Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills*

Michael Gechter† Nick Tsivanidis‡

June 27, 2023

Abstract

This paper investigates local spillovers from high-rise, market-rate buildings in developing country cities. We exploit a unique natural experiment that led most of Mumbai's 60 textile mills, which covered 15% of central city land, to be redeveloped into high rises during the 2000s. Using a unique set of newly digitized data, we document sizable local spillovers consistent with these new developments gentrifying their surrounding neighborhoods. To analyze the indirect and distributional effects, we develop and estimate a dynamic quantitative urban model with informality and slum evictions. The redevelopments benefit Mumbai's residents in aggregate by increasing residential housing supply and providing valuable commercial floorspace for formal firms, but inflict large costs on forcibly relocated slum dwellers. Making Mumbai's compensation policy for evicted slum residents significantly more generous would substantially reduce losses to the losers while mildly reducing aggregate surplus. This highlights an equity-efficiency trade-off inherent to slum redevelopment policy.

*We wish to thank David Atkin, Russell Cooper, Jonathan Eaton, Ben Faber, Pablo Fajgelbaum, Cecile Gaubert, Chang-Tai Hsieh, Dan Keniston, Kala Krishna, Abhay Pethe, Vidhadyar Phatak, Debraj Ray, Steve Redding, Jim Tybout and participants at many 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, Geetika Nagpal, Ricardo Romell, and Pronab Sen. We are grateful to our many research assistants for their important contributions to this project. Financial support from the DigitalGlobe Foundation and the Tata Centre for Development at University of Chicago is gratefully acknowledged. First Version: December 2017. A version with clickable links through the online appendix and supplemental material is available on the authors' websites as well as here.

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

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

1 Introduction

By 2050, over two billion people will move into cities in Asia and Africa (United Nations (2018)). India will account for the majority of this, where cities are projected to produce 70% of new jobs and GDP by 2030.[1] A huge amount of new housing will be built to accommodate this rapid urban growth and the rising incomes of residents. McKinsey (2010) estimates that India will need to build residential and commercial floorspace equivalent to a new Chicago every year, and large portions of this new construction will be market-rate high rises in desirable central areas. Among all residential real estate projects launched in India's seven largest cities in 2019, more than 52% were high rises with Mumbai and Delhi recording figures over 70% (Anarock (2019)).

While there is no doubt that this wave of high rise development will redraw these cities' skylines, the impacts on residents are unclear and fiercely debated. Economists typically argue that increasing floorspace supply lowers rents, allows productive cities to grow, and improves overall welfare and economic performance (Glaeser and Gyourko (2018); Hsieh and Moretti (2019)). However, many policymakers, affordable housing advocates, and journalists oppose new market-rate high rises. They argue that these cater to high earners and gentrify surrounding areas as local house prices and amenities rise, displacing previous residents (Florida, 2019). These same arguments are amplified in cities in developing countries, where slum residents nearby may be even more at risk from displacement if they can be easily evicted (Friedman (2014); Anand and Parussini (2017)).

This paper provides evidence on spatial spillovers from high rise developments in Mumbai, one of the developing world's largest megacities. Previous literature has focused on spatial spillovers from changes in housing quality in the US, emphasizing local impacts on house prices.[2] Yet increasing the stock of high-quality housing—especially through large changes induced by high rises—could lead to broader impacts beyond house prices and have indirect effects throughout the city if economic activity reallocates. This paper asks how new high rise developments affect the nature of economic activity in surrounding neighborhoods, summarized not only by impacts on house prices but also slum redevelopment, informal and formal employment, gentrification, and population density. We then assess the distribution of winners and losers, and how this is shaped by government policy determining how slum dwellers evicted from redeveloped slums must be

[1] Its largest city, Mumbai–which accounts for over 6% of national GDP–will itself have a GDP larger than many countries (McKinsey (2010)).

[2] See Owens et al. (2010); Autor et al. (2014); Hornbeck and Keniston (2017); Diamond and McQuade (2019).

compensated. However, we face three key challenges in answering these questions.

The first challenge is that decisions by developers to build a high rise in any location may be correlated with underlying trends in the desirability of its neighborhood. For example, the construction of modern high rises may gentrify surrounding neighborhoods, or instead, developers may build high rises in already gentrifying neighborhoods.[3]

This paper addresses this challenge by exploiting a unique natural experiment that led Mumbai's 60 textile mills, covering 15% of land in the central city (602 acres), to come onto the market for redevelopment at the same time during the 2000s. The industry had declined through the second half of the 20th century and the mills ultimately ceased operation in the 1990s. While this experience is shared by many cities across the world which typically redevelop industrial parcels into new residential and commercial properties befitting their modern economies, what made Mumbai unique was the outsized influence of the mill workers' labor unions.[4] These unions successfully lobbied the Maharashtra state government to prevent the mill owners from redeveloping these plots (other than at a huge cost), and as a result, they lay vacant during the 1990s.

This all changed with the passage of the Development Control Rules Regulation 58 (DCR 58) in 2001 which *de facto* lifted these restrictions. As described in Section 2, this was driven by a push by the same labor unions (who now believed the mill jobs would not come back) to allow redevelopment to commercial uses but reserve jobs for their children. The leadership of the state government passed DCR 58 to allow this, but with a subtle change to the regulation's language which unexpectedly made it so that the regulations surrounding mill redevelopment applied only to vacant area on mill sites. Since most land on the mill sites was built up, this led to the construction of high-end residential and commercial towers on a majority of sites after it was upheld by India's Supreme Court in 2006. The original law change was not pushed by local developers or mill owners, who would be more likely to respond to trends in local fundamentals.

A first look at satellite imagery from a handful of mills in Figure 1 shows the stark change in these neighborhoods, with large towers going up on each mill (outlined in blue) and nearby slums being cleared and converted into formal buildings.

The second challenge is that even with plausibly exogenous variation in high rise construction, it is rare to have sufficiently granular data in a developing country city to

[3]The Merriam-Webster dictionary defines gentrification as "a process in which a poor area (as of a city) experiences an influx of middle-class or wealthy people who renovate and rebuild homes and businesses and which often results in an increase in property values and the displacement of earlier, usually poorer residents."

[4]Examples include the Meatpacking district in New York, West Town in Chicago, Docklands and Hackney Wick in London, and the Cheonggyecheon restoration project in Seoul.

Figure 1: Apollo, Simplex and Hindoostan Mills

(a) 2000



(b) 2016



Note: Figure shows satellite images for the Apollo, Simplex and Hindoostan Mills in 2000 and 2016 (demarcated by the blue boundaries).

measure its impact. We address this by assembling new administrative datasets including i) digitized records from Maharashtra Department of Registrations of Stamps providing government-assessed floorspace values across more than 900 subzones between 1993 and 2013; ii) over 6,000 hand-drawn maps of census enumeration blocks we digitized in order to geolocate the universe of formal and informal establishments across 1998, 2005 and 2013 economic censuses; iii) population data from the 2001 and 2011 censuses obtained across 227 election wards. We then use machine learning to measure the changing educational composition of neighborhoods between 2001 and 2016 from the changes in the names of the residents who live there. Lastly, we use high-resolution remotely-sensed daytime satellite imagery to measure the location of slums over the same period.

With this database in hand, we proceed to measure spillovers from the mill redevelopments to nearby neighborhoods. We find patterns consistent with overall gentrification of neighborhoods near to but not containing mills: i) the price of formal sector floorspace rises; ii) the share of land with slums falls; iii) informal employment (more likely to occur in slums) falls; and iv) the average education of residents rises. Contrary to the idea that commercial high rises may increase productivity in nearby areas due to agglomeration economies, we do not find evidence that formal employment increases near mills that are redeveloped into commercial high rises, although they do drive large increases in formal employment on the redeveloped sites themselves. Contrary to the idea that residential high rises should increase population density by allowing more individuals to live on a parcel of land, we find that population density actually *falls* near mill sites. This is consistent with rich, higher-educated residents moving in and consuming more floorspace than the slum dwellers who lived there previously.[5]

Our key identifying assumption is that locations containing and nearby the textile mills would not have experienced differential trends in outcomes absent the DCR 58 policy change, conditional on controls and broader neighborhood fixed effects.[6] We provide four pieces of evidence consistent with this. First, we run an event study with formal sector floorspace prices where we have annual data spanning the policy. There is no significant pretrend in prices leading up to the policy change, but substantial growth thereafter. Second, we run a placebo check measuring proximity to both mills and other similarly-

[5]We provide evidence that slums have much higher population densities than locations with formal housing. While slums do not provide a lot of floorspace per unit of land, they do allow for many residents to live on a unit of land given the very small size of housing units in slums. See Table A3.

[6]Relative to the type of causal inference studied in Pollmann (2023) we do not claim that the mills were distributed quasi-randomly across central Mumbai, although their neighborhoods do have similar characteristics to others in the same region. Instead we make the assumption of parallel trends in outcomes without the policy change.

sized plots of industrial land as of 1991. If the results are being driven by a general process of gentrification of old industrial areas, then we should find the same spillovers from both types of industrial plots. We find no spillovers on outcomes from exposure to these industrial plots that looked like but were not textile mills. Third, we run a coarser event study examining the dynamics of employment across the three rounds of the economic census. We see no impacts of mill exposure on employment between 1998 and 2005, with all the effects coming between 2005 and 2013 (after the 2006 Supreme Court decision which ultimately drove the mill redevelopments). Fourth, we show no change in roads—the one public good we can accurately measure—near the mills. Had the government responded to the policy by providing more public goods in the area, this could have driven some of the price effects.

These results establish that the construction of high rises drove a pronounced reallocation of economic activity nearby. This could have aggregate implications which are soaked up in the constant term of our regressions, as well as indirect effects on other locations which receive residents and firms which are displaced. They also suggest a potential for rich distributional effects: while all residents may be (indirectly) affected by the reorganization of economic activity, some—such as those evicted from redeveloped slums—may be more acutely affected than others.

To shed light on these questions, we develop and estimate a quantitative spatial equilibrium model of a city. We extend recent work modeling spatial equilibrium (e.g. Ahlfeldt et al. (2015), Caliendo et al. (2019)) to the features relevant in our context. Our model is dynamic and features multiple groups of workers, informal and formal land use and firms, and slum evictions. Within each period, the distribution of workers across space and the share of land in each location developed formally or informally is taken as given and a standard static spatial equilibrium is attained. Across periods, residents make decisions about whether and where to move and developers make decisions about whether to redevelop slums to formal floorspace.

The core economics of the model comes from an endogenous amenity that increases with the average education level of residents nearby. We allow for the possibility that higher-educated workers are more willing to pay for amenities. When new high rises (of high amenity) are built, more higher-educated residents move in. This in turn increases amenities close by, raising formal sector floorspace prices and increasing the return to

developers from redeveloping slums in those locations. Developers convert slums with an elasticity that summarizes conversion costs. Informal employment, which has a comparative advantage in slums, also falls. These channels simply capture the main forces documented in the reduced-form analysis.

We directly model the slum redevelopment policy in Mumbai, whereby developers need to compensate a fraction of slum dwellers if they want to redevelop a slum.[7] Residents who live in redeveloped slums are forced to move from their current place of residence, therefore paying a moving cost. Those eligible are compensated by the developers (around 70%), but the remainder are not.

We estimate the model via indirect inference. Given a vector of structural parameters, we calibrate the model's unobservables to the pre-period equilibrium and simulate the impact of building high rises on the mill sites, holding all other model unobservables constant. We then run the same regressions on the model-generated data and search over parameters until these match their empirical counterparts. The identification assumption is, therefore, the same as in the empirics, namely that the change in economic activity surrounding the mills is orthogonal to trends in unobserved amenities and productivities in those neighborhoods.[8] [9] Our parameter estimates imply a large elasticity of amenities to the average education of residents of 0.88, in line with Tsivanidis (2022). We find a spatial decay of amenities of 0.19 which, while smaller than Ahlfeldt et al. (2015)'s estimate from Berlin, means spillovers decrease by 85% with 10 minutes of travel time and decay almost completely 20 to 30 minutes out. Lastly, we find a semi-elasticity of slum conversion to the relative profitability of formal development of 4.94, providing the first causal estimate of a slum conversion elasticity in the literature.

We then use the model to quantify the welfare consequences of the policy. Total societal surplus increases, accruing through total consumer surplus and developer profits. This comes from the increase in floorspace supply reducing the overall level of floorspace

[7] As discussed in Section 2, the compensation is a 25 m^2 housing unit the developer must build.

[8] Given the reallocation of economic activity we document, it is not clear from the reduced form what precisely is driving the spillovers. By incorporating these various channels of adjustment, such as residential sorting and slum redevelopment, the model identifies the strength of the underlying spillovers through what these channels are unable to explain by themselves. The model also allows for spatial linkages between locations through commuting and residential choices, which can complicate estimation of spillovers from the reduced form alone due to SUTVA violations.

[9] There may be alternative ways to identify the model's parameters. Under model misspecification, however, the expected value of at least one moment function will not be equal to zero, and parameter estimates will converge to the minimizers of the population objective function, which are known as pseudo-true parameter values. The pseudo-true parameter values in general depend on the identification/over-identification strategy. Since we have as a key goal the explanation of the pattern of reduced form spillover effects through the lens of our model, we prefer potentially pseudo-true parameter values which match these effects as closely as possible.

prices, and the new commercial towers providing valuable space for formal firms to produce. Yet these aggregate effects mask rich heterogeneity - there is a thin tail of relatively few residents who are forced to move by the policy and are hurt greatly. For example, while welfare for the low-educated slum residents whose slums are not redeveloped rises by 8% on average, those who are evicted and compensated lose 24% and those who are not compensated for eviction lose 35%.

Given these distributional consequences, we simulate alternative compensation policies to illuminate the equity-efficiency trade-off inherent to slum redevelopment policies. Were every slum resident in a redeveloped slum to be compensated three times Mumbai's actual compensation policy, then the average low-educated evicted slum dweller would no longer lose. But the behavioral response of developers to this higher compensation policy is to build less floorspace, attenuating some of of the efficiency gains from the policy: total consumer and developer surplus go down. Yet everyone still gains on average relative to the counterfactual without the mill redevelopment, suggesting that governments should grant generous compensation policies to those rehoused under slum redevelopment if they have even a mild dislike for inequality.[10]

Related Literature A body of previous work has studied the spillovers to neighborhoods from housing policies, focused primarily on impacts on housing prices. This studies the impacts of changing housing quality (Owens et al. (2010); Redding and Sturm (2016); Hornbeck and Keniston (2017), rent control (Autor et al. (2014)), low income housing (Diamond and McQuade (2019)), gentrification (Guerrieri et al. (2013)) and condominium construction (Boustan et al. (2023), Li (2022), Mast (2023)). We add to this literature by bringing evidence from a large and plausibly exogenous shock to the supply of high rise construction. This allows us to measure broader impacts on the organization of economic activity. Our model also captures the indirect effects from the reallocation of economic activity (e.g. due to displacement), which are missed by the regression-based approaches used in this literature.

This paper also relates to the literature studying the economics of slums (e.g. Field (2007), Field and Kremer (2008), Collins and Shester (2013), Barnhardt et al. (2016), Ali et al. (2016), Michaels et al. (2021), Franklin (2020), Takeuchi et al. (2008)). Harari and Wong (2021) study the impact of a slum improvement policy in Jakarta. Henderson et al. (2021) study the development of Nairobi as slums transition to formal construction as

[10]The government may withhold compensation to recent arrivals in order to dissuade future in-migration, as discussed in the context of withholding public goods from slums in Feler and Henderson (2011). This consideration is outside the scope of our model.

the city grows, using a framework that infers the costs of converting informal to formal land use from model-based residuals. We add to this body of work by using a natural experiment to measure the response of slum redevelopment to a shock to the desirability of formal land in an area.

Methodologically, this paper follows recent work using quantitative spatial models to analyze the welfare and aggregate consequences of local shocks, accounting for spatial linkages between locations (Ahlfeldt et al. (2015), Allen et al. (2015), Monte et al. (2018)). We relate most closely to the body of work within this literature using dynamic models (e.g. Caliendo et al. (2019), Balboni (2021)) and studying gentrification and residential sorting (Diamond et al. (2019); Tsivanidis (2022), Couture et al. (2021), Balboni et al. (2020), Almagro and Domínguez-Iino (2022), Almagro et al. (2023)). This paper adds to this work by developing a dynamic model with multiple worker groups, informal housing and employment, and slum evictions. This allows us to uncover a rich distribution of welfare effects, and assess the impacts of policies to compensate evicted slum dwellers that account for behavioral responses of developers. We also estimate the model via indirect inference to match the key reduced form moments, ensuring the model captures the key forces present in the data and increasing confidence in its' counterfactual predictions (Kline and Walters, 2019).

Lastly, this paper also contributes to papers studying the impact of land use regulations both in the US (Saiz (2010); Turner et al. (2014); Hsieh and Moretti (2019)) and in India (Bertaud and Brueckner (2005), Brueckner and Lall (2015), Duranton et al. (2019); Harari (2020)). Since much of this literature relies on cross-sectional variation across cities or neighborhoods, our key contribution is to combine a large change in land use regulations with rich data before and afterwards to precisely measure the regulation's impact on economic activity.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting, Section 3 describes the construction of our dataset, Section 4 provides our reduced form results, and Section 5 lays out our structural spatial equilibrium model and Section 6 its estimation. Section 7 provides our welfare analysis and Section 8 concludes.

2 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. Figure A1 shows the locations of the mills with respect to the district covering central Mumbai (Mumbai City district) and the pair of districts

covering the full extent of the city (Mumbai City and Mumbai Suburban districts). The industry declined after World War II, culminating in an 18-month strike spanning 1982 and 1983 when mill owners began to outsource production to lower-wage, favorably-regulated areas outside the city and make investments in other industries. In the following period, most mills closed or attempted to do so.[11] However, workers and city government officials wanted to see them reopen and return to their role as a stable employment source (Adarkar and Phatak, 2005).

The land occupied by the mills was owned by the government and leased to the mill owners at low rates (Adarkar, 2005; Menon, 2012) and prior to 1991, state law prevented textile mills from being redeveloped or used for any alternative purpose. By 1991, however, the government of Maharashtra had come to see the sale of mill land as a way of financing the mills' return to viability, while increasing the amount of public space and housing in the city. Therefore the state government amended the city's Development Control Rules by introducing Regulation 58 (DCR 58). The amendment stipulated that mill owners could sell land[12] which could then be redeveloped to alternative purposes, but that in exchange one-third of the land area sold would be 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. The land given over to the city would be credited to mill owners as development rights to land outside the city's main peninsula (known as transferable development rights (TDR), Menon (2012)). The funds gained from redevelopment were supposed to be spent only on the revival and modernization of the mills. However what amounted to a 67% tax on land sale did not prove attractive to mill owners, and very little development took place during the 1990s (D'Monte, 2006).

This all changed in 2001 when an amendment to DCR 58 was made by the Maharashtra state government that *de facto* lifted these restrictions. It was in fact a prominent mill workers' union Girni Kamgar Sangharsh Samiti who initially lobbied for an amendment to DCR 58. They wished to mandate that their members and their children would be employed in any new enterprises coming up on the mill site and that half of MHADA's share of the land be reserved for mill worker housing. However, as the initial amendment made its way through the state bureaucracy, the leadership of Maharashtra state government at the time modified the amendment so that the one-third rule applied only to open spaces on mill sites, the majority of which were covered in factory buildings.[13] The mill

[11] Government permission was required to close factories during this period.

[12] Commentators at the time and subsequently remarked on the curiosity of mill owners being allowed to sell land they were only renting (e.g. Srinivasan (2005), Menon (2012)).

[13] The leadership was later implicated in another real estate scandal, during the prosecution of which the

owners could therefore freely redevelop all covered spaces on mill sites (and one-third of any vacant space). 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 (Menon, 2012).

The push to change DCR 58 that led to the mill redevelopment therefore came from the labor unions rather than local developers or mill owners. This context plays a key role in our research design, since a concern is that the policy may have been driven by trends in local fundamentals (such as gentrification) that could have driven both the DCR 58 amendment and changes in economic activity around the mills.

Impact of the DCR 58 Amendment Table A2 shows that, prior to the regulatory change, mill lands were broadly similar to other parts of the city. Spatial units containing mills had lower floorspace prices and a smaller share of land taken up by slums, which is partially attributable to the fact that mill sites themselves did not house any slums. Within the geographic regions mill land occupied, spatial units containing mills actually did not differ much in terms of floorspace prices, rather having low population density which, again, is sensible given that no one lived on mill sites.

By 2016, however, 70% of the mill sites had either been redeveloped or were in process of being 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 (Emporis, 2018). The average of the 47 planned or completed high rise towers constructed on mill sites had about 88 floors, representing a dramatic increase in the density of formal floorspace in the mill neighborhoods.

Specifics of Mill Redevelopment 58% of the developments completed by 2016 were residential, with the remaining being commercial. The firms in commercial developments included restaurants and retail, as in the case where mills were converted to shopping malls, and many high-skill service sector establishments, in industries such as finance, real estate, and consulting. Table A1 shows that the main determinant of commercial development is the employment density of the surrounding area prior to redevelopment. In keeping with this, commercial developments are almost exclusively found on land zoned for industrial use. In the second column of the table, we see that areas, where mills were redeveloped, do not differ systemically from those where they were not.

Laws Governing Slum Redevelopment The redevelopment of slums is governed by the 1995 Slum Rehabilitation Act and administered by the Slum Rehabilitation Authority

Principal Secretary for Urban Planning was arrested and held in custody for over two months. The Chief Minister was also implicated.

(SRA). During the majority of the period covered by our data (up to 2017), if redevelopment occurred, residents who could prove occupancy prior to January 1, 1995, would be compensated with a 30-year *de jure* non-transferable lease for a residential tenement (Maharashtra Urban Development Department, 1995). In practice, however, it is easy to find discussions of the sale of the leases and rent of the tenements on major websites. News articles such as Nair (2018) report large fractions of occupants in SRA buildings to be illegal, which is in keeping with Barnhardt et al. (2016)'s findings in Ahmedabad. In the structural model outlined in Section 5 we therefore model compensation as an income transfer to eligible residents in the amount of the value of the tenement. Residents who could not establish occupancy prior to 1995 are not eligible for compensation.[14]

The new tenements are by regulation to be located in the same location as much as possible, constructed on the original site by the developer in exchange for TDR for increased building heights on other parts of the site, or on other developments. However, in cases where developers could argue in situ redevelopment is not possible, these new units can be placed elsewhere and are often located towards the outskirts of the city (Grey, 2023). Additionally, even residents to be rehoused in situ are relocated to transit camps or rental housing, again often on the outskirts, before development begins. The transit period can last 10 years, or indefinitely, as in cases where the in situ housing is never built (Suresh et al., 2013). The SRA was not forthcoming with records of what buildings are currently owned by compensated former slum dwellers so in our model we take a conservative approach and assume the compensation is not in situ and has value given by 25 m^2 of floorspace averaged over the distribution of floorspace prices in SRA buildings constructed to house former slum dwellers (see Figure A2 for the distribution across greater Mumbai).

For redevelopment to proceed 70% of slum residents eligible for compensation had to approve a plan. However, there are reports that the law was not perfectly enforced, for example with developers forging residents' signatures in order to pass the 70% threshold (Banerjee, 2012; Suresh et al., 2013).[15]

The SRA approach, where private developers undertake slum redevelopment subject to resident approval is not unusual internationally. For instance, Baker and McClain (2008) cites examples of similar policies in Cambodia, Thailand, and the Philippines.

[14]Proof of residency was established through a certified extract from a pre-1995 electoral roll (Maharashtra Urban Development Department, 1995).

[15]Table A3 describes characteristics of slum areas at baseline. Slum areas are extremely concentrated, with employment and population density increasing with slum share, and employment per resident doing the same. Firms are smaller in high-slum areas, and more likely to be engaged in manufacturing.

3 Data

Mill sites We produced our shapefile of mill sites by georeferencing a map from the Charles Correa Committee Report, a 1996 government-commissioned plan for the mill lands which was never implemented, reproduced in D'Monte (2006).[16]

Inferring socioeconomic status We create a novel index of socioeconomic status (SES) by exploiting the information available in the lists of voters' names by election ward made available by the Bombay Municipal Corporation (BMC) Election Office in 2002 and 2017. We mapped the election wards from each year based on the description of their bounding streets in BMC documents and harmonized data to the 2002 election ward data. As detailed in Appendix SM.1 (note: appendices with the prefix SM refer to supplementary material available on the authors' websites), the first step uses individual records from Mumbai District in the 2011 Socioeconomic and Caste Census (SECC) which provide the names, education and birthdates of everyone in Mumbai.[17] Last name is indicative of caste and religious group (Bhagavatula et al. (2019)) and membership in these groups is predictive of SES, so we classify classify each name's distance to a set of prototype names and then use Extremely Randomized Trees to predict education levels from this distance vector. We then feed in the voter lists to form an index of each election wards in the pre- and post-policy years. We refer to the resulting indices as the "education index."

Residential population Our data on residential population come from the 2001 and 2011 population censuses, which the BMC Election Office reports by election ward.

Floorspace prices Data on floorspace values come from annual official assessments produced by the city. These are published each year in the city's Ready Reckoner books for around 600 "subzones" across Mumbai and Mumbai suburban districts. The assessments are produced so that taxes on real estate transactions can be calculated based on the maximum of the sale price reported to authorities and the Ready Reckoner value for the subzone where the property is located. Anagol et al. (2022) provide evidence that the Ready Reckoner values are highly correlated with the validation data they use and reported prices are very often misreported. For historical values and subzone boundaries, we digitized paper copies of the Ready Reckoner book going back to 1993, except 1995, 2004, 2006 and 2009 which proved impossible to find.

[16] We cross-checked our shapefile against one created by the NGO Partners for Urban Knowledge, Action and Research (PUKAR) as part of a Ford Foundation-supported project chronicling the history of the mills.

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

Employment and firm characteristics Our employment data comes from the Fourth, Fifth, and Sixth Economic Census (EC) of India. These cover the universe of establishments (formal and informal) in 1998, 2005, and 2013 respectively. The process of geolocating the firm data at a fine enough scale for our neighborhood-level analysis is substantially involved. For the Sixth EC, we had a team of local research assistants manually geolocate thousands of hand-drawn maps of census enumeration blocks. For the Fourth and Fifth ECs, we infer the location of investigator units (IV Units), the administrative spatial unit one level up from enumeration blocks, based on the addresses of firms located in each. We describe our process in detail in Appendix SM.3. After aggregating 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 in Mumbai over the course of fifteen years.

The Sixth is our post-policy outcome wave of the EC, and unfortunately does not capture information about firm registration with the government. We will therefore proxy establishment formality throughout the rest of the paper by whether a firm employs 10 or more workers. We choose 10 workers as the threshold because, as described in detail in Amirapu and Gechter (2020), many important regulations only apply to firms with 10 or more workers. In the Fifth EC which records registration status, registration probability initially increases steeply with size but reaches an inflection point at around 10 workers.

Slum location The SRA of Mumbai conducted a survey of all slums in the city in 2016 and provide a map of the results which we digitized and geolocated. The SRA designates slum status based on the type of built structures in an area.[18] [19]

To measure slum locations in 2001, we adjust our 2016 SRA map by overlaying it on top of satellite images from 2001 to identify structures in our 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 that looked identical to those present in both years. Three research assistants and the authors went through this process independently and reached a consensus to maximize accuracy.

[18] SRA uses a topographical survey based on satellite images and LiDAR to map out the location of slums across the city. See Dhikle and Lakhena (2017) for a description.

[19] We validated the SRA data using satellite imagery from 2016 and the 2011 census enumeration block maps. The enumeration block maps 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. These categories roughly correspond to our understanding of the SRA's of slum and non-slum buildings, with pucca buildings having primarily stone, tile or concrete walls and roofs and kutcha buildings being made of all other materials such as unburnt bricks, bamboo, thatch and plastic/polythene.

Other Data The 2002-03 Mumbai Travel Survey (Baker et al., 2005) provides information on individual migration and commuting within Mumbai. We also digitized Eicher roadmaps of Mumbai made in 2003 and 2013 (the earliest and latest years available) to compute the area of land in each spatial unit taken up by main roads.

Spatial units To summarize, our data use three sets of spatial units. Population and SES status are measured at the election ward level, floorspace prices at the subzone level, and employment and industry composition at the IV unit level. We harmonize all election-ward level information to the 2002 election ward level by assuming uniform population density across later election wards. Our slum location shapefiles are arbitrarily geographically fine so we intersect them with the finest of the three spatial units, the IV Unit, to obtain the slum share of each IV Unit. The characteristics of all three spatial units are described in Table A4.

4 Reduced Form Results

4.1 Measuring treatment from mill redevelopment

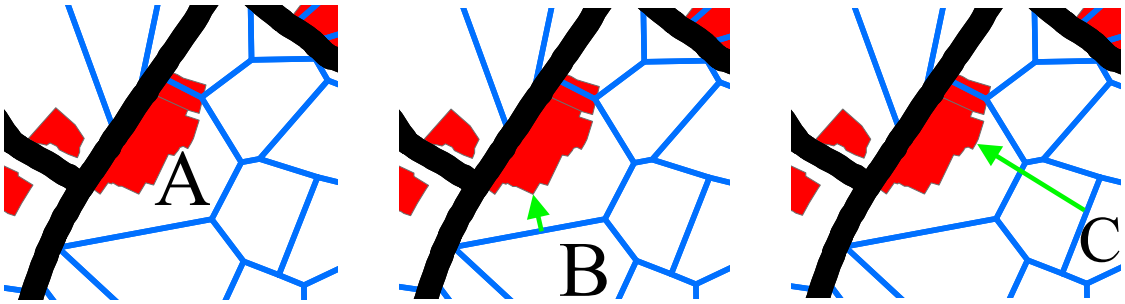
In the spirit of Autor et al. (2014) in our preferred specifications, we divide spatial units into four categories: (1) units containing mills, (2) units whose closest boundary to a mill site is within 500m of the site, (3) units whose closest boundary to a mill site is between 501 and 1000m of the site, and (4) all other units. Differential outcome growth in spatial units falling into category (1) within broad neighborhoods captures the direct effect of mills, along with some spillover onto parts of the spatial unit not covered by the mill site. Differential growth in spatial units in Category (2) captures a pure spillover effect, while those in category (3) are indicative of spatial decay in the spillover. Figure 2 illustrates the approach using IV Units as the spatial unit.

In addition to the simple measure used above, we create a continuous treatment effect measure that captures 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 away 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 for spatial unit i as

$$\text{MillExposure}_i = \sum_{j \notin i} d_{ij} \, \text{MillDensity}_j \tag{1}$$

$$\text{where } d_{ij} = \exp(-\kappa \cdot \text{dist}_{ij})$$

Figure 2: Measuring treatment from mill redevelopment



(a) IV Unit containing mill

(b) IV Unit in spillover region

(c) IV Unit farther in spillover region

Notes: Mill sites are represented by red areas and IV Unit boundaries are in blue (black lines represent city ward boundaries). The IV Unit containing the letter "A" in panel 2a belongs in category (1) because there is a mill site inside it. The IV Unit containing the letter "B" in panel 2b belongs in category (2) because it does not have a mill site inside it and the distance between its boundary and the closest mill site, represented by the green arrow, is 500m or less. The IV Unit containing the letter "C" in panel 2c has a greater distance to the nearest mill boundary because it does not contain a mill and the green line is larger than the one in panel 2b

and $\text{MillDensity}_i = \text{MillArea}_i/\text{Area}_i$. Locations indexed by j are a grid of 21,000 hexagon grid cells with 150 m^2 area each, which are consistent across the different geographies for locations i that depend on the particular outcome under consideration. This reduces noise introduced by the different size of spatial units in the geographies; if we just measure 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.

4.2 Investigating pre-trends

A threat to identification is that mill redevelopment, rather than generating spillover effects onto surrounding areas, was lobbied for and undertaken because of trends in the surrounding areas which continued post-redevelopment. This would tend to overstate the effects of the redevelopment itself. In the opposite direction, one could worry that the

amendment to DCR 58 was anticipated so that measuring effects after its passing would understate the impact of mill redevelopment.

To investigate these possibilities, we undertake an event study using formal floorspace price assessments the outcome we have available at an annual frequency. We run the following event study specification:

$$\ln r_{it} = \alpha_i + \sum_{\tau=1993}^{2013} \beta_\tau \mathbb{I}\{\tau = t\} \times \text{ContainsMill}_i + \gamma_{r(i)t} + \delta_t' X_i + \epsilon_{it} \tag{2}$$

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, ContainsMill_i is a dummy variable for whether subzone i intersects with a mill site, $\gamma_{r(i)t}$ is a region by year fixed effect, and X_i is a vector of controls including log distance to the CBD, log of the share of subzone i taken up by main roads in 2003, log spatial unit area, log initial slum share, log initial employment and population density, and the initial period education index.[20]

Figure 3a presents this event study graph and Figure 3b the exposure specification, which replaces ContainsMill_i in Equation 2 with a MillExposure_i from Equation 1 (standardized to have zero mean and unit standard deviation). The trends in the β_τ coefficients are in keeping with the account of the DCR 58 amendment in Section 2. There is no significant pretrend in floorspace prices in neighborhoods with mills in the years leading up to the sudden announcement of the DCR 58 law change in 2001. Prices begin to appreciate between 2001 and 2006 (more noticeably in 3b), during which the policy was announced but substantial uncertainty remained as it was contested in the courts. After the 2006 Supreme Court decision to uphold the law change, floorspace prices sustain a pronounced and gradually increasing appreciation. Taken together, these results suggest the lifting of redevelopment regulations on the mill sites was both unanticipated and exogenous to local economic conditions.

4.3 Main results

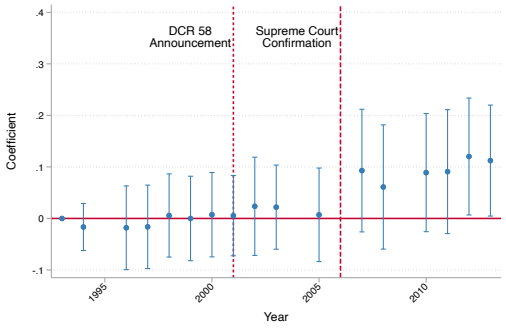
Our main specification, Equation 3, measures the long-difference-in-difference effect of belonging to each of the three categories of spatial units discussed in Section 4.1 on the outcomes captured by our data.

$$Y_{it} = \beta_1 \mathbb{I}\{\text{ContainsMill}\}_i \times \text{Post}_t + \beta_2 \mathbb{I}\{\text{1m-500m}\}_i \times \text{Post}_t + \beta_3 \mathbb{I}\{\text{501m-1km}\}_i \times \text{Post}_t$$

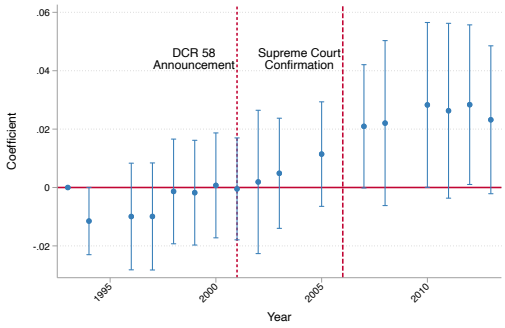
[20] "Initial" is the earliest year we have data available, so 1998 for employment, 2001 for slums and population, and 2002 for the education index.

Figure 3: No significant pretrends

(a) Event Study, Distance Bands



(b) Event Study, Exposure Measure



$$+ \alpha_i + \gamma_{r(i)t} + \delta_t' X_i + \epsilon_{it} \tag{3}$$

As in our event study specification, α_i is a spatial unit fixed effect, $\gamma_{r(i)t}$ is a time-specific effect associated with the region i belongs to.[21] X_i is a vector of controls allowed to have time-specific effects δ_t, which include those listed above as well as the log of assessed 1993 residential floorspace price. We omit lagged outcome values from the controls to guard against endogeneity generated by serially correlated errors. Each regression contains the earliest and latest available years, as described in the table notes.

The results are given in Table 1. We begin by examining the impact on land markets. The first column shows a pronounced impact on log formal sector house prices, implying a rise of 13% in locations containing a mill site and smooth decay with space thereafter. Locations whose boundary is within 500m of a mill site (but contain no mill site themselves) see a 9% increase in residential floorspace prices, decreasing to 5% 500m-1km away (albeit statistically indistinguishable from zero). In the second column, estimated by Poisson Pseudo Maximum Likelihood (Santos Silva and Tenreyro, 2006), we see that 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 23% in locations containing a mill and by 27% in locations within 500m of one. Note that mill sites contain no slums themselves: the total effect is a weighted average of the effect on mill sites (zero change) and areas outside (negative change), which attenuates the effect.

We then examine the impact on employment in the third and fourth columns, where outcomes are logged. Locations containing and near to mills experience large decreases

[21] 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 city wards. Using ward by year fixed effects instead gives us less power and more imprecise estimates, but we show the results are qualitatively similar in Table A10.

Table 1: Main Results

	Res Price	Slum Share	Inf Emp	Form Emp	Ed Index	Population
Contains Mill	0.130**	-0.226**	-0.413**	0.196	0.262**	-0.150**
	(0.059)	(0.102)	(0.186)	(0.330)	(0.126)	(0.059)
1m-500m	0.089*	-0.270***	-0.400**	0.202	-0.197	-0.117*
	(0.049)	(0.097)	(0.201)	(0.286)	(0.175)	(0.061)
500m-1000m	0.046	0.084	0.298	0.286	-0.157	-0.111
	(0.048)	(0.109)	(0.258)	(0.365)	(0.168)	(0.092)
N	1,050	894	486	438	446	434

Notes: Robust standard errors in parentheses. Dependent variables are logged with the exception of slum share which is estimated by Poisson Pseudo Maximum Likelihood, and education index which is in standard deviations of the 2002 education index. Regressions include own spatial unit fixed effects and allow for differential time trends by log distance to CBD, log spatial unit area, log of 1+initial period road density, initial slum share, log of initial employment density, log of initial population density, and initial education index. To guard against endogeneity generated by serially correlated errors we do not allow differential trends by lagged outcome values. We trim 1% tails of outcome variables. Each specification has two time periods, in most cases consisting of the earliest and latest years we have data available (slums: 2001 & 2016; employment 1998 & 2013; population 2001 & 2011; education 2002 & 2017). For house prices, we collapse to a pre-period (average of years available before but not including 2001) and a post-period (average of years available after but not including 2006). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

of 41% and 40% 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 A3), 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, so no strong evidence of agglomeration effects in the formal sector although the effects are noisy. We explore this finding in more detail in Section 4.5.

Lastly, we explore the impact of mill redevelopment on demographic composition in surrounding neighborhoods. The fifth column shows that our education index increases by 0.26 standard deviations of the 2002 measure in locations containing mills. The last column which shows results for log population finds that density falls by 15%. Election wards whose boundaries are between 1 and 500m of the closest mill sites experience a 12% drop in population density. The results are jointly 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. The difference in floorspace consumed per resident to produce this effect is quite dramatic, given the height of mill redevelopments.

Taken together, these results show the mill redevelopments had substantial impacts on surrounding neighborhoods suggestive of gentrification. Formal sector floorspace prices

rose, slums were cleared, informal employment fell, and the share of high-skilled residents increased. The first row of Table A5 provides analogous results using the alternative mill exposure measure given in Equation 1, yielding qualitatively identical conclusions.[22]

Addressing disamenities from mill sites While our interpretation of these results is that they reflect the impacts of an increase in economic activity on new high rises, it is possible that they capture the effect of removing the disamenity of nearby empty textile mills. We investigate this possibility in the second row of Table A5, which gives the exposure-measure version of our main results, controlling for a spatial unit's own share of land taken up by mill sites. This controls for the proximity to mills in a location and uses residual variation in exposure to mills in other nearby locations for identification. If the impacts we are measuring simply capture the removal of these disamenities, then adding this control should lead to a large decrease in the magnitude of the coefficients on mill exposure. We find the coefficient is largely invariant to this control, suggesting this is not what is driving our results. The third row adds a control for distance to the closest mill site. Results are slightly noisier but again qualitatively unchanged, showing that the nearest potential disamenity does not drive our results.

Real Estate Development Market Structure In a similar vein, one might think that large mill redevelopers may also redevelop nearby slums to remove disamenities for the residents of their new buildings. In Appendix SM.2 we show that not a single mill developer redevelops slums near mill sites. We also show the real estate development market is less concentrated than in other large Indian cities, suggesting collusion across developers is less likely to be at play.

4.4 Falsification

This section provides additional evidence supporting the causal nature of the mill redevelopment impacts. First, we conduct a placebo check comparing exposure to mill sites versus large industrial parcels similar to mill sites. If the process we document is driven by a process of gentrification of industrial areas, we should find the effects of exposure to these locations as well as mills. Second, we provide evidence of a lack of pre-trends in employment. Third, we show there is no increase in roads near the mill sites, supporting

[22]For reference we also provide regressions of outcomes on indicators for bands of distance from active new developments on former mill sites, instrumenting for these band indicators with indicators for bands of distance to any mill site in Appendix C.1. However, we prefer to use the model developed in Section 5 as a more detailed account of economic activity on redeveloped mill sites.

the idea that the effects are driven by the change in activity on the mill sites rather than the increased provision of public goods targeting these new areas.

Industrial Land Placebo Check To confirm our effects are not 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 10,000 m^2, the size of the smallest mill site, to ensure we are comparing similarly sized industrial plots. Table 2 shows the results, with Table A6 providing the corresponding results when using the exposure specification from Equation 1. In general the mill effects remain large and typically significant, while the point estimates for other industrial areas are always insignificant and much smaller than for mills.

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 3 incorporates the 5th EC 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 partially reverts by 2013. Taken together, the impacts overwhelmingly 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 the local government responded to the law change by investing in public goods near mill sites that may confound the true effects of the mills themselves. We regress the change in the share of IV Units taken up by main roads on our distance to mill measures as well as baseline controls. Table A8 shows no significant changes in road construction related to the mill sites.[23]

[23]Consistent with this finding, zoning (determining land use regulations, building heights, and so on) and plans for public good construction in Mumbai are undertaken in very discrete intervals. Zoning and public goods development plans in our period of study were determined by the 1991 Development Plan. It wasn't until 2016 that the Development Plan was updated once more, after our period of analysis.

Table 2: Changes Driven by Textile Mills, Not Industrial Land

	Res Price	Res Price	Slum Share	Slum Share	Inf Emp	Inf Emp
Contains Mill	0.117**	0.086	-0.256**	-0.204**	-0.436**	-0.487**
	(0.056)	(0.060)	(0.101)	(0.100)	(0.187)	(0.218)
1m-500m	0.101**	0.093*	-0.261***	-0.228**	-0.352*	-0.331
	(0.047)	(0.049)	(0.094)	(0.101)	(0.200)	(0.215)
500m-1000m	0.052	0.051	0.016	0.005	0.168	0.146
	(0.045)	(0.044)	(0.092)	(0.097)	(0.233)	(0.239)
Contains Mill Ind Land		0.047		-0.025		-0.146
		(0.043)		(0.127)		(0.276)
1m-500m Ind Land		-0.061		0.118		-0.407
		(0.047)		(0.102)		(0.263)
500m-1000m Ind Land		-0.015		0.070		-0.114
		(0.063)		(0.077)		(0.249)
N	1,106	1,106	882	882	496	496
	Form Emp	Form Emp	Educ	Educ	Res Pop	Res Pop
Contains Mill	0.121	0.106	0.262**	0.232*	-0.131**	-0.079
	(0.321)	(0.353)	(0.126)	(0.126)	(0.064)	(0.085)
<500m	0.075	0.086	-0.197	-0.228	-0.096	-0.071
	(0.298)	(0.328)	(0.175)	(0.175)	(0.066)	(0.068)
500m-1000m	0.463	0.503	-0.157	-0.202	-0.075	-0.045
	(0.371)	(0.368)	(0.168)	(0.169)	(0.101)	(0.112)
Contains Mill Ind Land		0.357		-0.235		-0.056
		(0.448)		(0.180)		(0.078)
<500m Ind Land		0.158		-0.111		0.066
		(0.446)		(0.228)		(0.097)
500m-1000m Ind Land		0.548		0.182		0.033
		(0.393)		(0.133)		(0.112)
N	460	460	446	446	450	450

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regressions include own spatial unit fixed effects and the same controls Table 1 (as before excluding lagged dependent variables). We trim 1% tails of outcome variables.

Table 3: No Pre-Trends in Employment

	Inf Emp	Form Emp
Contains Mill X 2005	0.002	-0.203
	(0.093)	(0.200)
Contains Mill X 2013	-0.448**	0.035
	(0.189)	(0.317)
1m-500m X 2005	0.061	0.341
	(0.096)	(0.213)
1m-500m X 2013	-0.462**	-0.072
	(0.196)	(0.306)
500m-1000m X 2005	0.171	-0.313
	(0.105)	(0.263)
500m-1000m X 2013	0.159	0.207
	(0.243)	(0.358)
N	752	722

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes own IV Unit fixed effects and allows for differential time trends by log distance to CBD. We trim 1% tails of outcome variables.

4.5 Lack of Agglomeration Spillovers

We dig deeper into the finding from Table 1 that formal employment does not respond strongly to mill redevelopments, which would be indicative of formal agglomeration effects outside the mills themselves.

First, we investigate whether the main null result masks heterogeneity in spillovers to formal employment nearby commercial mills (as theory would predict). Table 4 runs a version of our main regression, but splits each of the variables $\mathbb{I}\{\text{ContainsMill}\}_i$, $\mathbb{I}\{\text{1m-500m}\}_i$, and $\mathbb{I}\{\text{501m-1km}\}_i$ into Residential and Commercial counterparts based on the actual use of the new development.

The first column shows that locations with commercial mills do not show a decline in informal employment, while those with residential mills do, although the difference is somewhat imprecise. The negative effect on informal employment for IV Units whose boundaries are between 1 and 500m away from a mill site is almost identical across commercial and residential mills. The second column shows very large increases in formal employment in locations with mills that were redeveloped to commercial use, validating our classification of mill redevelopment and use. However, there is no significant change in formal employment for any other category.

Table 4: Heterogenous Employment Effects By Mill Use

	Inf Emp	Form Emp
Contains Mill Comm	0.337	1.678***
	(0.339)	(0.488)
Contains Mill Res	-0.456**	-0.323
	(0.227)	(0.395)
1m-500m Comm	-0.426	-0.066
	(0.270)	(0.452)
1m-500m Res	-0.493***	-0.030
	(0.188)	(0.307)
500m-1000m Comm	-0.070	0.225
	(0.261)	(0.447)
500m-1000m Res	0.120	-0.239
	(0.252)	(0.356)
N	486	438
p-value Contains Mill Comm==Res	0.08	0.01
p-value 1m-500m Comm==Res	0.86	0.95

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Contains Mill Comm is a dummy equal to if an iv unit contains a mill that was redeveloped to commercial use by 2008, while Contains Mill Res does the same for residential redevelopments. The other variables are defined analogously. Regressions include data from 1998 and 2013, and coefficients reported are for each variable interacted with a "Post" dummy for the year being 2013. The regression includes IV Unit fixed effects and the same controls as Table 1. We trim 1% tails of outcome variables.

Appendix C.3 tests for agglomeration spillovers onto nearby areas in the specific industries most common in the commercial redevelopments. As above, we do not find any positive evidence of agglomeration spillovers arising from the large increases in formal employment taking place on former mill sites. Our results no such spillovers are present over the time horizon we examine and given the sizes of IV Units (on average about 1/3 of a square kilometer). This does not rule out agglomeration effects *within* commercially redeveloped mills, consistent with evidence from Baum-Snow et al. (2021).

4.6 Robustness

We assess the sensitivity of our results to alternative specifications in robustness checks. Table A9 adds the full set of controls sequentially across three specifications. Coefficients are stable throughout. Table A10 uses 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. In Table A11, we shrink the size of distance bands measuring pure spillover effects. The results are quite similar to those in Table 1, and show evidence of spatial decay even within the 1 - 500m band for the residential floorspace price, slum, and informal employment outcomes. Lastly, Table A12 uses alternative decay rates when computing the mill exposure measure. The effects are economically and statistically stable throughout.

5 A Dynamic Urban Model with Informality

The previous results document substantial spillovers from the mill redevelopments but leave certain questions unanswered. First, as the regressions are based on relative comparisons across neighborhoods, they are silent on the effects on the level of economic activity such as floorspace prices or output. They are also silent on indirect effects on locations not directly exposed to the mills but affected by the reallocation and displacements the redevelopments induce. Second, it is unclear how to add up the various empirical effects to weigh the winners and losers from the redevelopment policy. These distributional effects could also be influenced by alternative policies to compensate slum dwellers, often a topic of heated debate in these cities.

We address these shortcomings by developing a structural model to estimate the spatial spillovers and conduct a quantitative analysis of the policy.[24] The model incorporates multiple spatial linkages through commuting, residential choices, and an amenity spillover that depends on a location's neighborhood composition and on those nearby.

The key economics of the redevelopment policy operates through this amenity spillover. When higher-educated workers are more willing to pay for amenities, they move into locations where high-quality formal high rises are built. In the presence of amenity spillovers, this increases amenities nearby, raising formal floorspace prices and leading developers to redevelop slums. This drives a pattern of neighborhood gentrification as the higher educated move into the new formal housing and low-educated slum residents move out. The model is dynamic, incorporating developer decisions on whether to redevelop slums and a rich distribution of welfare effects across individuals by their initial residential location and housing type, and whether or not they were evicted from a slum.

[24]The reduced form spillovers we estimate reflect both any direct spillover from the redevelopments as well as reallocation through commuting and residential choice linkages that depend on the particular geography of Mumbai. Estimating the direct spillover using a structural model that accounts for these other linkages provides an estimate that is more portable to other contexts with different geographies and commuting and residential choice linkages.

Setup. There are $i \in \{1, \ldots, I\}$ locations and $k \in \{I, F\}$ types of housing. Individuals of type $g \in \{1, \ldots, G\}$ choose across tuples (i, j, k) of where to live i, where to work j, and the type of housing to live in k. Time is discrete. For simplicity, we consider a two-period model, where $t = 0$ refers to the pre-policy period and $t = 1$ the post-policy period. We begin by considering the static equilibrium that takes residential populations and floorspace supplies $\{L_{Rikg}, H_{ik}\}$ as given and solves for labor and housing market equilibria that hold in any period, and omit time subscripts where it causes no confusion.

5.1 Static Module: Setup

Individuals. Indirect utility for worker $\omega \in [0, \bar{L}_g]$ of type g who has chosen to live in location i and housing type k and work in j is given by

$$U_{ijkg}(\omega) = \rho_g \ln u_{igk} + \ln w_{jg} - (1 - \beta) \ln r_{Rik} - \ln d_{ij} + \epsilon_j(\omega)$$

where u_{ikg} is the amenity for type-g individuals living in location i and housing type k, w_{jg} is the wage earned by type-g workers in location j, r_{Rik} is the price of type-k residential floorspace in i, $d_{ij} = \exp(-\delta_C t_{ij})$ is a commute disutility cost that depends exponentially on travel time t_{ij} and the parameter $\delta_C > 0$. $\epsilon_j(\omega)$ is an idiosyncratic taste of worker ω to work in j. The parameter ρ_g reflects type-g's preference weight on amenities. For example, if higher-educated individuals care more about amenities then $\rho_H > \rho_L$.

Within a period workers take their choice of residential location and housing type (i, k) as given and choose in which location j to work. Assuming $\epsilon_j \sim T1EV(1/\theta_g)$, the probability a worker commutes to j and her average income are

$$\pi_{j|ikg} = \frac{(w_{jg}/d_{ij})^{\theta_g}}{\sum_j (w_{jg}/d_{ij})^{\theta_g}}$$
$$\bar{y}_{ig} = E_g \sum_s \pi_{j|ikg} w_{jg}$$

where E_g is the number of effective units of labor type-g workers are endowed with. Expected utility of type g workers in (i, k) is therefore

$$\bar{U}_{ikg} = \ln \left(u_{igk}^{\rho_g} \tilde{y}_{ig} r_{Rik}^{\beta - 1} \right),$$

where $\tilde{y}_{ig} = \left[\sum_s (w_{sg}/d_{is})^{\theta_g} \right]^{1/\theta_g}$ captures location i's access to well-paid jobs.

Firms. The numeraire good is a CES aggregate of individual freely-traded varieties produced across the city (with the elasticity of demand σ_D). These varieties are differentiated by informal and formal sector $s \in \{I, F\}$, location and floorspace type. Firms in either sector can produce on either floorspace type (consistent with our data), but productivity A_{jsk} can differ by location, sector and floorspace type. This captures that formal firms may be more productive in formal buildings, for example. We make the Armington assumption so that there is a representative firm in each cell (j, s, k) producing under perfect competition using a Cobb-Douglas aggregate of labor and commercial floorspace

$$Y_{jsk} = A_{jsk} \left[\left(\sum_g \tilde{\alpha}_{gs}^{\frac{1}{\sigma_L}} L_{sgjk}^{\frac{\sigma_L - 1}{\sigma_L}} \right)^{\frac{\sigma_L}{\sigma_L - 1}} \right]^{\alpha} H_{Cjsk}^{1-\alpha}$$

Prices are equal to marginal costs and are given by

$$p_{jsk} = \frac{W_{js}^{\alpha} r_{Cjk}^{1-\alpha}}{A_{js}} \quad \text{where} \quad W_{js} = \left(\sum_g \tilde{\alpha}_{gs} w_{jg}^{1-\sigma_L} \right)^{\frac{1}{1-\sigma_L}},$$

where $\tilde{\alpha}_{gs}$ are shifters that affect demand from sector-s firms for type-g workers. H_{Cjsk} is the amount of commercial floorspace of type-k hired by firms from industry s in j, while L_{Csjkg} is the number of type-g workers they hire. Factor demands are given by

$$\begin{aligned} L_{Cjg} &= \alpha w_{jg}^{-\sigma_L} \sum_{sk} \tilde{\alpha}_{gs} W_{js}^{\sigma_L - 1} X_{jsk} \\ (1 - \vartheta_j) r_{Cjk} H_{Cjk} &= (1 - \alpha) \sum_s X_{jsk} \end{aligned} \tag{4}$$

where $X_{jsk} = \beta \left(W_{js}^{\alpha} r_{Cjk}^{1-\alpha} \right)^{1-\sigma_D} A_{jsk}^{\sigma_D - 1} \sum_{ik} Y_{ik}$ and the final good is the numeraire.

5.2 Static Module: Equilibrium

Land Market Clearing. Letting ϑ_{ik} denote the fraction of floorspace allocated to residential use, market clearing for residential floorspace implies

$$r_{Rik} = (1 - \beta) \frac{\sum_g \bar{y}_{ig} L_{Rikg}}{\vartheta_i H_{ik}}. \tag{5}$$

Floorspace is allocated to its highest-paying use, leading to a no-arbitrage condition on prices $r_{Rik} = r_{Cik} = r_{ik}$. Plugging this into the market clearing conditions for residential

and commercial floorspace yields

$$\vartheta_{ik} = \frac{(1-\beta)Y_{ik}}{(1-\beta)Y_{ik} + (1-\alpha)X_{ik}} \tag{6}$$

where $Y_{ik} = \sum_g \bar{y}_{ig} L_{Rikg}$ is total housing spending and $X_{ik} = \sum_s X_{isk}$ is total sales.[25]

Labor Market Clearing. Labor supply can be calculated simply from

$$L_{Cjg} = \sum_{i,k} \pi_{j|ikg} L_{Rikg} = \sum_{ik} \frac{(w_{jg}/d_{ij})^{\theta_g}}{\sum_s (w_{sg}/d_{is})^{\theta_g}} L_{Rikg}. \tag{7}$$

Labor market clearing requires that this equals labor demand in (4).

Endogenous Amenities. We allow for endogenous amenities u_{ikg} which depend on exogenous attributes $\bar{u}_{ikg}$ as well as the share of high-type individuals living close by

$$u_{ikg} = \bar{u}_{ikg} \left[\sum_j \exp\left(-\delta_U t_{ij}\right) \frac{L_{RjH}}{L_{Rj}} \right]^{\mu_U} \tag{8}$$

The parameter μ_U controls the overall size of these externalities and δ_U controls the rate at which they decay through space.[26] Since residential populations are taken as given, total amenities u_{ikg} are exogenous to the static equilibrium. Since we do not observe spatial spillovers from commercial mills, we do not allow for productivity spillovers in the baseline model and have A_{jsk} act as location characteristics exogenous to the model.

Static Equilibrium Definition. *Given a vector $\{L_{Rikg}, H_{ik}, u_{ikg}, t_{ij}\}$, a static equilibrium is a tuple $\{L_{Cjg}, w_{jg}, r_{ik}, \vartheta_{ik}\}$ such that: i) labor markets clear (4),(7); ii) floorspace markets clear (5) and iii) floorspace use satisfies land owner optimality (6).*

5.3 Dynamic Module: Floorspace Supply

In each location, there is a supply of developable land T_i with plots indexed by $\ell \in [0, T_i]$. At time t each plot is either developed formally or informally. Developers choose whether

[25] This condition can be rearranged to show that the market for commercial floorspace also clears at r_{ik}.

[26] The exogenous component $\bar{u}_{ikg}$ contain features of location i and housing type k that are valued by group g that are not perfectly summarized by its share of high-type residents (e.g. access to green space, noise, quality of buildings). One could also assume these spillovers depend on housing type - we did not have enough residual variation in the education share of residents moving into locations with mills to disentangle these channels empirically.

to redevelop it to an alternative use for $t+1$ using two available technologies.[27]

Formal Technology. The formal technology can construct multistory buildings using a Cobb-Douglas aggregator of land and capital $H_{Fi} = T_i^{1-\eta} K_i^{\eta}$. Capital is available at the same price across the city. We assume formal real estate in Mumbai is small relative to the overall capital market in India and take this price as given, and set it to one.

If developers redevelop a plot of land that was previously a slum, they have to compensate the slum dwellers who previously lived there. The compensation due is based on the actual policy in Mumbai. A fraction γ of residents are eligible for compensation.[28] Eligible slum dwellers are given a housing unit in a location which is a random draw from the observed distribution π_i^{SRA} of SRA housing units across locations (see Figure A2).[29] In Mumbai during our post-policy period, the size of the unit was 25 square meters which corresponds to the 34th percentile of the distribution of formal floorspace in the 2012 National Sample Survey on drinking water, sanitation, hygiene, and housing condition. As discussed in Section 2, recipients frequently rent or sell these units and live elsewhere. We, therefore, model this compensation policy as a transfer from developers to slum dwellers with value

$$y^{comp} = \sum_i \pi_i^{SRA} r_{iFt} h^{SRA}.$$

That is, the compensation is the average value of a 25 square meter housing unit h^{SRA} across the observed distribution of SRA units.

The developer's problem, if they use the formal technology, is to choose how many units of housing to build on a unit of land. Excluding the cost of land which will be added below, variable profits per unit of land are

$$\max_{k_i} \; \nu(\ell) r_{iFt+1} k_i^{\eta} - k_i - F_i \quad \text{where} \quad F_i = y^{comp} \gamma \frac{L_{RiIt}}{\lambda_{Iit} T_i}.$$

[27] We assume developers are small and behave competitively for two reasons. First, Appendix SM.2.2 shows that the environment in Mumbai looks broadly more competitive than several other large Indian cities. Second, while the assumption that development decisions are made independently across many small plots may seem extreme, it delivers a semi-log-linear expression for the probability a slum is redeveloped that depends on the relative price of formal floorspace in that location. We, therefore, view this more as a microfoundation to get us to this relationship; an alternative approach would be to assume a general functional form for this redevelopment probability that depends on relative prices and take a first-order approximation.

[28] In the policy this is based on tenure in the slum (see Section 2).

[29] As discussed in Section 2, while the Slum Rehabilitation Act specified these units should be built in the same location as much as possible, they were commonly built in different locations often at the outskirts of the city. They were often frequently rented out or sold, and so we model the transfer in the amount of the value of the tenement.

Here k_i^{η} are the units of housing built if k_i units of capital are used per unit of land, and $\nu(\ell)$ is an idiosyncratic profitability for formal construction in plot ℓ. Profits are therefore given by sales from floorspace construction, net of construction costs and the cost F_i of rehousing slum dwellers living on each unit of land in i at t (the term $L_{RiIt}/\lambda_{Iit}T_i$, times the fraction eligible for compensation γ). Solving this problem and letting $\tilde{\eta} \equiv \eta^{\eta/(1-\eta)}$ yields profits per unit of land

$$\pi_{iF|It}(\ell) = (1-\eta)\tilde{\eta} r_{iFt+1}^{\frac{1}{1-\eta}} \nu(\ell) - F_i.$$

Informal Technology If the informal technology is chosen, then single-story buildings are built at no cost, so profits are simply

$$\pi_{iI|It}(\ell) = r_{iIt+1}.$$

Redevelopment Choice. Developers choose to redevelop an informal plot to formal if $\pi_{iF|It}(\ell) > \pi_{iI|It}(\ell)$. We assume that $\nu(\ell)$ is drawn from an exponential distribution with shape κ, so that the conversion probability is given by

$$p_{F|Iit} = \exp\left(-\kappa \left(\frac{F_i + r_{Iit+1}}{(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}} \right) \right).$$

Developers are more likely to redevelop slums when the price of formal floorspace is high, relative to informal prices and the cost of compensating existing slum dwellers. Here κ captures features other than compensation costs that make it harder to redevelop slums, as it controls the sensitivity of slum conversions to the profitability from doing so.

As we only have two periods of data, we need to calibrate to a steady state of the model for the pre-policy period. To allow for both formal and informal land use in this steady state (as in the data), we allow for formal plots to become informal again. We repeat the conversion process above for formal plots, with the only difference being that formal plots do not face redevelopment costs. The probability a formal plot remains formal is

$$p_{F|Fit} = \exp\left(-\kappa \left(\frac{r_{Iit+1}}{(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}} \right) \right).$$

Land Use and Floorspace Supply. The share of formal land λ_{Fit+1} at $t+1$ is therefore

$$\lambda_{Fit+1} = \lambda_{Iit} p_{F|Iit} + \lambda_{Fit} p_{F|Fit} \tag{9}$$

and $\lambda_{Iit+1} = 1 - \lambda_{Fit+1}$. Floorspace supply is

$$H_{iFt+1} = \lambda_{iFt+1} T_i \tilde{\eta} r_{iFt+1}^{\frac{\eta}{1-\eta}} \tag{10}$$

$$H_{iIt+1} = \lambda_{iIt+1} T_i \tag{11}$$

where the first line use optimal capital demand $k_i = \tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}$ and housing density $h_i = k_i^{\eta}$.

Developer Profits and Land Prices. As shown in Appendix SM.3.3.1, average profits in plots that are developed informal and formal in i at time $t+1$ are

$$\Pi_{Fit+1} = \frac{1}{\kappa} \left[\kappa r_{Iit+1} + (1-\eta) \tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} \right] \qquad \text{and} \qquad \Pi_{Iit+1} = r_{Iit+1}.$$

Total developer profits in location i are then

$$\Pi_{it+1} = T_i \times \left[\lambda_{Fit+1} \Pi_{Fit+1} + \lambda_{Iit+1} \Pi_{Iit+1} \right].$$

In each location we assume there is free entry amongst developers so that the price of land is bid up to developer profits.

5.4 Dynamic Module: Residential Choice

Residents start period t in a location-housing type combination and decide where to live next period. The value function of worker ω of type g who lives in (i,k) at t is

$$W_{ikg}(\omega) = \bar{U}_{ikgt} + \beta^D \Bigg[\psi_{ikt} \underbrace{\max_{rs \neq ik} \left\{ \bar{U}^E_{rsgt+1} - \ln \mu_{ikrsg} + \varepsilon_{M,rs}(\omega) \right\}}_{\text{Evicted}} + (1 - \psi_{ikt}) \underbrace{\max_{rs} \left\{ \bar{U}_{rsgt+1} - \ln \mu_{ikrsg} + \varepsilon_{M,rs}(\omega) \right\}}_{\text{Not Evicted}} \Bigg].$$

The first term $\bar{U}_{ikgt}$ is the flow utility from their (i,k,g) state at time t. The term in the bracket represents expected utility in period $t+1$. With probability ψ_{ik} the resident is evicted and forced to move to another location $rs \neq ik$. With probability $1 - \psi_{ik}$ they are not evicted and can choose to live in all locations (including their current one). We parameterize the probability of eviction as depending on the net change in slums

through[30]

$$\psi_{ikt} = \begin{cases} \mathbb{I}\left\{\lambda_{iIt+1} < \lambda_{iIt}\right\}\left[1 - \lambda_{iIt+1}/\lambda_{iIt}\right] & k = I \\ 0 & k = F \end{cases}.$$

This satisfies i) $\psi_{ikt} \in [0,1]$; ii) if there is no change in slums then $\psi_{ikt} = 0$ and iii) if $\lambda_{iIt+1} = 0$ then $\psi_{ikt} = 1$. There is no eviction from formal housing, so $\psi_{iFt} = 0$.

Workers of type-g also face a moving cost μ_{ikrsg} to move from (i,k) to (r,s), consisting of a fixed cost F_g and a variable cost that depends on travel time at rate δ_g^M

$$\mu_{ikrsg} = \exp\left(F_g \times \mathbb{I}\left\{ik \neq rs\right\} + \delta_g^M t_{ir}\right).$$

While most quantitative models incorporate the variable component, the fixed cost allows for the event of moving to be costly regardless of the move's distance. We also allow for moving to be especially costly for the poor by letting these costs differ by worker group.

Residents also have idiosyncratic preference shocks $\varepsilon_{M,rst+1}$ for moving to each location-housing type pair (r,s) in $t+1$. Assuming these are drawn from a T1EV distribution with shape ν,[31] expected welfare $W_{ikg} \equiv E_\omega[W_{ikg}(\omega)]$ is given by

$$W_{ikg} = \bar{U}_{ikgt} + \beta^D\left[\psi_{ikt}\ln\Omega^E_{ikgt+1} + (1-\psi_{ikt})\ln\Omega^{NE}_{ikgt+1}\right],$$

$$\text{where } \Omega^E_{ikgt+1} = \left[\sum_{rs\neq ik}\left(b^E_{rsgt+1}/\mu_{ikrsg}\right)^\nu\right]^{1/\nu}$$

$$\text{and } \Omega^{NE}_{ikgt+1} = \left[\sum_{rs}\left(b_{rsgt+1}/\mu_{ikrsg}\right)^\nu\right]^{1/\nu}.$$

Here we have defined the mean utility terms as $b_{rsgt+1} = \exp(\bar{U}_{rsgt+1}) = u_{igk}^{\rho_g}\tilde{y}_{ig}r_{Rik}^{\beta-1}$ and $b^E_{rsgt+1} = u_{igk}^{\rho_g}\tilde{y}^E_{ig}r_{Rik}^{\beta-1}$ where $\tilde{y}^E_{ig} = \left[\sum_s((w_{sg}+\gamma y^{comp})/d_{is})^{\theta_g}\right]^{1/\theta_g}$ is access to jobs accounting for the expected income that evicted residents receive. The terms Ω^E_{ikgt+1} and

[30] An alternative is to tie eviction probabilities to the gross change in slums. However, this requires the measurement of eviction probabilities in the steady state, which is not observed without an additional year of data before the policy was implemented. It also substantially increases the time it takes to invert the model to solve for unobservables while evaluating the model moments (current estimation takes around two days on a high-performance computing cluster). Mean utilities can no longer be independently estimated from other unobservables, as informal prices now factor into the residential choice equation via the steady-state redevelopment probability. Given that slum dwellers would likely only face eviction during a net change in slums, and considering these practicalities, we opted for this particular parameterization.

[31] This shape parameter is constant across groups. In estimation, this enters multiplicatively with the group-specific migration cost parameters F_g, δ_g^M. Allowing ν to vary by group would therefore lead to the same migration costs since it is νF_g and $\nu\delta_g^M$ which control the residential choice moments in the model which we match (exactly) to the data.

Ω^{NE}_{ikgt+1} are the expected utility that individuals of type g starting in location i and housing-type k received if they are evicted and not evicted, respectively. The only difference is that those evicted are forced to move (they cannot choose to live in pair (i, k)), and with probability γ are compensated.

Standard results on T1EV distributions imply that the number of individuals of type-g moving from i, k to r, s is

$$L_{rs|ikgt} = \left[(1 - \psi_{ikt}) \left(\frac{b_{rsgt+1}}{\mu_{ikrsg}\Omega^{NE}_{ikgt}} \right)^{\nu} + \psi_{ikt}\mathbb{I}\{ik \neq rs\} \left(\frac{b_{rsgt+1}}{\mu_{ikrsg}\Omega^{E}_{ikgt}} \right)^{\nu} \right] L_{Rikgt} \tag{12}$$

so that the number of residents in r, s at $t + 1$ is simply $L_{Rrst+1} = \sum_{ik} L_{rs|ikgt}$.

Dynamic Equilibrium Definition. *Given an initial distribution $\{L_{Rikgt}, \lambda_{ikt}\}$ in the pre-policy period, a dynamic equilibrium in the post-policy period is given by a tuple $\{L_{Rikgt+1}, \lambda_{ikt+1}, H_{ikt+1}, u_{ikgt+1}\}$ such that i) residential populations are given by $L_{Rrst+1} = \sum_{ik} L_{rs|ikgt}$ with $L_{rs|ikgt}$ given by (12); ii) land use shares λ_{ikt+1} follow (9); iii) floorspace supplies are given by (10) and (11); amenities u_{ikgt+1} satisfy (8) and iv) the static equilibrium conditions hold in $t + 1$.*

5.5 Modeling the Mumbai Mills Redevelopment

We model the DCR 58 law change as opening up the mill locations at $t + 1$ which at t were closed off from economic activity. Formally, $i \in \mathcal{L}_t$ are the locations with economic activity at t. Then $i \in \mathcal{L}_{Mills}$ are the set of locations i such that $t_{rit} = \infty \forall r \in \mathcal{L}_t$ and $\bar{u}_{rkg} = A_{rks} = 0 \; \forall r, k, g, s$. Then in $t + 1$, $t_{rit} < \infty \; \forall \; r \in \mathcal{L}_t$, mills with commercial activity draw A_{Mill} productivity in formal sector employment and formal commercial floorspace, and mills with residential activity draw $\bar{u}_{Mill}$ amenity in formal residential floorspace.[32] Since new amenities and productivities are drawn only in the formal sector, these sites are fully specialized in formal activity as in the data.[33] The set of active locations at $t + 1$ are therefore $\mathcal{L}_{t+1} = \mathcal{L}_t \cup \mathcal{L}_{Mills}$.

[32] This assumes the effects of mill sites are due to actual economic activity on the redeveloped sites rather than, for instance, anticipation of the activity. See Appendix C.2 for evidence this is the case.

[33] If we instead considered mills to be an expansion of land in locations, we would not see full specialization on mills and would need to assume mills inherit the same amenities and productivities. While the removal of the mills could improve amenities directly due to removing an eyesore, our empirical results at the end of Section 4.3 show that controlling for the share of mill land within a location or the distance to its closest mill (capturing the eyesore effect) does little to the point estimate of the impact of exposure to mills in nearby locations. This suggests this channel is not first order in the data.

Welfare Consequences of the Mill Redevelopment. We measure the welfare consequence of the policy as the change in welfare $W'_{ikg} - W_{ikg}$ if the only thing that happened between t and $t+1$ was the development of the mill sites. We focus here on the welfare change for those evicted without compensation to develop intuition, but analogous expressions apply to those evicted with compensation and those who are not evicted (see Appendix SM.3.3.2).

The change in welfare for those who are evicted without compensation is

$$\Delta W_{ikg}^{E,NoComp} = \frac{1}{\nu} \ln \left(\sum_{rs \in \mathcal{L}_{t+1} \setminus ik} \pi_{rs|ikg}^{L} \left(\hat{u}_{ikgt}^{\rho_g} \hat{\tilde{y}}_{rgt} \hat{r}_{Rrst}^{\beta-1} \right)^{\nu} + \pi_{ikg}^{Mill,E,NoComp} \times \left(\frac{\Omega_{ikgt+1}^{E,NoComp}}{\Omega_{ikgt}^{NE}} \right)^{\nu} \right)$$

where $\Omega_{ikgt+1}^{E,NoComp}$ is the average post-period utility of those evicted from (i,k) without compensation.

The first term captures the change in mean utilities at each location active at t, weighted by the share of workers choosing that location in the pre-period. Eviction hurts slum dwellers by forcing them to move: they lose the value from staying in their current location, $\pi_{ik|ikg}^{L} \left[\hat{u}_{rsgt}^{\rho_g} \hat{\tilde{y}}_{rgt} \hat{r}_{Rrst}^{\beta-1} \right]^{\nu}$, and are forced to pay moving costs. This loss will be high when moving is infrequent (i.e. $\pi_{ik|ikg}^{L}$ is high) and when mean utility rises in their previous place of residence. The second term captures the variety effect of adding new locations for residence on mill sites, reflected in the fraction of individuals in i,k,g choosing them π_{ikg}^{Mill} and their average change in welfare. The contribution of this term will be small for evicted slum dwellers if few of them choose to live in redeveloped mills.

6 Model Estimation

We estimate the model's key parameters through indirect inference. For a vector of parameters, we first calibrate to the pre-period equilibrium and feed in the mill redevelopment shock holding all other model unobservables constant. We then compute model moments, most of which consist of our empirical regression coefficients. We then search for parameters that minimize the distance between model and data moments. The identification assumption is therefore that the changes in economic activity surrounding mill sites are caused by exposure to the redevelopment on mill sites and not by differential trends in unobservables in these locations. This is true by construction in the model moments, and is the same exclusion restriction we provided evidence for in Section 4. [34]

[34] As discussed in Appendix D.1, we harmonize our datasets to one common geography (hex cells). These constitute locations i in the model. When computing moments, we aggregate back up to spatial

Model Unobservables. In order to recover the model unobservables from available data, we make one further parametric assumption that $A_{jsk} = A_{js}\iota_{sk}$. That is, productivity at the location-sector-floorspace level has a location-sector specific component and a sector-floorspace type-specific component. We normalize $\iota_{IF} = \iota_{II} = \iota_{FI} = 1$ and therefore have only ι_{FF} as a free parameter which characterizes how much more productive formal sector firms are in formal floorspace.

We assume the model is in steady state in the pre-policy period t. Appendix D.2 shows that given i) a vector $(\alpha, \beta, \theta_g, \delta_C, \kappa, \eta, E_g, \sigma_L, \tilde{\alpha}_{gs}, \delta_U, \mu_U, \rho_g, \delta_g^M, F_g, \iota_{FF})$ of structural parameters and ii) data from the pre-policy period $\{L_{Rikg}, L_{Cis}, \lambda_{ik}, r_{iF}, t_{ij}, T_i\}$, then a vector $(w_{ig}, r_{Ii}, b_{ikg}, u_{ikg}, \bar{u}_{ikg}, X_{js}, A_{jsk}, \tilde{H}_{ik}, \xi_{ik}, \vartheta_{ik})$ of unobservables can be recovered. With these in hand, we proceed to estimating the model's structural parameters.

Calibrated Parameters. The parameters $(\sigma_L, \sigma_D, \delta_C, \eta, \alpha, \beta, \tilde{\alpha}_{gs}, \iota_{FF}, \gamma)$ are calibrated to external and internal moments without solving the model. Since many are fairly standard, we relegate them to Table A13 and their discussion to Appendix D.3. That section also reports the estimation of θ_g from a gravity equation for commute flows. We find that low-educated workers have more elastic commute flows with $\theta_H = 2.69, \theta_L = 3.6$, consistent with a value around 3 reported in the literature.

Indirect Inference. We estimate the remaining parameters to match moments on residential moves (identifying moving costs F_g, δ_g^M) and our reduced form regression coefficients (identifying $\kappa, \mu_U, \delta_U, \bar{A}_{Mill}, \bar{u}_{Mill}, \rho_g$). We begin by normalizing ρ_L so that $E[\rho_g] = 1$, i.e. the average valuation of amenities is one.[35]

Moving costs are identified by two sets of moments. First, F_g is identified by the fraction of people who move within 15 years (one period in the model) since a higher fixed cost of moving makes moving less likely. Second, the variable cost of making distant moves δ_g^M is identified by the odds ratio of the probability of moving within the same neighborhood to the probability of moving to a different neighborhood (conditional on moving). This is because a higher variable cost makes it more likely that residents make shorter moves and stay within the same neighborhood. See Appendix D.4 for details.[36]

units from the empirical analysis. For example, house price regressions in the model aggregate first to the subzone level before running the regression, population regressions aggregate to election wards, etc. This way we run precisely the same regression in the model as in the empirics.

[35]The overall size of the spillover μ_U cannot be separately identified from the scale of ρ_g, which motivates this normalization. We prefer this normalization, rather than ignoring μ_U altogether, since it gives an estimate of the overall scale of spillovers that can be benchmarked to other contexts.

[36]Since these moments do not require us to solve the full model, we estimate migration parameters separately. This substantially speeds up estimation of the remaining parameters that require solving the

The results are shown in Table 5. The model is able to exactly match all four moments. High-educated workers move more frequently (45% have moved in the past 15 years versus 33% of low-educated workers), so we estimate a higher fixed cost of moving for the low-educated. Since the high-educated are more likely to make long-distance moves (between neighborhoods) than the low-educated, we find a slightly lower variable migration cost for high-educated workers of 0.011 compared to 0.013. Our fixed cost estimate is large, equivalent to the cost of moving around two and a half hours away. This shows the importance of allowing for fixed moving costs when considering slum relocation, since being evicted even to neighborhoods close by can impose large costs.

Table 5: Relocation Cost Parameters

Panel A: Model and Data Moments

Moment	**Param Id'd**	**Model**	**Data**
Fraction Moved in Last 15 Years, High-Skill	F_H	0.454	0.454
Fraction Moved in Last 15 Years, Low-Skill	F_L	0.338	0.338
Odds Ratio Move In vs Outside Nbhd, High-Skill	δ_H^M	0.622	0.623
Odds Ratio Move In vs Outside Nbhd,Low-Skill	δ_L^M	0.782	0.784

Panel B: Parameter Estimates

Parameter	**Description**	**Estimate**	**Standard Error**
F_H	Fixed Moving Cost, High-Skill	1.401	0.0312
F_L	Fixed Moving Cost, Low-Skill	1.477	0.0197
δ_H^M	Variable Moving Cost, High-Skill	0.011	0.0004
δ_L^M	Variable Moving Cost, Low-Skill	0.013	0.0003

Notes: Bootstrapped standard computed over 400 draws of moments from a non-parametric bootstrap where we resample the data 400 times with replacement. See Appendix D.4 for details.

With these relocation cost estimates in hand, we estimate the remaining parameters to match our empirical regression coefficients as shown in Table 6. While these parameters are jointly estimated and depend on each other, we discuss the sources of variation used to identify each one. The precise algorithm is detailed in Appendix D.4.

The productivity of mills redeveloped to commercial use $\bar{A}_{Mill}$ is defined relative to average productivity in formal land across the city. This is identified by the increase in formal employment on commercially redeveloped mill sites in Table 4, since greater relative productivity on these sites increases their employment growth in the post-policy period. Given the very large increase of 1.678 log points in the data, we estimate that commercial mills are about 20 times more productive than the average formal building

whole model, since mean utilities (and therefore amenities) can be solved for once rather than repeatedly for each draw of the parameters estimated in the indirect inference step.

in the city.[37] This will be important when considering the welfare consequences of the policy, since it opened up a lot of very productive sites for production by formal firms.

The amenity on mills redeveloped to residential use $\bar{u}_{Mill}$ is defined relative to formal locations nearby (within a 20-minute walk of a mill site).[38] This is identified by the rental price of mill properties relative to other properties within the neighborhood, as a higher $\bar{u}_{Mill}$ increases rents on mill sites all else equal.[39] This is a 30% premium in the data, which we match closely. The model matches this by giving residential mills only 60% of the formal amenity than its adjacent properties. This is because the geographies differ, with neighborhoods in the 99acres data being larger than a 20-minute walking radius from mills and the locations surrounding mills having higher formal amenities relative to these broader neighborhoods. However this parameter is imprecise, meaning we cannot reject residential mills having higher (or lower) amenities than adjacent formal properties.

Conditional on $\bar{u}_{Mill}$ being pinned down by this moment, the overall size of the amenity spillover μ_U is identified through the increase in formal floorspace prices in locations containing residential mills. The spatial decay δ^U is identified by the estimated spillover of how much these prices increased in locations near to but not including residential mills.[40] The estimate of the spillover of 0.88 is very close to the 0.81 Tsivanidis (2022) finds in Bogotá using a similar model, and about one quarter of Diamond (2016)'s estimate for the US.[41] The estimated spatial decay of 0.19 is smaller than Ahlfeldt et al. (2015)'s estimate for Berlin, but spillovers still decay fairly quickly in travel time. They decrease by about 85% with 10 minutes of travel time and decay almost completely 20 to 30 minutes out.

The sorting parameter ρ_H is identified by the increase in the share of high-educated workers in election wards containing mills, given that the shock to amenities in mills common across worker groups. A higher value of ρ_H will, all else equal, lead to greater relative inflows of high-educated residents into these neighborhoods since they value the new amenities more. Our estimate of $\rho_H = 1.034$ suggests high-types are more willing

[37] An alternative would be to have commercial mills draw a multiple of productivity nearby, rather than productivity across the whole city, which might lead to a smaller estimate. It would not matter for our quantification, since the change in formal employment would be the same.

[38] That is, residential mills draw an amenity equal to $\bar{u}_{Mill}$ times the average amenity in formal housing within a 20-minute walk of individual mill (i.e. it can vary by mill).

[39] This data comes from the 99acres website in 2021, which lists properties available for rent in the formal sector. Other properties in the neighborhood are defined as those in Mumbai South and Mumbai Harbor, the two of nine neighborhoods on 99acres containing mills.

[40] Note we use distance to residential mills rather than all mills in general as reported in Table 1, since this directly identifies the residential amenities which directly change education-mix of residents nearby.

[41] See Appendix SM.4 for the derivation of the comparison with Diamond (2016) as well as with more standard functional forms such as Ahlfeldt et al. (2015).

Table 6: Model Fit and Estimated Parameters

Panel A: Model and Data Moments

Moment	Param Id	Model	Data
β in $\ln \hat{L}_{i,\text{ Formal}} = \beta \mathbb{I}\{\text{ContainsMillComm}\}_i + ... + \epsilon_i$	$\bar{A}_{Mill}$	1.664	1.678
β_1 in $\ln \hat{r}_{Fi} = \beta_1 \mathbb{I}\{\text{ContainsMill}\}_i + \beta_2 \mathbb{I}\{\text{1m-500m}\}_i + \beta_3 \mathbb{I}\{\text{500m-1km}\}_i + \epsilon_i$	μ_U	0.236	0.236
β_2 in $\ln \hat{r}_{Fi} = \beta_1 \mathbb{I}\{\text{ContainsMill}\}_i + \beta_2 \mathbb{I}\{\text{1m-500m}\}_i + \beta_3 \mathbb{I}\{\text{500m-1km}\}_i + \epsilon_i$	δ_U	0.063	0.063
β_1 in $\text{EdIndex}_i = \beta_1 \mathbb{I}\{\text{ContainsMill}\}_i + \beta_2 \mathbb{I}\{\text{1-500m}\}_i + \beta_3 \mathbb{I}\{\text{500m-1km}\}_i + \epsilon_i$	ρ_H	0.315	0.262
β_1 in $\ln \hat{\lambda}_{Ii} = \beta_1 \mathbb{I}\{\text{ContainsMill}\}_i + \beta_2 \mathbb{I}\{\text{1m-500m}\}_i + \beta_3 \mathbb{I}\{\text{500m-1km}\}_i + \epsilon_i$	κ	-0.140	-0.156
$r_{F,Mill}/r_{F,NearMills}$	$\bar{u}_{Mill}$	1.256	1.311

Panel B: Parameter Estimates

Parameter	Description	Estimate	Standard Error
$\bar{A}_{Mill}$	Productivity of commercial mill sites	22.909	3.895
μ_U	Elasticity of amenities to nbhd composition	0.886	0.235
δ_U	Spatial decay of endogenous amenity	0.190	0.058
ρ_H	Relative preference weight on amenities for H-skill	1.034	0.122
κ	Slum conversion elasticity	4.947	1.354
$\bar{u}_{Mill}$	Amenity on residential mill sites	0.631	0.575

Notes: Bootstrapped standard computed over 400 draws of moments from a normal distribution with mean and standard deviation given by the mean and standard error of our empirical point estimates. See Appendix D.4 for details.

to pay for amenities as one might expect.

Lastly, the slum conversion parameter is identified by the reduction in slums in locations with mills.[42] We find a semi-elasticity of slum conversion to relative profits of $\kappa = 4.947$, suggesting relatively low conversion costs.[43] While our model separately accounts for the cost of compensating slum dwellers through the fixed cost F_i so this should not be reflected in this estimate, this parameter to some extent reflects the institutional context of Mumbai.

[42] The model and data moment here are using OLS rather than PPML, which took too long to be computationally feasible. Each evaluation of the objective could take minutes at some parameter vectors; estimation takes almost two days on a high-performance computing cluster.

[43] One may be concerned that mill developers are undertaking slum redevelopment themselves to clear disamenities around their high-priced new buildings. However, in Appendix SM.2.1 we show that there is no overlap in the identities and mill and slum redevelopers.

7 Welfare Implications

7.1 Quantifying the Impact of the Mill Redevelopment

Aggregate and Distributional Effects. To compute the consequences of DCR 58, we follow the same exercise as in the indirect inference procedure and allow the mill sites to be redeveloped but hold all other unobservables at their pre-period values. This isolates the impact of the policy from other changes that occurred to the city during the 2000s.

Table 7 shows the average welfare changes across worker categories. Overall, workers who are not evicted gain a lot from the policy. High-educated workers gain about 21.4% in welfare terms while the low-educated gain by 6.7%. The fact that high-type workers benefit the most reflects that they value amenities more, have lower relocation costs, and receive relatively higher wages in commercially-redeveloped mills due to the high demand they face from formal firms there.

Table 7: Welfare Effects

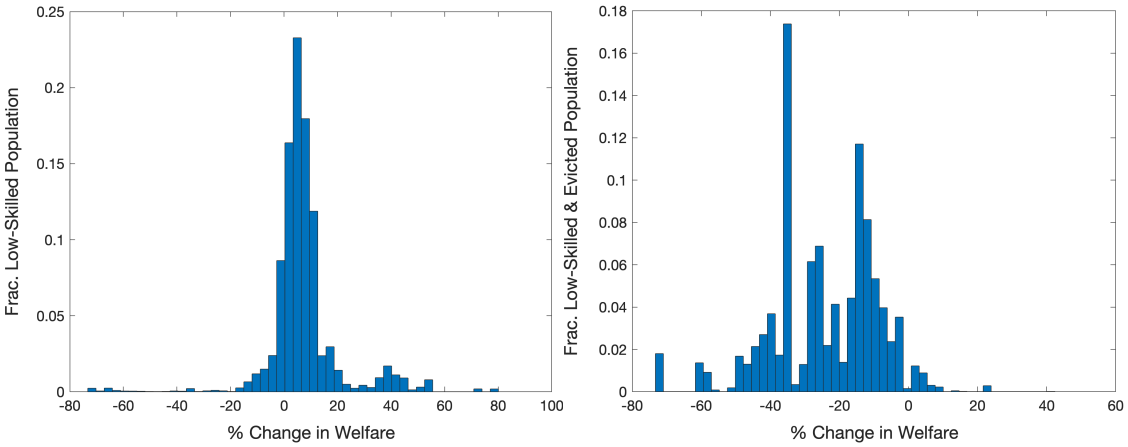
	Formal	**Informal**			
		Not Evicted	Evicted Average	Evicted Compensated	Evicted Not Compensated
High Skill	0.214	0.168	0.019	0.039	-0.030
Low Skill	0.067	0.083	-0.245	-0.202	-0.352

Notes: Results show % changes in welfare divided by 100.

The same is not true for the evicted. Low-educated workers evicted from their slums experience an average drop in welfare of 24.5%, ranging from 35.2% for those not compensated to 20.2% for those compensated. High-educated workers who are evicted are hurt much less, due to the three differences with low-educated workers mentioned above.

These numbers reflect the average welfare effects across workers in different initial states and mask the underlying distribution of welfare impacts. Figure 4 shows this distribution for low-educated workers. Panel (a) shows a rich heterogeneity of welfare effects, where the average is around 6% but many gain by more. A long and thin left tail shows there are some who are hurt a lot by the policy. Panel (b) zooms in on these individuals who are evicted and shows most of them experience drops in welfare of 20% or more, with some even below 60%. Since only about 2% of slum residents are evicted due to the policy, this is not a large number of individuals, but those who are experience very large drops in welfare.

Figure 4: Distribution of Welfare Effects for the Low-Educated



(a) All Workers (b) Evicted

Decomposing the Source of the Effects. Given the large increases in welfare, Table 8 uses the results from Section 5.5 to decompose the effects allowing each margin to change sequentially. Eviction alone leads to very large drops in welfare of 28.2% for high-educated workers and 52.2% for the low-educated. High-type workers benefit a lot from the adjustment of amenities in the second row, while low-educated workers are little affected. This is driven by our estimate of ρ_H above 1. The largest change comes when income is allowed to adjust in the third row. This increases welfare by around 20 percentage points for both worker types. This highlights how the very high productivity of commercial towers on commercially-redeveloped sites drives large aggregate gains throughout the city. In aggregate, house price response hurts workers slightly since they appreciate close to the mills.[44] The last row, which incorporates the variety effect from adding the new locations to the model, leads to a mild increase in welfare.

Isolating the Impact of Commercially Redeveloped Mills. One feature of the mill redevelopment that makes it unique is that both commercial and residential high rises were constructed. The previous discussion highlights the outsize impact of the commercial high rises. Often the policy debate surrounds residential towers alone, so to isolate their impact we run a counterfactual where we re-run the model setting $\bar{A}^{Mill} = 0$.

The results are shown in Table 9. In line with the findings of the importance of increasing incomes in the decomposition, we find that without this rise in productivity on commercially redeveloped sites the welfare effects are around one-quarter of the effects in

[44]Table A14 shows that formal floorspace prices fall and informal floorspace prices rise as their supply rises and falls respectively with the development of the mills and the conversion of slums nearby.

Table 8: Welfare Decomposition

	Formal	**Informal**	
		Not Evicted	Evicted
High Skill			
Eviction	-	-	-0.282
Amenities	0.039	0.016	-0.191
Income	0.241	0.179	0.032
House Prices	0.212	0.167	0.009
New Locations (Mills)	0.214	0.168	0.019
Low Skill			
Eviction	-	-	-0.522
Amenities	0.038	0.010	-0.432
Income	0.087	0.092	-0.236
House Prices	0.066	0.082	-0.254
New Locations (Mills)	0.067	0.083	-0.245

Note: Table decomposes welfare effect by adding adjustment of each margin sequentially. Last row of each panel corresponds to full welfare effect. Averages across the same 3 initial states as in Table 7 are reported.

the baseline. The losses to the evicted are slightly larger now that there are no labor market benefits from the productive commercial mills. Figure A5 shows that the increase in floorspace supply lowers rents far from the sites, highlighting the aggregate effects captured by the model but absorbed in the intercept of the regressions.

Table 9: Welfare Effects: Residential Mills Only

	Formal	**Informal**			
		Not Evicted	Evicted Average	Evicted Compensated	Evicted Not Compensated
High Skill	0.040	0.031	-0.070	-0.052	-0.113
Low Skill	0.039	0.026	-0.281	-0.245	-0.369

Notes: Results show % changes in welfare divided by 100.

7.2 Equity vs Efficiency: Increasing Compensation Policies

The previous results suggest that underneath aggregate welfare gains from the mill redevelopments there are residents who were evicted from slums and hurt greatly by the

Table 10: Alternative Compensation Policies

	Observed Compensation	Everyone Compensated	Everyone Compensated 3x
Δ Welfare Informal & Evicted (High-skill)	0.019	0.033	0.137
Δ Welfare Informal & Evicted (Low-skill)	-0.245	-0.208	0.034
Consumer Surplus	0.032	0.032	0.031
Land Surplus	0.085	0.084	0.052
Total Surplus	0.118	0.116	0.083

Notes: The first column shows results from the baseline results with the observed compensation policy (80.56% of slum residents compensated). Second column increases the fraction of residents compensated to 1. Third column compensates all slum dwellers and increases the compensation amount by a factor of 3. First two rows show % changes in welfare on average for evicted slum dwellers (divided by 100). The last three rows show surplus as a fraction of post-period total GDP in the baseline model. Consumer Surplus is measured in equivalent variation terms. Land Surplus is equal to developer profits. GDP is defined as the total output in the city as well as the income of landlords.

policy. Could the city increase its compensation policy towards its evicted slum dwellers, and still leave everyone at least as well off?

Table 10 begins by showing in column 1 the breakdown of total consumer and land owner surplus as a fraction of Mumbai's GDP under the observed compensation policy. The policy led to an increase in total surplus worth about 11.8% of Mumbai's GDP, about one quarter of which accrued to residents and the rest to land owners. The first two rows repeat the original figures for the average welfare change amongst the evicted.

Columns 2 and 3 then increase the compensation developers are required to pay slum residents when they redevelop a slum. The second column requires everyone to be paid (i.e. $\gamma = 1$). The third column requires everyone to be compensated three times the observed amount of compensation. Under this latter option, even the evicted are made (on average) better off. However, we see that aggregate consumer and land owner surplus falls with total surplus falling 4 percentage points. This is because developers face higher costs (decreasing their surplus), and as a result build less housing leading to higher prices than otherwise and lower consumer surplus. This highlights an equity-efficiency trade-off implied by the model for these sorts of redevelopment policies. However, given the acute losses we find for those evicted it seems this is a worthwhile trade-off to make for policymakers with even a mild dislike for inequality within the city. After all, society is still better off in aggregate in absolute terms under this more generous option.

7.3 Robustness and Additional Counterfactuals

Sensitivity to Key Parameters. Since our approach estimates migration costs from voluntary moves, we could be underestimating the moving costs experienced by evicted slum dwellers if the costs involved in evictions are higher. Individuals may lose access to social networks (e.g. Barnhardt et al. (2016)) and lose other forms of capital. Unfortunately, we lack data on those relocated from slums to measure this directly and there is scant quasi-experimental evidence to use as a benchmark. One study from the US find that individuals evicted from their homes in the US lose 8-14% of their income, smaller than our baseline result of a 36.1% drop in real income (Collinson et al. (2022)), suggesting we may not be underestimating costs of eviction too much.

To assess the sensitivity of our results to this possibility, Appendix D.5 increases moving costs for evicted slum dwellers by 10% and 50% under the main counterfactual. Welfare losses for the evicted rise as a result, but the main insight remains that more generous compensation policies for the evicted reduce these losses while also reducing aggregate surplus from the policy (which nonetheless remains positive).

Appendix D.5 shows that increasing the housing supply elasticity η from the low value we calibrate to leaves the results essentially unchanged.

Slum Redevelopment Counterfactual. Since the mill redevelopment policy is specific to Mumbai, Appendix D.5 conducts a more general counterfactual in which formal housing amenities rise by 20% in a location with the median-sized slum. This provides insight into what happens if an area near an existing slum gentrifies. The results have the same qualitative flavor as those from the main mill counterfactual: large parts of the slum are redeveloped, those who are evicted experience large welfare losses, but those losses are attenuated as the compensation policy becomes more generous.

8 Conclusion

This paper provides evidence on spatial spillovers from market-rate high rise development in one of the developing world's largest megacities. We find effects on surrounding neighborhoods consistent with gentrification: formal house prices rise, slums are converted, informal employment falls, the average education of residents rises and population density falls. To assess aggregate and distributional effects, we develop and estimate a dynamic spatial model. We find a large elasticity of amenities to the education-composition of residents, and a spatial decay smaller than developing country estimates. This implies policies that seek to build modern high rises in valuable central areas of developing cities can have transformational impacts on surrounding neighborhoods.

A key policy question as developing city grow is whether to redevelop or upgrade slums and how to compensate residents. The observed policy in Mumbai of giving (some fraction) of evicted slum dwellers a housing unit in a potentially different location is common to other countries. A key question, then, is how large the compensation to slum dwellers should be. We find that even with this policy, the slum redevelopment generated by the increase in amenities caused by the inflow of high-educated residents into the high rises leads to large losses for those who are evicted. A policy that is over three times as generous as the actual one can leave these workers better off on average, at a cost to aggregate surplus as developers respond by building less formal floorspace. This highlights an equity-efficiency trade-off in slum redevelopment policy. However, given the acute losses for the evicted and the fact the impact on total surplus is still positive after such a policy, it may be one that many governments should think is worth making.

References

Adarkar, N. (2005). 600 Acres of Mill Land: For the Public or the Privileged. Technical report.

Adarkar, N. and V. K. Phatak (2005). Recycling Mill Land: Tumultuous Experience of Mumbai. *Economic and Political Weekly 40*(51), 5365–5368.

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.

Ali, D. A., M. Collin, K. Deininger, S. Dercon, J. Sandefur, and A. Zeitlin (2016). Small price incentives increase women's access to land titles in Tanzania. *Journal of Development Economics 123*, 107–122.

Allen, T., C. Arkolakis, and X. Li (2015). Optimal city structure. *Working Paper*.

Almagro, M., E. Chyn, and B. A. Stuart (2023). The welfare and distributional consequences of neighborhood change: Evidence from Chicagoâs public housing demolition.

Almagro, M. and T. Domínguez-Iino (2022). Location sorting and endogenous amenities: Evidence from Amsterdam.

Amirapu, A. and M. Gechter (2020). Indian Labor Regulations and the Cost of Corruption: Evidence from the Firm Size Distribution. *Review of Economics and Statistics 102*(1), 34–48.

Anagol, S., V. Balasubramaniam, T. Ramadorai, and A. Uettwiller (2022). A Bad Bunch: Asset Value Under-Reporting in the Mumbai Real Estate Market. *SSRN Electronic Journal*.

Anand, S. and G. Parussini (2017). Mumbai's slums make way for luxury residential towers. *The Wall Street Journal*. Available at: https://www.wsj.com/articles/mumbais-slums-make-way-for-luxury-residential-towers-1485259208.

Anarock (2019). Share of high-rises at all-time high of 52% in 2019, villas at all-time low of 2%. Available at: https://anarock.com/blogs/high-rises-villas.

Autor, D. H., C. J. Palmer, and P. A. Pathak (2014). Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts. *Journal of Political Economy 122*(3), 661–717.

Baker, J., R. Basu, M. Cropper, S. V. Lall, and A. Takeuchi (2005). Urban Poverty and Transport: The Case of Mumbai. *World Bank Policy Research Working Paper*.

Baker, J. L. and K. McClain (2008). Private Sector Involvement in Slum Upgrading. *Global Urban Development 4*(2).

Balboni, C. (2021). In harm's way? infrastructure investments and the persistence of coastal cities.

Balboni, C., G. Bryan, M. Morten, and B. Siddiqi (2020). Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies. *Preliminary Draft*.

Banerjee, S. (2012, jun). Last stand at Sion Koliwada – where the spectre of demolition looms. *The Hindu*.

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 9*(1), 1–32.

Baum-Snow, N., R. Pavan, and N. Gendron-Carrier (2021). Local Productivity Spillovers. *Working Paper*.

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. *Working Paper*.
Boustan, L., R. A. Margo, M. M. Miller, J. Reeves, and J. Steil (2023). JUE insight: Condominium development does not lead to gentrification. *Journal of Urban Economics 133*, 103524. Special Issue: JUE Insight Shorter Papers.
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 G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, Chapter 21, pp. 1399–1455. Elsevier.
Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the China trade shock. *Econometrica 87*(3), 741–835.
Card, D. (2009). Immigration and inequality. *American Economic Review 99*(2), 1–21.
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.
Collins, W. J. and K. L. Shester (2013). Slum Clearance and Urban Renewal in the United States. *American Economic Journal: Applied Economics 5*(1), 239–273.
Collinson, R., J. E. Humphries, N. Mader, D. Reed, D. I. Tannenbaum, and W. van Dijk (2022). Eviction and Poverty in American Cities. *Quarterly Journal of Economics, forthcoming*.
Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2021). Income growth and the distributional effects of urban spatial sorting.
de Chaisemartin, C. and X. D'Haultfoeuille (2018). Fuzzy Differences-in-Differences. *Review of Economic Studies 85*, 999–1028.
Dhikle, S. and R. Lakhena (2017). GIS Based Slum Information Management System. In *17th Esri India User Conference*.
Diamond, R. (2016). The determinants and welfare implications of US Workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524.
Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development. *Journal of Political Economy 127*(3), 1063–1117.
Diamond, R., T. McQuade, and F. Qian (2019). Who benefits from rent control? The equilibrium consequences of San Francisco's rent control expansion.
D'Monte, D. (Ed.) (2006). *Mills for sale: the way ahead*. Mumbai: Marg Publications.
Duranton, G., S. E. Ghani, A. G. Goswami, and W. Kerr (2019). Why is intermediating houses so difficult? Evidence from iBuyers. *Available at SSRN 2584246*.
Emporis (2018). www.emporis.com.
Feler, L. and J. V. Henderson (2011). Exclusionary policies in urban development: Under-servicing migrant households in Brazilian cities. *Journal of Urban Economics 69*(3), 253–272.
Field, E. (2007). Entitled to Work: Urban Property Rights and Labor Supply in Peru. *Quarterly Journal of Economics 122*(4), 1561–1602.
Field, E. and M. Kremer (2008). Impact Evaluation for Slum Upgrading Interventions. *Mimeo*.
Florida, R. (2019). How housing supply became the most controversial issue in urbanism. *CityLab*. Available at: https://www.bloomberg.com/news/articles/2019-05-23/why-urbanists-are-arguing-about-housing-supply.
Franklin, S. (2020). Enabled to work: The impact of government housing on slum dwellers in South Africa. *Journal of Urban Economics 118*.
Friedman, U. (2014). 70% of India has yet to be built. *The Atlantic*. Available at: https://www.theatlantic.com/international/archive/2014/06/70-percent-of-india-has-yet-to-be-built/373656/.
Geurts, P., D. Ernst, and L. Wehenkel (2006). Extremely randomized trees. *Machine Learning 63*, 3–42.
Glaeser, E. and J. Gyourko (2018). Housing constraints and spatial misallocation. *Journal of Economic Perspectives 32*(1), 3–30.
Grey, D. (2023, Jan). India's biggest slum faces wrecking ball as residents fear change. *Al Jazeera*.
Guerrieri, V., D. Hartley, and E. Hurst (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics 100*, 45–60.
Harari, M. (2020). Cities in bad shape: Urban geometry in India. *American Economic Review 110*(8), 2377–2421.
Harari, M. and M. Wong (2021). Slum upgrading and long-run urban development: Evidence from Indonesia.
Henderson, V., T. Regan, and A. J. Venables (2021). Building the City: From Slums to a Modern Metropo-

lis. *Review of Economic Studies 88*, 1157–1192.
Hornbeck, R. and D. Keniston (2017). Creative destruction: Barriers to urban growth and the great Boston fire of 1872. *American Economic Review 107*(6), 1365–1398.
Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.
Kline, P. and C. R. Walters (2019). On Heckits, LATE, and Numerical Equivalence. *Econometrica 87*(2), 677–696.
Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady 10*, 707–710.
Li, Q. and J. Racine (2004). Cross-Validated Local Linear Nonparametric Regression. *Statistica Sinica 14*, 485–512.
Li, X. (2022). Do new housing units in your backyard raise your rents? *Journal of Economic Geography 22*(6), 1309–1352.
Maharashtra Urban Development Department (1995). Development Control Regulations for Greater Mumbai.
Mast, E. (2023). JUE Insight: The effect of new market-rate housing construction on the low-income housing market. *Journal of Urban Economics 133*.
McKinsey (2010). India's urban awakening: Building inclusive cities, sustaining economic growth. Available at: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/Urban%20awakening%20in%20India/MGI_Indias_urban_awakening_executive_summary.pdf.
Menon, M. R. (2012, jun). Mumbai's Mill Workers : A Small but Significant Victory. *NewsClick*.
Michaels, G., D. Nigmatulina, F. Rauch, T. Regan, N. Baruah, and A. Dahlstrand (2021). Planning ahead for better neighborhoods: Long-run evidence from Tanzania. *Journal of Political Economy 129*(7), 2112–2156.
Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018). Commuting, migration and local employment elasticities. *Journal of Public Economics 108*(12), 3855–3890.
Morris, C. N. (1983). Parametric Empirical Bayes Inference : Theory and Applications. *Journal of the American Statistical Association 78*(381), 47–55.
Nair, A. (2018, sep). SRA houses 40 per cent illegal tenants. *The Asian Age*.
Owens, R., E. Rossi-Hansberg, and P.-D. Sarte (2010). Housing Externalities. *The Journal of Poitical Economy 118*(3), 485–534.
Patel, N. (2011, mar). Battle over Mumbai's slums. *The Guardian*.
Pollmann, M. (2023). Causal Inference for Spatial Treatments. *Working Paper*.
Redding, S. J. and D. M. Sturm (2016). Estimating Neighborhood Effects: Evidence from War-Time Destruction in London. *Working Paper*.
Reif, J. (2022). Strgroup: Stata module to match strings based on their levenshtein edit distance.
Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics 25*(3), 1253–1296.
Santos Silva, J. M. C. and S. Tenreyro (2006). The Log of Gravity. *The Review of Economics and Statistics 88*(November), 641–658.
Srinivasan, P. (2005, apr). Landgrabbers Inc. *Business Today*.
Suresh, J. H., S. Suradkar, A. Bhide, A. C. Deshpande, and S. Singh (2013). Interim Report of People's Commission on Irregularities and Illegalities in SRA Projects. Technical report, People's Commission on Irregularities and Illegalities in SRA Projects, Mumbai.
Takeuchi, A., M. Cropper, and A. Bento (2008). Measuring the welfare effects of slum improvement programs: The case of Mumbai. *Journal of Urban Economics 64*(1), 65–84.
Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: Evidence from Bogotá's TransMilenio.
Turner, M. A., A. Haughwout, and W. van der Klaauw (2014). Land Use Regulation and Welfare. *Econometrica 82*(4), 1341–1403.
United Nations (2018). 68% of the world population projected to live in urban areas by 2050, says UN. Available at: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

Online Appendix for "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills"

A Appendix Tables

Table A1: Mill land use determinants

	Comm	Dev
ln Area	-0.124	0.051
	(0.139)	(0.115)
log Dist CBD	-0.091	0.253
	(0.515)	(0.280)
ln Emp Density	1.001*	0.322
	(0.521)	(0.236)
ln Res Density	0.008	0.500
	(0.931)	(0.457)
ln Res Price	-0.329	0.193
	(0.766)	(0.381)
N	36	60

Notes: Data is from Emporis (2018) and supplemented with data our RAs collected from the web on current building use. Column 1 is at the development-level and includes developments which were active by 2016. Note that a development may contain multiple towers, so the number of towers is higher. Column 2 is at the mill-level, and the outcome is whether the mill was developed by 2016. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A2: Baseline characteristics of treated and untreated areas

	No controls			Region FE	Ward FE
	No Mill mean	With Mill mean	diff	diff	diff
Log Price Res	8.71	8.33	-0.382***	0.011	0.151
Log Population Density	-2.75	-3.01	-0.264	-0.292*	-0.205
Log Employment Density	-4.11	-4.39	-0.280	-0.034	0.067
Ratio of pop/emp	21.62	19.95	-1.670	-3.748	-2.173
Log Distance to CBD	8.29	8.57	0.280	0.024	0.066
Slum share	0.05	0.04	-0.012	-0.020*	-0.023*

Table A3: Slum characteristics

	Ln(Emp Dens)	Ln(Pop Dens)	Ln(Emp/Pop)	Ln(Avg Est Size)	Manuf share
Slum share	0.648***	0.232***	0.424***	-0.133***	0.169***
	(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 A4: 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 A5: Main Results - Exposure Measure

	Res Price	Slums	Informal Emp	Formal Emp	Educ	Population
Mill Exposure	0.031**	-0.235***	-0.262***	0.008	0.168***	-0.045**
	(0.014)	(0.064)	(0.077)	(0.131)	(0.045)	(0.019)
N	1,106	882	496	460	446	450
Mill Exposure (ctrl own share)	0.027*	-0.210***	-0.195**	0.022	0.145***	-0.067***
	(0.015)	(0.072)	(0.089)	(0.149)	(0.047)	(0.025)
N	1,106	882	496	460	446	450
Mill Exposure						
(ctrl own share + closest mill dist)	0.027	-0.191**	-0.219*	0.036	0.218***	-0.046
	(0.019)	(0.075)	(0.132)	(0.222)	(0.065)	(0.028)
N	1,106	882	496	460	446	450

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes own spatial unit fixed effects and same controls as the baseline specification (again omitting lagged dependent variables as controls).

Table A6: Industrial Land Placebo, Multilateral Exposure Measure

	Res Price	Res Price	Slum Share	Slum Share	Inf Emp	Inf Emp	Form Emp	Form Emp	Educ	Educ	Res Pop	Res Pop
Mill Exposure	0.031**	0.061	-0.235***	-0.494***	-0.262***	-0.277	0.008	0.406	0.168***	0.276***	-0.045**	-0.052
	(0.014)	(0.042)	(0.064)	(0.126)	(0.077)	(0.175)	(0.131)	(0.318)	(0.045)	(0.081)	(0.019)	(0.059)
Industrial Land Exposure		-0.033		0.291***		0.016		-0.410		-0.123		0.008
		(0.044)		(0.106)		(0.156)		(0.282)		(0.095)		(0.064)
N	1,106	1,106	882	882	496	496	460	460	446	446	450	450

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes own spatial unit fixed effects and same controls as the baseline specification (again excluding lagged dependent variables as controls).

Table A7: Residential Floorspace Price Effects, Heterogeneity by Active Status

	Res Price
Contains Mill	-0.113
	(0.082)
Contains X Active	0.343***
	(0.105)
1m-500m	0.059
	(0.081)
1m-500m X Active	0.038
	(0.084)
500m-1000m	0.052
	(0.055)
500m-1000m X Active	-0.009
	(0.070)
N	1,050

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Results from a regression of formal residential floorspace price on indicators for distance-to-nearest mill site bands, interacted with indicators for subzones containing an active new development. Regression includes own spatial unit fixed effects and same controls as the baseline specification. Top 1% tails of outcomes are trimmed. Robust standard errors in parentheses.

Table A8: No Correlated Changes in Roads

	dlog(Main Roads)
Contains Mill	-0.088
	(0.116)
1m-500m	0.111
	(0.124)
500m-1000m	-0.115
	(0.125)
N	352

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes IV Unit fixed effects and allows for differential time trends by region, log distance to CBD and log IV Unit area.

Table A9: Adding In Controls

	(1)	(2)	(3)
Res Price			
Contains Mill	0.221***	0.150**	0.130**
	(0.060)	(0.058)	(0.059)
1m-500m	0.115**	0.085*	0.089*
	(0.047)	(0.048)	(0.049)
N	1,070	1,070	1,050
Slums			
Contains Mill	-0.241**	-0.233**	-0.226**
	(0.108)	(0.111)	(0.102)
1m-500m	-0.218**	-0.218**	-0.270***
	(0.087)	(0.089)	(0.097)
N	954	954	894
Emp Inform			
Contains Mill	-0.564***	-0.723***	-0.413**
	(0.215)	(0.191)	(0.185)
1m-500m	-0.530***	-0.767***	-0.400**
	(0.203)	(0.200)	(0.200)
N	486	486	486
Emp Form			
Contains Mill	0.146	0.134	0.196
	(0.308)	(0.320)	(0.329)
1m-500m	0.153	0.128	0.202
	(0.263)	(0.276)	(0.285)
N	438	438	438
Education			
Contains Mill	0.287**	0.191	0.262**
	(0.138)	(0.129)	(0.126)
<500m	-0.156	-0.201	-0.197
	(0.179)	(0.181)	(0.175)
N	448	448	446
Population			
Contains Mill	-0.074	-0.119**	-0.150**
	(0.073)	(0.058)	(0.059)
1m-500m	-0.025	-0.120*	-0.117*
	(0.075)	(0.070)	(0.061)
N	434	434	434
FE	Region	Region	Region
logDistCBD	X	X	X
logArea		X	X
logRoadDens		X	X
Slum Share Init		X	X
LogEmpDens Init			X
LogPopDens Init			X
LogResPrice Init			X
500m-1000m X Post	X	X	X

Table A10: Finer Fixed Effects

	(1)	(2)
Res Price		
Contains Mill	0.221***	0.139*
	(0.060)	(0.080)
1m-500m	0.115**	0.047
	(0.047)	(0.084)
N	1,070	1,070
Slums		
Contains Mill	-0.241**	-0.179
	(0.108)	(0.116)
1m-500m	-0.218**	-0.173
	(0.087)	(0.112)
N	954	954
Emp Inform		
Contains Mill	-0.564***	-0.379
	(0.215)	(0.303)
1m-500m	-0.530***	-0.436
	(0.203)	(0.287)
N	486	486
Emp Form		
Contains Mill	0.146	0.383
	(0.308)	(0.452)
1m-500m	0.153	0.322
	(0.263)	(0.365)
N	438	438
Education		
Contains Mill	0.287**	0.381*
	(0.138)	(0.202)
1m-500m	-0.156	-0.046
	(0.179)	(0.188)
N	448	448
Population		
Contains Mill	-0.074	-0.017
	(0.073)	(0.099)
1m-500m	-0.025	-0.022
	(0.075)	(0.095)
N	434	434
FE	Region	Ward
logDistCBD	X	X
500m-1000m X Post	X	X

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression is same diff-in-diff specification as the baseline, with log distance to CBD as control along with spatial unit fixed effects.

Table A11: Alternative Distance Bands

	Res Price	Population	Slum Share	Inf Emp	Form Emp	Ed Index
Contains Mill	0.122**	-0.131**	-0.231**	-0.504***	0.121	0.297**
	(0.059)	(0.056)	(0.101)	(0.176)	(0.308)	(0.119)
1m-250m	0.084	-0.085	-0.365***	-0.592***	0.288	-0.147
	(0.064)	(0.074)	(0.123)	(0.212)	(0.322)	(0.111)
250m-500m	0.078	-0.129**	-0.143	-0.363	-0.068	-0.199
	(0.063)	(0.054)	(0.111)	(0.284)	(0.354)	(0.475)
N	1,050	434	894	486	438	446

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes own spatial unit fixed effects and same controls as baseline specification (also excluding lagged dependent variables as controls).

Table A12: Alternative Exposure Decays

	Res Price	Slums	Informal Emp	Formal Emp	Educ	Population
Mill Exposure, Decay=1	0.032**	-0.237***	-0.278***	0.037	0.152***	-0.044**
	(0.014)	(0.064)	(0.077)	(0.134)	(0.048)	(0.019)
N	1,116	894	500	464	450	454
Mill Exposure, Decay=0.7	0.020**	-0.159***	-0.020***	0.002	0.096***	-0.029**
	(0.009)	(0.045)	(0.006)	(0.010)	(0.033)	(0.013)
N	1,116	894	500	464	450	454
Mill Exposure, Decay=1.5	0.061**	-0.428***	-0.470***	0.092	0.290***	-0.080**
	(0.029)	(0.113)	(0.130)	(0.230)	(0.081)	(0.037)
N	1,116	894	500	464	450	454

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. Regression includes own spatial unit fixed effects and same controls as baseline specification (also excluding lagged dependent variables as controls)..

Table A13: Parameters Calibrated Before Solving Model

Parameter	Description	Value	Data Source
Panel A: Externally Calibrated			
σ_L	EoS between H&L skill labor	1/0.7	Card (2009)
σ_D	Elasticity of demand	4	Within range reported in Feenstra et. al. (2014)
δ_C	Relationship between commute times and commute costs	0.013	Ahlfeldt et al. (2015) and Tsivanidis (2022)
ν	Residential Choice Elasticity	3	Tsivanidis (2022)
η	Housing Supply Elasticity	0.60	Saiz (2010), matching the most inelastic city (Miami).
Panel B: Calibrated to Indian Moments			
α	Labor cost share for firms	0.596	Labor's share of national income in India in 2000 (Penn World Tables)
β	Consumer expenditure share on housing	0.25	Avg. expenditure share on housing in Mumbai from NSS 55 & 61.
$\tilde{\alpha}_{gs}$	Share of wage bill from F & I firms going to H & L-skill workers. FH, FL, IH, IL in order.	0.44 0.56 0.17 0.83	NSS Sch 10 employment by education and firm size.
ι_{FF}	Productivity of formal relative to informal floorspace for formal firms	2.24	1998 Economic Census of India. See Appendix D.3 for details.
θ_g	Commuting Elasticity. θ_H, θ_L in order	2.69 3.60	Mumbai Travel Survey. See Appendix D.3 for details.
γ	Fraction of slum dwellers, eligible for compensation	0.70	Interpolated value in 2005 using NSS 58 & 69.
E_g	Wage premium of high-educated workers	0.357	NSS Schedule 1.2 round 49. Solved as part of wage inversion algorithm

Table A14: Impacts on Floorspace Prices

	% Change
Formal Floorspace Prices	-0.496
Informal Floorspace Prices	0.332
Land Owner Surplus (Total)	6.505
Land Owner Surplus (Excluding Mills)	-2.016

Notes: First two rows show changes in population-weighted average floorspace prices, using initial population as weights. Second two rows show changes in land owner surplus, which is given by land values (equal to developer profits).

Table A15: Slum Redevelopment Counterfactual

	Baseline	All Compensated	All Compensated 1.5x
Formal, High Skill	0.012	0.012	0.012
Formal, Low Skill	0.003	0.003	0.003
Informal, High Skill & Not Evicted	-0.029	-0.029	-0.029
Informal, Low Skill & Not Evicted	-0.033	-0.033	-0.033
Informal, High Skill & Evicted	-3.618	-0.089	5.516
Informal, Low Skill & Evicted	-16.381	-9.572	0.811
Pr(Slum Redev) Treated Location	93.424	93.132	92.656

Notes: Results show % changes in welfare and the probability the slum in the treated location is redeveloped in the counterfactual where formal amenities are increased by 20% in the same location as the median-sized slum. Third column shows results where all slum residents are compensated, and by 150% the value of the observed compensation policy.

Table A16: Robustness to Key Parameters

	Baseline	35% Larger η	10% Larger F	10% Larger F, 3XComp	50% Larger F	50% Larger F, 3XComp
Informal, Evicted: High Skill	-0.010	-0.012	-0.100	0.014	-0.410	-0.296
Informal, Evicted: Low Skill	-0.274	-0.277	-0.399	-0.089	-0.709	-0.400
Consumer Surplus	0.023	0.023	0.024	0.022	0.025	0.023

Notes: Results show welfare impacts across alternative parameter values. Column 1 shows the baseline result. Column 2 shows the results if the housing supply elasticity is increased by 35%, moving from the Saiz (2010) estimates for Miami to Chicago. Columns 3-6 show results from models which increase the fixed cost of moving by 10% and 50%, with the observed compensation and the counterfactual from the main draft where everyone is compensated 3 times as much.

B Appendix Figures

Figure A1: Mill locations

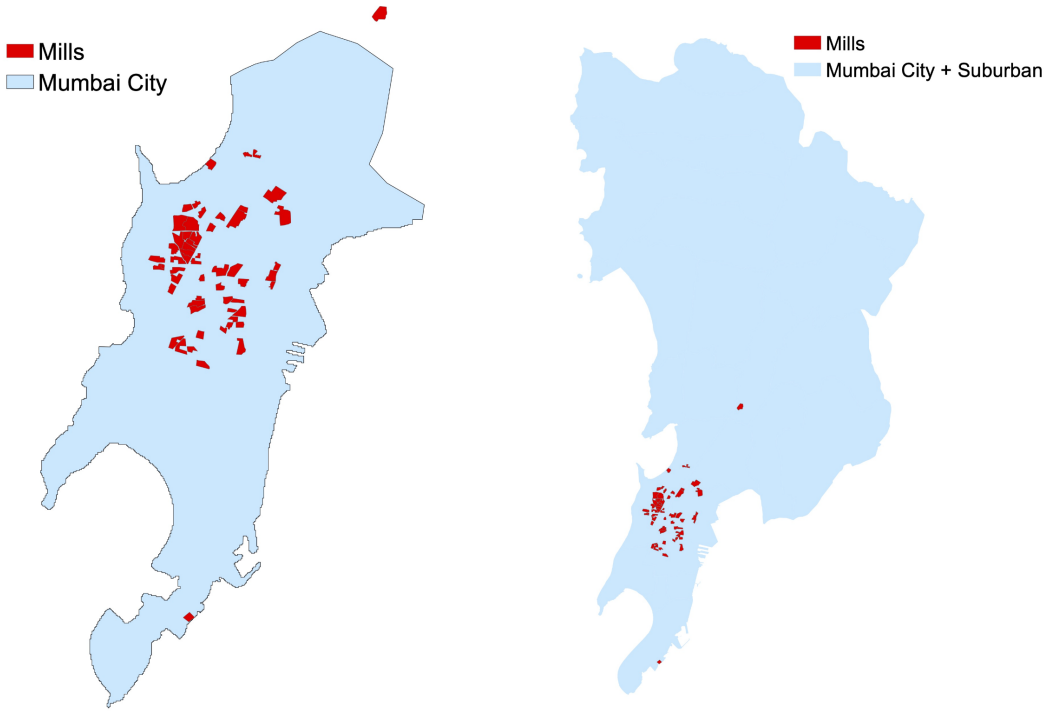


Figure A2: Distribution of SRA housing

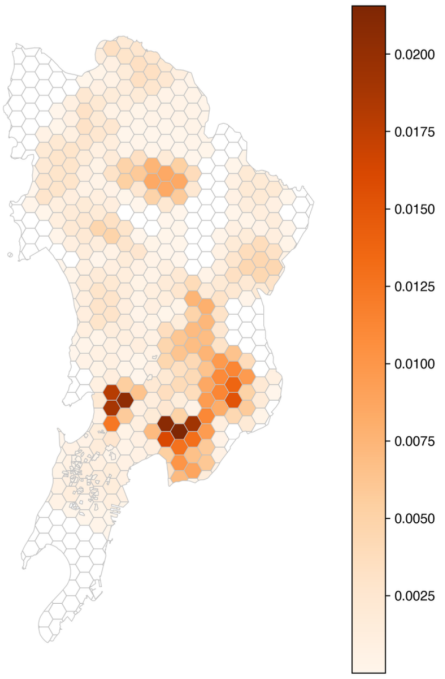
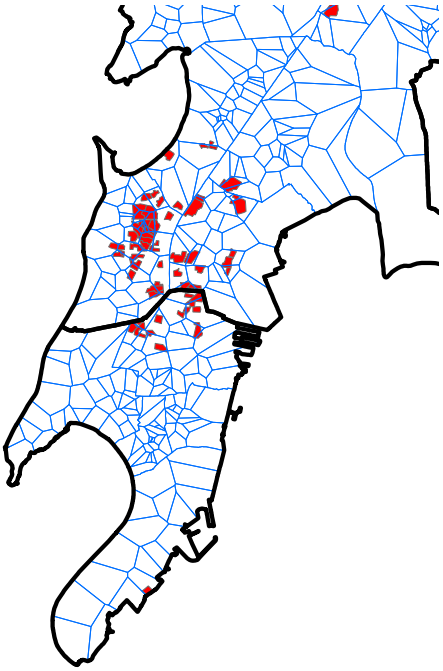


Figure A3: Regions, IV Units, and Mills



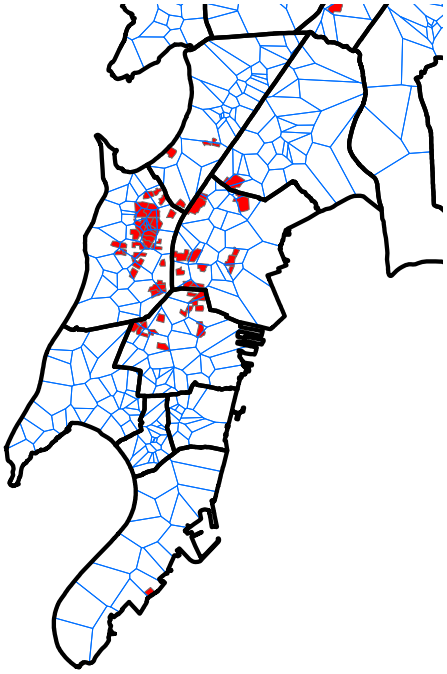
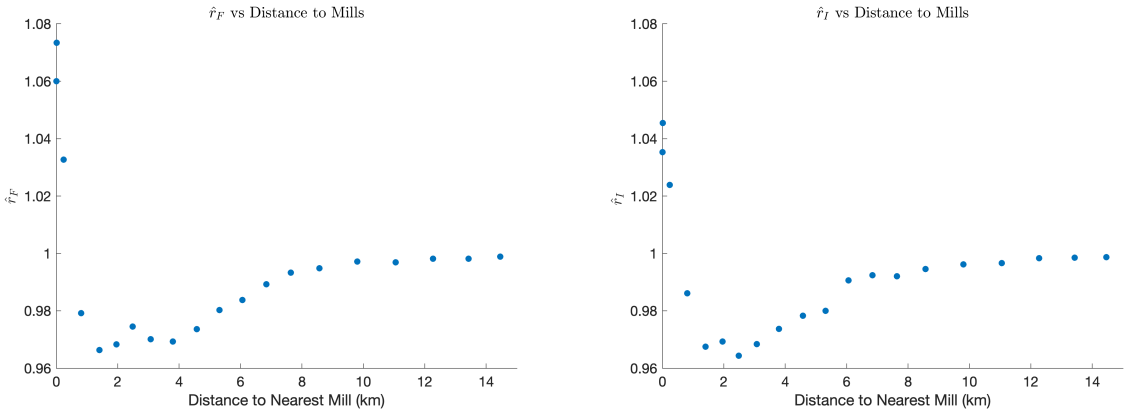


Figure A4: Wards, IV Units, and Mills

Figure A5: Formal and Informal Floorspace Price Gradients: Residential Mills Only Counterfactual



C Additional Empirical Results

C.1 IV Estimation of Mill Redevelopment Effects

In this appendix, we regress outcomes on indicators for bands of distance to an active new development on a former mill site. We present both OLS and instrumental variables regressions, where in the latter we instrument for the regressors of interest using the same distance to mill site band indicators as in Equation 3 in the main text. Outcomes and controls are the same as in the main text, with the exception of slum share where we run linear instrumental variables regressions with 0.01 + the log of the slum share of an IV Unit as the outcome as opposed to the PPMLE regressions on slum share in the main text.[45] Table A17 below shows the results. As expected, the IV coefficients scale up the reduced form coefficients from Table 1. The OLS results do not differ dramatically.

C.2 Anticipation Effects on Floorspace Prices

Our main reduced form results assumes the effects of mill sites are due to actual economic activity on the redeveloped sites rather than, for instance, anticipation of the activity. Table A7 displays results from the following regression for floorspace prices, which we would expect to reflect capitalization of any anticipation effects,

$$\begin{aligned} Y_{it} =& \beta_1 \, \mathbb{I}\{\text{ContainsMill}\}_i \times \text{Post}_t + \beta_2 \, \mathbb{I}\{\text{ContainsDevelopedMill}\}_i \times \text{Post}_t + \beta_3 \, \mathbb{I}\{\text{1m-500m}\}_i \times \text{Post}_t \\ &+ \beta_4 \, \mathbb{I}\{\text{1m-500mDeveloped}\}_i \times \text{Post}_t + \beta_5 \, \mathbb{I}\{\text{501m-1km}\}_i \times \text{Post}_t \\ &+ \beta_6 \mathbb{I}\{\text{501m-1kmDeveloped}\}_i \times \text{Post}_t + \alpha_i + \gamma_{r(i)t} + \delta_t' X_i + \epsilon_{it}, \end{aligned}$$

which adds an interaction for having a redeveloped mill site where the new development is active to the main specification in Equation 3. β_2, β_4, and β_6 illustrate the differential effect of having a developed mill site compared to the effect of an undeveloped mill site. The results are noisy since spatial units with developed mill sites will often have undeveloped mill sites as well, but the coefficient pattern is consistent with results primarily being associated with economic activity. Note that developed status is endogenous so these results should be thought of as descriptive. See also Appendix C.1 where we estimate instrumental variable regressions for the effect of an active redeveloped mill site.

[45] These are fuzzy DID regressions, as described in de Chaisemartin and D'Haultfoeuille (2018). Because no observations in the control group are ever treated, de Chaisemartin's and D'Haultfoeuille's WALD-DID and WALD-TC estimands are the same and are what we estimate.

Table A17: IV Estimates of Mill Redevelopment Effects

	House Price	House Price	Slum Share	Slum Share	Inform Emp	Inform Emp
Contains Mill	0.230***	0.152*	-0.333**	-0.316**	-0.388*	-0.358
	(0.071)	(0.082)	(0.132)	(0.155)	(0.210)	(0.291)
1m-500m	0.060	0.094	-0.182**	-0.155	-0.548***	-0.668**
	(0.049)	(0.058)	(0.081)	(0.118)	(0.186)	(0.281)
500m-1000m	0.030	0.080	0.028	0.051	0.020	0.631
	(0.055)	(0.084)	(0.108)	(0.161)	(0.250)	(0.560)
N	1,050	1,050	1,166	1,166	486	486
	Form Emp	Form Emp	Educ	Educ	Pop	Pop
Contains Mill	0.265	0.222	0.407***	0.442***	-0.060	-0.191**
	(0.383)	(0.493)	(0.140)	(0.162)	(0.060)	(0.096)
1m-500m	0.121	0.205	-0.119	-0.170	-0.158**	-0.110
	(0.291)	(0.380)	(0.119)	(0.207)	(0.063)	(0.085)
500m-1000m	-0.203	0.565	-0.209	-0.239	0.061	-0.195
	(0.357)	(0.750)	(0.215)	(0.260)	(0.069)	(0.187)
N	438	438	446	446	434	434
Specification	OLS	IV	OLS	IV	OLS	IV

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. OLS and instrumental variables regressions of outcomes on bands of distance to an active new development on a former mill site. In IV specifications we instrument for the bands of distance to an active new development indicators with indicators for bands of distance to any mill site. Robust standard errors in parentheses. Regression includes own spatial unit fixed effects and allows for differential time trends by log distance to CBD, log spatial unit area, log initial period road density, initial slum share, log of initial employment density, log of initial population density, and initial education index. We omit lagged outcome values from the controls to guard against endogeneity generated by serial correlation in errors. 1% tails of outcome variables are dropped.

C.3 Agglomeration Spillover Analysis by Industry

In this appendix, we check for evidence of agglomeration spillovers by industry. Since mills redeveloped to commercial use tend to be occupied by service sector firms, if there to be an employment response to agglomeration spillovers we would expect to see it in other service sector firms. Table A18 reproduces the analysis from column 3 of Table 4, separately for firms falling into four industries. The first column reports effects for firms in high-skill service industries, where we see the largest increase in employment on mill sites themselves, as predicted. However, we see no increase in formal employment in high-skill services nearby (in the 1 - 500km away group of IV Units) or slightly farther away (in the 501 - 1km away group) from either commercial or residential sites. In manufacturing, we see no significant increase in formal employment in IV Units containing

mill sites or near them, as we would expect.

Restaurants and retail also have a large presence in commercial mills so we see an increase in formal employment in the industry in IV Units containing commercially-redeveloped mills. We see a decrease in restaurant and retail employment in IV Units containing residentially-redeveloped mills. In IV Units near mill sites, we see no significant increase in formal restaurant and retail employment regardless of mill redevelopment type. Finally, for firms in all other industries we do not see significant changes in employment either on mill sites or nearby.

Table A18: Formal Employment Effects By Mill Use and Industry

	HSS	Manuf	Rest&Retail	Other
Contains Mill Comm	2.499***	-0.092	1.398***	0.836
	(0.567)	(0.822)	(0.464)	(0.591)
Contains Mill Res	-0.407	-0.733	-0.658**	0.447
	(0.382)	(0.570)	(0.328)	(0.442)
<500m Comm	0.056	-1.183	-0.109	-0.531
	(0.451)	(0.800)	(0.416)	(0.570)
<500m Res	0.158	0.308	0.026	0.256
	(0.347)	(0.405)	(0.284)	(0.380)
500m-1000m Comm	-0.022	-1.454***	0.059	-0.066
	(0.487)	(0.556)	(0.398)	(0.433)
500m-1000m Res	-0.593	0.773	-0.430*	0.242
	(0.572)	(0.576)	(0.256)	(0.543)
N	312	286	372	404

Notes: Outcome is log employment growth in high-skilled services, manufacturing, restaurants and retail, and other industries. Contains Mill Comm and Contains Mill Res are dummies for whether a location contains a mill that was redeveloped commercial or residential by 2016 respectively, with other dummies defined analagously. Regression includes own spatial unit fixed effects and allows for differential time trends by log distance to CBD, log spatial unit area, log initial period road density, initial slum share, log of initial employment density, log of initial population density, and initial education index. We omit lagged outcome values from the controls to guard against endogeneity generated by serial correlation in errors. 1% tails of outcome variables are dropped. Robust SEs reported in parentheses.

In Table A19, we undertake the same analysis for informal employment by industry. We do not see any positive informal employment responses to mill redevelopment regardless of redevelopment type or industry. Instead we see decreases in informal employment, particularly in IV Units containing or close to residentially-redeveloped mills.

Table A19: Informal Employment Effects By Mill Use and Industry

	HSS	Manuf	Rest&Retail	Other
Contains Mill Comm	0.201	0.241	0.261	0.339
	(0.489)	(0.405)	(0.320)	(0.335)
Contains Mill Res	-1.000***	-0.543	-0.544***	-0.383*
	(0.313)	(0.352)	(0.207)	(0.208)
<500m Comm	-0.075	-0.829*	-0.388	-0.185
	(0.397)	(0.425)	(0.276)	(0.267)
<500m Res	-0.885***	-0.297	-0.603***	-0.504***
	(0.255)	(0.264)	(0.196)	(0.190)
500m-1000m Comm	-0.033	-0.135	-0.183	0.092
	(0.338)	(0.376)	(0.271)	(0.261)
500m-1000m Res	-0.542	0.368	0.072	-0.043
	(0.504)	(0.318)	(0.265)	(0.248)
N	450	488	498	500

Notes: Outcome is log employment growth in high-skilled services, manufacturing, restaurants and retail, and other industries. Contains Mill Comm and Contains Mill Res are dummies for whether a location contains a mill that was redeveloped commercial or residential by 2016 respectively, with other dummies defined analagously. Regression includes own spatial unit fixed effects and allows for differential time trends by log distance to CBD, log spatial unit area, log initial period road density, initial slum share, log of initial employment density, log of initial population density, and initial education index. We omit lagged outcome values from the controls to guard against endogeneity generated by serial correlation in errors. 1% tails of outcome variables are dropped. Robust SEs reported in parentheses.

D Model Appendix

D.1 Data Preparation

Our datasets come in different geographies which we need to transform to one unified set of locations for the model. We do this by first creating a grid of 600 hex cells that cover Mumbai. We then intersect these with the mill sites (which constitute a distinct set of locations in the model). The set of locations $\mathcal{L}_t$ in the model is therefore the complement of these hex cells and the mill sites.

We then project our data down to the hex cell level. For population and employment, we intersect each IV unit and election ward with the hex cell layer. We then distribute each unit's population and employment down to the hex cells it intersects with based on the relative area of these intersections. For floorspace prices, we follow a similar process but instead assign each hex cell an area-weighted average of the floorspace prices of the subzones it intersects with. Slum area is measured directly by intersecting the slum

shapefile with the hex cells, and land areas are similarly computed directly in ArcGIS.

Given the large number of 600^{600} pairwise travel times in the model, we compute t_{ij} by sampling a random 1000 pairs and calculate the travel time between them in google maps. We then project travel times on distance, and use the fitted model to predict travel times for each pair in the model.

D.2 Recovering Model Unobservables

Suppose we are given a vector $(\alpha, \beta, \theta_g, \delta_C, \kappa, \eta, E_g, \sigma_L, \tilde{\alpha}_{gs}, \delta_U, \mu_U, \rho_g, \delta_g^M, F_g, \iota_{FF})$ of structural parameters, along with data from the pre-policy period $\{L_{Rikg}, L_{Cis}, \lambda_{ik}, r_{iF}, t_{ij}, T_i\}$. This sub-section shows how the vector $(w_{ig}, r_{Ii}, b_{ikg}, u_{ikg}, \bar{u}_{ikg}, X_{js}, A_{jsk}, H_{ik}, \xi_{ik}, \vartheta_{ik})$ of model unobservables can be recovered.

Steady State Equations. We assume a steady state in the pre-period where $x_{it} = x_{it+1} = x_i$ for any variable x_i. In such a steady state, the equations from the static equilibrium are simply

$$
\begin{aligned}
L_{Cjg} &= \alpha w_{jg}^{-\sigma_L} \sum_{sk} \tilde{\alpha}_{gs} W_{js}^{\sigma_L - 1} X_{jsk} \\
r_{Rik} &= (1-\beta) \frac{\sum_g \bar{y}_{ig} L_{Rikg}}{\vartheta_i H_{ik}} \\
\vartheta_{ik} &= \frac{(1-\beta) Y_{ik}}{(1-\beta) Y_{ik} + (1-\alpha) X_{ik}} \\
L_{Cjg} &= \sum_{ik} \frac{(w_{jg}/d_{ij})^{\theta_g}}{\sum_s (w_{sg}/d_{is})^{\theta_g}} L_{Rikg} \\
u_{ikg} &= \bar{u}_{ikg} \left[\sum_j \exp\left(-\delta_U t_{ij}\right) \frac{L_{RjH}}{L_{Rj}} \right]^{\mu_U}
\end{aligned}
$$

The equations from the dynamic equilibrium collapse to

$$
\begin{aligned}
\lambda_{Fi} &= \lambda_{Ii} \exp\left(-\kappa \left(\frac{F_i + r_{Ii}}{(1-\eta)\tilde{\eta} r_{Fi}^{\frac{1}{1-\eta}}} \right) \right) + \lambda_{Fi} \exp\left(-\kappa \left(\frac{r_{Ii}}{(1-\eta)\tilde{\eta} r_{Fi}^{\frac{1}{1-\eta}}} \right) \right). \\
L_{Rrsg} &= \sum_{ik} \left(\frac{b_{rsg}}{\mu_{ikrsg} \Omega_{ikg}} \right)^{\nu} L_{Rikg} \\
H_{iF} &= \lambda_{iF} \xi_{iF} \tilde{T}_i \tilde{\eta} r_{iF}^{\frac{\eta}{1-\eta}} \\
H_{iI} &= \lambda_{iI} \xi_{iI} \tilde{T}_i
\end{aligned}
$$

since $\psi_{ik} = 0 \; \forall ik$ when $\lambda_{Iit} = \lambda_{Iit+1}$.

Note that we have replaced land supply $T_i = \xi_{ik}\tilde{T}_i$ for use k, where $\tilde{T}_i$ is land area as observed in the data and ξ_{ik} is a residual that captures the quality of land in location i for use k. We need to introduce the residuals ξ_{ik} because there are two equations for $H_i k$ that simultaneously need to hold, namely $H_{ik} = (1-\beta)\frac{\sum_g \bar{y}_{ikg} L_{Rikg}}{\vartheta_{ik} r_{ik}}$ and $H_{ik} = \lambda_{ik}\xi_{ik}\tilde{T}_i\tilde{\eta}_k r_{ik}^{\eta_k}$. Without the additional residual, there is no guarantee the right hand side of the first will equal the right hand side of the second, but adding this additional residual capturing the quality of land for different types of development in each location allows the model to exactly match the data.

Model Inversion. We now show how this system of equations can be inverted to solve for the model's unobservables in the initial equilibrium.

First, we solve for wages by equating supply and demand for labor

$$\sum_s \frac{\tilde{\alpha}_{gs}(w_{jg}/E_g)^{-\sigma_L}}{\sum_h \tilde{\alpha}_{hs}(w_{jh}/E_h)^{-\sigma_L}} L_{Cjs} = E_g \sum_{ik} \frac{(w_{jg}/d_{ij})^{\theta_g}}{\sum_j (w_{jg}/d_{ij})^{\theta_g}} L_{Rikg}$$

Given data L_{Rikg}, L_{Cjs} and parameters $\tilde{\alpha}_{gs}, d_{ij}, \theta_g, \sigma_L, E_g$ this is a system of $G \times I$ equations in as many unknowns. The solution is only to scale, as the system is homogeneous of degree zero in wages.

Next, we use the steady-state relationship for residential population flows

$$L_{Rrsg} = \sum_{ik} \frac{(b_{rsg}/\mu_{ikrsg})^{\nu}}{\sum_{lm}(b_{lmg}/\mu_{iklmg})^{\nu}} L_{Rikg}$$

to solve for a unique vector of (to-scale) mean utilities b_{rsg} given data on residential populations L_{Rikg}, and parameters ν, μ_{ikrsg} where migration costs, in turn, depend on data on travel times $t_i j$ and parameters F_g, δ_g^M.[46]

We then recover r_{iI}, ξ_{ik}, H_{ik} jointly in the following procedure. Since we observe formal sector floorspace prices, the steady state equation for land use shares (also observed) provides a system of equations in r_{iI} given F_i. Since $F_i = y^{comp}\gamma\frac{L_{RiIt}}{\lambda_{Iit}T_i}$ and $T_i = \xi_{ik}\tilde{T}_i$ where $\tilde{T}_i$ is data on observed land areas, we solve for r_{iI} and ξ_{ik} jointly through the following algorithm

- Guess ξ_{ik0}
- Given ξ_{ikt}

[46] Proof of uniqueness comes from Ahlfeldt et al. (2015) since this system has the same structure as that of the system for wages in that paper.

- Compute $F_i = y^{comp}\gamma\frac{L_{RiIt}}{\lambda_{Iit}T_i}$.
- Solve for r_{Ii} in

$$\lambda_{Fi} = \lambda_{Ii}\exp\left(-\kappa\left(\frac{F_i + r_{Ii}}{(1-\eta)\tilde{\eta}r_{Fi}^{\frac{1}{1-\eta}}}\right)\right) + \lambda_{Fi}\exp\left(-\kappa\left(\frac{r_{Ii}}{(1-\eta)\tilde{\eta}r_{Fi}^{\frac{1}{1-\eta}}}\right)\right).$$

- Obtain floorspace areas from

$$\tilde{H}_{iF} = \lambda_{iF}E\left[\epsilon_F|\text{Choose } F\right]T_i\tilde{\eta}r_{iF}^{\frac{\eta}{1-\eta}}$$
$$\tilde{H}_{iI} = \lambda_{iI}T_i$$

  and compute the update $\tilde{\xi}_{ik} = (1-\beta)\frac{\sum_g \bar{y}_{ig}L_{Rikg}}{\vartheta_i\tilde{H}_{ik}r_{Rik}}$
- Update $\xi_{ikt+1} = \chi\xi_{ikt} + (1-\chi)\xi_{ikt+1}$ for update weight $\chi \in (0,1)$.

- Continue until $||\xi_{ikt} - \xi_{ikt+1}||_\infty < \epsilon^{tol}$.

Note that the average productivity shock on redeveloped slums $E\left[\epsilon_F|\text{Choose } F\right]$ is given in Section SM.3.3.1.

With (to-scale) values for $w_{jg}, b_{rsg}, r_{ik}, \xi_{ik}$, we obtain amenities u_{igk} using that $b_{rsg} = u_{igk}^{\rho_g}\tilde{y}_{ig}r_{Rik}^{\beta-1}$ and $\tilde{y}_{ig}$ is known given the values for wages. Given values for δ_U, μ_U, exogenous amenities $\bar{u}_{ikg}$ are recovered using values for L_{Rikg}, u_{ikg} and the definition of amenities in (8).

Lastly, sales X_{jsk} are recovered from $X_{js} = \frac{L_{Cjs}}{\alpha W_{js}^{\sigma_L - 1}\sum_g (w_{jg}/T_g)^{-\sigma_L}\tilde{\alpha}_{gs}}$ using that from labor demand a fraction α of costs accrue to labor and revenues equal costs. While sales in hand, we finally find that

$$A_{js} = \left(\frac{X_{js}}{W_{js}^{\alpha(1-\sigma_D)}\beta\sum_{ik'}Y_{ik'}\sum_k\left(r_{Cjk}^{1-\alpha}/\iota_{sk}\right)^{1-\sigma_D}}\right)^{\sigma_D - 1}$$

by inverting the expression for goods demand from households, and using that $A_{jsk} = A_{js}\iota_{sk}$.

D.3 Calibration

ι_{FF} **.** To derive an expression for ι_{FF} in terms of data and known parameters, begin by writing employment in j, s, k as

$$\begin{aligned} L_{Cjsk} &= W_{js}^{\sigma_L - 1} X_{jsk} \alpha \sum_g (w_{jg}/T_g)^{-\sigma_L} \tilde{\alpha}_{gs} \\ &= \beta \left(W_{js}^{\alpha} r_{Cjk}^{1-\alpha} \right)^{1-\sigma_D} A_{jsk}^{\sigma_D - 1} E \times W_{js}^{\sigma_L - 1} \alpha \sum_g (w_{jg}/T_g)^{-\sigma_L} \tilde{\alpha}_{gs} \end{aligned}$$

Relative employment of formal firms in location j in formal floorspace is then

$$L_{CjFF}/L_{CjFI} = \left(r_{CjF}/r_{CjI}\right)^{-(1-\alpha)(\sigma_D - 1)} \iota_{FF}^{\sigma_D - 1}$$

Taking the relative difference of this object between formal and informal firms and setting $\iota_{IF} = 1$ then yields

$$\begin{aligned} \iota_{FF}^{\sigma_D - 1} &= \left(\frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}} \right) \\ \Rightarrow (\sigma_D - 1) \ln \iota_{FF} &= \ln \frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}} \\ \Rightarrow (\sigma_D - 1) \ln \iota_{FF} &= E \left[\ln \frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}} \right] \\ \Rightarrow \iota_{FF} &= \exp \left(E \left[\ln \frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}} \right]^{\frac{1}{\sigma_D - 1}} \right) \end{aligned}$$

where the third line follows from taking expectations of both sides.

The object $E\left[ln\frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}}\right]$ we estimate approximately from the following regression

$$\ln L_{Cjs} = \alpha + \beta_1 \text{SlumShare}_j + \beta_2 \text{Formal}_s + \beta_3 \text{SlumShare}_j \times \text{Formal}_s + \epsilon_{js}.$$

We have that

$$\begin{aligned} E\left[\ln L_{CjF} | \text{SlumShare}_j = 0\right] - E\left[\ln L_{CjF} | \text{SlumShare}_j = 1\right] &= -(\beta_2 + \beta_3) \\ E\left[\ln L_{CjI} | \text{SlumShare}_j = 0\right] - E\left[\ln L_{CjI} | \text{SlumShare}_j = 1\right] &= -\beta_2. \end{aligned}$$

Define $L_{CjFF} \equiv E\left[\ln L_{CjF} | \text{SlumShare}_j = 0\right]$.

$$\begin{aligned} E\left[\ln \frac{L_{CjFF}/L_{CjFI}}{L_{CjIF}/L_{CjII}}\right] \approx & \left(E\left[\ln L_{CjF} | \text{SlumShare}_j = 0\right] - E\left[\ln L_{CjF} | \text{SlumShare}_j = 1\right]\right) \\ & - \left(E\left[\ln L_{CjI} | \text{SlumShare}_j = 0\right] - E\left[\ln L_{CjI} | \text{SlumShare}_j = 1\right]\right) \\ = & -\beta_3. \end{aligned}$$

So then

$$\iota_{FF} \approx \exp\left(-\beta_3^{\frac{1}{\sigma_D - 1}}\right).$$

We find that $\beta_3 = -0.524$ (0.309), so that with $\sigma_D = 4$ we have $\iota_{FF} \approx 2.2394$.

The key intuition behind the approach is that we see formal employment dropping faster in locations with more slums than we do for informal employment, so the calibration infers that formal land must be relatively more productive for formal plants.

θ_g. Summing across housing types and taking logs of the conditional commute shares in Section 5.1 delivers the gravity equation

$$\ln(\pi_{j|ig}) = \gamma_{jg} + \delta_{ig} - \theta_g \delta_C t_{ij} + \varepsilon_{ij}$$

where ε_{ij} is measurement error. We estimate this via PPML to account for zeros in the commuting data using the 2019 Mumbai Travel Survey and find coefficients on time for -0.0468 (0.004) and -0.0349 (0.002) for low- and high-educated workers respectively across 1,152 origin-destination-education cell observations. Dividing by δ_C gives the results in the table.

D.4 Estimation

Step 1: Migration Parameters.The moments to identify the migration parameters are

$$\begin{aligned} \Pr(\text{Move}|ikg) &= 1 - \left(b_{ikg}/\Omega_{ikg}\right)^{\nu} \\ \frac{\Pr\left(\text{Same Nbhd}|\text{Move}, ig\right)}{\Pr\left(\text{Diff Nbhd}|\text{Move}, ig\right)} &= \frac{\sum_{r \in N(i),k} \sum_{h \in \{I,F\}} \frac{\left(b_{rhgt}/\exp\left(F_g + \delta_g^M t_{ir}\right)\right)^{\nu}}{\Omega_{ikgt}^{\nu}}}{1 - \sum_{r \in N(i),k} \sum_{h \in \{I,F\}} \frac{\left(b_{rhgt}/\exp\left(F_g + \delta_g^M t_{ir}\right)\right)^{\nu}}{\Omega_{ikgt}^{\nu}}} \end{aligned}$$

where $N(i)$ are locations which are in the same neighborhood as i, defined as those within 3km of location i from the Mumbai Travel Survey.[47]

These moments depend on both structural parameters F_g, δ_g^M as well as mean utilities b_{ikg}. Since these mean utilities can be solved uniquely to scale from the steady state condition for residential populations, we solve this as a constrained optimization problem

$$\min_{\Theta} (m(\Theta) - \hat{m})' W (m(\Theta) - \hat{m})$$
$$\text{s.t. } L_{Rikg} = \sum_{rs} \frac{(b_{ikg}/\mu_{rig})^{\nu}}{\sum_{r's'}(b_{r's'g}/\mu_{ir'g})^{\nu}} L_{Rrsg}$$

where $\Theta = (F_g, \delta_g^M, b_{ikg})$, $m(\Theta)$ represents the vector of neighborhood-population-weighted averages of the expressions above (2 for each group), $\hat{m}$ their empirical counterpart, and W is the identity matrix.

Step 2: Remaining Parameters via Indirect Inference. After estimating the migration parameters F_g and δ_g^G in the first step, we then jointly estimate the remaining parameters $\mathbf{\Theta} = \left\{\bar{A}_{mill}, \mu_U, \delta_U, \rho, \bar{u}_{mill}, \kappa\right\}$ using indirect inference. The moments to identify these structural parameters are given in Table 6.

The procedure is as follows:

- For a given parameter vector $\mathbf{\Theta}$, we solve for pre-period unobservables. We then simulate how the city would have changed if the only thing that had happened was the redevelopment of the mills. As described in Section 5.5, we do this by moving mill locations from having infinite travel costs to all other locations to having those from the post-period, and given commercial mills productivity $\bar{A}_{Mill}$ and residential mills amenities $\bar{u}_{Mill}$. We do not endogenize the construction of the mill redevelopments, simply setting them equal to the 99th percentile of building density across the city (and assume half the land on mill sites is used to construct the new towers, approximately what we observe in the data).[48] We use the model to compute the post-policy equilibrium, if this was the only change experienced by the city.

- We then use the pre- and post-policy period data from the model (i.e. observed, and counterfactual) to compute the moments from the model as outlined in Table 6.

[47] Following discussions with the team that implemented the survey, this is how "neighborhood" was defined.

[48] We do however account for the cost of building the mill redevelopments when computing land owner surplus from the policy.

Our distance function to be minimized between model and data moments is given by

$$\Omega(\boldsymbol{\Theta}) = \sum_i \frac{\left| m_i^{\text{model}}(\boldsymbol{\Theta}) - m_i^{\text{data}} \right|}{\frac{1}{2}\left| m_i^{\text{model}}(\boldsymbol{\Theta}) \right| + \frac{1}{2}\left| m_i^{\text{data}} \right|} \tag{13}$$

where $\mathbf{m}^{\text{model}}(\boldsymbol{\Theta})$ and m_i^{data} are model and data moments.

- Since we cannot establish convexity of our objective function 13, we would ideally use a global optimizer to solve it. Unfortunately, given the length of time it takes to evaluate (up to 10 minutes at some parameter values), this was not computationally feasible. Instead we use the following two-step algorithm to search for a global optimum
  - In the first step, we evaluate objective function over a grid of 20,000 points across the parameter space. We use a Halton sequence to define the points, which is more efficient than an evenly-spaced grid.
  - In the second step, we take the points with the ten lowest objective values and start ten local optimizations using these points as starting values. The optimization follows an interior-point algorithm. Our final estimate is that which yields the lowest objective value across all these runs. Reassuringly, the solutions are fairly stable across these runs.

Standard Errors. We then compute standard errors using a mixture of non-parametric and parametric bootstrap procedures. The standard errors for the migration parameters and the remaining model parameters are jointly calculated. The procedure is as follows:

- First, we draw B empirical moments:
  1. For the migration moments, which are the fraction of movers and odds of moving in or outside the neighborhood, we use a non-parametric bootstrap, resampling the data B times with replacement and computing moment values each time.
  2. For the remaining parameters, the identifying moments are regression coefficients. For each of the identifying moments (except $\bar{u}$), we perform a parametric bootstrap and randomly draw values from their estimated asymptotic distribution, a multivariate normal $\mathcal{N}\left(\hat{m}, \hat{\boldsymbol{\Sigma}}\right)$, where $\hat{m}$ are the point esti-

mates from Section 4 and $\hat{\Sigma}$ their estimated covariance matrix.[49]

- For each of the B empirical moment vectors obtained in this way, we minimize objective function 13 using local optimization with starting values equal to our point estimates of the parameters. The standard errors are the standard deviation of the resulting parameter values across these B runs.

D.5 Supplementary Model Results

Slum Redevelopment Counterfactual. Since the mill redevelopment is specific to the context of Mumbai, we run a counterfactual that relates to the redevelopment of an existing slum. Specifically, we increase amenities in formal housing $\bar{u}_{iFg}$ in the location with the median-sized slum. This counterfactual therefore provides insight into what happens if an area near an existing slum gentrifies, and how this interacts with slum redevelopment policy.

The results are shown in Table A15. The first six rows show the average welfare effects for different groups, while the the last row shows the probability the slum redevelops $p_{F|Ii}$ in the treated location. The first column shows the impacts of the change in amenities under the observed compensation policy for existing slum dwellers. The second column computes the effects if all slum residents are given the observed compensation (i.e. increasing the fraction of residents eligible for compensation to one), while the third column compensates all slum dwellers 150% of the observed compensation.[50]

The results have the same qualitative flavor as those from the main mill counterfactual. Increasing formal amenities in this location causes slums to become more likely to be redeveloped leading to substantial welfare losses for slum residents who are evicted. Increasing the generosity of compensation for relocated slum dwellers reduces this negative effect on those evicted, while also making it less likely the slum is redeveloped (and reducing average welfare, by amounts so small they are smaller than 3 significant digits).

Robustness to Alternative Parameters. Our baseline calibration sets η to match the housing supply elasticity in the most inelastic city in the US (Miami) from Saiz (2010). The second column of Table A16 shows sensitivity of the main welfare results to calibrating this to the value of Chicago, a city with a fairly elastic supply 35% higher than

[49] We do not have a way of measuring sampling error for the moment identifying $\bar{u}$ so we consider it to be observed without error. We do however include θ_g in this procedure.

[50] This is all that ends up being necessary to make evicted slum dwellers no worse off, as opposed to 3 times compensation for the main mill counterfactual.

Miami. The results barely move, showing that our main insights are fairly robust to the calibration of η.

One drawback of our analysis is that we lack data which follow those relocated from slums. It is possible that residents who are relocated from slums face moving costs higher than individuals who move voluntarily. Since our approach measures migration costs from voluntary moves, this means we could be underestimating the moving costs experienced by evicted slum dwellers.

Before exploring the sensitivity of our results to alternative values of eviction costs, it is worth first benchmarking our results to existing estimates. Unfortunately, little quasi-experimental evidence exists looking at labor market outcomes for those evicted from slums. One estimate is from Collinson et al. (2022) which finds evictions lower earnings by between 8% and 14% in American cities.[51] Our baseline estimation yields an average welfare loss of 36% in real income (which includes more than only income loss), suggesting (using this one, imperfect benchmark) that our results do not seem too attenuated in terms of the welfare losses we find for those evicted.

That said, columns 3 and 4 of Table A16 increase the fixed cost of relocation by 10% under the observed and a three-times-as-generous compensation policy for those relocated under slum redevelopment, while columns 5 and 6 do so by 50%. Compared to the baseline results in column 1, increasing relocation costs substantially reduces welfare for those evicted from slums. Welfare losses for the low-skilled who are evicted rise to 40% and 70% from 30% in the baseline. However, the main insight remains: more generous compensation reduces the welfare losses to the evicted, while also reducing the overall gains in consumer surplus.

[51] Suresh et al. (2013) highlight specific channels for additional losses from forced relocations such as loss of immobile capital investments in enterprises and housing. Patel (2011) documents the destruction of housing investments in Mumbai during our period of study.

Supplementary Material for "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills"

SM.1 SES measure construction

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 227 election wards spread across Mumbai and Mumbai Suburban districts, with an average population of about 60,000 (see Table A4). 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.

The SECC asks respondents for each household member's birthdate and highest level of education completed with the following ordered categories: illiterate, less than primary, primary, middle, secondary, higher secondary, and graduate or higher. For voting-age individual i in the 2011 SECC, we treat the answer to the education variable as the ordered response variable $education_index_i$ taking values from 0 to 6. Our goal is then to estimate the regression function $f(\cdot)$ in

$$education_index_i = f(last_name_i) + \epsilon_i \tag{14}$$

With $\hat{f}$ measuring the predicted education level associated with a last name, we 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. Gentrification of a ward will be measured by individuals with names associated with high education levels in the SECC moving in, and/or individuals with names associated with low education levels moving out.

To estimate $f(\cdot)$ in the SECC data, we need a away of encoding the names numerically. 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 (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. Levenshtein distance is commonly used in fuzzy string matching and to correct misspellings, in popular packages like Reif (2022)'s `strgroup` Stata package.

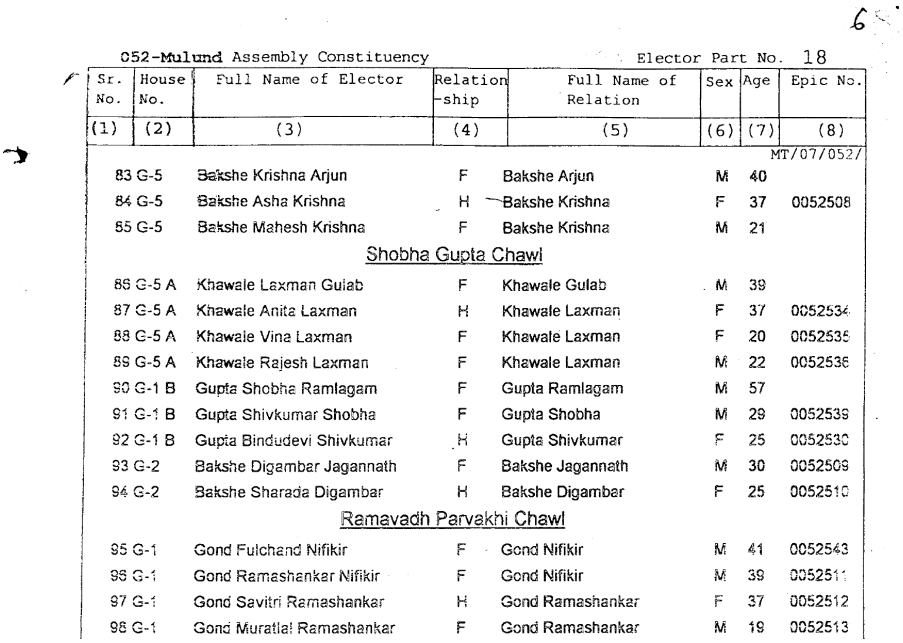
Using Levenshtein distance to the prototypes means that misspellings of names or mistakes in optical character recognition (OCR) will be encoded similarly to the correct name, which is important since the posted voter lists are scans (see Figure SM6). It also means that our regression pools information across similar names, providing mean-squared-error advantages akin to those in Empirical Bayes estimation of, for instance, teacher effects (Morris, 1983; Chetty et al., 2011).

The dimension of the encoded name 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 now specify our method of estimating $f(\cdot)$ in equation (14. We use Extremely Randomized Trees (Geurts et al. (2006)), a faster-to-compute variant of Random Forests.

We select 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 Figure (SM6) panel (b)).

Figure SM6: 2002 voter list example pages

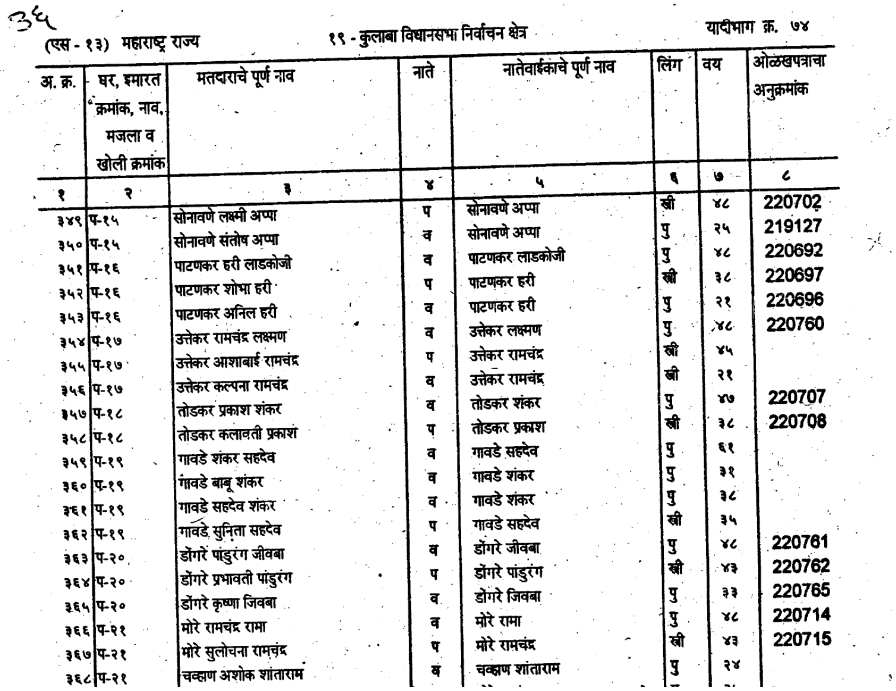
(a) English-language list



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

(b) Marathi-language 2002 voter list



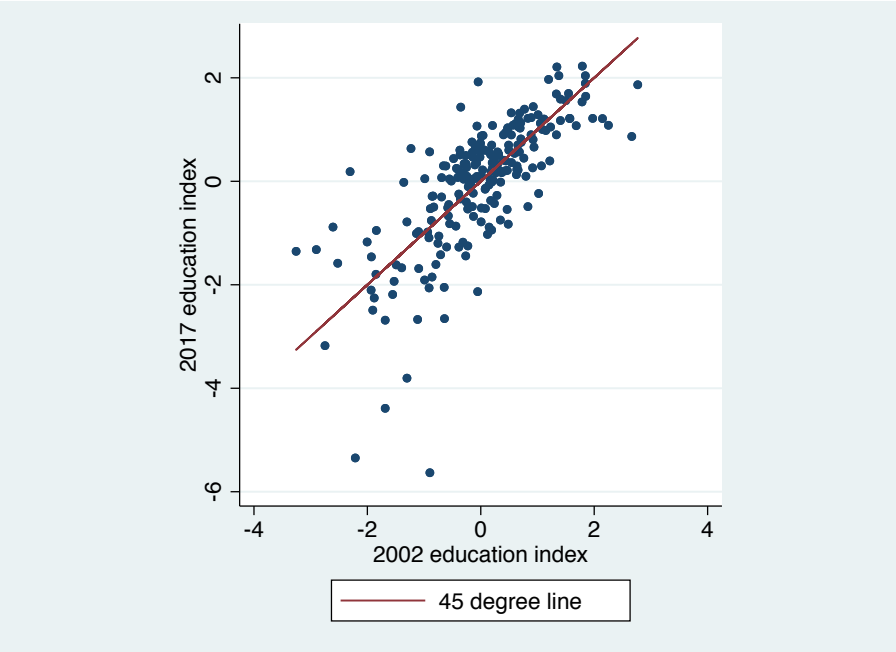
(एस - १३) महाराष्ट्र राज्य — १९ - कुलाबा विधानसभा निर्वाचन क्षेत्र — यादीभाग क्र. ७४

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	

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.

We validate our election-ward-level average education indices in two ways. First, we check their stability over time. Each point in Figure SM7 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 Mumbai was gentrifying over time.

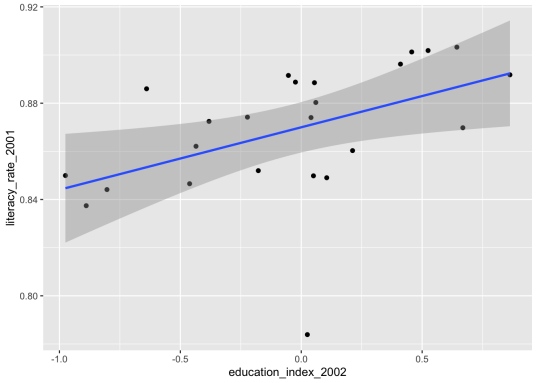
Figure SM7: Education index over time



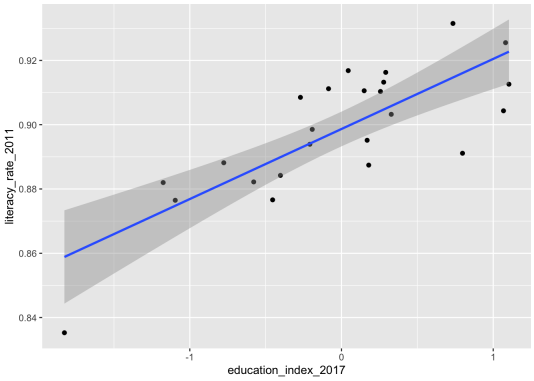
Second, we compare literacy rates, which are available for 24 city wards from the 2001 and 2011 population censuses, with the average values of the education index at the city ward level from 2002 and 2017. The results are shown in Figure SM8 with the same normalization as in Figure SM7. The indices in both 2002 and 2017 correlate well with the corresponding literacy rates.

Figure SM8: 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



SM.2 Real Estate Developer Market Structure

SM.2.1 Mill Redevelopers and Slum Redevelopers are Different

In this appendix, we investigate whether developers of mill sites participate in slum redevelopment to internalize the externalities slums may impose on the mill redevelopments. Developers might do this by redeveloping slums bordering the luxury apartments they have built on former mill sites so apartment buyers would not have to see slums from their windows and thus would be willing to pay higher prices. This kind of behavior would lead us to overestimate the causal slum redevelopment elasticity with respect to the price of formal floorspace in a location.

To check into this possibility we identify all plots that border one of the mill sites and were re-developed from a slum to a new development between 2001 and 2016. For

each new building built on these re-developed sites, we look up the identity of the developer through a Google search. In most cases, the name of the building is indicative of its developer (e.g., building: Ahuja Towers—developer: Ahuja Constructions, Amann Marina—Amann Group). Then, for each of the redeveloped sites *near* mills, we compare their developers with the developers of the adjacent mill sites, also identified using a Google search.

The results are in Table SM20 below. This table shows the full set of developers, with entries in the "Redeveloped these mills" column for mill redevelopers and entries in the "Redeveloped slums near these mills" column for slum redevelopers. Rows for slum redevelopers are highlighted. The table shows there is no overlap between developers of mills and nearby slums so the concern raised at the beginning of this appendix does not apply.

Table SM20: Developers of Mills and Surrounding Slums Differ

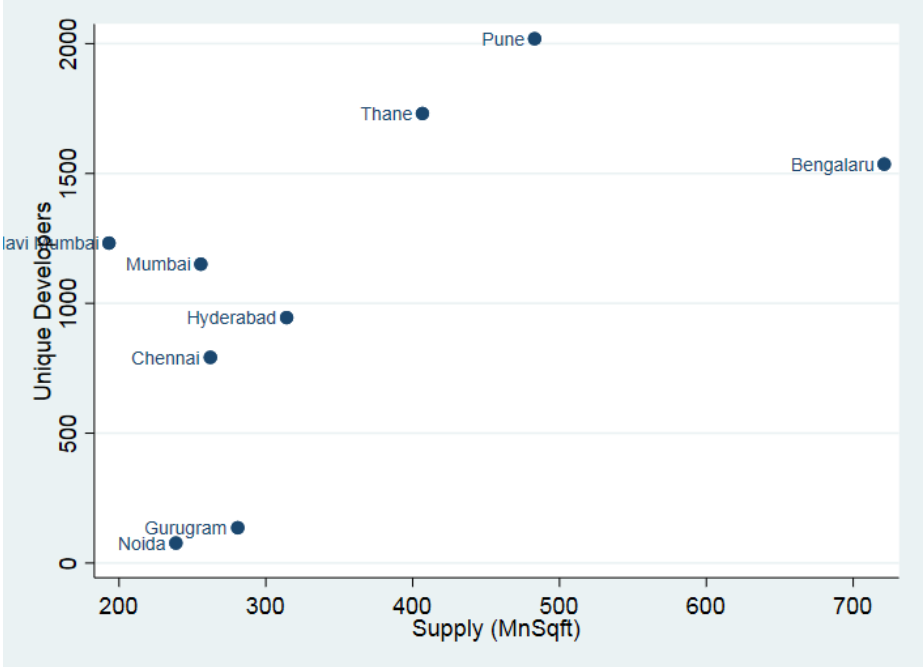
Developer	**Redeveloped these mills**	**Redeveloped slums near these mills**
Aditya Birla Group	Century Spinning & Weaving Mills	
Ahuja Constructions		Century Spinning & Weaving Mills; Standard Mills no.1
Amann Group		Bharat Mills; Raghuvanshi Mills; Piramal Spinning & Weaving Mills; Phoenix Mills
Ansals Buildwell Ltd.	Shree Ram Mills	
Ashford Group	Matulya Mills	
Bombay Realty	Bombay Dyeing (Spring Mills)	
Chaitanya Group	Standard Mills no.1	
Chandak Group		Bharat Mills; Raghuvanshi Mills; Piramal Spinning & Weaving Mills; Phoenix Mills
Gomata Janta CHSL		Prakash Cotton Mills
Indiabulls	Bharat Mills; Elphinstone Mills (South)	
Jaguar Buildcon Private Limited	Raghuvanshi Mills	
Lodha Group	Apollo Mills (South)	
Lokhandwala Infrastructure		Bharat Mills; Raghuvanshi Mills; Apollo Mills (South); Piramal Spinning & Weaving Mills; Shakti Mills; Sitaram Mills; Phoenix Mills
Mahalaxmi Developers		Century Spinning & Weaving Mills

Maharashtra Housing and Area Development Authority	Century Spinning & Weaving Mills	
Marathon Group	Piramal Spinning & Weaving Mills	
Neumec Group		Bharat Mills; Raghuvanshi Mills; Piramal Spinning & Weaving Mills; Phoenix Mills
Omkar Realtors & Developers		Century Spinning & Weaving Mills
Peninsula Land Ltd	Gokuldas Morarjee Mills no.2; New Great Eastern Spinning & Weaving Mills	
Phoenix Mills	Phoenix Mills	
Raheja Universal	Prakash Cotton Mills	
Royal Developers		India United Mills no.5; Mafatlal Mills nos.1 New Hind Textile Mills; New Great Eastern Spinning & Weaving Mills
Ruparel Realty		Bombay Dyeing (Spring Mills)
Sheth Developers	Standard Mills no.1	
Shree Naman Group		Elphinstone Mills (South)
Shree Ram Urban Infrastructure Ltd.	Shree Ram Mills	
Skylark Buildcon Pvt. Ltd. And Vrinda Enterprises		Century Spinning & Weaving Mills; Standard Mills no.1
The Cynergy Group	Standard Mills no.1	

SM.2.2 Market Concentration

Figure SM9 below shows that the environment in Mumbai looks more competitive than most other large Indian cities in that there are more developers per million square feet of built floorspace than expected based on the relationship in other cities. We are grateful to Santosh Anagol, Vimal Balasubramaniam, Tarun Ramadorai, and Antoine Uettwiller for generating this figure on our behalf from the Propequity data discussed in Anagol et al. (2022).

Figure SM9: Number of Unique Developers vs. Floorspace Supply in Several Large Indian Cities



SM.3 Geolocation of Economic Census Data

SM.3.1 Fourth and Fifth Economic Census

The Fourth (1998) and Fifth (2005) ECs have the same set of spatial units, which we were able to locate using the Fifth EC's directory of addresses for establishments with more than 10 workers. We clean and geolocate these addresses using the Google Maps API, then associate them with geographic identifiers present in the microdata using an iterative algorithm that ultimately provides us with a set of polygons representing investigator units (IV Units), which are the second-smallest spatial units recorded in the used in the EC. As

shown in Table A4, they are quite small, covering only about 15,000 persons and dividing Mumbai and Mumbai Suburban districts into 825 spatial units. Figure A4 shows a map of IV Unit boundaries in blue, with city wards superimposed in black and mill sites in red. In the rest of this subsection we describe our algorithm for locating IV Units in the 5th EC.

The EC5 directory of establishments is an auxiliary dataset recording an establishment's name, address, 4-digit industry code (NIC code) and an interval containing its number of employees. EC enumerators were instructed to record an establishment's address if it employed 10 or more workers. We first geocode all addresses in Mumbai and Mumbai Suburban Districts using Google's Places and Geocoding APIs. If we also had each directory establishment's IV Unit ID, we would have coordinates known to lie within that IV Unit, which allow us to approximate the IV Unit's boundary using Voronoi polygons. Unfortunately the IV Unit ID is only available in the EC microdata and there is no establishment ID common to the microdata and directory, so we use the following methods to match directory establishments to establishments in the microdata in order to generate an approximate centroid for each IV Unit in ECs 4 and 5.

SM.3.1.1 Ward X NIC X Employment Bins with a Single IV Unit

The EC microdata record each establishment's city ward, employment level (unbinned), and NIC code. The first step of our algorithm considers the set of unique combinations of ward, NIC code, and employment intervals (as defined in the directory) present in the EC microdata, which we will refer to as ward X NIC X employment bins. We categorize both microdata and directory firms into these bins.

We then subset to those bins in which all microdata establishments share the same IV Unit. Since the directory establishments are a subset of the microdata establishments, all directory firms in such a bin must lie within that single IV Unit. However, since address quality and geocoding performance varies within the directory, some of the directory firm coordinates associated with a bin may lie outside the true boundaries of the bin's IV Unit.

To identify outlying coordinates, we first average the longitude and latitude coordinates of each of the directory firms associated with a single IV Unit to arrive at an approximate centroid for that IV unit. As a measure of a set of coordinates being "too far" from its IV Unit centroid, we choose 2 times $AvgNeighborDistance$, the average distance between centroids of IV Units next to each other in the sequence of IV Unit IDs. We know from the procedures followed in the EC that IV Unit 1 is next to IV Unit 2, which is next to IV Unit 3, and so on. After excluding the coordinates we consider too far from their approximate IV Unit centroids according to this measure, we average the

remaining coordinates and update the approximate centroids.

SM.3.1.2 Improving Approximate Centroid Accuracy

We now investigate bins where all microdata establishments belong to IV Units that have been assigned centroid coordinates in the previous step. We try to match directory firms without an IV Unit to one of the microdata establishments in the bin. For these unmatched establishments, we compute the distance from the establishment's coordinates to the centroid of each IV Unit in the bin. Establishments less than 1.5 times $AvgNeighborDistance$ from a **single** IV Unit are assigned to that IV Unit. We then recompute IV Unit approximate centroids.

We re-run this procedure (find unmatched establishments, assign if they are close enough to a single centroid, recompute centroids) as many times as the maximum number of firms in a bin, by which point the centroids are stable. The purpose of assigning additional directory firms to IV Units who already have centroid coordinates is improving the accuracy of the approximate centroid by, for example, having establishment locations from around the IV Unit rather than a single establishment on the IV Unit boundary.

SM.3.1.3 Assign Directory Firms far from IV Unit Centroids

With this set of IV Unit centroids in hand, we now examine bins that contain microdata establishments belonging to one or more IV Units assigned centroid coordinates in the previous steps and microdata establishments belonging to **one** IV Unit that has not been assigned coordinates yet. Within these bins, we examine the directory firms that have not yet been assigned an IV Unit. We now assess whether any of those directory firms are sufficiently far from the existing IV Unit centroids so that they should be considered to belong to a separate IV Unit. If so, we assign that directory firm's coordinates to the one IV Unit in the bin without coordinates.

To implement this idea, we first update $AvgNeighborDistance$ based on having run the procedures in SM.3.1.1 and SM.3.1.2. We then calculate the distance of each directory firm without an IV Unit to the centroids of the IV Units in the bin. If any of these firms is more than 1.4 times $AvgNeighborDistance$ from all the centroids, we consider that it belongs to the IV Unit in the bin that has not yet been assigned coordinates. We then update our estimates of all IV Unit centroid coordinates.

With more IV Units assigned centroid coordinates through the procedure described in the previous paragraph, some additional bins may now have onyl a single IV Unit without assigned coordinates. We therefore repeat the procedure until reaching an iteration when

no new centroid coordinates are assigned.

We finish this step of the procedure by (1) removing coordinates too far from their IV Unit centroid and recomputing the centroid coordinates as in SM.3.1.1 and (2) re-running the procedure described in SM.3.1.2.

SM.3.1.4 Imputation based on IV Unit ID sequence I

We now use the sequence of IV Unit IDs to impute a candidate set of coordinates for IV units that have not been located to this point. For each unlocated IV Unit, we take an average of the longitude and latitude coordinates of located IV Units in the same city ward which are either neighbors of the unlocated Unit in the ID sequence, or neighbors of the Unit's neighbors. We produce a measure of the radius around the candidate centroid that could contain a firm which is actually in the IV Unit. When the candidate centroid coordinates are not based on both neighbors in the IV Unit ID sequence, there is more uncertainty about the true centroid's location and we use a larger radius. To capture this we define a value

$$err_u = \frac{2\sum_{v\in\{-2,2\}} \mathbb{I}\{\text{IV Unit located}\}_{u+v} + \sum_{v\in\{-1,1\}} \mathbb{I}\{\text{IV Unit located}\}_{u+v}}{\sum_{w=-2}^{2} \mathbb{I}\{\text{IV Unit located}\}_{u+v}}$$

where u denotes the ID of the unlocated IV Unit.

Starting from the set of directory establishments not yet assigned an IV Unit, we keep establishments farther than 1.4 times $AvgNeighborDistance$ from all located IV Units. We then compute the distance of the remaining establishments to the candidate centroid coordinates of unlocated IV Units from the previous paragraph. We assign a directory establishment to IV Unit u if the establishment's coordinates place it less than $err_u \times 1.5 \times AvgNeighborDistance$ from u's candidate centroid. We then generate centroid coordinates for the IV Units with newly-assigned directory establishments. We subsequently re-run the procedure described in the first part of this paragraph until reaching an iteration when no new IV Units are assigned directory establishments.

Now we update the candidate centroid coordinates for IV Units which still have not been assigned any directory establishments and re-run the procedure from the first two paragraphs of this subsection until reaching an iteration when no candidate centroid coordinates for IV Units without directory establishments update. As in SM.3.1.3, we then (1) un-assign directory establishments with coordinates too far from their assigned IV Unit's centroid and recompute IV Unit centroid coordinates following the method from SM.3.1.1 and (2) re-run the procedure from SM.3.1.2. Since some directory firms have

the same coordinates due to imperfect geolocation by the Google APIs we use, we drop any IV Units with duplicate centroids at this stage.

SM.3.1.5 Robustness

The last column of Table SM21 shows the number of IV Units geolocated at each step using the constants described so far. The top row shows that we can assign centroid coordinates to just over half (426) the 818 IV Units in Mumbai District using only ward X NIC X employment bins with a single IV Unit as described in subsections SM.3.1.1 and SM.3.1.2. We can generate coordinates for about 200 more IV Units using directory establishments located far from the IV Unit centroids assigned in the previous step, as described in SM.3.1.3. Finally, we locate 29 more IV Units using the IV Unit ID sequence as described in SM.3.1.4 for a total of 652 IV Units located.

The remaining four columns show number of IV Units located under alternative specifications of tuning parameters (constants). The fourth and fifth column show the implications of changing the threshold for being far enough from the centroids of previously-located IV Units in a bin with one unlocated IV Unit from $1.4 \times AvgNeighborDistance$ from section SM.3.1.3 to $0.7 \times AvgNeighborDistance$ and $2.1 \times AvgNeighborDistance$, respectively. Comfortingly, increasing the distance threshold results in only a 30% reduction in IV Units located. Reducing the distance threshold has a bigger effect, increasing the number of IV Units located by 56%, but assuming that a firm only $0.7 \times AvgNeighborDistance$ away from already-located centroids does not lie in any of these is quite extreme. Neither alternative has much of an effect on IV Units located downstream using the method described in SM.3.1.4. Columns 2 and 3 show the effects of changing the 1.5 in $err_u \times 1.5 \times AvgNeighborDistance$ from section SM.3.1.4 to 0.75 and 2.25, respectively. Neither change has a big effect on the total number of IV Units located.

Table SM21: Sensitivity to Constants in IV Unit Geolocation

method	ivcount_deflate_closeness	ivcount_inflate_closeness	ivcount_deflate_distance	ivcount_inflate_distance	ivcount_original
SM.3.1.1 and SM.3.1.2	426	426	426	426	426
Adding SM.3.1.3	197	197	308	128	197
Adding SM.3.1.4	15	52	22	26	29
Total	**638**	**675**	**756**	**580**	**652**

SM.3.1.6 Wrapping Up: Imputation Based on IV Unit ID Sequence II

All imputation of IV Unit location up to this point has gone through the geolocated coordinates of directory of establishment firms. To locate the final 173 IV Units, we use the sequence of IV Unit IDs more directly. We fit separate local constant regressions with

second-order Gaussian kernels for IV Unit longitude and latitude with unit ID as one regressor and IDs of 98 city sub-wards as the other. We use the cross-validation method of Li and Racine (2004) to select the bandwidths for each variable. For longitude they are 5.4 for Unit ID and 0.5 for subward ID. For latitude, the bandwidths are 6.5 for Unit ID and 0.45 for subward ID. We impute the predicted values of longitude and latitude from these regressions for the unlocated IV units, plugging in their IV Unit and subward IDs. In practice the small bandwidths for subward ID mean that our imputation procedure only uses information on longitude and latitude from within the Unit's subward.

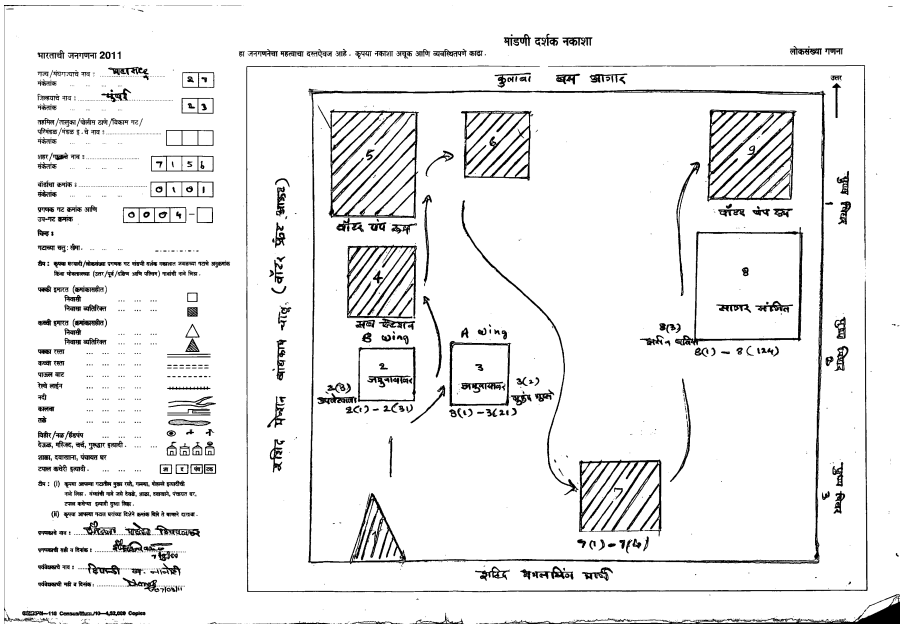
With our preferred centroid estimates in hand for each IV Unit we generate the a map of their extents using Voronoi polygons within the boundaries of the city wards the IV Units are know to lie in. Figure A4 shows the resulting polygons, with blue boundaries, in Mumbai district.

SM.3.2 Sixth Economic Census

For the Sixth EC, we were able to purchase enumeration block maps covering Mumbai district, where most mills are located (see Figure A1), and a part of Mumbai Suburban district (block maps used in earlier Economic Census rounds have been destroyed). The enumeration block maps are hand-drawn by the enumerator assigned to each block during the survey. Enumeration blocks are the smallest geographic unit in the Economic Census. Each map contains a sketch of the roads (labelled), landmarks and buildings within the block (see Figure SM10 below for an example). Since almost all maps are written in Marathi, 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.

Out of 7,338 enumeration blocks in Mumbai district, there are 1,105 for which we either could not purchase maps or the quality of the maps was not sufficient to establish coordinates. To geolocate these blocks we use the strategy from SM.3.1.6, running nonparametric regressions of block longitude and latitude on enumeration block ID and subward ID (again, we know that enumeration blocks with neighboring ID numbers were geographical neighbors as well). The bandwidths for latitude are 1.3 for enumeration block ID and 0.5 for subward ID. For longitude they are 3.4 and 0.21. So on both coordinates the information being used to relocate enumeration blocks is quite local. Finally, there are two subwards containing stretches of unlocated enumeration block IDs where such local interpolation is not possible. In these two subwards we perform global B-spline regressions which fit the located enumeration block ID relationship with longitude and latitude well, and impute predicted values for the 261 unlocated enumeration blocks.

Figure SM10: Example 2011 Census Enumeration Block map



For the enumeration blocks in Mumbai Suburban district for which we could not purchase maps, we follow the same algorithm laid out in subsections SM.3.1.1, SM.3.1.2, SM.3.1.3, and SM.3.1.4 using the same preferred constants, and with the additional advantage of having known locations for the blocks we have mapped. In contrast to the 2005 EC, the 2013 EC's directory of establishments records a 3-digit NIC code, smaller intervals of number of employees and an address for firms having 8 or more workers in 2013. Taken together these differences make the CEB location process easier than the IV Unit location. We do not need a the further imputation step from SM.3.1.6 in Mumbai Suburban district.

The final step is to aggregate the census enumeration blocks to the level of the larger 2005 IV Units to harmonize the Sixth Economic Census data with that of earlier years. This aggregation also reduces the role of any measurement error in enumeration block location.

SM.3.3 Model Derivations

SM.3.3.1 Developer Profits

This section derives the expressions for developer profits from the text. On informal plots, $\Pi_{Ii} = r_{Ii}$.

On formal plots that were formal at t, $\Pi_{F|Fi} = E\left[\epsilon_F \tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} | \epsilon_F > r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right]$. So we need to know $E\left[\epsilon_F | \epsilon_F > r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right]$. We have that

$$\begin{aligned} E\left[\epsilon_F | \epsilon_F > r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right] &= \int_{r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \epsilon \frac{dF(\epsilon)}{1-F(\epsilon)} \\ &= \int_{r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \exp\left(\kappa r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right) \epsilon\kappa \exp\left(-\kappa\epsilon\right) d\epsilon \\ &= \exp\left(\kappa r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right) \int_{r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \epsilon\kappa \exp\left(-\kappa\epsilon\right) d\epsilon \end{aligned}$$

Integration by parts implies

$$\int_a^{\infty} \epsilon\kappa e^{-\kappa\epsilon} d\epsilon = \left[-\frac{(\kappa\epsilon+1)}{\kappa} e^{-\kappa\epsilon}\right]_a^{\infty}$$

So

$$\int_{r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \epsilon\kappa \exp\left(-\kappa\epsilon\right) d\epsilon = \frac{(\kappa r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} + 1)}{\kappa} \exp\left(-\kappa r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right)$$

since $\lim_{\epsilon\to\infty} \frac{(\kappa\epsilon+1)}{\kappa} e^{-\kappa\epsilon} = 0$. Therefore,

$$E\left[\epsilon_F | \epsilon_F > r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right] = \frac{1}{\kappa(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}} \left[\kappa r_{Ii} + (1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right]$$

Therefore

$$\Pi_{F|Fi} = (1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} E\left[\epsilon_F | \epsilon_F > r_{Ii}/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right] = \frac{1}{\kappa}\left[\kappa r_{Ii} + (1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right]$$

On formal plots that were informal,

$$\Pi_{F|Ii} = E\left[\epsilon_F(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} - F_i | \epsilon_F > r_{Ii}/((1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}} - F_i)\right].$$

The same algebra delivers

$$E\left[\epsilon_F | \epsilon_F > (r_{Ii} + F_i)/(1-\eta)\tilde{\eta} r_{Fit+1}^{\frac{1}{1-\eta}}\right]$$

$$
\begin{aligned}
&= \int_{(r_{Ii}+F_i)/(1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \epsilon \frac{dF(\epsilon)}{1-F(\epsilon)} \\
&= \int_{(r_{Ii}+F_i)/(1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}}^{\infty} \exp\left(\kappa(r_{Ii}+F_i)/(1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}\right) \epsilon\kappa \exp\left(-\kappa\epsilon\right) d\epsilon \\
&= \frac{1}{\kappa(1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}} \left[\kappa(r_{Ii}+F_i) + (1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}\right]
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\Pi_{F|Ii} &= E\left[\epsilon_F | \epsilon_F > (r_{Ii}+F_i)/((1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}\right](1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}} - F_i \\
&= \frac{1}{\kappa}\left[\kappa(r_{Ii}+F_i) + (1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}\right] - F_i \\
&= \frac{1}{\kappa}\left[\kappa r_{Ii} + (1-\eta)\tilde{\eta}r_{Fit+1}^{\frac{1}{1-\eta}}\right]
\end{aligned}
$$

Note that while developers that convert informal plots to formal face a higher fixed cost to do so which lowers average profits, they select into redevelopment based on the idiosyncratic profitability ϵ_F. This shock has to be larger for redevelopers of informal plots to make redevelopment profitable, and with the exponential distribution these two effects cancel out so that average profits on formal plots are independent of the initial state (i.e. $\Pi_{F|Ii} = \Pi_{F|Fi} = \Pi_F$). This delivers the expressions in the text.

SM.3.3.2 Welfare

For workers who are not evicted

$$
\begin{aligned}
\Delta W_{ikg}^{NE} &= \ln\left(\frac{\Omega_{ikgt+1}^{NE}}{\Omega_{ikgt}^{NE}}\right) \\
&= \frac{1}{\nu}\ln\left(\frac{\sum_{rs\cup\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}}\right) \\
&= \frac{1}{\nu}\ln\left(\frac{\sum_{rs}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}} + \frac{\sum_{\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}}\right) \\
&= \frac{1}{\nu}\ln\left(\sum_{rs}\frac{(b_{rsgt}/\mu_{ikrsg})^{\nu}}{\sum_{r's'}(b_{r's'gt}/\mu_{ikr's'g})^{\nu}}\hat{b}_{rsgt+1}^{\nu} + \frac{\sum_{\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}}\right)
\end{aligned}
$$

Since the last term simplifies to

$$\frac{\sum_{\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}} = \frac{\sum_{\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs\cup\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}} \times \frac{\sum_{rs\cup\mathcal{L}^{ResMills}}(b_{rsgt+1}/\mu_{ikrsg})^{\nu}}{\sum_{rs}(b_{rsgt}/\mu_{ikrsg})^{\nu}}$$
$$= \pi_{ikg}^{Mill} \times \left(\frac{\Omega_{ikgt+1}^{V}}{\Omega_{ikgt}^{V}}\right)^{\nu},$$

we arrive at

$$\Delta W_{ikg}^{V} = \frac{1}{\nu}\ln\left(\sum_{rs}\frac{(b_{rsgt}/\mu_{ikrsg})^{\nu}}{\sum_{r's'}(b_{r's'gt}/\mu_{ikr's'g})^{\nu}}\left[\hat{u}_{rs}^{\rho_g}\hat{\tilde{y}}_{rgt}\hat{r}_{Rrst}^{\beta-1}\right]^{\nu} + \pi_{ikg}^{Mill} \times \frac{\Omega_{ikgt+1}^{V}}{\Omega_{ikgt}^{V}}\right)$$

Similar derivations apply for the expressions for those who are evicted with and without compensation to arrive at

$$\Delta W_{ikg}^{E,NoComp} = \frac{1}{\nu}\ln\left(\sum_{rs\in\mathcal{L}_{t+1}\setminus ik}\pi_{rs|ikg}^{L}\left(\hat{u}_{ikgt}^{\rho_g}\hat{\tilde{y}}_{rgt}\hat{r}_{Rrst}^{\beta-1}\right)^{\nu} + \pi_{ikg}^{Mill,E,NoComp} \times \left(\frac{\Omega_{ikgt+1}^{E,NoComp}}{\Omega_{ikgt}^{NE}}\right)^{\nu}\right)$$

$$\Delta W_{ikg}^{E,Comp} = \frac{1}{\nu}\ln\left(\sum_{rs\in\mathcal{L}_{t+1}\setminus ik}\pi_{rs|ikg}^{L}\left[\hat{u}_{rsgt}^{\rho_g}\hat{\tilde{y}}_{rgt}^{Comp}\hat{r}_{Rrst}^{\beta-1}\right]^{\nu} + \pi_{ikg}^{Mill,E,Comp} \times \left(\frac{\Omega_{ikgt+1}^{E,Comp}}{\Omega_{ikgt}^{NE}}\right)^{\nu}\right)$$

where $\hat{\tilde{y}}_{rgt}^{Comp} \equiv (\tilde{y}_{rgt+1} + y^{comp})/\tilde{y}_{rgt}$ and $\Omega_{ikgt+1}^{E,Comp}, \Omega_{ikgt+1}^{E,NoComp}$ are the average utilities of residents in (i,k) who are evicted with and without compensation.

Note that those who choose mill sites get the same change in utility as everyone starting in the same state (i,k,g) due to the selection on idiosyncratic draws. Any difference in average utility on redeveloped mill sites is offset by the average idiosyncratic preference draw for those choosing to live there.

SM.4 Benchmarking the Spillover Estimates

To benchmark our estimates of the amenity externality, we abstract away from the spatial spillovers to focus on the size of the externality rather than its spatial decay. The functional forms used in Ahlfeldt et al. (2015), Diamond (2016) and our paper are

$$u_i = \bar{u}_i L_i^{\mu^A} \qquad \text{(Ahlfeldt et al. (2015)))}$$
$$u_i = \bar{u}_i \left(L_{Hi}/L_{Li}\right)^{\mu^D} \qquad \text{(Diamond (2016))}$$
$$u_i = \bar{u}_i \left(L_{Hi}/L_i\right)^{\mu^M} \qquad \text{(This paper and Tsivanidis (2022))}$$

where $L_i = L_{Hi} + L_{Li}$ is total population across high-education and low-education groups. We index the functional forms by A(RSW), D(iamond) and M(umbai) to distinguish them in what follows.

To compare our estimates with Diamond (2016), we take a first order approximation of our functional form to get

$$\begin{aligned} d\ln u_i &= d\ln \bar{u}_i + \mu^M \left[d\ln L_{Hi} - \pi_{Hi} d\ln L_{Hi} - (1-\pi_{Hi}) d\ln L_{Li}\right] \\ &= d\ln \bar{u}_i + \mu^M (1-\pi_{Hi}) d\ln \left(L_{Hi}/L_{Li}\right). \end{aligned}$$

where $\pi_{Hi} = L_{Hi}/L_i$ is the share of high types in location i. This is the same as Diamond (2016) with μ^D replaced by $\mu^M(1-\pi_{Hi})$. She finds $\mu^D \approx 2.6$. With $\pi_{Hi} = 0.3$ on average, using $\mu^M \approx 0.88$this implies $E\left[\mu^M(1-\pi_{Hi})\right] = 0.88 \times 0.7 = 0.62$. So our estimate is about one quarter of Diamond (2016)'s estimate.

To compare our estimates with Ahlfeldt et al. (2015), note that the change in the endogenous amenity component is (to a first order)

$$\begin{aligned} &\mu^A d\ln L_i && \text{(Ahlfeldt et al. (2015))} \\ &\mu^M \underbrace{(1-\pi_{Hi}) d\ln \left(L_{Hi}/L_{Li}\right)}_{\approx d\ln(L_{Hi}/L_i)} && \text{(This paper and Tsivanidis (2022))} \end{aligned}$$

So a change in the distribution of residential population will have the same effect on endogenous amenities in either model when

$$\begin{aligned} \mu^A d\ln L_i &= \mu^M d\ln \left(L_{Hi}/L_i\right) \\ \Leftrightarrow d\ln \left(L_{Hi}/L_i\right) &= \frac{0.15}{0.88} \times d\ln L_i = 0.17 \times d\ln L_i \end{aligned}$$

where we have plugged in $\mu^A = 0.15$ and $\mu_M = 88$. The change will be the same when the high-skill share increases by about 20% of every percentage change in total population. For example, if the average skill share is 0.3 and the population doubles in an average neighborhood, the high-skill share would increase by 20% or from 0.3 to 0.36.

While it is hard to directly compare these models as one allows demographic composition of neighborhoods to change while the other does not, the intent of this exercise is to understand whether our estimate of $\mu_M = 0.88$ is unreasonably large compared to Ahlfeldt et al. (2015)'s estimate. If one thinks that seeing the skill share of a neighborhood moving from 0.3 to 0.36 as one doubles the population density of a neighborhood, then this doesn't appear too unreasonable.