, how do the aggregate impacts of this class of policies weigh up against distributional consequences, and can governments design policies that balance any trade-off? Unfortunately, the lack of both high-resolution spatial data in developing country cities combined with large-scale quasi-experiments that shock the supply of formal sector housing has meant that the answer to these questions remains an empirical challenge.

This paper combines new spatially-disaggregated data with a remarkable change in land use policy in urban India to make three contributions to our understanding of the spatial spillovers from urban renewal in the developing world. First, we exploit a sudden and unanticipated law change in one of the world's largest megacities: Mumbai. This led 60 textile mills that lay vacant during the late 1980s and 1990s, but took up over 600 acres (or 15%) of central city land, to come onto the market for redevelopment during the 2000s. Second, we digitize a number of new administrative datasets and leverage machine learning techniques to generate new high-resolution data on changes in economic activity. This includes the first application of a deep convolutional neural network to measure changes in slum cover from daytime satellite imagery, and frontier dimension reduction methods to evaluating changing SES of neighborhoods from the changing composition of residents' last names. Third, we use these data to evaluate the spatial spillovers from the policy and use a quantitative urban model to disentangle the source of the linkages and assess aggregate and distributional impacts.

We find evidence supporting sizeable spatial spillovers from new, formal developments on former mill sites that transformed surrounding neighborhoods. First and foremost, formal sector house

prices rise both in and nearby locations with former mill sites. Second, the share of slums in these neighborhoods falls in response to the increased value of alternate, formal land use. Commensurate with evidence we provide on slums being dense hives of both residence and employment, informal employment activity also falls without compensating increases in formal employment density. Population density also falls, with the caveat that our results are somewhat noisier for this outcome. Third, the average education of residents rises. These patterns paint an overall picture of a gentrification process driven by the new construction that affects not only the treated locations themselves, but also the neighborhoods that surround them. [Note: results of our structural estimation exercise are forthcoming.]

Mumbai's textile mills 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 Section 3, this changed in the 2000s when an amendment to the city's development control regulations suddenly lifted these restrictions and allowed mill owners to redevelop their sites without hinderance. This decision was announced in 2001, contested in the courts, and ultimately upheld by India's Supreme Court in 2006.

Evaluating the effects of policies on outcomes such as land use, property prices, residence, employment and demographic composition in developing country cities is typically complicated by a lack of data at small-scale geographies. We confront this challenge by constructing a number of new datasets. First, we digitize paper records from the Maharashtra Department of Registrations of Stamps that provide values of formal residential and commercial floorspace across 600 subzones in the city between 1993 and 2013. Second, we locate the universe of formal and informal establishments in the city using the 1998, 2005 and 2013 Economic Censuses. Third, we obtain population totals in 2001 and 2011 across 227 election wards obtained from the Bombay Municipal Corporation Election Office.

We generate new data on slum cover and SES neighborhood composition over the past twenty years using frontier techniques from the machine learning literature. First, we provide the first application of a deep convolutional neural network (CNN) to measure slums from high-resolution daytime satellite imagery. We provide the CNN with a training dataset we obtain for a single year consisting of a satellite image of suburban Mumbai and a shapefile of identified slums. The network learns the features of the image that best predict the slum status of pixels, and we use the trained network to predict the location of slums in Mumbai in 2001 and 2016. We obtain about twice the quality of performance with our method than we do when using prior frontier methods in remote sensing for slum measurement (R^2 of 0.94 vs 0.45 for predicted vs actual slums on the testing set). Second, we use a dimension reduction technique (as yet unused in the economics literature) based on the number of edits required to transform an arbitrary last name to a set of prototype last names.

This technique allows us to estimate a predictive mapping between last names and education levels which is robust to misspellings and pools information across names to improve performance. This mapping can then be applied to publically-available lists of voters participating in highly-localized elections every five years to infer changes in the socioeconomic status of voters.

We begin our analysis by using our only dataset available at an annual frequency–formal sector house prices–to assess the presence of pre-trends in neighborhoods with textile mills. We find no impact of mill sites on house price growth in the years leading up the announcement of the change to Mumbai's development control regulations in 2001, a steady rise between then and the Supreme Court decision in 2006, and a sustained positive impact thereafter. This patterns suggests both that the policy change was in fact unanticipated and was not driven endogenously by local economic trends (for example if mill owners lobbied government to repeal the law to allow them to benefit from growing neighborhood house prices). Section 3 provides a historical account of the DCR 58 amendment, which corroborates these empirical patterns in the data.

Our main specification then examines the impact of the mill redevelopment on locations that contain mills as well as spillovers onto nearby locations not directly affected. We find an increase in formal sector house prices, a reduction in slum cover, and an increase in the average education of residents. The redevelopments therefore appear to drive a process of gentrification in surrounding neighborhoods. We find smaller slums are more likely to convert in response to the same formal sector house price shock, suggesting convex adjustment costs.

Informal activity also decreases both within and nearby locations with mills, consistent with the high employment density within slums. In contrast, formal employment density is much less affected: while we find it (intuitively) increases in spatial units that contain mills that redevelop into commercial buildings, there is no significant increase in locations near any type of mill (commercial or otherwise). Standard theories of agglomeration would suggest that in productivity (and hence employment) should rise in locations experiencing exogenous shocks in employment density. Either there are limited spatial productivity spillovers within formal employment, or perhaps these only show up in the long-run and aren't captured within the 7 years between the policy change and our post-period data. In either case close to a decade on, it doesn't seem like there are transformative formal sector employment impacts from new redevelopments.

We run a number of falsification and robustness checks to provide additional support for the causal nature of our results. First, we run a placebo check to show there is no there is no additional effect of proximity to large industrial buildings (after accounting for distance to textile mills). This addresses that our effects could be driven by secular trends in gentrification to convert old industrial neighborhoods into high-end residences. Second, we use the panel dimension of our employment data to show there is no growth in employment between 1998 and 2005 in the lead up to the Supreme Court decision. Instead, all of the effects occur between 2005 and 2013. Third, we use data on the

one public expenditure for which we were able to construct a spatially-disaggregated panel dataset and show that the area of main roads was not changing near textile mills over the period. This addresses that even if the DCR 58 law change was exogenous to pre-existing economic trends, the policy response may not have been, but it does not appear the government coordinated policy to support the redeveloping neighborhoods. We do not document any contemporaneous changes in zoning or land use policy either. Fourth, we show robustness to alternative specifications, control variables, and a multilateral treatment effect measure that captures the full intensity of exposure to mills that can occur in two-dimensional space.

The last part of the paper develops a quantitative urban model with formal and informal land and labor markets to disentangle the source of observed spatial spillovers, quantify the aggregate and distributional impacts of the redevelopments, and characterize alternative counterfactual policies surrounding urban redevelopments and slum resident compensation. Low- and high-skill workers decide where to live and work. Formal and informal firms produce across the city using labor and commercial floorspace. The key new feature of our model is that it features both formal and informal buildings. Land owners in each location choose across these two construction technologies according to their relative price: an increase in the relative price of formal housing in a location should lead to a decrease in the share of slums, with an elasticity determined by the size of conversion costs across uses. [Note: results for this section are forthcoming.]

2 Related Literature

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

Within a large body of work examining the size and nature of spillovers,[1] a smaller strand has sought sources of exogenous variation to estimate these forces.[2] Regarding housing externalities, Rossi-Hansberg, Sarte, and Owens (2010) examine spillovers from urban revitalization policies on house prices in Richmond, Virginia. Hornbeck and Keniston (2014) study the impact of new housing construction induced by Boston's Great Fire on land values nearby. Regarding agglomeration externalities, Ahlfeldt, Sturm, Redding, and Wolf (2015) estimate the size of these spillovers leveraging the shock to density delivered by the division and reunification of Berlin. We contribute to these strands of literature by estimating the size of these spillovers using a unique natural experiment within a developing country city, and provide new evidence on their implications in the presence of slums.

In development, we contribute to the literature on slums in developing countries (see Brueckner and Lall (2015) and Marx, Stoker, and Suri (2013) for reviews). Most have focused on titling (Field

[1] see Rosenthal and Strange (2019), Duranton and Puga (2020) for recent, relevant reviews as well as Diamond and Mcquade (2015) and Redding and Sturm (2016) and the literature cited therein.

[2] Our paper also relates to work using historical natural experiments to examine spatial spillovers in a dynamic setting through path dependence (e.g. Davis and Weinstein 2002, Bleakley and Lin 2012 and Kline and Moretti 2014).

(2007)), slum upgrading (Field and Kremer 2008, Harari (2016)) and relocation from slums to public housing (Barnhardt, Field, and Pande (2016)). More closely related is Henderson, Regan, and Venables (2016) 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 estimate the size of slum conversion costs by estimating the response of slums to formal house price shock induced by our policy change rather than inferring them from model-based residuals. 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.

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,[3] 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.[4] 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 of papers in geography using this approach to identify slum cover. 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) for slum identification, making use of insights from Yuan (2017) to our problem.[5] The CNN can be thought of as generating spatial features tailored to the identification of slums (see Mallat (2016) on this interpretation, Goodfellow, Bengio, and Courville (2016) for a general primer, and Géron (2019) for a hands-on guide), and we see a large increase in predictive performance from its use.

We also relate to the work on the impact of land regulations, most of which focuses on developed

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

[4]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 et al. (2016) use LiDAR images to measure building height in addition to manually tracing building footprints.

[5]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.

countries (e.g. Ihlanfeldt 2007; Zhou, McMillen, and McDonald 2008; Saiz 2010; Turner, Haughwout, and Van Der Klaauw 2014). 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 and Brueckner (2005); Brueckner and Lall (2015); Harari (2016)). 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). We combine a natural experiment with new sources of data to identify the contribution of distortionary land use regulations on city structure.

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 III, Rossi-hansberg, and Sarte (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 primarily by developing a model featuring both informal and formal housing supply to fit our developing country context.

3 Institutional Context

Mumbai's Textile Mills and DCR 58 Mumbai's 60 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.[6] However, the industry declined after World War II as the city increasingly to employment in services. The coup de grace for most of the mills came during an 18-month strike spanning 1982 and 1983, leading most to cease operation.

While many cities experience redevelopment of old industrial neighborhoods as the relative productivity of alternative land use changes (e.g. New York's meatpacking district), strict regulation by the state government prevented Mumbai's mill owners from doing so. Prior to 1991, state law prevented textile mills being redeveloped or used for any alternative purpose. In 1991, as part of the broad reforms taking place in India, the state government amended the law by introducing Regulation 58 of the city's Development Control Rules (DCR 58). This stipulated that mills could be redeveloped to alternative purposes, so long as one-third of the land was given to 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. What amounted to a 66% tax on land did not prove attractive to mill owners, and next to no development took place during the 1990s.

This all changed in 2001, when an amendment to DCR 58 was made by Maharashtra state govern-

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

ment that de facto lifted these restrictions. This amended terms of the one third rule to apply only to open spaces on mill sites, the majority of which were covered in factory buildings. The mill owners could therefore freely redevelop all covered spaces on mill sites (and one third of any vacant space). It was in fact a prominent mill workers' union Girni Kamgar Sangharsh Samiti who initially lobbied for the change. They saw the jobs in the mills were not coming back as the industry had shifted out to cheaper, peri-urban areas, and lobbied to facilitate the redevelopment of mill sites into other commercial purposes so long as their members would be employed there. However, as an initial amendment made its way through the state bureaucracy, a bureaucrat Ramanand Tiwari (later arrested as part of a different corruption scandal) added the amendment about DCR 58 only applying to vacant spaces on mill sites.

Despite passing in 2001, the DCR 58 amendment was contested in the courts. The case went all the way to the Indian Supreme Court, which finally ruled in their favor in 2006 and allowed large-scale redevelopment to proceed.

Impact of the DCR 58 Amendment Analysis of satellite imagery suggests that the Supreme Court decision in 2006 rather than the initial amendment in 2001 was key in driving large-scale redevelopment. Today 70% of mill sites have been redeveloped. 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.[7] 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.

Laws Governing Slum Redevelopment The redevelopment of slums is governed by the 1995 Slum Rehabilitation Act. This regulation covers all official or "notified" slums, but non-notified slums are

[7]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.

also eligible if they fullfil the conditions laid out in the Act and are approved by the Slum Rehabilitation Authority. If redevelopment occurs, then residents who can prove occupancy prior to January 1 1995 will be compensated with a residential tenement with area of 225 square feet.[8] These new tenements are located in the same location as much as possible, constructed on the original site by the developer in exchange for transferable development rights (TDR) to increased building heights on other parts of the site. However, in cases where this in situ redevelopment is not possible these new units can be placed elsewhere and are often located towards the outskirts of the city. Residents who cannot establish occupancy prior to this date are not eligible for compensation. Moreover, there are reports that the law was not perfectly enforced, for example with developers forging residents signatures in order to pass the 70% threshold. [*Notes: Citations forthcoming. Updates to the paper will incorporate this context more formally into the assumptions of the model.*]

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. A second is our use of machine learning techniques new to economics to generate a novel, spatially-disaggregated measure of socioeconomic status from text data on names.

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.[9] The

[8] Residence is established through documents issued by the Maharashtra government ("photo passes"), or additional documentation such as ration cards, tax and utility bills, and birth certificates.

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

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[10] recommends reducing the dimension of the problem by performing specific, known calculations to extract a set of "spatial features"[11] 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 a 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 and non-linear transformations called a Deep Neural Network (DNN) 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 (see Athey and Imbens (2019) for an accessible introduction written for economists). By directly using the explanatory variables, the DNN avoids the information loss resulting from aggregation to pre-specified spatial features. The specific implementation we use is known as a Convolutional Neural Network (CNN) where the linear transformations take the form of increasingly abstract representations of the image and can be thought of as "filters" (see Géron (2019) for more details).

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.[12] Our data come from two sources. First, the Slum Rehabilitation Authority of Mumbai

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

[11] 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.

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

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.[13] 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.[14] 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[15] 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 footprint and the space taken up by primary, secondary, tertiary, and unclassified roads on OpenStreetMap (Table 1 gives the average characteristics of these blocks). 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 District. 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 A.1 plots the results of this exercise, with the points representing blocks and the fitted line the result of regressing

[13] 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.

[14] 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."

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

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).[16] 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[17], 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 A.2 demonstrates the approach. Panel (a) shows a part of the training area containing large slum (small, low structures) and non-slum (taller buildings) 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

[16]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.

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

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 A.1, 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.

For earlier years, we lack primary data sources on the location of slums. In ongoing work, we use the model trained on 2016 data to predict slums in 2001. At present, we instead adjust our 2016 SRA map by overlaying it on top of images from 2001 and 2005 to identify structures in our training sample that 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.[18]

4.2 Inferring Socioeconomic Status

To assess gentrification, we would ideally use a rich, spatially-disaggregated measure of socioeconomic status (SES) available over a long time horizon. Unfortunately to the best of our knowledge, no such data have been collected in India. We therefore create our own novel index of SES by exploiting the information available in the lists of voters' names by Election Ward which the Bombay Municipal Corporation (BMC) Election Office makes available on its website.[19] Every five years, each election ward elects a representative to the Municipal Corporation's legislature, and BMC publishes lists of the names of voters participating in the ward's election. There are about 227 election wards spread across Mumbai and Mumbai Suburban districts, with an average population of about 60,000 (see Table 1). To bracket the Supreme Court's upholding of the change to DCR 58, we choose 2002 and 2017 as our pre-change and post-change elections, respectively.

Since last name is indicative of caste and religious group (Bhagavatula, Bhalla, Goel, and Vissa (2019)) and membership in these groups is predictive of SES, we could apply a mapping from all the last names in an election ward to an index of SES to impute an average SES level for the ward. To generate this mapping, we use individual records from Mumbai District in the 2011 Socioeconomic

[18] In supplementary results available upon request, we address whether measurement error could be driving our results. For each year, we construct the prediction error as the difference between the predicted slums and actual slums (per the 2016 SRA map, or RA-adjusted 2001 map) and regress this on our mill measure. We find no significant effect, suggesting any measurement error introduced by differing image specifications across years is not confounding our results.

[19] See Olivetti and Paserman (2015) and Arkolakis, Lee, and Peters (2019) for two other recent examples of using names as data in economics.

and Caste Census (SECC), collected in conjunction with that year's Population Census.[20] The SECC asks respondents for each household member's birthdate and highest level of education completed with ordered categories illiterate, less than primary, primary, middle, secondary, higher secondary, and graduate or higher. We treat this as the ordered response variable $education\_index_i$ taking values from 0 to 6 and fit the (heuristically-specified) regression

$$education\_index_i = f(last\_name_i) + \epsilon_i \tag{1}$$

for individuals of voting age in 2011. We can then apply the estimated $\hat{f}(\cdot)$ to each name in an election ward-level voter list and use the within-ward average as a measure of SES in that ward-year combination.

To run (1) on the SECC data, we have to decide how to encode the names numerically. One might initially think of using a dummy variable for each name such that $\hat{f}(last\_name)$ would be the average value of the education index for individuals in the SECC with last name $= last\_name$.[21] However, this approach has two problems. First, the voter lists are image files. Figure (A.3) shows two examples from the 2002 elections. Therefore we will use Optical Character Recognition (OCR) software to convert the images to digital text. Occasional small errors in digitization are inevitable with this kind of software. So we want our encoding to be robust to the possibility of small misspellings. Second, particularly for infrequently-observed names, we can lower the prediction error of $\hat{f}(\cdot)$ by pooling information across names, which is known as shrinkage and underlies popular methods such as Empirical Bayes estimation, for example of teacher value-added (Morris, 1983; Chetty, Friedman, and Rockoff, 2011).

We therefore encode last names using "similarity encoding," recently suggested by Cerda, Kégl, and Varoquaux (2018). Cerda et al. (2018) propose encoding strings (last names for us) by identifying a number of prototypes and encoding the rest of the strings in terms of their distance to the prototypes. In principle, one can use any distance function between strings. We use the Levenshtein ratio which is a function of the number of edit operations needed to transform one name into another (the Levenshtein distance, Levenshtein (1966)), normalized by the sum of the number of characters in the two names. Following Cerda et al. (2018), so that the result can be interpreted as a positive measure of similarity, the ratio is subtracted from 1 and the all edit operations in the Levenshtein distance are given a weight of 1 except for the "replace" operation which is given a weight of 2:

$$sim_{levratio}(last\_name_1, last\_name_2) = 1 - \frac{d_{lev}(last\_name_1, last\_name_2)}{|last\_name_1| + |last\_name_2|}$$

where $|\cdot|$ indicates the number of characters in a name, and $d_{lev}(\cdot,\cdot)$ is our Levenshtein distance.

[20] We are grateful to Paul Novosad and Sam Asher for sharing these data.

[21] This is known as the "one-hot" encoder in the machine learning literature.

Levenshtein distance is commonly used in fuzzy string matching and to correct for misspellings, in popular packages like `fuzzywuzzy` (https://github.com/seatgeek/fuzzywuzzy), which has about 4,700 citing repositories and 270 citing packages on Github as of this writing.

The dimension of the encoding is equal to the number of prototype last names - dramatically smaller than the number of unique last names (252,617 in the SECC data we use), which would be the dimension of the matrix of last name dummies. With our encoding method set, we just need to specify our method of estimating $f(\cdot)$ in equation (1). We use Extremely Randomized Trees (Geurts, Ernst, and Wehenkel (2006)), a variant on Random Forests.

Implementation We selected the number of prototypes, the method for estimating $f(\cdot)$ from a set of alternatives, and the hyperparameters for each of the alternative $f(\cdot)$ estimators to minimize cross-validated prediction error on the SECC data. This procedure selects Extremely Randomized Trees with a maximum tree depth of 16 and 200 prototypes. We perform OCR using Google's Cloud Vision API, whose primary asset is its ability to digitize text in Marathi, the local language of Maharashtra which all 2017 voter lists and the large majority of 2002 voter lists are written in (see (A.3) panel (b)). We use Cerda et al. (2018)'s companion python package, `dirty_cat`, to compute similarity to each prototype name, and follow Cerda et al. (2018) in selecting the most common last names as prototypes.

Validation We validate our election-ward-level average education indices in two ways. First, we check their stability over time. Each point in Figure 4 represents an election ward, with the average education index in 2002 on the X-axis and the average education index in 2017 on the Y-axis. Both education indices have been normalized by subtracting the mean value at the election ward level in 2002 and dividing by the 2002 election-ward-level standard deviation. Reassuringly, the ranking of the election wards is quite stable over time with a Spearman rank correlation coefficient of 0.83. The mass around the 45 degree line shows that most changes are small, with more positive changes than negative ones which reflects the fact that average education levels were increasing over time.

Second, we compare literacy rates, which are available for 24 city wards from the 2001 and 2011 population censuses, with the average values of the education index at the city ward level from 2002 and 2017. The results are shown in Figure 5 with the same normalization as in Figure 4. The indices in both 2002 and 2017 correlate well with the corresponding literacy rates.

4.3 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 Fourth, Fifth and (newly available) Sixth Economic Census

(EC) of India. This covers the universe of establishments in 1998, 2005, and 2013 respectively. For the Sixth EC, 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. We hired a team of local enumerators to manually provide longitude and latitude coordinates for each of these blocks with two enumerators assigned to each block. We investigated discrepancies and took the average longitude and latitude when no preferred coordinates emerged.

Unfortunately for the Fourth and Fifth ECs, block maps have been destroyed. However, in conjunction with the Fifth EC, the Ministry of Statistics and Program Implementation publishes a directory of addresses for formal establishments with more than 10 workers. We clean and geolocate these addresses using the Google Maps API.[22]Using the information in the directory (industry code, employment in bins, city ward) we can imperfectly match these formal establishments to entries in the EC microdata. Since the microdata entries retain their geographic identifiers, when we match a from the directory to a row in the microdata we learn at least one set of longitude-latitude coordinates associated with that identifier.

Based on this idea, we devised an iterative algorithm that ultimately provides us with a set of polygons representing investigator units (IV units), which are the spatial units just larger than an enumeration block. As shown in Table 1, they are fortunately still quite small, covering only about 15,000 persons and dividing Mumbai and Mumbai Suburban districts into 825 spatial units. After we have aggregated our enumeration block level data from the Sixth EC up to the IV unit level, we have a unique spatial panel that covers the universe of firms (formal and informal) over the course of almost 20 years.

We construct data on residential population using the 2001 and 2011 censuses. As before, geographic coverage is typically only available by town. However, we found that we could access population in both years from the BMC Election Office at the election ward level.[23]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 600 "subzones" across Mumbai and Mumbai suburban districts. We digitized paper copies of the assessments going back to 1994.[24] While these official assessments are partially based on transaction records, we are currently collecting plot-level transaction

[22]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. We experimented with 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.

[23]The EO combine 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.

[24]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.

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.[25] 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.

4.4 Spatial Units

To summarize, our data use four sets of spatial units. Floorspace prices are at the subzone level, employment results at the IV unit level, and population and SES status at the election ward level. Population results use 2012 election wards to stick close to the 2001 and 2011 population census years, while SES status uses 2002 wards to allow for the long difference between 2002 and 2017. All spatial units are described in Table 1 (election ward characteristics are similar across years).

5 Reduced Form Results

This section presents our reduced form analysis of the impacts of the Mumbai Mill redevelopment. We begin with an event study to diagnose the presence of pre-existing trends in house prices around the DCR 58 law change that lifted the restrictions on mill redevelopment. We then present our main specifications, explore heterogeneous effects, and conclude with a number of falsification and robustness checks.

5.1 Event Study

We begin by using the one outcome available at an annual frequency–formal sector house prices–to shed light on pre-existing trends in neighborhoods with textile mills affected by the DCR 58 law change. We run the following event study specification:

$$\ln r_{it} = \alpha_i + \sum_{\tau=1993}^{2013} \beta_\tau \, \mathbb{I}\{\tau = t\} \times \text{ShareMill}_i + \gamma_{r(i)t} + \delta_t' X_i + \epsilon_{it}$$

where $\ln r_{it}$ is the log of residential floorspace prices in subzone i in year t, α_i is a subzone fixed effect, $\mathbb{I}\{\tau = t\}$ is a dummy for whether τ equals t, $\text{ShareMill}_i = \text{MillArea}_i/\text{Area}_i$ is the share of land taken up by mills in subzone i, $\gamma_{r(i)t}$ is a region by year fixed effect,[26] and X_i is a vector of controls

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

[26] Region takes on one of four values for being in the South, Center, North West or North East of the city. It is a union of the 24 adminstrative wards in the city. Using ward ward by year fixed effects instead gives us less power and more

for baseline characteristics (e.g. log distance to CBD, log area).

The β_τ coefficients capture the impact on house prices of having a greater share of mills in year τ relative to the baseline year. Our main concerns are that (i) the law change may have been anticipated, leading us to underestimate the size of its impacts and (ii) it may have been endogenous to local economic trends, for example if house prices were growing in these areas and mill owners lobbied lawmakers to change the law to allow them to benefit. Both would show up through increases in β_τ in the years leading up to DCR 58.

Figure 6 presents the event study graph. The trends in the β_τ coefficients tell a similar story to the anecdotal accounts the DCR 58 law change summarized in Section X. First, there is no significant pretrend in house prices in neighborhoods with mills in the years leading up to the sudden accouncement of the DCR 58 law change in 2001. Second, prices begin to appreciate between 2001 and 2006 during which the policy was accounced but substantial uncertainty remained as it was contested in the courts. Third, after the 2006 Supreme Court decision to uphold the law change, house prices sustain a pronounced and gradually increasing appreciation. Taken together, this results suggests the lifting on redevelopment regulations on the mill sites was both unanticipated and exogenous to local economic conditions.

5.2 Main Results

Our primary aim is to estimate the impact of the redevelopment of mill sites on economic activity, distinguishing between the direct effect on millsites themselves and spillovers on surrounding neighborhoods. Our main specification uses a simple and transparent regression to examines how outcomes change with distance to the closest millsite. However, since two locations the same distance from their closest mill may have different exposure to redeveloped sites (e.g. if one is surrounded by mills on three sides), we also run a second specification that uses a multilateral exposure measure to capture this variation and typically increases precision.

Distance Bands. Our main specification runs the following regression:

$$\begin{aligned} Y_{it} = {} & \beta_1 \, \mathbb{I}\{\text{ContainsMill}\}_i \times \text{Post}_t + \beta_2 \, \mathbb{I}\{\text{1m-500m}\}_i \times \text{Post}_t + \beta_3 \, \mathbb{I}\{\text{501m-1km}\}_i \times \text{Post}_t \\ & + \alpha_i + \gamma_{r(i)t} + \delta_t' X_i + \epsilon_{it} \end{aligned} \tag{2}$$

where Y_{it} is an outcome for location i in year t, Post_t is a dummy for whether t is in the post-period after 2006, and the indicator variables $\mathbb{I}\{\cdot\}$ are dummies for whether i contains any mill, or whose boundary is 1-500m or 501-1000m away from the boundary of the closest mill site respectively. Note that the coefficient on $\mathbb{I}\{\text{ContainsMill}\}_i$ potentially conflates both direct and spillover effects since no

imprecise estimates, but we show the results are qualitatively similar in Section X.

geographic unit contains exclusively mills. However, the latter two dummies capture pure spillover effects since they turn on for locations that contain no mills.

Table 2 presents the results. We begin by examining the impact on land markets. First, column (1) shows a pronounced impact on formal sector house prices that rise 13% in locations that contain a mill site, and decay smoothly with space thereafter: locations without whose boundary is within 500m of a mill site (but contain no mill site themselves) see an 8% increase in house prices, decreasing to 5% 500m-1km away (albeit statistically indistinguishable from zero). Second, in column (3) slum share falls in precisely the same bands, commensurate with the increasing return to redeveloping slums to formal housing in locations where formal house prices are rising. Slum share falls by 22% in locations that contain a mill and by 25% in locations within 500m of one. The smaller coefficient on $\mathbb{I}\{\text{ContainsMill}\}_i$ is consistent with the fact that mill sites contain no slums themslelves: the total effect is a weighted average of the effect on mill sites (zero change) and areas outside (negative change), which attenuates the effect. Once we adjust for this, the effect is monotonic.[27] Indeed the coefficient on $\mathbb{I}\{\text{ContainsMill}\}_i$ here exclusively captures spillovers.

We then examine the impact on employment in columns (4) and (5). Locations containing and near to mills experience large decreases of 48% and 41% in employment taking place in informal establishments respectively. This is intuitive given that informal employment is more likely to take place in slums (see Table A.1), so that when slums are converted the informal establishments which were located there are also displaced. In contrast, we see no major impacts on formal sector employment in column (5), although the effects are noisy. We will revisit employment impacts more thoroughly in Section 5.3.

Lastly, we explore the impact of mill redevelopment on demographic composition in surrounding neighborhoods. Column (6) shows the education index by 0.27 standard deviations in locations containing wards, while population density falls by 12% in column (2). This is consistent with a process of gentrification where richer and more-educated residents who consume more housing floorspace per capita sort into these neighborhoods, displacing the previous poor residents who lived in high density slums. There are no significant spillover effects for either outcome, although the geographic units (election wards) used in both regressions are the largest of all datasets (see Table 1) and are therefore more prone to co-mingling the own and spillover effect in the first row. The next section will sharpen both outcomes by generating multilaterial exposure measures that exploit the full spatial heterogeneity in exposure to mills.

Taken together, these results show the mill redevelopments had substantial impacts on surrounding neighborhoods suggestive of gentrification. Formal sector house prices rose, slums were cleared,

[27] In particular we construct the adjusted share of slums on non mill land as $\widetilde{SlumShare}_i = \text{SlumArea}_i/(\text{Area}_i - \text{MillArea}_i)$. When we do this, we find a monotonic relationship with coefficients of -0.262 (0.105) and -0.243 (0.094) on $\mathbb{I}\{\text{ContainsMill}\}_i \times \text{Post}_t$ and $\mathbb{I}\{\text{1m-500m}\}_i \times \text{Post}_t$ respectively. We retain the simpler measure in the results since it is more transparent.

informal employment fell, and the share of high-skilled residents increased.

Multilaterial Mill Exposure Measure. In addition to the simple measure used above, we create a continuous treatment effect measure that captures the overall exposure to mills nearby. For example, two mills both 500m away from their closest mill may have very different exposure to redevelopments if one is surrounded by new mills 500m on all four sides while the other is 500m from a single mill on one side. To capture this, we define the measure of mill exposure as

$$\text{MillExposure}_i = \sum_{j \notin i} \pi_{ij} \, \text{MillDensity}_j$$
$$\text{where } \pi_{ij} = \frac{\exp(-\kappa \cdot \text{dist}_{ij})}{\sum_s \exp(-\kappa \cdot \text{dist}_{is})}$$

and $\text{MillDensity}_i = \text{MillArea}_i / \text{Area}_i$. Locations indexed by j are a grid of 21,000 hexagon grid cells with 150m2 area each, which are consistent across the different geographies for locations i that depend on the particular outcome in consideration. This reduces noise introduced by the different size of spatial units in the geographies; if we just distance from location i to all other locations j in the geography, it would be as if the mill density was all located at the centroid of the respective spatial unit, introducing larger measurement error for more aggregated geographies. Since our aim is to use this to estimate spillovers, we exclude cells that are within location i so that the measure only captures exposure to mills outside a location itself. We measure distance in km and, for simplicity, set $\kappa = 1$ implying a location 1km away gets 22% of the weight of somewhere 0km away. We vary this decay rate in robustness checks.

The benefit of this measure is that (i) it is independent of the size of geography used and (ii) it captures the full spatial variation in exposure to mills on all sides of each spatial unit. It can be thought of the average mill density surrounding a location i, and can be formally derived from first order approximations to functions where outcomes depend on housing quality nearby that decays exponentially with distance (e.g. Ahlfeldt et. al. 2015).[28]

Table 3 reports results from the same regression as (2) but replacing the treatment effect measures with $\text{MillExposure}_i \times \text{Post}_t$. MillExposure_i is normalized so that the coefficient in each column can be interpreted as the impact of increasing exposure by one sd on each outcome. The first row shows similar qualitative effects (yet more precise) as with the baseline distance to boundary measure: be-

[28]For example, suppose house prices in i depend on some neighborhood attributes u_i as well as the quality of housing nearby Q_j that decay exponentially with distance so that $r_i = f\big(u_i, \sum_j \exp(-\kappa \cdot \text{dist}_{ij}) Q_j\big)$. A log-linear approximation then yields that

$$d \ln r_i = \alpha \sum_j \frac{\exp(-\kappa \cdot \text{dist}_{ij}) Q_j}{\sum_s \exp(-\kappa \cdot \text{dist}_{is}) Q_s} d \ln Q_j + (1-\alpha) d \ln u_i$$

The MillExposure_i measure approximates the first term by setting $Q_j = 1 \; \forall j$ in the share terms and proxying the percentage change in housing equality with the density of mills $d \ln Q_j \propto \text{MillArea}_j / \text{Area}_j$.

ing exposed to more mills nearby leads (i) house prices and the education of residents to rise, (ii) the share of slums, informal employment density and residential population density to fall and (iii) no effect on formal employment density. While the mill exposure measure only considers distance mills located outside a polygon itself, this may still be correlated with the presence of mills inside it. The second row therefore adds a dummy for whether the location contains a mill (interacted with the post-period dummy); the spillover effects captured by the coefficient on MillExposure_i remain important.

Instrumental Variable Specifications. The previous results report reduced form effects of exposure to mill sites, not all of which were necessarily developed. Tables A.2 and A.3 in the Appendix report IV specifications which use the distance to all mill sites to instrument for the distance to redeveloped mill sites, and the mill exposure measure to all mill sites defined above to instrument for the mill exposure measure to redeveloped mill sites. These simply scale the reduced form results in Tables 2 and 3 by the appropriate first stage coefficients. House price and slum coefficients tend to be slightly attenuated in the IV distance to boundary specification, suggesting developers chose to redevelop more valuable mill sites (we provide more evidence on this in Appendix X). For other outcomes coefficients seem roughly stable across OLS and IV.

5.3 Heterogeneity: Employment and Slum Impacts

Employment Impacts. Our previous results show little impact on formal employment either in or around mill areas, albeit with some imprecision. Standard theories of agglomeration would suggest that in productivity (and hence employment) should rise in locations experiencing exogenous shocks in employment density. Table 4 tests for whether the reduced form effects from the previous section mask such heterogeneity by exploring whether effects differ by whether a mill was redeveloped for residential or commercial purposes. [Note: use is endogenous, instrument forthcoming.]

First, consider the impact on formal employment. Column 5 shows that locations containing commercially-developed mills experience a large (approximately 135%) increase in formal employment density. This doubles up as a nice validation check: locations we record as redeveloping as commercial indeed experience large increases in formal sector employment in the economic census. However, the remaining rows show no significant increases in formal employment in proximate locations. Either there would appear to be small spillovers from formal employment, or perhaps these effects are long-run phenomena we are unable to pick up with our data,[29] but formal sector agglomeration spillovers do not jump out of our data. Column 6 shows no substantial impact on formal employment density close to residential mills. Taken together, there are large increases in formal

[29] Mills are classified as residential/commercially redeveloped if they had opened in that use by 2008, and our post-period for employment is 2013.

employment density in locations containing commercially-redeveloped mills but nowhere else.

Columns 3 and 4 explore changes in informal employment density by distance to each mill type. Column 4 shows a similar pattern to the reduced form results: locations containing or close to residentially-redeveloped mills experience large falls in informal employment density of around 50%. Column 3 shows a much more muted effect on informal employment in places with commercial redevelopments, but a similar effect nearby. This is suggestive of within-location spillovers between formal and informal sectors, as the reduction in informal employment is more muted in places which experiences large increases in formal employment. [Note: statistical comparison forthcoming.] Finally, columns 1 and 2 show the net effect of these forces: total employment in general falls in locations containing and nearby mill sites, except in places containing commercial mill sites where formal employment increases.

Table 5 explores how these effects varied across sectors. Locations containing mills saw large increases in formal employment in establishments in high-skilled services and retail industries. In locations nearby mills, reductions in employment were largest in informal manufacturing, followed by informal retail and formal manufacturing.

Slum Impacts. The nature of adjustment costs for slums is ambiguous. It could be that larger slums are harder to convert due to their larger voting power, increased incentives for coordination through local interest and representation groups, or an increased salience within minds of city residents that makes their clearance harder to fly under the radar. However, the return to developing large slums may be higher if, for example, developers need large plots to construct formal buildings.

Table 6 examines how the impact of proximity to mills on the share of land taken up by slums varied by initial slum area. In general, column 1 shows the reduction in slum share was greatest for initially small slums, although this heterogeneity only holds in locations closest to mills. This could be a purely mechnical effect: if all slums lost the same area, then areas with initially smaller slums would experience larger reductions in the slum share. Column 2 shows this does not appear to be the case, since initially smaller slums tend to lose a smaller area of slums too. Taken together, this suggests adjustment costs to redevelop slums are increasing in slum size and it is smaller slums that are more vulnerable to clearance.

5.4 Falsification Checks

This section provides three additional pieces of evidence supporting the causal nature of the redevelopment impacts we document. First, we show the effects are caused by distance to the mill sites themselves rather than distance to large industrial buildings. This rules out that our effects are capturing secular trends in genrification that often see large industral buildings becoming redeveloped (e.g. the meatpacking district or waterside parts of Brooklyn in New York). Second, we use an addi-

tional 2005 wave of the economic census to show the employment impacts occur only after DCR 58 was upheld by the Supreme Court. This establishes an absence of pretrends in employment growth, supporting the patterns of Figure 6 in a seperate dataset. Third, we show there is no substantial increase in roads near the mill sites. This was the only public good for which spatially disaggregated panel data was available. One might worry that, even if this policy change was exogenous to pre-existing trends in economic activity, the government may have provided complementary investments in public goods to maximize the gains from this opportunity which would lead us to overestimate the true effect of the mill redevelopment itself. At least for the one public good we could gather data on, this does not appear to be the case.

Industrial Land Placebo Check. To examine whether our effects are driven simply by proximity to large industrial buildings, we use the 1991 Existing Land Use map of Mumbai to identify areas used for industrial purposes as of the most recent year available before DCR 58. We select sites larger than 10,000m2, the size of the smallest mill site, to ensure we are comparing similarly sized industrial plots. Table 7 shows the results. While we are sometimes underpowered to identify the seperate impacts, but overall the mill site effects tend to be robust to controlling for distance to other large industrial sites. House prices do tend to grow mildly in locations with industrial sites, but there are no spillover effects. Other outcomes–slums, informal and formal employment, and skill mix of residents–are completely unaffected by distance to industrial sites.

Employment Pre-Trend Check. Our baseline specifications examine employment growth between 1998 and 2013 using the 4th and 6th rounds of the economic census. Table 8 incorporates the 5th economic census from 2005 and extends the baseline specification to replace post-period dummies with year dummies to examine how the effects of proximity to mills on employment growth evolve between 1998, 2005 and 2013. We find precise zero differential employment growth in informal establishments in neighborhoods close to mills between 1998 and 2005, but sharp falls of around 50% between 2005 and 2013. As in the main specifications there is little movement in formal employment, except for a mild increase by 2005 that evaporates by 2013. Taken together, the impacts ovewherlmingly show up between 2005 and 2013 with little employment change in areas closer to mills in the lead up to DCR 58 between 1998 and 2005.

Changes in Roads. Finally, we examine whether local government responded to the law change by investing in public goods near millsites that may have confounded the true effects of the mills themselves. We were able to access Eicher maps of Mumbai made in 2003 and 2013 (the earliest and latest years available). These provide detailed maps of the city which color code roads by type. We digitized these to compute the area of land in each iv unit taken up by main roads, and regress this

on our distance to mill measures as well as baseline controls. Table 10 shows no significant changes in road construction related to the mill sites.

While this data on roads was the only available fine-grained spatial data on public goods, there are other reasons to think pubic policy would have been sluggish to respond. Mumbai's development plans which determine land use regulations (such as zoning and height restrictuions) are very infrequently updated: the first was sanctioned in 1967, the second in 1991, and the third in 2018. No major changes to other to related laws, such as the rent control act, occurred over the period.

5.5 Robustness Checks

We assess the sensitivity of our results to alternative specifications in robustness checks. Table A.5 adds the full set of controls sequentially across three specifications. The primary outcomes for which we see more precise effects in the previous tables–house prices, slums, informal employment and education–are stable throughout. Table A.7 examines how the results vary with finer spatial fixed effects at the ward level. We have less power to identify effects using only variation within wards: most coefficients fall around 20-30% with increases around twice as large for standard errors. The end result is qualitatively similar to when we include region fixed effects only, but less precise. Lastly, Table A.9 uses alternative decay rates when computing the mill exposure measure. The effects are economically and statistically stable throughout.

6 Model

To explain and quantify the forces described above, this section presents a general equilibrium model of a city. The key difference with existing quantitative urban models (e.g. Ahlfeldt et al. (2015)) is that we incorporate informal housing and informal employment. We note here that this section is currently under heavy revision, and that estimation results are forthcoming.

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 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.

6.1 Workers

There is a fixed number of high- and low-skilled residents in the city denoted by $\bar{L}_g$ for $g \in G = \{H, L\}$.[30] 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.[31] Standard results imply that the mass of workers living and working in different locations is given by

$$L_{Rig}^k = \lambda_g^U (u_{ig}^k (r_{Ri}^k)^{\beta-1})^\theta \Phi_{Rig} \tag{3}$$

$$L_{Fjg} = \lambda_g^U w_{jg}^\theta \Phi_{Fig} \tag{4}$$

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.[32] Overall worker welfare $\bar{U}_g$ is given by

$$\bar{U}_g = \gamma \left[\sum_{i,j,k} \left(u_{ig}^k w_{jg} (r_{Ri}^k)^{\beta-1} / d_{ij} \right)^\theta \right]^{1/\theta}. \tag{5}$$

Average income of residents of i is determined by the probability of commuting to different em-

[30] 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).

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

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

ployment 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

$$r^k_{Ri} = (1-\beta)\frac{\sum_g \bar{w}_{ig} L^k_{Rig}}{H^k_{Ri}} \tag{6}$$

6.2 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

$$L_{Fjk} = \left(\frac{\alpha A_{jk}}{w_j}\right)^{\frac{1}{1-\alpha}} H^k_{Fj} \tag{7}$$

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

$$r^k_{Fj} = (1-\alpha) A_{jk}^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w_j}\right)^{\frac{\alpha}{1-\alpha}} \tag{8}$$

6.3 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

$$\begin{aligned} \vartheta_{ki} = 1 \quad & r^k_{Ri} > (1-\xi_i) r^k_{Fi} \\ \vartheta_{ki} \in (0,1) \quad & r^k_{Ri} = (1-\xi_i) r^k_{Fi} \\ \vartheta_{ki} = 0 \quad & r^k_{Ri} < (1-\xi_i) r^k_{Fi} \end{aligned} \tag{9}$$

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.

6.4 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 $H^F_i = T_i^{1-\eta} K_i^{\eta}$, 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 prob-

lem for an owner with profit shock vector ϵ is given by

$$\max\left\{\pi_i^F\epsilon^F, \pi_i^I\epsilon^I, \epsilon^V\right\}.$$

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

$$\lambda_{Fi} = \frac{\tilde{\tau}_i(r_i^F)^{\frac{\kappa}{1-\eta}}}{1+\tilde{\tau}_i(r_i^F)^{\frac{\kappa}{1-\eta}}+(r_i^I)^\kappa} \tag{10}$$

$$\lambda_{Ii} = \frac{r_{Ii}^\kappa}{1+\tilde{\tau}_i(r_i^F)^{\frac{\kappa}{1-\eta}}+(r_i^I)^\kappa} \tag{11}$$

$$\lambda_{Vi} = 1-\lambda_{Fi}-\lambda_{Ii}. \tag{12}$$

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

$$H_i^F = \lambda_{Fi}T_i\tilde{\eta}(r_i^F)^{\frac{1}{1-\eta}} \tag{13}$$

$$H_i^I = \lambda_{Ii}T_i \tag{14}$$

6.5 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 (4) is consistent with demand for labor by firms (7),

2. **Floorspace Market Clearing** The market for residential floorspace for each housing type clears (6) and its price is consistent with residential populations (3), firms earn zero profits (8) and floorspace shares are consistent with landowner optimality (9),

3. **Land Use and Floorspace Supply** The share of land allocation to formal and informal use in each location (10-11) and the supply of floorspace on each type of land use (13-14) is consistent with landowner optimality.

4. **Closed City** Populations add up to the city total, i.e. $\bar{L}_g = \sum_{i,a} L_{Riag}\ \forall g$.

6.6 Introducing Spillovers in Amenities and Productivity

A long literature points to the importance of spillovers in cities.[33] 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

$$u_{ig}^{k} = \bar{u}_{ig}^{k} \left[\sum_{j} d_{ij}^{-\delta_U} (H_{Fi}/T_i) \right]^{\mu_{U.g}} \tag{15}$$

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.[34]

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

$$A_{jk} = \bar{A}_{jk} \left[\sum_{s} d_{js}^{-\delta_A} (L_{Fs}/T_s) \right]^{\mu_A} \tag{16}$$

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.

[33] This idea dates back at least to Adam Smith (1776), and was articulated more fully in Marshall (1890).

[34] 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.

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

To be added.

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.

Arkolakis, C., S. K. Lee, and M. Peters (2019). European Immigrants and the United States' Rise to the Technological Frontier. *Working Paper*, 1831524.

Athey, S. and G. W. Imbens (2019). Machine Learning Methods Economists Should Know About. (March).

Barnhardt, S., E. Field, and R. Pande (2016). Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India. *American Economic Journal: Applied Economics, forthcoming*.

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

Bhagavatula, S., M. Bhalla, M. Goel, and B. Vissa (2019). Cultural Diversity in Corporate Boards and Firm Outcomes.

Brueckner, J. K. and S. V. Lall (2015). Cities in Developing Countries. Fueled by Rural-Urban Migration, Lacking in Tenure Security, and Short of Affordable Housing. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 1399–1455. Elsevier B.V.

Cerda, P., B. Kégl, and G. Varoquaux (2018). Similarity encoding for learning with dirty categorical variables. *Machine Learning 107*(8), 1477–1494.

Chetty, R., J. Friedman, and J. Rockoff (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.

Diamond, R. and T. Mcquade (2015). Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low Income Property Development. *Mimeo*, 1–59.

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

Donaldson, D. and A. Storeygard (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives 30*(4), 171–198.

Duranton, G. and D. Puga (2020). The Economics of Urban Density. *Journal of Economic Perspectives, Forthcoming*.

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*.

Field, E. (2007). Entitled to Work: Urban Property Rights and Labor Supply in Peru. *Quarterly Journal of Economics 122*(4), 1561–1602.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras & Tensorflow*.

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

Geurts, P., D. Ernst, and L. Wehenkel (2006). Extremely randomized trees. *Machine Learning 63*, 3–42.

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.

Harari, M. (2016). Cities in Bad Shape : Urban Geometry in India. *Working Paper*.

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

Hornbeck, R. and D. Keniston (2014). Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872. *NBER 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).

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady 10*, 707–710.

Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A 374*.

Marx, B., T. Stoker, and T. Suri (2013, nov). The Economics of Slums in the Developing World. *Journal of Economic Perspectives 27*(4), 187–210.

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

Morris, C. N. (1983). Parametric Empirical Bayes Inference : Theory and Applications. *Journal of the American Statistical Association 78*(381), 47–55.

Olivetti, C. and M. D. Paserman (2015). In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review 105*(April), 1–31.

Owens III, R., E. Rossi-hansberg, and P.-D. Sarte (2017). Rethinking Detroit. pp. 1–45.

Redding, S. J. and D. M. Sturm (2016). Estimating Neighborhood Effects: Evidence from War-Time Destruction in London. *Working Paper*.

Rosenthal, S. S. and W. Strange (2019). How Close Is Close? The Spatial Reach of Agglomeration Economies Stuart. *Journal of Economic Perspectives, Forthcoming*.

Rossi-Hansberg, E., P.-D. Sarte, and R. Owens (2010). Housing Externalities. *The Journal of Poitical Economy 118*(3), 485–534.

Tsivanidis, N. (2017). Commuting Technologies, City Structure and Urban Inequality: Evidence from Bogotá's TransMilenio. *Working Paper*.

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

Figures

Figure 1: Apollo, Simplex and Hindoostan Mills

(a) 2000

(b) 2016

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

(a) SRA Slums

(b) Predicted Slums

Figure 3: Validation Sample: SRA vs CNN Slums

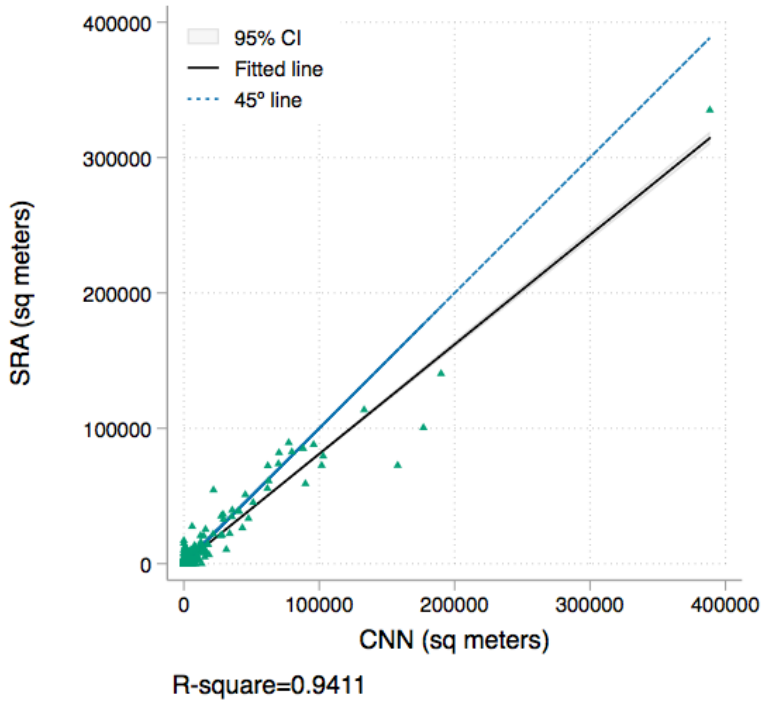


Figure 4: Education index over time

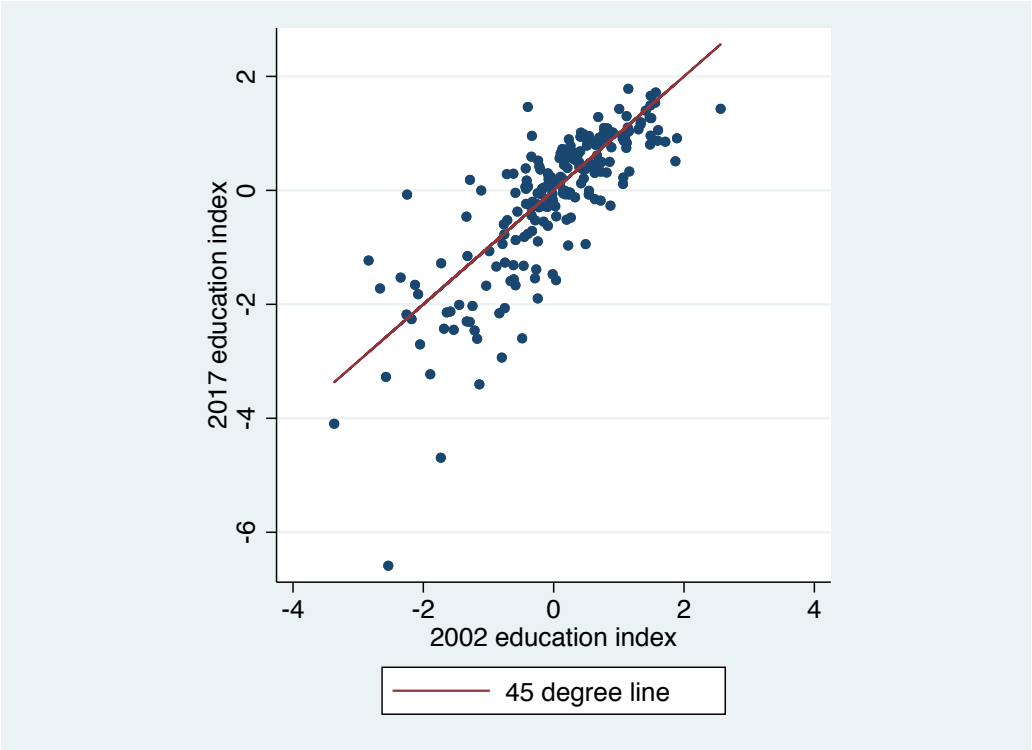
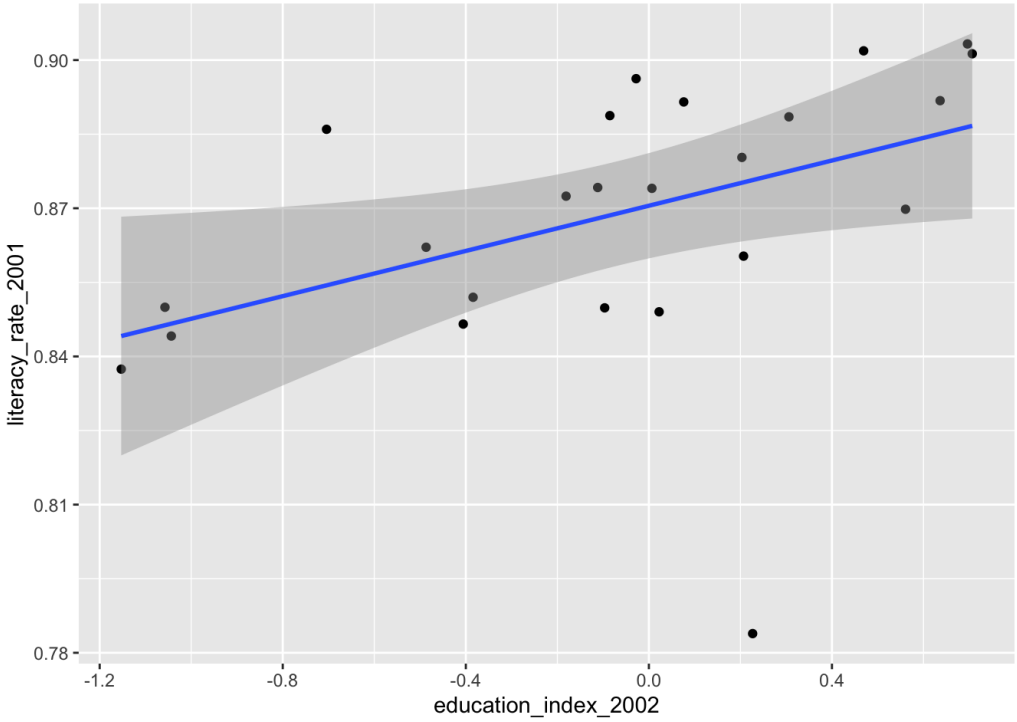
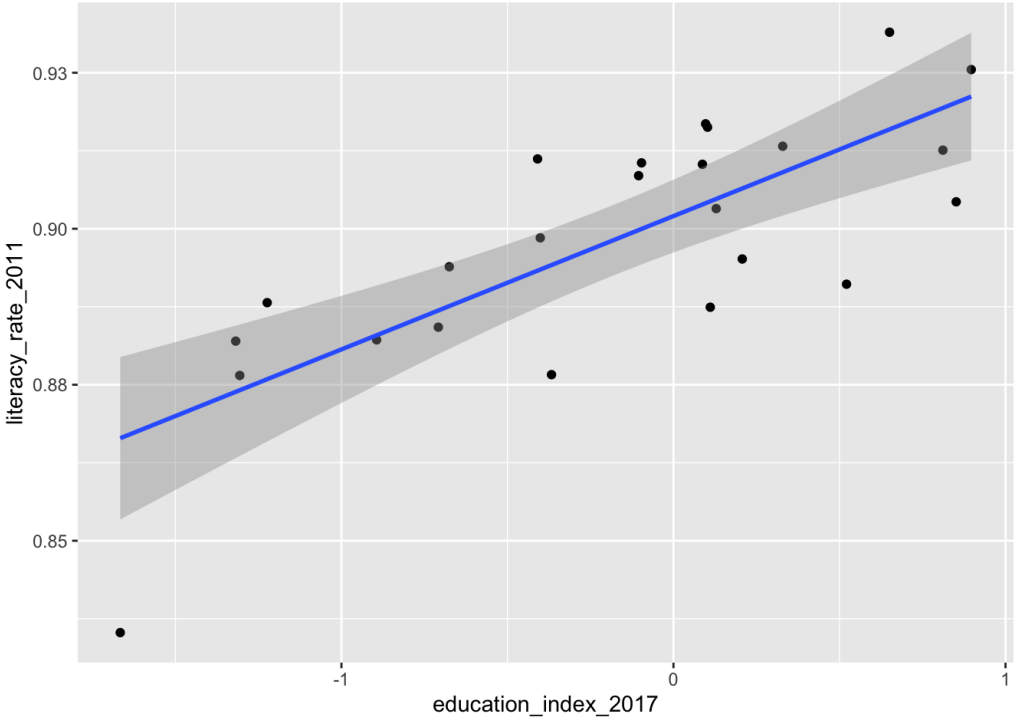


Figure 5: Ward-level education index vs. literacy rate from population census

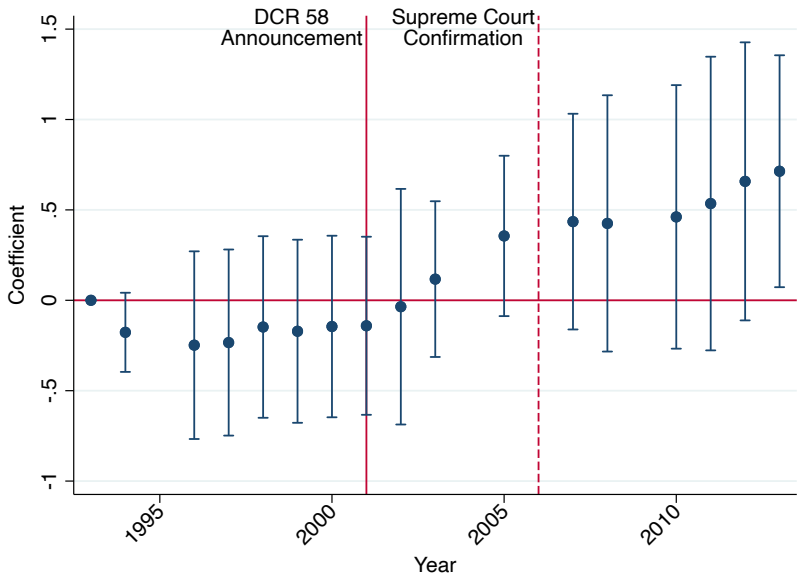


(a) 2002 education index, 2001 population census literacy rate



(b) 2017 education index, 2011 population census literacy rate

Figure 6: Event Study: House Price Growth Around DCR 58



Tables

Table 1: Characteristics of Geographic Units

	Election Wards	Subzones	IV Units
Number of units	227	574	818
Area(m2)	1,811,829	825,937	367,691
Population	54,812	18,816	14,833

Table 2: Main Results

	Res Price	Population	Slum Share	Inf Emp	Form Emp	Ed Index
Contains Mill	0.131**	-0.127**	-0.221**	-0.480**	0.318	0.270**
	(0.058)	(0.064)	(0.104)	(0.189)	(0.332)	(0.126)
1m-500m	0.081*	-0.098	-0.247***	-0.415**	0.181	-0.223
	(0.047)	(0.068)	(0.095)	(0.203)	(0.306)	(0.177)
500m-1000m	0.052	-0.069	0.095	0.214	0.742*	-0.156
	(0.045)	(0.100)	(0.111)	(0.252)	(0.444)	(0.198)
N	1,116	454	894	500	464	424

Note: Outcomes are log residential house price, log population density, slum share, log employment in informal and formal establishments, and standardized education index. All specifications include location fixed effects, region by period fixed effects, and full controls interacted with period fixed effects. These controls consist of (i) exogenous characteristics (log distance to CBD, log area, log road density (plus one)) in all specifications and (ii) economic activity in the pre-period, excluding the lagged value of the outcome variable. More precisely, house price specifications control for log initial employment density, log initial population density and initial slum share; slum regressions control for log initial house prices, log initial employment density and log initial population density; employment regressions control for log initial population density, log initial house prices and initial slum share; population and education index regressions control for log initial employment density, log initial house prices and initial slum share. The slum share regression is estimated via PPML to account for zeros in the data; the remainder are estimate via OLS. Precise years of pre- and post-period vary by dataset: house prices (pre:1993-2000 average, post: 2007-2013 average), population (pre: 2001, post: 2011), slums (pre: 2001, post: 2016), employment (pre: 1998, post: 2013), education index (pre:2002, post:2016). Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3: Main Results, Mill Exposure Measure

	House Prices	Slums	Informal Emp	Formal Emp	Educ	Population
Mill Exposure	0.041***	-0.259***	-0.283***	0.042	0.150***	-0.047**
	(0.013)	(0.060)	(0.080)	(0.132)	(0.047)	(0.020)
N	1,142	954	500	464	424	454
Mill Exposure	0.035**	-0.270***	-0.312***	0.021	0.135**	-0.054**
(ctrl Contains Mill)	(0.015)	(0.066)	(0.103)	(0.170)	(0.060)	(0.025)
N	1,142	954	500	464	424	454

Note: See Table 2 for full description, specification is the same except for treatment effect measure. Second run additionally includes a dummy for whether a location contains any mill interacted with a post-period dummy. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4: Employment Impacts: Heterogeneity by Mill Use

	All Emp	All Emp	Inf Emp	Inf Emp	Form Emp	Form Emp
Contains Mill	-0.160	-0.594**	-0.276	-0.510**	1.359**	-0.187
	(0.462)	(0.266)	(0.379)	(0.213)	(0.542)	(0.409)
1m-500m	-0.616**	-0.550**	-0.578**	-0.566***	0.094	0.064
	(0.297)	(0.216)	(0.263)	(0.187)	(0.415)	(0.294)
500m-1000m	-0.326	-0.262	-0.342	0.004	0.291	-0.382
	(0.267)	(0.311)	(0.240)	(0.246)	(0.416)	(0.402)
N	500	500	500	500	464	464
Mill Type	Comm	Res	Comm	Res	Comm	Res

Note: Outcome is at the formality X location level, except for column 1 which is the change in total employment at the location level. Controls and fixed effects otherwise the same as in Table 2. For each outcome of log employment in all, informal and formal establishments, odd (even) columns run baseline regressions against distance bands to commercially- (residentially-) redeveloped mills. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 5: Employment Impacts: Heterogeneity by Sector

	All Emp	Form Emp	Inf Emp
Contains Comm Mill X			
HS Services	1.412**	2.187***	0.496
	(0.569)	(0.355)	(0.444)
Manufacturing	-0.272	0.371	-0.414
	(0.457)	(0.527)	(0.375)
Retail	-0.116	1.263**	-0.090
	(0.465)	(0.495)	(0.412)
Other	-0.262	1.061**	-0.411
	(0.467)	(0.434)	(0.368)
<500m Comm Mill X			
HS Services	0.242	0.098	0.022
	(0.381)	(0.349)	(0.334)
Manufacturing	-1.485***	-0.872**	-1.324***
	(0.366)	(0.415)	(0.321)
Retail	-0.703**	0.044	-0.621**
	(0.348)	(0.447)	(0.295)
Other	-0.524	0.160	-0.573**
	(0.321)	(0.375)	(0.277)
500m-1000m Comm Mill X			
HS Services	0.605	0.450	0.283
	(0.398)	(0.419)	(0.328)
Manufacturing	-0.880**	-0.545	-0.734**
	(0.352)	(0.362)	(0.315)
Retail	-0.665**	0.227	-0.458*
	(0.303)	(0.378)	(0.259)
X Other	-0.235	0.077	-0.264
	(0.281)	(0.315)	(0.249)
N	1,996	1,668	1,993

Note: Outcome is at the industry X formality X location level, except for column 1 which is the change in total employment at the industry X location level. Each specification include location X sector fixed effects, region X sector X Post-period fixed effects, and baseline controls. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 6: Slum Impacts: Heterogeneity by Area

	Slum Share	Slum Area
containsMills X		
Q1	-0.593**	-0.463**
	(0.266)	(0.211)
Q2	-0.417*	-0.323
	(0.247)	(0.199)
Q3	-0.069	-0.026
	(0.084)	(0.100)
1m-500m X		
Q1	-0.299	-0.177
	(0.190)	(0.173)
Q2	-0.233*	-0.147
	(0.140)	(0.127)
Q3	-0.186	-0.168
	(0.119)	(0.132)
500m-1000m X		
Q1	0.119	-0.162
	(0.270)	(0.236)
Q2	-0.025	-0.262
	(0.159)	(0.199)
Q3	0.018	0.034
	(0.143)	(0.126)
N	878	878
Tercile Dummies X Post	X	X
Full Controls X Post	X	X
FE	Region	Region

Note: Q are dummies for tercile of the initial slum area distribution, Q3 being locations with the largest initial slums. Full interactions are included in controls. Each row is the coefficient on the variable interacted with Post dummy. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 7: Falsification Check: Distance to Large Industrial Sites

	Res Price	Res Price	Slum Share	Slum Share	Inf Emp	Inf Emp	Form Emp	Form Emp	Educ	Educ
Contains Mill	0.154***	0.104*	-0.233**	-0.172	-0.467**	-0.548**	0.197	0.237	0.270**	0.257**
	(0.058)	(0.059)	(0.111)	(0.105)	(0.191)	(0.228)	(0.329)	(0.383)	(0.126)	(0.126)
<500m	0.092**	0.073	-0.218**	-0.181*	-0.496**	-0.519**	0.017	0.071	-0.223	-0.231
	(0.046)	(0.048)	(0.089)	(0.096)	(0.199)	(0.215)	(0.311)	(0.351)	(0.177)	(0.172)
500m-1000m	0.060	0.054	0.092	0.095	0.242	0.228	0.550	0.590	-0.156	-0.186
	(0.045)	(0.043)	(0.108)	(0.111)	(0.256)	(0.267)	(0.430)	(0.453)	(0.198)	(0.199)
Contains Mill Ind Land		0.076*		-0.052		0.227		0.083		-0.183
		(0.042)		(0.129)		(0.264)		(0.452)		(0.178)
<500m Ind Land		-0.053		0.097		-0.003		-0.004		-0.202
		(0.048)		(0.086)		(0.247)		(0.415)		(0.254)
500m-1000m Ind Land		-0.008		0.076		0.246		0.310		0.122
		(0.062)		(0.073)		(0.244)		(0.408)		(0.127)
N	1,142	1,142	954	954	500	500	464	464	424	424

Note: Controls, fixed effects and first three rows are the same as in the main tables. Rows 4 to 6 contain the same distance measure but instead classify based on the distance from a polygon's boundary to the closest large industrial use site per the 1991 existing land use map of Mumbai. These sites are defined as those from that map larger than 10000m2, approximately the minimum size of mill sites themslves. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 8: Falsification Check: Employment Growth Pre- and Post-DCR 58

	Inf Emp	Form Emp
Contains Mill X 2005	-0.009	-0.127
	(0.093)	(0.206)
Contains Mill X 2013	-0.468**	0.166
	(0.190)	(0.331)
0m-500m X 2005	0.051	0.356*
	(0.096)	(0.215)
0m-500m X 2013	-0.489**	-0.048
	(0.197)	(0.309)
500m-1000m X 2005	0.168	-0.161
	(0.104)	(0.289)
500m-1000m X 2013	0.241	0.491
	(0.256)	(0.431)
N	758	727

Note: Relative to the baseline specification, data from 1998, 2005 and 2013 are included. Year dummies are therefore used in place of post-period dummies. Standard errors clustered at the location-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 10: Falsification Check: Changes in Roads 2003-2013

	dlog(Main Roads)
Contains Mill	-0.088
	(0.116)
1m-500m	0.111
	(0.124)
500m-1000m	-0.115
	(0.125)
N	352
FE	Region
Dist CBD, Area Ctrls	X

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: Road data is from 2003 and 2013 Eicher maps, digtized and intersected with iv units. Regression runs the log change in area taken up by main roads in each IV unit in Mumbai District against the distance to boundary dummies, controlling for region fixed effects, log distance to CBD and log area. Standard errors clustered at the location-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

A Additional Tables

Table A.1: Characteristics of Slums

	Log(Emp Density)	Log(Pop Density)	Log(Emp/Pop)	Log(Avg Establishment Size)	Manuf share
Slum share	0.650***	0.219***	0.438***	-0.117***	0.168***
	(0.069)	(0.031)	(0.071)	(0.027)	(0.012)
N	6,525	6,489	6,489	6,525	6,525

Note: Slum share indicates the share of buildings that are slums. All specifications include region fixed effects and controls for log distance to CBD and area quintiles. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.2: IV-Distance to Boundary

	House Price	House Price	Pop Dens	Pop Dens	Slum Share (PPML)
Contains Mill	0.229***	0.131**	-0.033	-0.127**	-0.347***
	(0.070)	(0.058)	(0.065)	(0.064)	(0.125)
<500m	0.050	0.081*	-0.141**	-0.098	-0.309***
	(0.048)	(0.047)	(0.067)	(0.068)	(0.095)
500m-1000m	0.050	0.052	0.121	-0.069	0.013
	(0.055)	(0.045)	(0.076)	(0.100)	(0.117)
N	1,116	1,116	454	454	894

	ln(Slum Share)	ln(Slum Share)	Inform Emp	Inform Emp	Form Emp	Form Emp
Contains Mill	-0.332**	-0.234**	-0.471**	-0.480**	0.259	0.318
	(0.132)	(0.093)	(0.216)	(0.189)	(0.392)	(0.332)
<500m	-0.180**	-0.090	-0.536***	-0.415**	0.128	0.181
	(0.081)	(0.084)	(0.184)	(0.203)	(0.293)	(0.306)
500m-1000m	0.031	0.035	0.030	0.214	-0.195	0.742*
	(0.106)	(0.094)	(0.235)	(0.252)	(0.397)	(0.444)
N	1,166	1,166	500	500	464	464
Spec	OLS	IV	OLS	IV	OLS	IV

Note: See main tables for sull explanation of specification, this is the same specification as the reduced form. IVPoisson regression for slums did not converge, so log(0.01+slum share) included instead. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.3: IV-Mill Exposure

	House Prices	Slums (PPML)	Slums (Logged)	Informal Emp	Formal Emp	Educ	Population
OLS	0.039***	-0.270***	-0.123***	-0.310***	0.060	0.155***	-0.051***
	(0.013)	(0.063)	(0.039)	(0.078)	(0.131)	(0.044)	(0.019)
N	1,142	894	1,166	500	464	424	454
IV	0.040***		-0.110***	-0.302***	0.059	0.154***	-0.049***
	(0.013)		(0.036)	(0.078)	(0.132)	(0.045)	(0.019)
N	1,142		1,166	500	464	424	454

Note: See main tables for sull explanation of specification, this is the same specification as the reduced form. IVPoisson regression for slums did not converge, so log(0.01+slum share) included instead. Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.5: Robustness: Gradual Inclusions of controls

	(1)	(2)	(3)
Contains Mill	0.219***	0.155***	0.132**
	(0.059)	(0.058)	(0.058)
¡500m	0.117**	0.092**	0.081*
	(0.046)	(0.046)	(0.047)
500m-1000m	0.087*	0.059	0.051
	(0.047)	(0.045)	(0.045)
N	7,386	7,386	7,220
Population			
Contains Mill	-0.048	-0.091	-0.127**
	(0.077)	(0.063)	(0.064)
¡500m	-0.000	-0.103	-0.098
	(0.079)	(0.074)	(0.068)
500m-1000m	0.006	-0.044	-0.069
	(0.120)	(0.113)	(0.100)
N	454	454	454
Slums			
Contains Mill	-0.241**	-0.233**	-0.221**
	(0.108)	(0.111)	(0.104)
¡500m	-0.218**	-0.218**	-0.247***
	(0.087)	(0.089)	(0.095)
500m-1000m	0.086	0.092	0.095
	(0.109)	(0.108)	(0.111)
N	954	954	894
Emp Form			
Contains Mill	0.129	0.197	0.318
	(0.317)	(0.329)	(0.331)
¡500m	-0.035	0.017	0.181
	(0.300)	(0.311)	(0.305)
500m-1000m	0.277	0.550	0.742*
	(0.418)	(0.430)	(0.442)
N	464	464	464
Emp Inform			
Contains Mill	-0.665***	-0.798***	-0.621***
	(0.228)	(0.199)	(0.181)
¡500m	-0.542***	-0.781***	-0.519***
	(0.204)	(0.201)	(0.199)
500m-1000m	-0.229	-0.059	0.080
	(0.311)	(0.262)	(0.247)
N	500	500	500
Education			

Table A.7: Robustness: Alternative Fixed Effects

	(1)	(2)
Contains Mill	0.219***	0.140*
	(0.059)	(0.074)
¡500m	0.117**	0.052
	(0.046)	(0.075)
500m-1000m	0.087*	0.054
	(0.047)	(0.048)
N	7,386	7,386
Population		
Contains Mill	-0.048	-0.018
	(0.077)	(0.100)
¡500m	-0.000	-0.028
	(0.079)	(0.096)
500m-1000m	0.006	-0.065
	(0.120)	(0.119)
N	454	454
Slums		
Contains Mill	-0.241**	-0.179
	(0.108)	(0.116)
¡500m	-0.218**	-0.173
	(0.087)	(0.112)
500m-1000m	0.086	0.109
	(0.109)	(0.115)
N	954	954
Emp Form		
Contains Mill	0.129	0.620
	(0.317)	(0.456)
¡500m	-0.035	0.367
	(0.300)	(0.380)
500m-1000m	0.277	0.461
	(0.418)	(0.491)
N	464	464
Emp Inform		
Contains Mill	-0.665***	-0.416
	(0.228)	(0.305)
¡500m	-0.542***	-0.403
	(0.204)	(0.288)
500m-1000m	-0.229	-0.314
	(0.311)	(0.354)
N	500	500
Education		

Table A.9: Alternative Decay Rates in Mill Exposure Measure

	House Prices	Slums	Informal Emp	Formal Emp	Educ	Population
Weighted Share, Decay=1	0.031**	-0.255***	-0.298***	-0.298***	0.156***	-0.049**
	(0.015)	(0.062)	(0.077)	(0.077)	(0.046)	(0.019)
N	1,116	894	500	500	424	454
Weighted Share, Decay=1.25	0.030**	-0.238***	-0.264***	-0.264***	0.149***	-0.046**
	(0.015)	(0.057)	(0.069)	(0.069)	(0.042)	(0.018)
N	1,116	894	500	500	424	454
Weighted Share, Decay=1.5	0.029*	-0.230***	-0.245***	-0.245***	0.148***	-0.044**
	(0.015)	(0.054)	(0.065)	(0.065)	(0.039)	(0.018)
N	1,116	894	500	500	424	454

Note: Standard errors clustered at the loocation-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

B Additional Figures

Figure A.1: Validation Sample: SRA vs SPFEAS Slums

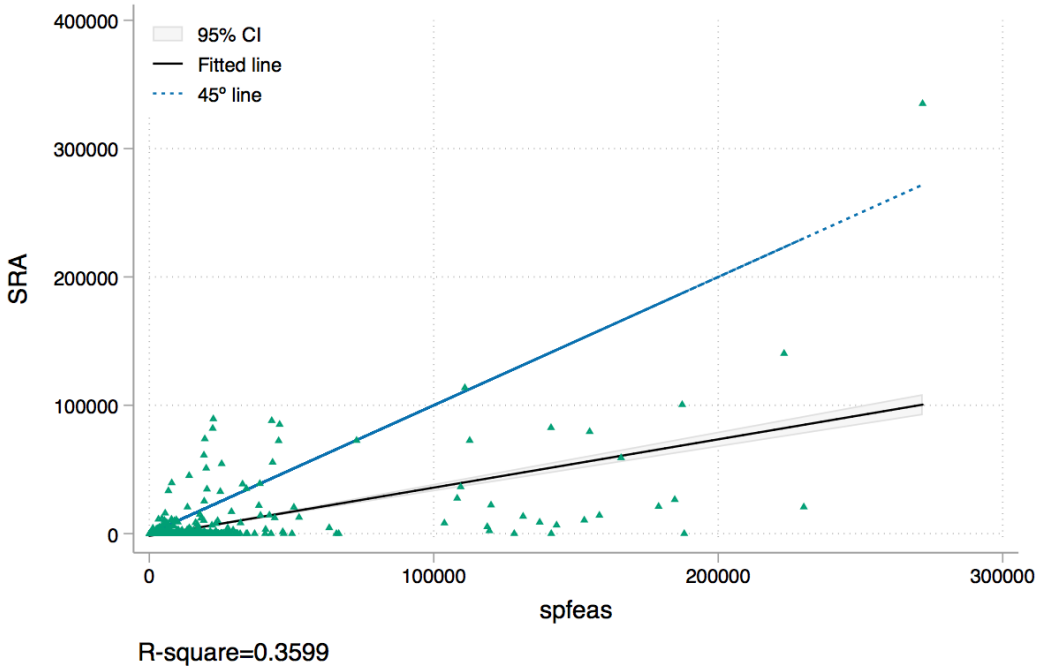


Figure A.2: Generating signed distance to a slum boundary

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

(b) Signed distance from a slum boundary

Figure A.3: 2002 voter list example pages

6

052-Mulund Assembly Constituency — Elector Part No. 18

Sr. No.	House No.	Full Name of Elector	Relation-ship	Full Name of Relation	Sex	Age	Epic No.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
							MT/07/052/
83	G-5	Bakshe Krishna Arjun	F	Bakshe Arjun	M	40	
84	G-5	Bakshe Asha Krishna	H	Bakshe Krishna	F	37	0052508
85	G-5	Bakshe Mahesh Krishna	F	Bakshe Krishna	M	21	
		Shobha Gupta Chawl					
86	G-5 A	Khawale Laxman Gulab	F	Khawale Gulab	M	39	
87	G-5 A	Khawale Anita Laxman	H	Khawale Laxman	F	37	0052534
88	G-5 A	Khawale Vina Laxman	F	Khawale Laxman	F	20	0052535
89	G-5 A	Khawale Rajesh Laxman	F	Khawale Laxman	M	22	0052536
90	G-1 B	Gupta Shobha Ramlagam	F	Gupta Ramlagam	M	57	
91	G-1 B	Gupta Shivkumar Shobha	F	Gupta Shobha	M	29	0052539
92	G-1 B	Gupta Bindudevi Shivkumar	H	Gupta Shivkumar	F	25	0052530
93	G-2	Bakshe Digambar Jagannath	F	Bakshe Jagannath	M	30	0052509
94	G-2	Bakshe Sharada Digambar	H	Bakshe Digambar	F	25	0052510
		Ramavadh Parvakhi Chawl					
95	G-1	Gond Fulchand Nifikir	F	Gond Nifikir	M	41	0052543
96	G-1	Gond Ramashankar Nifikir	F	Gond Nifikir	M	39	0052511
97	G-1	Gond Savitri Ramashankar	H	Gond Ramashankar	F	37	0052512
98	G-1	Gond Muratlal Ramashankar	F	Gond Ramashankar	M	19	0052513

(a) English-language list

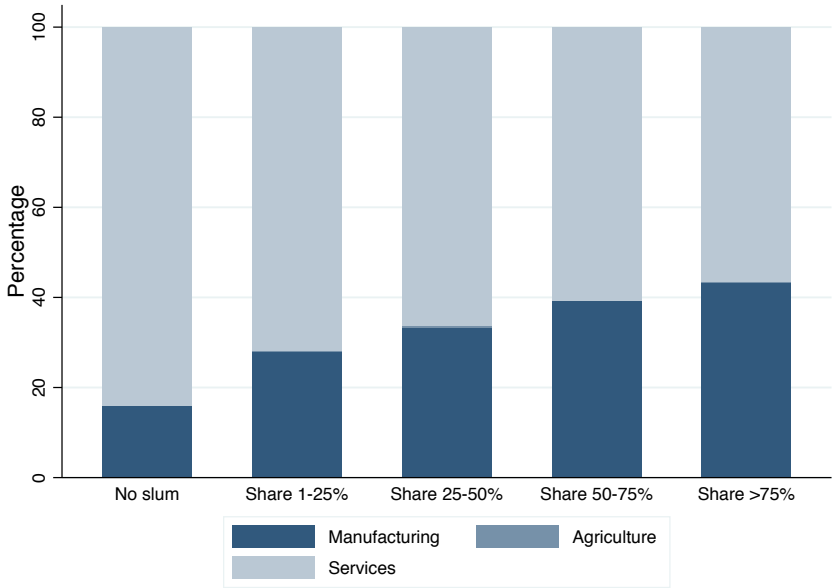
३५

(एस - १३) महाराष्ट्र राज्य — १९ - कुलाबा विधानसभा निर्वाचन क्षेत्र — यादीभाग क्र. ७४

अ. क्र.	घर, इमारत क्रमांक, नाव, मजला व खोली क्रमांक	मतदाराचे पूर्ण नाव	नाते	नातेवाईकाचे पूर्ण नाव	लिंग	वय	ओळखपत्राचा अनुक्रमांक
१	२	३	४	५	६	७	८
३४९	प-१५	सोनावणे लक्ष्मी अप्पा	प	सोनावणे अप्पा	स्त्री	४८	220702
३५०	प-१५	सोनावणे संतोष अप्पा	व	सोनावणे अप्पा	पु	२५	219127
३५१	प-१६	पाटणकर हरी लाडकोजी	व	पाटणकर लाडकोजी	पु	४८	220692
३५२	प-१६	पाटणकर शोभा हरी	प	पाटणकर हरी	स्त्री	३८	220697
३५३	प-१६	पाटणकर अनिल हरी	व	पाटणकर हरी	पु	२१	220696
३५४	प-१७	उत्तेकर रामचंद्र लक्ष्मण	व	उत्तेकर लक्ष्मण	पु	४८	220760
३५५	प-१७	उत्तेकर आशाबाई रामचंद्र	प	उत्तेकर रामचंद्र	स्त्री	४५	
३५६	प-१७	उत्तेकर कल्पना रामचंद्र	व	उत्तेकर रामचंद्र	स्त्री	२१	
३५७	प-१८	तोडकर प्रकाश शंकर	व	तोडकर शंकर	पु	४७	220707
३५८	प-१८	तोडकर कलावती प्रकाश	प	तोडकर प्रकाश	स्त्री	३८	220708
३५९	प-१९	गावडे शंकर सहदेव	व	गावडे सहदेव	पु	६१	
३६०	प-१९	गावडे बाबू शंकर	व	गावडे शंकर	पु	३१	
३६१	प-१९	गावडे सहदेव शंकर	व	गावडे शंकर	पु	३८	
३६२	प-१९	गावडे सुनिता सहदेव	प	गावडे सहदेव	स्त्री	३५	
३६३	प-२०	डोंगरे पांडुरंग जीवबा	व	डोंगरे जीवबा	पु	४८	220761
३६४	प-२०	डोंगरे प्रभावती पांडुरंग	प	डोंगरे पांडुरंग	स्त्री	४३	220762
३६५	प-२०	डोंगरे कृष्णा जिवबा	व	डोंगरे जिवबा	पु	३३	220765
३६६	प-२१	मोरे रामचंद्र रामा	व	मोरे रामा	पु	४८	220714
३६७	प-२१	मोरे सुलोचना रामचंद्र	प	मोरे रामचंद्र	स्त्री	४३	220715
३६८	प-२१	चव्हाण अशोक शांताराम	व	चव्हाण शांताराम	पु	२४	

(b) Marathi-language 2002 voter list

Figure A.4: Industry Composition of Slums



Note: Figure shows the share of employment in aggregate industries in locations (iv units) according to the share of area accounted for by slums. Since our spatial units for employment datasets do not perfectly overlap with slums, there is no way for us to compare employment in locations with 100% slums vs those with 0% slums. This nevertheless shows the positive relationship which is linearly extrapolated to 100% slums in the specification in Table A.1.