Distant Shocks, Migration, and Housing Supply in India

Arnab Dutta* Sahil Gandhi† Richard K. Green*

May 18, 2021

Abstract

It is hard to estimate housing supply elasticities. India is a particularly useful country to study housing supply because it is large and has a variety of housing typologies. We use a panel instrumental variable framework to estimate the supply elasticity of non-durable, durable, and vacant residential housing units in urban India. We use two migration-inducing exogenous events — negative rainfall shocks and a highway upgrade program — occurring in a distant state as demand shifters for local urban housing markets. We apply the Rosen-Roback spatial equilibrium setting to show that both the negative rainfall shocks and the highway upgrade program in a distant state increase inter-state migration. This increase leads to higher population and household growth, and therefore, higher demand for housing in local urban markets. Our findings are three-fold. First, we estimate the long-term supply elasticity of durable housing in urban India to be 1.64. This estimate is substantially lower than the long-run housing supply elasticity estimates of 6-13 for metropolitan areas in the United States seen in the literature. Second, we find that the supply elasticity of non-durable housing is -0.55. Negative supply elasticity of non-durable housing is consistent with the existence of urban gentrification through the demolition and upgrading of slums. And finally, we estimate the elasticity of vacant residential housing unit supply to be 2.63. We posit that a relatively higher vacant housing unit elasticity indicates speculative building by developers.

JEL Classification: J61, R23, R31

Keywords: Housing, Migration, India

*University of Southern California

†CSEP India (Formerly Brookings Institution India)

1 Introduction

Indian cities were home to 377 million people, or roughly 10.4% of the global urban population in 2011 (Census of India, 2011; Desa *et al.*, 2014). Indian Census data indicate that India's urban population grew by roughly 91 million between 2001 and 2011. A third of this growth can be attributed to rural-urban migration. Prior academic literature indicates that internal migration has been historically low in India (Munshi and Rosenzweig, 2016). However, this seems to be changing as the number of internal migrants moving into Indian cities went up by almost 70% between the 1990s and the 2000s (see figure 4b). Recent studies have found that the Information Technology (IT) boom of the late 1990s and the early 2000s partly explains this growth in the internal movement of Indians (Ghose, 2019). Theory and evidence elsewhere suggest that the burgeoning urban population, along with an increase in internal migration into urban areas, is likely to be accompanied by a surge in housing demand (Molloy *et al.*, 2011). But is the change in the supply of housing in Indian cities enough to meet this rising demand?

In this paper, we use a panel instrumental variable framework to estimate the supply elasticities of non-durable, durable, and vacant residential housing units in urban India.[1] We use two migration-inducing exogenous events — negative rainfall shocks and a highway upgrade program — occurring in a distant state as demand shifters for local urban housing markets. We apply the Rosen-Roback spatial equilibrium setting to show that both the negative rainfall shocks and the highway upgrade program in a distant state increase inter-state migration (Rosen, 1979; Roback, 1982). Increased inter-state migration leads to higher population and household growth, and therefore, higher demand for housing in local urban markets. As an example, let's say that we want to estimate the housing supply elasticity of urban areas in the relatively wealthy Indian state of Maharashtra. Now, consider the poorest state of Bihar as the distant state where there is a drought and a highway upgrade program. These events in Bihar will likely spur migration between Bihar and Maharashtra. The urban population in Maharashtra will change in response to the Bihar-Maharashtra migration, thereby acting as a demand shock to Maharashtra's urban housing. We show that the negative rainfall shocks and the highway upgrade program in distant states act as demand shifters for durable, non-durable, and vacant residential houses in local urban markets.

Our findings are three-fold. First, we estimate the supply elasticity of durable housing in

[1] Durable houses are made of materials like concrete, stone, and bricks. Non-durable houses have walls and roofs made of grass, thatch, bamboo, plastic, and mud and are typically found in slums and squatter settlements. Hereon, we refer to informal, slums, and non-durable houses interchangeably.

urban India to be roughly 1.64. This estimate is substantially lower than the long-run housing supply elasticity estimates of 6-13 for metropolitan areas in the United States (Malpezzi and Maclennan, 2001). Our estimate is also very close to the *short-run* supply elasticity of 1.75 obtained by Saiz (2010) for the average metropolitan area in the United States. Needless to say, short-run supply elasticities by their very nature are smaller than long-run elasticities. Second, we find that the supply elasticity of non-durable housing is -0.55. The negative supply elasticity of non-durable housing is counterintuitive, suggesting that as non-durable housing rents increase, the supply of non-durable residential housing units decreases. This is consistent with urban gentrification that occurs in two ways in Indian cities. First, a simultaneous increase in rents paid by slum dwellers and land values around slums attract real estate developers. Slums are cleared to construct durable residential and commercial real estate space (Bhan, 2009). And second, slums are upgraded through various government and non-government programs that convert non-durable units to durable ones (Rains *et al.*, 2019; Rains and Krishna, 2020). Finally, we estimate the elasticity of vacant residential housing unit supply to be 2.63, which is larger than the elasticity of durable housing unit supply. We posit that developers are engaging in speculative construction with the expectation of higher demand as market rents go up (Gandhi *et al.*, 2021a).

Prior academic literature on housing supply has predominantly focused on developed countries like the United States and has underscored the role of regulations and natural land constraints like hilly terrains in reducing the supply elasticity of housing in metropolitan areas (Baum-Snow and Han, 2019; Green *et al.*, 2005; Saiz, 2010). Similar regulatory constraints also exist in developing countries like India. The land and housing markets in Indian cities are heavily regulated with floor-area-ratio (FAR) restrictions, urban land ceilings, and stringent rent control laws.[2] Studies have indicated that these regulations impose significant building costs on developers (Bertaud and Brueckner, 2005; Brueckner and Sridhar, 2012; Gandhi *et al.*, 2021b). Therefore, durable housing supply elasticity estimates almost surely reflects land-use policy decisions.

Informal housing in India accounts for a large share of the housing stock and plays an important role in filling the supply gap left by the formal housing market. Niu *et al.* (2021) underscored the role of informal housing markets in reducing urbanization costs in Chinese cities by providing low-income migrants with cheaper housing. In India, roughly 17% of urban households live in slums or non-durable houses. Hence, a non-durable housing supply elasticity estimate is important for understanding supply dynamics in India.

We make three contributions to the existing literature on housing supply. First, we

[2]The Urban Land (Ceiling and Regulation) Act of 1976 required firms and individuals to sell vacant land beyond a certain size to the government at low prices (Sridhar, 2010).

provide a policy-relevant estimate of durable housing supply elasticity in the context of India. Second, to the best of our knowledge, this is the first paper that provides a non-durable housing supply elasticity estimate in a developing country. The closest attempt at estimating a non-durable housing supply elasticity has been made by Niu *et al.* (2021). However, they calculate a proxy for informal housing elasticity using the share of village areas on the edges of cities over the urban built-up area. By contrast, we use direct observations for informal or non-durable housing. Finally, to the best of our knowledge, this is the first paper to estimate the supply elasticity of *vacant* houses.

The rest of the paper is organized as follows. In section 2 we describe the data used for analysis and present some stylized facts about housing and migration in India in section 3. Section 4 provides a theoretical discussion of the Rosen-Roback spatial equilibrium setting applied in this paper to explain the mechanisms through which distant state shocks act as demand shifters in local housing markets. Section 5 presents the empirical implementation. We present the results and robustness checks in section 6 and conclude in section 7.

2 Data

Multiple publicly available sources of data are used for the analysis in this paper. We use the National Sample Survey Organization (NSS) datasets to compute average rent and consumption values. For data on migration, urban population, urban households, urban household size, number of rooms in an urban house, urban housing units, and urban surface area, we use data from the Census of India.[3] Rainfall data published by the India Meteorological Department (IMD) is taken from the Open Government Data platform. We also obtain consumer price index data from the Labor Bureau of India. We construct a panel dataset corresponding to Census years 2001 and 2011. In this section, we describe the source of data for each type of variable used in the analysis. We also discuss the process of district boundary matching across years to get consistent geographical boundaries for hypothetical district-regions across years.

2.1 Migration

Data on migration in India are sparse. The two most comprehensive sources of migration data are the Census D-series tables and the household-level survey on employment and migration conducted by the NSS during 2007-08. However, since the empirical framework

[3]All Census tables used in this paper are publicly available and can be downloaded from the following weblink: https://censusindia.gov.in/DigitalLibrary/Tables.aspx

used in this paper relies on aggregated data, migration variables from the NSS survey are not applicable here. Instead, we obtain inter-state aggregate migration figures based on the D-3 tables obtained for Census years 2001 and 2011. Using Census migration data has a couple of caveats. First, the empirical framework requires data on inter-regional migration flows, which is not available at finer geographic scales like districts.[4] We use state-level migration flows as an imperfect substitute for inter-district flows. And second, the Census migration tables do not identify heterogeneous movement patterns like seasonal and temporary migration, which limits the scope of analysis. However, since Census is the only publicly available source of aggregated migration data, we resort to this as the best possible option.

The key details in the D-3 tables from Census 2001 and 2011 include the number of individuals who moved into a given state by gender, reason for movement, the time at which individuals moved (less than a year ago, 1-4 years ago, 5-9 years ago, etc.), and the state of last residence. The number of in-migrants into a state from another state is disaggregated into the number of individuals who moved into urban and rural areas. These details allow us to construct a matrix of (i, j) state pairs with in- and out-migration flows from j to i and from i to j. We aggregate all individuals who moved in the decade leading up to the respective Census years. Table 1 presents the summary statistics for urban in-migrants. Note that the summary statistics for urban out-migration would be the same since urban out-migration at state i is urban in-migration at j and vice versa.

2.2 District Boundary Matching

We use state-level data for migration, rainfall, and highway variables and district-level data for all other variables. While some variables like urban population, households, and housing units are obtained from the Census, variables like rent and consumption are obtained from the NSS. The district boundaries vary by years and the sources of data. First, between 2001 and 2011, the number of districts in India reported in Census went up from 593 to 640.[5] And second, the NSS datasets from rounds 55, 58, 66, and 69 report district boundaries that are not up-to-date with the district boundaries reported by the latest Census year.

The inconsistency of district boundaries across time and data sources necessitates the construction of hypothetical district boundaries. This is done by combining all contiguous districts affected by changes in administrative boundaries across time. The approach for the construction of the hypothetical district boundaries is two-fold. First, we identify districts in all datasets and years that had any change in boundary in the time period of study.

[4]District-level datasets only report the state of last residence instead of the district of last residence. The latter would be necessary to construct inter-district migrant flow variables.

[5]10 districts in 2001 and 3 districts in 2001 did not have any urban areas as defined by Census.

Next, we carve out the largest geographical area comprising the contiguous districts that had an overlap or a change in boundary and create a new hypothetical district boundary. We construct 479 hypothetical district-regions that are consistent across time and datasets used in the analysis. All Census and NSS datasets are mapped to these 479 hypothetical district boundaries.[6] Note that owing to missing rent and income data for some districts, part of the analysis uses only a subset of these 479 hypothetical districts.

2.3 Population, Households, Residential Houses, and Surface Area

We use the Primary Census Abstract (PCA) tables, the housing H-series, and the household HH-series tables from the Census for a number of variables, including urban population, urban housing units, urban households, and the urban surface area. We use the same tables for obtaining both state-level and district-level variables. The PCA tables are used to obtain the urban population and the number of urban households in both 2001 and 2011. For the urban surface area, we use the PCA table for 2011. We calculate the urban surface area for a state or a district from the town-level surface area given by the town directory tables in 2001.

The Census defined three categories of residential houses in table H-4 in 2001. Permanent residential houses are made of both durable walls and roofs. Temporary residential houses are made of both non-durable walls and roofs. And, semi-permanent residential houses are made of either a non-durable wall and a durable roof or a durable wall and a non-durable roof.[7] In 2011, Census reported the number of residential houses by the material of roof and walls in the H-4 table. Based on the definition of permanent, temporary, and semi-permanent residential houses in 2001, we construct the number of permanent, temporary, and semi-permanent houses for 2011 as well.[8] For our analysis, we make two changes to Census's definitions. First, we redefine temporary to include both semi-permanent and temporary. This new categorization makes it easier for us to interpret the supply elasticity estimates. And second, we use the terms "durable" and "non-durable" instead of "permanent" and "temporary" respectively. We feel that our terminology communicates the nature of these

[6] All files and documents related to the construction of consistent hypothetical district boundaries across time and datasets can be obtained from authors on request.

[7] Based on Census's definition in 2001, durable walls are made of G.I., metal, asbestos sheets, burnt bricks, stone, or concrete, and durable roofs are made of tiles, slates, G.I., metal, asbestos sheets, bricks, stone, or concrete. Non-durable walls are made of grass, thatch, bamboo, plastic, polythene, mud, unburnt brick, or wood, and non-durable roofs are made of grass, thatch, bamboo, wood, mud, plastic, or polythene.

[8] In 2011, the number of houses with walls built from stone were disaggregated into the number of stone walls that are packed with mortar and those that are not. The number of houses with roofs made of tiles were disaggregated into hand-made tiles and machine-made tiles. We add the newly formed categories to calculate the number of houses with stone walls and tile roofs, consistent with the 2001 categorization.

housing units more accurately than those used by the Census.

In 2001, Census provided the median number of rooms in a household at the state and district-level in table H-6. However, the 2011 Census only provides the number of households with zero, one, two, three, four, five, and six and more rooms in table HH-4. We calculate the approximate median number of rooms at the district-level by ordering the number of households in a district, in ascending order of the number of rooms occupied by households. The mean household size is taken from Census table HH-1 in both 2001 and 2011. And finally, the number of vacant residential houses is taken from Census table H-1 in both 2001 and 2011.

State-level summary statistics for urban population, urban households, and the urban surface area can be found in table 1. District-level summary statistics for urban population, urban households, urban non-durable, durable, and vacant residential houses, mean urban household size, the median number of rooms in an urban household, and the urban surface area are presented in table 2.

2.4 Rents and Consumption

The National Sample Survey Organization conducts household surveys every year. Each round of survey, spanning six months to a year, has a central theme that determines the type of questions asked in the surveys. One of the most common and frequent survey themes is consumption expenditure. The large-scale thick round consumption expenditure surveys are conducted over nationally representative samples of at least 100,000 households focusing on details related to expenditure on a comprehensive list of commodities.[9] A less frequent theme of survey is that of housing conditions, conducted once every five to ten years.[10] The housing conditions surveys are conducted over similarly large samples of around 100,000 households and cover details that include the material of dwelling occupied by households, whether households live in rented homes or not, and the amount of rent paid by renter households. We use the consumption expenditure round surveys to calculate district and state-level average monthly per capita expenditure of households. The housing conditions

[9]Commodities include food, condiments, education and health expenses, and expenditure on durables like footwear, clothing, bed, and repair and maintenance of homes. Between 1983-2012, there have been eight thick round consumption surveys — rounds 38, 43, 50, 55, 61, 64, 66, and 68 corresponding to 1983, 1987, July 1993-June 1994, July 1999-June 2000, July 2004-June 2005, July 2007-June 2008, and July 2011-June 2012. All other consumption surveys during intermediate years are thin rounds consisting of smaller samples of households.

[10]There have been four housing conditions surveys conducted by the NSS thus far. These are rounds 49, 58, 65, and 69, which correspond to January 1993-June 1993, July 2002-December 2002, July 2008-June 2009, July 2009-June 2010, and July 2012-December 2012.

round surveys are used to calculate the district-level average rent paid by renter households.[11]

There were no thick round consumption expenditure or housing conditions surveys at the time of Census enumeration during 2001 and 2011. Hence, we use the most comprehensive thick survey round of consumption and housing data, with the widest coverage, that is closest to the Census years 2001 and 2011. For housing conditions, we use round 58 (July 2002-December 2002) for 2001 and round 69 (July 2012-December 2012) for 2011. Both housing conditions round surveys were started six months after the end of the previous Census enumeration year. Since completion of housing construction projects takes a long time in Indian cities, it is unlikely that the number of housing units would have changed dramatically during that time period (Gandhi *et al.*, 2021b). We deflate the nominal rent values reported by renter households in the surveys using the industrial worker consumer price index (CPI) series, provided by the Labor Bureau of India, with base year 2001.[12] Then, we calculate the district-level mean of log real urban rents reported by surveyed renter households in urban areas of the district.

For consumption expenditure, we use round 55 (July 1999-June 2000) for 2001 and round 66 (July 2009-June 2010) for 2011. Both these surveys are completed six months before the following Census enumeration year. These are the only two thick round surveys that are closest to the corresponding Census years.[13] Since consumption is likely to be very volatile across years, there could be a bias induced by the measurement error in the lagged consumption variable. This is unlikely to be a concern since we intend to use decadal and cross-sectional variation in district and state-level consumption expenditure as a measure of income differences across time and space. The calculated log average of consumption expenditure of all surveyed households in a district or a state during a year will reduce the volatile component. We deflate the reported mean monthly per capita consumption expenditure values using the industrial worker consumer price index (CPI) series, provided by the Labor Bureau of India, with the base year 2001.

Summary statistics for the mean monthly inflation-adjusted rent and the mean monthly per capita consumption expenditure at the district-level can be found in table 2. Summary statistics for the state-level mean monthly per capita consumption expenditure are given in table 1.

[11] All NSS datasets can be downloaded after creating an account on the website of the Ministry of Statistics and Programme Implementation: http://microdata.gov.in/nada43/index.php/catalog/central/about

[12] Data on the industrial worker CPI series can be found on the Labor Bureau of India's website: http://www.labourbureau.gov.in/LBO_indexes.htm

[13] The NSS consumption expenditure survey round 68 (July 2011-June 2012) started six months after 2011 ended and hence is equally close to 2011 as round 66. However, we choose round 66 over round 68 for 2011 since round 66 predates the Census year like round 55, so that measurement error in both years can be attributed to a lag in the variable.

2.5 Rainfall and Highway

In the empirical analysis, we use two state-level exogenous events as instrumental variables. These are a rainfall shock and a highway upgrade program. The rainfall shock is measured as the number of months when rainfall was less than 80% of the long-term normal in a state in the decade leading up to a Census year. The highway upgrade variable is measured as a dummy variable that takes the value of one if the state in question was included in the National Highways Development Project Phase I, also known as the Golden Quadrilateral (GQ) highway upgrade program, after 2001. Fourteen states and union territories were recipients of the GQ program.[14] Further details on the GQ program can be found in Ghani *et al.* (2016).

For data on rainfall deviation, we use the publicly available data from the Open Government Data (OGD) portal of the central government of India.[15] This dataset is originally sourced from the IMD, and it reports the percentage deviation of rainfall from the long-term average on a monthly basis between 1901 and 2015. The original data from OGD provides rainfall departure percentages for each of the 36 meteorological subdivisions in India.[16] We map these meteorological subdivisions to state boundaries and recalculate the rainfall departure values at the state-level. Then, we compute the number of months, over the decade leading up to the Census years, when rainfall was less than 80% of the long-term normal. The IMD uses this cutoff to designate regions as rainfall deficient. Summary statistics for the number of months with absolute rainfall less than 80% of the long-term normal in the decade leading up to the Census years can be found in table 1.

3 Stylized Facts

There is a dearth of academic literature studying the relationship between migration and housing in India. In this section, we provide some key stylized facts about the urban housing markets in India. The discussion is primarily around the growth in quantity and quality of urban housing. We also explain the potential role of migration in causing demand shifts in urban housing. We end this section by making a case for inter-state migration as a demand shock for housing in urban India.

[14] The 12 states included in the GQ program were Andhra Pradesh, Bihar, Gujarat, Haryana, Jharkhand, Karnataka, Maharashtra, Odisha, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal. In addition, the union territories of Dadra & Nagar Haveli and Delhi were also part of the program.

[15] Please visit the link: https://data.gov.in/

[16] The meteorological subdivisions are constructed based on the commonality of weather conditions in a geographical region and are somewhat consistent with the boundaries of smaller states. Larger states consist of more than one and up to three meteorological subdivisions.

3.1 Housing in Urban India

India's urban population, second only to China, grew from roughly 286 million in 2001 to 377 million in 2011. Yet, there is very little academic literature on the urban housing markets of India. Demand-side and affordability studies were conducted almost two decades ago (Tiwari and Parikh, 1998; Tiwari *et al.*, 1999). Recent studies have focused on regulatory hurdles and litigations that impede new housing construction and rental housing supply in Indian cities (Gandhi *et al.*, 2021b,a). Here, we provide some broad stylized facts about urban housing in India.

There are roughly 78 million urban residential durable and non-durable housing units for the 79 million urban households in India, indicating a shortage of one million (see figure 1a).[17] However, the number of urban housing units grew substantially between 2001 and 2011. Figure 1b shows that the number of urban residential housing units grew by 50% during the decade 2001-2011. A large portion of this growth comes from the 61% increase in durable housing units. This is consistent with the significant growth in real estate and housing construction documented in the policy literature (Joshi *et al.*, 2006). The stunted growth of non-durable units suggests that slums are either being cleared or upgraded to durable units. Note also that the share of non-durable housing units in the overall housing stock fell from 21% in 2001 to 15% in 2011, consistent with the slower growth in non-durable housing as seen in figure 1b.

But is the relatively higher increase in durable urban housing units an indication of better provision of housing services over time? Figure 2a and figure 2b show that both durable and non-durable urban housing unit increases are associated with a marginal decrease in the corresponding per capita floor area of such units. This indicates that even with a growth in the number of housing units, the consumption of housing services as measured by the per capita floor area consumption might actually fall over time. This argument is strengthened by the fact that the mean per capita floor area in durable houses fell from 81 square feet in 2001 to 77 square feet in 2011 as seen in figure 3. The per capita floor area for non-durable houses increased from 48 square feet to 52 square feet during the same decade. These findings are consistent with the theoretical implications of clearing or upgrading slums discussed by Olsen (1969). Recent literature suggests that slums are upgraded and change over time in Indian cities (Rains *et al.*, 2019; Rains and Krishna, 2020). As slums are cleared or upgraded, market prices for slum housing increase which results in non-slum owners letting

[17] The total number of households in urban India is roughly 81 million of which 79 million are non-institutional. "A group of unrelated persons who live in an institution and take their meals from a common kitchen is called an Institutional Household. Examples of Institutional Households are boarding houses, messes, hostels, hotels, rescue homes, jails, ashrams, orphanages, etc." (Census of India, 2011).

their housing units depreciate until such units provide the same quantity of housing services as slums so that they can earn profits from the higher market prices for slum housing services. Thus, clearing or upgrading slums can lead to inferior quality of non-slum housing. We get back to this point in section 6.3 while discussing the negative supply elasticity of non-durable housing.

Another striking aspect of urban Indian housing markets is the large increase in the number of vacant houses. Between 2001 and 2011 the number of vacant units increased by 5 million or 83%. Gandhi *et al.* (2021a) show that stringent rent control laws that favor tenants reduce the incentives for owners to lease their homes to tenants, thereby increasing the number of vacant houses. We return to this again in section 6.3 when we discuss the elasticity of supply of vacant houses.

3.2 Internal Migration

A large body of literature examines the historically low rates of migration in India. Low mobility has been attributed to rural caste-based informal insurance networks and lower provision of public goods due to fiscal federalism (Bhavnani and Lacina, 2017; Munshi and Rosenzweig, 2016). In fact, the most common migration in India is that of women moving from their birthplaces to their spouses' homes after marriage (Rosenzweig and Stark, 1989). However, recent literature suggests that this pattern is changing, as more Indians are moving internally as a result of the IT boom of the late 1990s and the early 2000s (Ghose, 2019).

Figure 4a and figure 4b show that the number of inter-city and rural-urban migrants in Indian cities, moving in the decade leading up to Census years, increased by almost 70% from 37 million in 2001 to 62 million in 2011. Such migrants moving in the decade leading up to Census years constituted a sixth of the urban population in India in 2011, up from one-eighth in 2001 (see figure 5a). The contribution of rural-urban migrants to urban population also went up from 21% in 2001 to 33% in 2011. Most of this growth in internal migration in India owes primarily to people moving within state boundaries. However, the share of rural-urban migrants in urban population growth is still substantially lower than it is in other developing countries. For instance, rural-urban migrants' contribution to China's urbanization during 1978-1999 was 75% (Zhang and Shunfeng, 2003). Although, this does not paint a complete picture of mobility and urbanization in India because a large number of Indians move seasonally between one to six months for employment before going back to their homes (Imbert and Papp, 2020). More research on the heterogeneity of migrants will have to be conducted to fully ascertain the role of migrants in the urbanization of India.

One concern about using inter-state migration shocks as demand shifters to identify housing supply is that such migrants constitute only a fraction of the urban population. Figure 5a indicates that only 4% of the urban population in India in 2011 were inter-state migrants. In fact, for every person moving across state boundaries, there are three others who move within states. These low rates of inter-state mobility are consistent with the literature on the role of various state-level policies such as quotas for state residents in the education system that exclude individuals from other states (Ghose, 2019; Kone *et al.*, 2018). Despite these hindrances, inter-state in-migration to cities grew at a faster pace than intra-state migration (see figure 4b). Moreover, inter-state migrants' contribution to urban growth almost doubled between 2001 and 2011. These figures indicate that while it is true that inter-state migrants occupy a significantly lower share of urban housing, the growth over the 2001-2011 decade is large enough to cause a shift in the housing demand from the decade of 1991-2001. This argument is further strengthened by the significant and positive relationship between the number of in-migrants and the number of durable and non-durable housing units in urban areas, seen in figure 6a and figure 6b. In section 6.1 and section 6.2, we discuss several regression results that indicate the strength of inter-state migration-inducing shocks in explaining local housing demand.

4 Theoretical Framework

We use the Rosen-Roback spatial equilibrium framework to analyze the effect of distant region shocks on inter-regional mobility and local housing demand. A shock that affects rents and incomes in a distant region induces spatial disequlibrium, thereby inducing mobility. Such mobility affects local housing demand if net inward mobility to the local region is non-zero. Therefore, distant region shocks that affect rents and incomes in the distant region act as demand shifters that can be used to estimate local housing supply in the local markets. In this section, we provide an analytical discussion of these effects.

4.1 Spatial Equilibrium

The spatial equilibrium framework is derived from Roback (1982). Consider an economy with a local region i where we are interested in estimating the housing supply elasticity, and a distant region j that have exogenous shocks to its economy. The number of individuals occupying regions i and j are n_i and n_j respectively. We assume that each individual is equivalent to a household in either region. In both locations, individuals earn w and derive utility from housing services h, a numeraire good c, and location-specific amenities a. Individuals can only transact h and c in the market. Amenities a are exogenously given in a location at any given point in time. The market-clearing rent for housing services is r. The user-cost model relates r to the market-clearing house price p through the equation $r = p(\iota + \tau + \lambda + \pi)$. Here, ι is the cost of capital, τ is the property tax rate, λ is the rate of depreciation, and π is the rate of expected appreciation (Poterba, 1984). The fact that market-clearing rent for housing services is an appropriate measure of market-clearing price for housing as a composite commodity is well established in the literature (Brueckner *et al.*, 1987; Mills, 1967).

The representative individual's utility maximization problem at i can be written as follows:

$$\max_{h_i, c_i} U_i(h_i, c_i) + a_i \quad \text{s.t.} \quad h_i r_i + c_i = w_i \tag{1}$$

Here, $U(.)$ is a strictly quasiconcave utility function such that equation (1) results in an interior solution. The resulting demand for housing services at i is $h_i^d(r_i, w_i)$. Hence, the aggregate demand for housing services at i can be written as follows:

$$H_i^D = n_i h_i^d(r_i, w_i) \quad \text{where} \quad h_i^d(r_i, w_i) > 0 \tag{2}$$

As long as the number of households and the total population at i are monotonically related, relaxing the assumption that each individual in the economy is equivalent to a household does not alter the model mechanisms. Hence, through the rest of this section, we assume that each individual constitutes a household. In the empirical implementation, we analyze the implications of spatial disequilibrium on both the total population and households at i.

The implied indirect utility obtained by the representative individual at i is $V_i(r_i, w_i, a_i)$. At equilibrium, the values of r and w adjust such that, given the location-specific amenities in every region, the indirect utility is equal across both regions i and j. The spatial equilibrium is characterized as follows:

$$V_i(r_i, w_i, a_i) = V_j(r_j, w_j, a_j) = \bar{V} \tag{3}$$

At this equilibrium, there are no gains to mobility between i and j.

4.2 Spatial Disequilibrium, Mobility, and Local Housing Demand

Now consider a shock z_j at the distant region j that does not affect amenities a_j but changes rent r_j or income w_j, or both, thus changing the utility V_j of individuals at j.[18] The shock z

[18] The shock z can also potentially affect amenities a. The ultimate impact of z on mobility m between i and j will be determined by the implicit function of mobility $m(V(r(z), w(z), a(z)))$ in z. Hence, assuming

could be a negative shock like a drought or a positive shock like a highway upgrade program. Because z_j affects rent r_j and income w_j, it follows that $V_j(r_j(z_j), w_j(z_j), a_j)$ is an implicit function of z_j. Hence, in response to z_j, we have a state of spatial disequilibrium as follows:

$$\tilde{V}_j = V_j(z_j) = V_j(r_j(z_j), w_j(z_j), a_j) \neq \bar{V} = V_i \tag{4}$$

Since there are gains to mobility because of the difference in V_i and V_j, the shock z_j will induce mobility between j and i until r and w adjust in both i and j, so that $\tilde{V}_j = \tilde{V}_i = \tilde{V}$. In other words, a shock affecting rents and incomes at a distant region j induces movement between the distant and the local regions so that rents and incomes change in both locations until spatial equilibrium is restored and there are no gains to moving. This proposition is consistent with past literature on the effects of regional labor and housing market shocks on inter-regional mobility in the United States (Molloy *et al.*, 2011; Saks and Wozniak, 2011).

We characterize mobility m between regions i and j as the vector (m_{ji}, m_{ij}). m_{ji} represents the number of individuals moving from j to i and m_{ij} denotes the number of individuals moving from i to j. In other words, m_{ji} represents in-migration from the distant region j to the local region i and m_{ij} represents out-migration from local region i to the distant region j. This framework is consistent with the bi-directional movement of individuals across regions observed in the data. Spatial equilibrium implies that the net movement between two regions in equilibrium should be equal to zero. A disequilibrium induced by a shock will cause net migration into the region where utility is higher.

Suppose that $m_{ij}(.)$ and $m_{ji}(.)$ are two distinct functions of the indirect utilities V_i and V_j. At the spatial equilibrium, we have $m_{ji}(\bar{V}, \bar{V}) = m_{ij}(\bar{V}, \bar{V})$ implying that net movement between i and j is zero. Now, in response to the shock z_j the indirect utility at the distant state j changes from $\bar{V}$ to $V_j(z_j)$. The resulting migration functions can be written as follows:

$$m_{ij}(\bar{V}, V_j(z_j)) = m_{ij}(z_j); \quad m_{ji}(\bar{V}, V_j(z_j)) = m_{ji}(z_j) \tag{5}$$

Equation (5) implies that both in- and out-migration are implicit functions of the shock z_j. We further make the following assumptions:

$$m'_{ij}(z_j),\ m'_{ji}(z_j) \geq 0 \tag{6}$$

$$\text{If } m'_{ij}(z_j) = 0 \text{ then, } m'_{ji}(z_j) > 0 \tag{7}$$

$$\text{If } m'_{ji}(z_j) = 0 \text{ then, } m'_{ij}(z_j) > 0 \tag{8}$$

that a remains exogenous in the shock system does not change much. Note however, that while mobility will respond to $r(z)$, $w(z)$, and $a(z)$, and r and w will change in response to such mobility, a will not. In other words, in this shock system $a(z)$ changes only in response to z.

Equation (6) implies that in- and out-migration are weakly increasing functions of the shock z_j. The additional assumptions given by equation (7) and equation (8) imply that the shock z_j should affect movement in at least one direction between i and j. These assumptions restrict the universe of shocks z to only those that have a non-zero effect on net mobility across regions.

Both in-migration m_{ji} into j from i and out-migration m_{ij} from i to j will affect the number of individuals n_i living in i through the function $n_i(m_{ji}, m_{ij})$. Therefore, the effect of the distant region shock z_j on local population n_i can be written as follows:

$$\frac{dn_i}{dz_j} = \frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) + \frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) \tag{9}$$

Equation (9) implies that the net effect of the distant region shock on local population is the sum of shock-induced in- and out-migration effects on local population. We make the two following assumptions on the effect of in- and out-migration on population at i:

$$\frac{\partial n_i}{\partial m_{ji}(z_j)} \geq 0 \quad \text{and} \quad \frac{\partial n_i}{\partial m_{ij}(z_j)} \leq 0 \tag{10}$$

Equation (10) implies that the number of individuals n_i at i weakly increases in response to in-migration and weakly decreases in response to out-migration. The weak inequality follows from the fact that the natural rate of growth component in population changes is a major factor and can act as a countering force to both in- and out-migration effects on the local population. Since we do not explicitly model the natural rate of growth component in dn_i, we allow for the possibility of population changes to be independent of migration.

Proposition 1. *Under the assumptions given by equations (6) to (8) and equation (10),* $\frac{dH_i^D}{dz_j} \lesseqgtr 0$ *if and only if* $\left|\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j)\right| \lesseqgtr \left|\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j)\right|$

Proposition 1 indicates that the aggregate demand for local housing services H_i^d respond to migration-inducing shocks at the distant region j. The direction of change in aggregate demand for housing services at i depends on the relative magnitude of the in-migration and out-migration effects on local population resulting from the shock z_j. To see this, let us first write the effect of the shock z_j on the aggregate demand for local housing services H_i^d, as follows:

$$\frac{dH_i^D}{dz_j} = \frac{\partial H_i^D}{\partial n_i} \frac{dn_i}{dz_j} h_i^d(r_i, w_i) = \frac{\partial H_i^D}{\partial n_i} \left(\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) + \frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) \right) h_i^d(r_i, w_i) \tag{11}$$

The last expression in equation (11) is derived by substituting equation (9) after differentiating the aggregate demand H_i^D given by equation (2) with respect to z_j.

The fact that $\frac{dH_i^D}{dz_j} \lesseqgtr 0$ implies $\left|\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j)\right| \lesseqgtr \left|\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j)\right|$ directly follows from equation (11) and the inequality $\frac{\partial H_i^D}{\partial n_i} > 0$ derived from equation (2). Now, to see the if condition, note first that $\frac{\partial n_i}{\partial m_{ji}}$, $m'_{ji}(z_j)$ and $m'_{ij}(z_j)$ are all weakly positive and $\frac{\partial n_i}{\partial m_{ij}}$ is weakly negative. Hence, we have $\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) \geq 0$ and $\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) \leq 0$. If $\left|\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j)\right| = \left|\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j)\right|$, then $\frac{dH_i^D}{dz_j} = 0$ trivially follows from equation (11). Now, if $\left|\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j)\right| \neq \left|\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j)\right|$, then there are three possibilities. First, we can have $\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) > 0$ and $\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) = 0$, in which case equation (11) implies $\frac{dH_i^D}{dz_j} > 0$. The second possibility is where $\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) = 0$ and $\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) < 0$, in which case $\frac{dH_i^D}{dz_j} < 0$ follows from equation (11). And finally, we can have $\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j) > 0$ and $\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j) < 0$, in which case we have $\frac{dH_i^D}{dz_j} > 0$ if $\left|\frac{\partial n_i}{\partial m_{ji}} m'_{ji}(z_j)\right| > \left|\frac{\partial n_i}{\partial m_{ij}} m'_{ij}(z_j)\right|$.

Proposition 1 implies that a distant region migration-inducing shock acts as a demand shifter in local housing markets. The driving mechanisms behind the distant region shock effect on local housing demand can be described as follows. First, a shock at the distant point affects rents and incomes in that region. This, in turn, changes the indirect utility in the distant region, thereby inducing a state of spatial disequilibrium in the economy. The resulting difference in utilities across the two regions implies gains to mobility. Individuals move across regions. This movement causes a change in the local population and households, thus affecting local housing demand.

4.3 Local Housing Supply

Now, let us consider the total housing stock H_i^S in the region i that is supplied through a competitive market. H_i^S is a function of r_i. At the market equilibrium, we have $H_i^S = H_i^D(r_i, w_i, n_i)$. In other words, the market equilibrium implies that the housing supply equals the aggregate quantity of housing services demanded within a region. Let us assume that the supply function is log-linear. Then the reduced form for the inverse supply function at i can be written as follows:[19]

$$\log(r_i) = \chi + \frac{1}{\eta_i} \log(H_i^S) \tag{12}$$

where the housing supply elasticity at i is η_i. Since housing supply is never perfectly elastic, η_i is a finite real number greater than zero.[20]

Estimating η_i in equation (12) presents a classic endogeneity problem since we only

[19]Throughout the paper, the log function is used to denote the natural log of its argument.

[20]See Green *et al.* (2005) for a discussion on imperfect housing supply elasticities and the various reasons for why that is the case in the context of a monocentric city model.

observe market equilibrium values of r_i and H_i^S. Hence, we need exogenous demand shifters that can trace the slope $1/\eta_i$ of the inverse supply curve. Proposition 1 shows that exogenous shocks z_j incident upon a distant region can act as a demand shifter at i if the shock z_j induces net non-zero mobility between i and j. We can write the reduced form effect of z_j on the aggregate demand for housing services as follows:

$$\log(H_i^D) = \kappa_0 + \kappa_1 z_j \tag{13}$$

Proposition 1 implies that κ_1 could be either negative or positive, and its sign depends on the relative magnitude of the in- and out-migration effects of the shock z_j. The predicted $\log(H_i^D)$ obtained after estimating the parameters κ_0 and κ_1 in equation (13) is an exogenous demand shock which can be substituted in equation (12) to estimate η_i. If the shock-induced migration affects construction wages, then κ_1 might include supply-side factors as well, a concern we address in the empirical section.

We estimate supply elasticities for non-durable houses, durable houses, and vacant houses. Since these categories of housing represent different markets, their slopes will be different. The reduced form supply equations for each of these three separate housing markets are given as follows:

$$\log(r_i^N) = \chi^N + \frac{1}{\eta_i^N}\log(H_i^N) \tag{14}$$

$$\log(r_i^Q) = \chi^Q + \frac{1}{\eta_i^Q}\log(H_i^Q) \tag{15}$$

$$\log(r_i) = \chi^V + \frac{1}{\eta_i^V}\log(H_i^V) \tag{16}$$

where H_i^N, H_i^Q, and H_i^V are the supply of non-durable houses, durable houses, and vacant houses, respectively; r_i^N and r_i^Q are the rents for non-durable houses and durable houses. Since we do not know the number of vacant durable and the number of vacant non-durable houses, we hypothesize that the supply of overall number of vacant houses H_i^V is a function of the prevailing overall market rent r_i. The demand shock given by equation (13) acts as a demand shifter for all of these housing segments with different magnitudes of effects. Hence, we use equation (13)'s demand shifter to estimate the inverse elasticities η_i^N, η_i^Q, and η_i^V.

5 Empirical Implementation

The theoretical framework discussed in section 4 explains that distant region shocks affect local population, and hence, local housing demand. The driving channel of effect is the migration flow of individuals across regions. In this section, we first discuss the empirical framework for estimating the effect of a distant state shock on inter-state mobility and the resulting effects on local urban population and household growth. Next, we provide the estimating equations to analyze the effect of the distant state shocks on local urban residential housing units. These two estimation exercises are meant to test whether Proposition 1 holds true and provide empirical evidence for the mechanisms through which the distant state shocks act as demand shifters in the local housing market. Finally, we provide the housing supply estimation model using the distant state shocks as local urban housing demand shifters. We end this section with a discussion on the instruments.

5.1 Distant Shocks, Migration, and Urban Population Growth

In this section, we discuss the estimating equations for the effect of an exogenous shock z_j at a distant region j on migration inflows from j to the local region i and outflows from i to j. Then we present the empirical model to estimate the effect of such migration response on population growth in region i. Here, the urban area in each Indian state represents the local region i, and both rural and urban areas in a different state represent the distant region j. For this estimation exercise, we propose a panel instrumental variable framework. The first-stage regressions provide the predicted growth of migration flows due to the distant region shocks. The second-stage regressions estimate the effect of such predicted growth of migration flows on population and household growth in region i.

To implement the empirical model, we use an unbalanced panel of 1,108 and 1,125 state and union territory pairs, using data from the Census of India, for 2001 and 2011 respectively.[21] We estimate the effect of shocks incident upon urban and rural areas in distant states on in-migration from distant states to the local state and out-migration from the local state to the distant states. The two Census years are indexed by $t = \{2001, 2011\}$. To empirically model the effect of an exogenous shock z_{jt} incident at j in Census year t on migration inflows m_{jit} from j to i, and outflows m_{ijt} from i to j, we estimate the following pair of first-stage equations:

$$\log(m_{jit}) = \alpha_0 + \alpha_1 z_{jt} + \alpha_2 x_{it} + \ddot{\theta_{ij}^{m}} + \ddot{\omega_t^{m}} + \varphi \tag{17}$$

$$\log(m_{ijt}) = \gamma_0 + \gamma_1 z_{jt} + \gamma_2 x_{it} + \tilde{\theta_{ij}^{m}} + \tilde{\omega_t^{m}} + \varrho \tag{18}$$

where m_{jit} and m_{ijt} are the number of in-migrants and out-migrants at i during the decade leading up to the Census year t; $\ddot{\theta_{ij}^{m}}$ and $\tilde{\theta_{ij}^{m}}$ represent fixed effects for (i, j) pairs of local states

[21]Hereon, we use the terms states and union territories interchangeably.

and distant states; $\ddot{\omega}_t^m$ and $\tilde{\omega}_t^m$ represent time fixed effects; x_{it} is a vector of controls that include the log of mean monthly per capita consumption as a proxy for income, the urban surface area, and the urban surface area squared at i; the vector $z_{jt} = \{R_{jt}, G_{jt}\}$ consists of two exogenous variables used as instruments. R_{jt} represents the number of months in the last decade with rainfall levels less than 80% of the long-term normal at j. G_{jt} is a dummy variable that represents a state's inclusion in the National Highways Development Project I or the Golden Quadrilateral (GQ) highway upgrade program after 2001. 14 states were recipient of the GQ highway upgrade program.

After estimating the parameters in equation (17) and equation (18), we predict the in- and out-migration flows $\widehat{\log(m_{jit})}$ and $\widehat{\log(m_{ijt})}$ respectively. Now, suppose that the total number of urban households at i is given by g_i. We estimate the effect of shock-induced in- and out-migration flows $\widehat{\log(m_{jit})}$ and $\widehat{\log(m_{ijt})}$ on urban population and urban household growth in the state i. The second-stage equations can therefore be written as follows:

$$\log(n_{it}) = \delta_0 + \delta_1 \widehat{\log(m_{jit})} + \delta_2 \widehat{\log(m_{ijt})} + \delta_3 x_{it} + \theta_{ij}^n + \omega_t^n + \vartheta \tag{19}$$

$$\log(g_{it}) = \sigma_0 + \sigma_1 \widehat{\log(m_{jit})} + \sigma_2 \widehat{\log(m_{ijt})} + \sigma_3 x_{it} + \theta_{ij}^g + \omega_t^g + \varepsilon \tag{20}$$

where θ_{ij}^n and θ_{ij}^g represent fixed effects for (i, j) pairs of local and distant states; ω_t^n and ω_t^g represent time fixed effects. The control vector x_{it} is the same as in equation (17) and equation (18).

Note that the regressions given by equations (17) to (20) do not include any control variables for the distant state j. The motivation behind including the urban surface area of i as a control variable is that several settlements are reclassified and declassified as Census towns each Census year. This changes the urban area across Census years.[22] Since we have aggregated data for the urban area in a region, controlling for the urban area allows us to mitigate any effect on migration, urban population, and urban households that can be attributed to the change in the urban area itself. We also control for urban area squared to account for the non-linear relationship between the urban area and the outcome variables.

Even though we include the mean per capita consumption as a proxy for income at i as a control, we exclude consumption at j from the vector of controls. We will explain the reason behind including consumption at i as a control and then discuss the reasons for excluding consumption at j. But first, note that equation (17) and equation (18) do not represent conventional gravity models. Most gravity models in the literature are used to study the effect of origin and destination factors in determining mobility. Besides the geographical

[22] Census towns are areas without an urban administrative body, but with urban-like features with at least 5,000 people, a population density of at least 400 persons per sq. Km. and with at least 75% of the male workforce employed in non-agricultural activities.

distance between two regions, several economic and demographic variables measured at the origin (push factors) and the destination (pull factors) are included in these models. The panel framework discussed above controls for many time constant heterogeneous variables like geographical and linguistic distance, preexisting social networks between regions, and climate, that are included in conventional gravity models.

The role of consumption controlled at i is different from that of gravity models. We posit that a state of spatial disequilibrium induced by distant shocks changes mobility patterns between regions. Such mobility affects the population at i. While section 4.2 explicitly models the effect of changing population on demand for housing at i, it does not say anything about labor market effects of mobility in region i. If the supply of laborers at i changes in response to the shock-induced mobility, then we would expect the labor market equilibrium at i to reflect that. The resulting change in incomes will also have an effect on housing demand at i. To capture this general equilibrium effect on housing demand through the labor market equilibrium changes resulting from shock-induced migration, we control for consumption as a proxy for income at i.

There could be a housing supply effect of migration as well that is incident through labor demand in the construction industry. In other words, if employment and incomes in the construction industry change in response to migration, then the shock-induced migration can also have housing supply effects. This can be a threat to the identification strategy, particularly because a large number of migrant workers in India move for construction employment. However, such migrants are more likely to be seasonal workers moving for one to six months in a given year. We address this concern in robustness checks, where we present estimates for the effect of distant state shocks on long-term migrants who moved at least a year before the Census is conducted.

There are two reasons for not controlling for consumption at j. First, equation (5) indicates that mobility is an implicit function of the shock z_j. This is because z_j affects the indirect utility V_j by changing rent r_j and income w_j, thereby inducing a state of spatial disequilibrium that spurs mobility between i and j. Hence, any effect of the shock on mobility that is incident through income and rent at j will be captured by the shock itself. Second, if we control for consumption at j in equation (19) and equation (20), we are essentially saying that changes in consumption in a distant state j can affect the urban population in state i through a channel that is independent of migration. This is not plausible in a spatial equilibrium framework. However, it is still possible that a national level shock results in a heterogeneous income effect on states. In such a scenario, not controlling for consumption at j would pose a threat to the identification strategy given by equations (17) to (20) if the distant state shock z_{jt} is correlated with such an income shock. If z_{jt} is correlated with the error term in equations (17) to (20), then the coefficient estimates without controlling for consumption at j would be biased. To see if this is true, we run the same regressions given by equations (17) to (20) and find that the coefficient estimates are similar to those obtained without controlling for consumption at j. These additional results can be seen in appendix tables A.7 to A.14.

5.2 Urban Population Growth and Housing Demand

In the previous section 5.1, we discussed the estimating equations for analyzing the effect of exogenous migration on urban population and household growth. In this section, we provide the empirical model to estimate the effect of exogenous urban population and urban household growth on urban housing demand. Here, exogenous urban population and urban household growth are determined by migration-inducing shocks in a distant region. As in section 5.1, we use a panel instrumental variable model for estimating the effect of exogenous urban population and household growth on housing demand. Here, i consists of urban areas in a district of the local state, and j represents both rural and urban areas in a distant state. We use an unbalanced panel of (i, j) district-state pairs for the two Census years 2001 and 2011. For 2001, we have 220 × 34 district-state pairs, and for 2011 we have 208 × 34 pairs.

First, we estimate the effect of distant shocks z_{jt}, incident upon j in Census year t on urban population and household growth at i. The first-stage equations can therefore be written as follows:

$$\log(n_{it}) = \lambda_0 + \lambda_1 z_{jt} + \lambda_2 y_{it} + \tilde{\theta_{ij}^{n}} + \tilde{\omega_t^{n}} + \mu \tag{21}$$

$$\log(g_{it}) = \tau_0 + \tau_1 z_{jt} + \tau_2 y_{it} + \tilde{\theta_{ij}^{g}} + \tilde{\omega_t^{g}} + \nu \tag{22}$$

where z_{jt} is the same as in equation (17) and equation (18); $\tilde{\theta_{ij}^{n}}$ and $\tilde{\theta_{ij}^{g}}$ represent fixed effects for (i, j) pairs of local districts and distant states; $\tilde{\omega_t^{n}}$ and $\tilde{\omega_t^{g}}$ represent time fixed effects. In addition to the variables included in the control vector x_{it}, y_{it} also includes the mean urban household size and the median number of rooms in a household at i.

From equation (21) and equation (22) we first estimate the predicted values of urban population growth $\widehat{\log(n_{it})}$ and urban household growth $\widehat{\log(g_{it})}$. Suppose that the equilibrium number of housing units in district i in Census year t is given by H_{it}. Then, to estimate the effect of $\widehat{n_{it}}$ and $\widehat{g_{it}}$ on H_{it} at district i, we have the following second-stage equations:

$$\log(H_{it}) = \beta_0 + \beta_1 \widehat{\log(n_{it})} + \beta_2 y_{it} + \tilde{\theta_{ij}^{H}} + \tilde{\omega_t^{H}} + \rho \tag{23}$$

$$\log(H_{it}) = \pi_0 + \pi_1 \widehat{\log(g_{it})} + \pi_2 y_{it} + \ddot{\theta_{ij}^{H}} + \ddot{\omega_t^{H}} + \phi \tag{24}$$

where $\tilde{\theta}_{ij}^{H}$ and $\ddot{\theta}_{ij}^{H}$ represent fixed effects for (i, j) pairs of local districts and distant states; $\tilde{\omega}_t^{H}$ and $\ddot{\omega}_t^{H}$ represent time fixed effects; y_{it} is as defined in equation (21) and equation (22). The coefficients β_1 and π_1 provide an estimate for the effect of migration-inducing distant exogenous shock-driven urban population growth on urban housing demand. We estimate three separate coefficient pairs of β_1 and π_1, one each for non-durable, durable, and vacant residential housing units.

5.3 Demand Shifters and Housing Supply Estimation

In this section, we propose an empirical framework to estimate the inverse supply elasticity of urban housing at the local region i. As defined in section 5.2, i consists of urban areas in a district of the local state, and j represents both rural and urban areas in a distant state. Estimating the slope of an inverse supply curve from equilibrium price and quantities would provide inconsistent estimates because of endogeneity. To address this, we need demand shifters. Proposition 1 in section 4 shows that shocks at a distant state j, that affect rent and income at j, act as demand shifters for housing in district i. As in section 5.2, we use an unbalanced panel of (i, j) district-state pairs for the two Census years 2001 and 2011 for estimating the inverse supply function. For 2001, we have 220×34 district-state pairs, and for 2011 we have 208×34 pairs.

We begin by estimating the effect of exogenous shocks z_{jt}, incident at j on the log number of housing units H_{it} in i, in Census year t. The first-stage equation for the demand shifter can be written as follows:

$$\log(H_{it}) = \Psi_0 + \Psi_1 z_{jt} + \Psi_2 x_{it} + \theta_{ij}^{H} + \omega_t^{H} + \xi \tag{25}$$

where H_{it} is as defined in equation (23) and equation (24); x_{it} is as defined in equations (17) to (20); z_{jt} is the same as in equation (17) and equation (18); θ_{ij}^{H} represent a vector of fixed effects for (i, j) pairs of local districts and distant states; ω_t^{H} represent time fixed effects. We run three different models using equation (25), one each for non-durable, durable, and vacant houses. Using the notations introduced in section 4, we denote H_{it}^{N}, H_{it}^{Q}, and H_{it}^{V} to represent the number of non-durable houses, durable houses, and vacant houses in district i in Census year t respectively. Substituting the parameters estimated for each housing type in the three versions of equation (25), we get the predicted values $\widehat{\log(H_{it}^{N})}$, $\widehat{\log(H_{it}^{Q})}$, and $\widehat{\log(H_{it}^{V})}$ for non-durable, durable, and vacant houses respectively.

We use the predicted $\widehat{\log(H_{it}^{N})}$, $\widehat{\log(H_{it}^{Q})}$, and $\widehat{\log(H_{it}^{V})}$ as demand shocks in the empirical counterpart of the reduced form inverse supply functions given by equations (14) to (16), and rewrite them as follows:

$$\log(r_i^N) = \Sigma_0 + \Sigma_1 \widehat{\log(H_{it}^N)} + \Sigma_2 x_{it} + \theta_{ij}^N + \omega_t^N + \zeta \tag{26}$$

$$\log(r_i^Q) = \Theta_0 + \Theta_1 \widehat{\log(H_{it}^Q)} + \Theta_2 x_{it} + \theta_{ij}^Q + \omega_t^Q + \epsilon \tag{27}$$

$$\log(r_i^V) = \Phi_0 + \Phi_1 \widehat{\log(H_{it}^V)} + \Phi_2 x_{it} + \theta_{ij}^V + \omega_t^V + \varepsilon \tag{28}$$

where $\log(r_i^N)$, $\log(r_i^Q)$, and $\log(r_i)$ are the mean of log rents for non-durable houses, durable houses, and for the entire sample of non-durable and durable houses combined; θ_{ij}^N, θ_{ij}^Q, and θ_{ij}^V represent fixed effects for (i, j) pairs of local districts and distant states; ω_t^N, ω_t^Q, and ω_t^V represent time fixed effects. The demand shocks $\widehat{\log(H_{it}^N)}$, $\widehat{\log(H_{it}^Q)}$, and $\widehat{\log(H_{it}^V)}$ traces the slopes Σ_1, Θ_1, and Φ_1 of the inverse supply functions for non-durable, durable, and vacant houses respectively. In other words, from equations (14) to (16) and equations (26) to (28) we have,

$$\eta^N = \frac{1}{\Sigma_1} \tag{29}$$

$$\eta^Q = \frac{1}{\Theta_1} \tag{30}$$

$$\eta^V = \frac{1}{\Phi_1} \tag{31}$$

where η^N, η^Q, and η^V are the supply elasticities of non-durable houses, durable houses, and vacant houses respectively.

There are two things to note here. First, contrary to existing literature on housing supply estimation, we do not control for construction cost in the empirical inverse supply equations (26) to (28). This is because we do not have any data on the construction cost at the district-level in India.[23] Second, a possible concern may arise owing to the various rent control laws present in Indian states that prohibit landlords from increasing rents (Harari, 2020). However, this is unlikely to be a cause for concern since our analysis uses a panel estimation framework. There were no new rent control laws enacted after 2001. Amendments to the preexisting rent control laws did not have provisions that could affect rents paid by tenants (Gandhi *et al.*, 2021a). Hence, the rent control law effects would be mostly absorbed by the time invariant heterogeneous fixed effects.

[23]The Construction Industry Development Council database provides monthly construction cost indexes for the largest cities in India since 2007. This data is not applicable in this paper because the time period of our analysis intersects with this data partly and we conduct a district-level analysis instead of at the city-level.

5.4 Discussion on Instruments

The empirical models given by previous sections 5.1 to 5.3 are meant to test the hypothesis that distant state shocks act as local housing demand shifters by inducing migration across regions. Equations (17) to (20) will be used to estimate the effect of in- and out-migration in local urban regions on urban population and household growth. Since there are two endogenous independent variables, we require two instruments to identify equations (17) to (20). We propose negative rainfall shocks and a national highway upgrade program occurring at a distant state as instruments for the growth rate of local urban migration, urban population and households, and the number of urban housing units. Below, we discuss the validity of these two instruments.

Negative rainfall shocks act as negative income shocks in most parts of India owing to the largely rainfall-dependent agricultural practices. Hence, rainfall levels less than 80% of the long-term normal induce drought-like conditions in several regions and are deemed to be unfavorable for agricultural output. There is a body of literature examining this relationship between rainfall shocks and agricultural output and its subsequent impact on migration (Jayachandran, 2006; Morten, 2019; Rosenzweig and Udry, 2014). Rainfall shocks have been used as an instrument to study civil conflict and dowry deaths in India (Sarsons, 2015; Sekhri and Storeygard, 2014). Bhavnani and Lacina (2017) constructed an instrument from negative rainfall shocks to estimate the effect of inter-state migration flows on fiscal federalism in India. Consistent with their use of a rainfall shock instrument, and the definition used by the IMD to designate regions as rainfall deficient, we measure the rainfall shock as the number of months when absolute rainfall was less than 80% of the long-term normal.[24]

The validity of rainfall shocks at a distant region as instruments can be argued on two fronts. First, a negative rainfall shock in a distant state is a strong predictor of the growth of local urban migration, urban population and households, and the number of urban housing units as seen in the first-stage regression results given by table 3, table 5, and table 8. Second, the exogeneity assumption implies that a negative rainfall shock occurring in one state should be sufficiently unexpected and uncorrelated with unobserved factors that affect urban population and household growth and the number of urban housing units in another state. One way this can be violated is if there is a spatial correlation in rainfall shocks occurring in neighboring states. To rule out this possibility, we conduct robustness checks by running the regressions given by the empirical models discussed in sections 5.1 to 5.3, using rainfall shocks in non-contiguous states as instruments. These robustness results, to be discussed in section 6.4, are roughly unchanged from the models using rainfall shocks

[24] The complete list of all weather event definitions used by the IMD can be downloaded from the following weblink: https://www.imdpune.gov.in/Weather/Reports/glossary.pdf

in all other states as the instrument. Another potential concern is that the rainfall shocks could spillover into neighboring states as income shocks. However, any such spillover effects of income shocks can only be driven by the migration of individuals and firms from one state to the other. Since negative rainfall shocks predominantly affect agricultural incomes, we would not expect firms to move in response to such income shocks, especially given the high sunk cost of setting up businesses in India. Therefore, we can argue that negative rainfall shocks meet the exclusion restrictions for an instrument.

The National Highways Development Project Phase I (NHDP I) or the Golden Quadrilateral (GQ) highway project was introduced as a highway upgrade program by the Central government of India in 2000, and it came into effect in 2001. In most cases, the project was undertaken to upgrade preexisting national highways connecting the four largest metropolitan cities in India — Delhi, Mumbai, Kolkata, and Chennai — from two lanes to four lanes. These highways ran through 14 states and union territories. The Golden Quadrilateral project has been documented as a positive economic shock since it affected firm relocation along the highway in the states through which it passed (Ghani *et al.*, 2016).

Inclusion of a state in the GQ program post-2001 is the second distant state shock in the empirical framework discussed in sections 5.1 to 5.3. We expect two countervailing effects of the highway upgrade program in a state. First, due to firm relocation along the highways, we would expect to see a growth in employment in the program states. And second, the firm and employment growth will also lead to a positive income shock in program states. These two effects would have a subsequent impact on the mobility of individuals between states. Based on conventional models of mobility, the employment effect would induce movement to the states which were part of the GQ. However, the income effect itself consists of two additional opposing forces. First, higher incomes at the present state of location would reduce outward mobility as predicted by the Harris-Todaro models of rural-urban migration (Harris and Todaro, 1970; Todaro, 1969). And second, higher incomes would also spur movement out of the state of location because higher incomes insure individuals against risky migration outcomes (Morten, 2019; Munshi and Rosenzweig, 2016). The net income effect on mobility hinges on the relative strength of these two factors.

We argue that a distant state's inclusion in the GQ program post-2001 is a valid instrument for growth of urban migration, urban population and households, and the number of urban housing units. First, a distant state's inclusion in the GQ program after 2001 is a significant predictor of local urban migration, urban population and households, and the number of urban housing units as seen in the first-stage regressions presented in table 3, table 5, and table 8. Second, to satisfy the exogeneity assumption, the inclusion of one state in the GQ program post-2001 should be exogenous to the urban population, households, and housing units in another state. This can be violated if the inclusion of one state in the GQ program is correlated with the inclusion of another state. This is unlikely to be the case since these highways were constructed on trade routes that were built during ancient and colonial times. For instance, the National Highway II (NH2) was constructed on portions of the Grand Trunk Road that was first built by the emperor Chandragupta Maurya during the 3rd century BCE and later redeveloped under the rule of emperor Sher Shah Suri, the Mughals, and the British Raj (Elisseeff, 2000; Thapar, 2015). Another concern is that neighboring states have a higher probability of being on ancient trade routes, thereby inducing a correlation between the inclusion of contiguous states in the GQ program. In the panel IV framework, such time constant state border effects would be eliminated. The GQ upgrade program across contiguous states might also potentially affect housing supply if better contiguous-state road networks lead to higher trading, and thus, reduced prices of construction material. Robustness checks with distant non-contiguous states' inclusion in the GQ program after 2001 as the instrument yield similar results as discussed in section 6.4.[25] Third, the only channel other than the migration of individuals through which the GQ program in one state post-2001 can affect the urban population, households, and housing units in another state is through the simultaneous relocation of firms. However, it is unlikely that the relocation of firms producing non-housing goods would have any effect on the population, households, and housing units that is independent of individual migration. If real estate firms or developers relocate across state boundaries in response to the GQ program, then the housing supply would also be affected by the GQ program in a different state, thus violating the exogeneity assumption. But, given the heterogeneity in property rights and laws across Indian states due to individual state governments' jurisdiction over land and real estate, it is unlikely that developers from one state would relocate to another in response to the GQ program.[26] Hence, a distant state's inclusion in the GQ program post-2001 is a valid instrument for urban migration, population, households, and housing units in the local state.

6 Results

The previous section 4 and section 5 lay down the theoretical and the empirical framework to analyze the effect of distant region shocks on local urban housing demand due to mobility in

[25] Even though we expect construction material to be traded across neighboring states, it's unlikely that trading of construction material happens across non-contiguous states.

[26] In some cases, states might directly prohibit non-residential individuals from property ownership or construction of houses. For instance, Karnakata and Sikkim allow individuals to own land and construct houses only upon providing state domicile certificates.

spatial disequilibrium. This allows for the estimation of local urban housing supply elasticity using the distant state shocks as demand shifters. In this section, we first estimate the effect of the distant state shocks on inter-state mobility. Next, we estimate the effect of exogenous shock induced-mobility on local urban population and household growth. Then, we estimate the effect of exogenous urban population and household growth on local urban housing demand. We use the same distant state exogenous shocks as housing demand shifters to estimate the local urban housing supply elasticity. Additional identification concerns are addressed with robustness checks. We end this section with a discussion on state-level urban durable and vacant housing supply elasticities.

6.1 Effect of Shock-induced Migration on Population Growth

We begin by estimating the effect of two shocks incident upon distant region j on the log of in-migration over the last decade from j to local region i and log of out-migration over the last decade from i to j. As defined in section 5.1, the urban area in each Indian state represents the local region i and both rural and urban areas in another state represents the distant region j. The first shock is measured as the number of months when absolute rainfall was less than 80% of the long-term normal in j. The second shock is the inclusion of state j in the Golden Quadrilateral (GQ) highway upgrade program after 2001. We first run fixed-effects regressions using the empirical model given by equation (12) and equation (13). These are the first-stage regressions for equation (14) and equation (15). We use an unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011. The results are presented in table 3.

We observe a number of things here. First, while the rainfall shock at j is a strong predictor of in-migration from j to i (columns 1 and 2 in table 3), it does not have any effect on out-migration from i to j (columns 3 and 4 in table 3),. Controlling for consumption and urban surface area at i, one additional month of rainfall levels less than 80% of the long-term normal at j increases in-migration from j to i by 1.5%. This is consistent with the literature that negative rainfall shocks in a region spur outward mobility from the region (Rosenzweig and Udry, 2014; Bhavnani and Lacina, 2017). Second, the highway upgrade at j is a positive and significant predictor of in-migration from j to i and out-migration from i to j. Controlling for consumption and urban surface area at i, a distant state's inclusion into the GQ program post-2001 increases in-migration from j to i by 27% and out-migration from i to j by 35%. This is consistent with theory and literature suggesting that the employment effect of firm relocation along the highway would increase movement toward states included in the GQ program, and the higher insurance due to the wealth effect would spur movement outward from those states (Bartik, 1993; Morten, 2019). Note that the effect of the GQ program in j has a higher effect on out-migration from i to j. This indicates that the employment effect is a stronger predictor of migration than the wealth effect. The results indicate a net flow of individuals to states that were included in the GQ program post-2001. Third, consumption at i has no association with in-migration to i, but it is negatively correlated with out-migration from i. This indicates that higher per capita consumption might be incentivizing individuals to stay in their current locations instead of moving. This finding is contrary to the exogenous wealth effect of the distant state's inclusion in the GQ program, as seen through higher in-migration from j to i. And finally, the urban surface area has a non-linear relationship with both in- and out-migration at i.

Next, we estimate the effect of urban in-migration and out-migration on urban population and household growth at i, using the panel IV regressions given by equation (14) and equation (15). The two IVs used for the log of urban in- and out-migration are the number of months when rainfall was less than 80% of the long-term normal and a dummy variable indicating the inclusion of the distant state j in the GQ highway upgrade program after 2001. The results presented in table 4 are the second-stage regressions for the effect of in- and out-migration at i on urban population and households.

We notice three things in table 4. First, an increase in urban in-migrants leads to an increase in both the urban population and households. Controlling for consumption and urban surface area at i, a 1% increase in urban in-migrants from j to i increases urban population by roughly 0.96% and urban households by 1.13%. A higher effect of in-migration on households compared to population could possibly be due to smaller migrant household size.[27] Second, we also find that urban out-migration and urban consumption at i has no relationship with either the urban population or households. The relationship between consumption and population growth is consistent with a portion of the large body of literature with mixed findings (Greenwood, 1997). Finally, we see that the urban population increases at a decreasing rate in response to an increase in urban surface area. This indicates that the urban population density decreases with surface area, consistent with literature on monocentric city models (Duranton and Puga, 2015).

6.2 Effect of Exogenous Population Growth on Housing Units

In this section, we first estimate the effect of two exogenous shocks incident upon j on urban population and household growth in i. As defined in section 5.2, i consists of urban areas in

[27] The National Sample Survey data on employment and migration indicates that the average household size of migrant households moving to urban areas is 4.4 while that of non-migrant households living in urban areas is 5.4.

a district of the local state, and j represents both rural and urban areas in a distant state. The first shock is measured as the number of months when rainfall was less than 80% of the long-term normal at state j. The second shock is the inclusion of state j in the Golden Quadrilateral (GQ) highway upgrade program after 2001. Then we estimate the effect of urban population and household growth, as predicted by the exogenous distant state shocks, on the number of urban non-durable, durable, and vacant housing units at i. The empirical model is a panel IV framework with equation (16) and equation (17) as the first-stage fixed-effects regressions and equation (18) and equation (19) as the second-stage regressions. We use an unbalanced panel of 220 $\times$ 34 and 208 $\times$ 34 (i, j) district-state pairs for the two Census years 2001 and 2011, respectively.

The first-stage fixed effects regression results are presented in table 5. There are five things to note in these results. First, both the rainfall shock and the highway upgrade program at the distant state j lead to an increase in urban population and households in district i. Controlling for urban consumption, urban surface area, mean urban household size, and the median number of rooms in an urban household, an additional month of rainfall levels less than 80% of the long-term normal at state j leads to an increase in urban population by 0.9% and urban households by 1.3%. The inclusion of state j in the GQ program after 2001 increases urban population by 11% and urban households by 14%. This is consistent with findings in table 4 that the shocks had a positive effect on in-migration from j to i and that such in-migration led to growth in urban population and households. Since out-migration from i to j does not have any effect on urban population or households (see table 4), the out-migration inducing force of employment effect due to the highway upgrade program at j is eliminated. Second, contrary to the previous finding in table 4 that higher consumption at i is not associated with out-migration from i to j or in-migration from j to i, here we find that a 1% higher consumption at i is associated with a decline in urban population by 0.03% and urban households by 0.04%. This is perhaps due to the fact that consumption is an imperfect proxy for income and does not say much about wealth which is a bigger predictor of mobility, and hence, population growth (Munshi and Rosenzweig, 2016). Third, consistent with previous findings in table 4, we see here that both the urban population and household growth are increasing and concave in urban surface area. Fourth, one additional person in an average urban household in i is associated with a lower urban population by 8% and lower urban households by 14%. While the decline in households logically follows from the negative relationship between households and household size, the negative association between household size and population growth is not immediately obvious. It could be that the growth in the elderly population is contributing to an increase in urban household size but lowering fertility rates which cause a decline in the urban population. And finally, we see that an additional room in a house on average has no relationship with urban population growth, but is associated with an increase of 2% in urban households.

Now we discuss the second-stage regression results for the effect of distant state shock-induced urban population growth on the number of urban non-durable, durable, and vacant housing units, based on equation (18) that are presented in table 6. Here as well, we note five things. First, a distant state shock-induced growth in urban population has a positive effect on all types of urban housing units. Controlling for urban consumption, urban surface area, mean urban household size, and the median number of rooms in urban households, a 1% increase in urban population leads to a 0.08% increase in urban non-durable houses, 1.8% increase in urban durable houses, and 2.4% increase in urban vacant houses. A higher response of durable housing units compared to non-durable ones is consistent with the idea that poorer migrants moving into non-durable houses have a lower endowment, and hence, lower demand for housing services. The significant increase in vacant houses is in part due to out-migrants' contribution to higher vacant houses and developers engaging in speculative building with the expectation that future demand would be higher in response to migrant in-flows. Second, a 1% higher consumption predicts lower non-durable houses by 0.3% and vacant houses by 0.2% but has no association with the number of durable houses. Third, the urban surface area has a non-linear relationship with all three types of housing. While non-durable houses increase at a decreasing rate, both durable and vacant houses decrease at an increasing rate as urban surface area increases. This indicates that as Indian cities grow in surface area, non-durable houses crowd out durable and vacant houses. One reason behind this could be that while durable houses can be built with multiple floors, non-durable houses are restricted to one or two floors. Fourth, one additional person in a household on average is associated with 8% fewer non-durable houses, 4% fewer durable houses, and has no association with the number of vacant houses. This is consistent with the fact that as the average urban household size increases, there is lower demand for non-durable and durable houses, but since vacant houses are unoccupied, we do not observe any relationship with mean household size. And finally, an additional room in a house on average is associated with 15% fewer non-durable houses, 2.5% fewer vacant houses, and 4.2% higher durable houses.[28]

The estimates for the effect of distant state shock-induced urban household growth on the number of urban non-durable, durable, and vacant housing units, based on equation (19) are presented in table 7. The direction of effects are similar to that reported in table 6. First, a

[28]There could be concerns of endogeneity here since the median number of rooms in an urban household in district i could be correlated with the number of non-durable, durable, or vacant housing units. To remove such concerns, we ran regressions without the median number of rooms as a control and found the results to be largely similar. These results can be found in appendix table A.15-table A.17.

1% increase in urban households due to a distant state shock increases the number of urban non-durable houses by 0.06%, durable houses by 1.3%, and vacant houses by 1.8%. Second, 1% higher consumption is associated with 0.3% fewer urban non-durable houses and 0.2% fewer urban vacant houses but a 0.03% higher durable houses. Third, the urban surface area has a positive concave relationship with non-durable houses but a decreasing relationship with durable and vacant houses. Fourth, an increase in the average urban household size by one member is associated with 8% fewer non-durable houses and 4% fewer vacant houses. There is no association between urban household size and durable houses. And finally, one additional room in a house on average is associated with 15% fewer non-durable houses, 5% fewer vacant houses, and 2.4% more durable houses. The coefficients of mean household size and the median number of rooms are consistent with that seen in table 6.

Even though the coefficient estimates obtained from table 7 are of the same sign as those seen in table 6, the magnitude are different. The effect of urban population growth on urban housing units' growth is about 30% lower than the effect of urban households' growth. The lower magnitude of effect of household growth without controls is consistent with the fact that urban households grew faster than urban population, and hence, had a lower magnitude of impact than that of urban population.[29] However, even after controlling for household size, the magnitudes of effect of household growth is lower than that of population growth, albeit with a reduced difference than that obtained without controls. This is due to the scale effect of smaller magnitude of change in household size across districts and Census years relative to changes in household and population growth. The reduction in the difference between household and population effect on durable houses is lower than that for non-durable houses. This is possibly because the fall in household size of durable housing occupants is slower than that of non-durable housing occupants.

6.3 Housing Supply Elasticity Estimates

In the previous section 6.1 and section 6.2, we discussed empirical estimates for the effect of distant state shocks on mobility, urban population, urban households, and urban housing units. These results helped explain the mechanisms through which negative rainfall shocks and the Golden Quadrilateral (GQ) highway upgrade program status of distant states can affect local urban housing demand. We build on this structure to estimate the supply elasticity of local urban housing using distant state rainfall shocks and the distant state's GQ highway upgrade program inclusion status as demand shifters in a panel IV empirical framework. We estimate three supply elasticity figures, one each for three different types

[29] Urban households grew at the decadal rate of 45% whereas urban population grew at the rate of 32% (Census of India, 2011).

of urban housing units — non-durable, durable, and vacant. The first-stage regressions are given by equation (20) where we regress the log number of housing units in a district i on two exogenous shocks at a distant state j outside the state in which district i is located. The first shock is measured as the number of months when rainfall was less than 80% of the long-term normal in state j. The second shock is the inclusion of state j in the GQ highway upgrade program after 2001. Next, we run the second-stage regressions based on equations (21) to (23) where we regress the average log rent of non-durable, durable, and all types of houses on the predicted log of non-durable, durable, and vacant housing units from the first-stage regression. We use an unbalanced panel of 220 $\times$ 34 and 208 $\times$ 34 (i, j) district-state pairs for the two Census years 2001 and 2011, respectively.

The first-stage regression results are presented in table 8. We note three things in these results. First, both the rainfall shock and the GQ highway upgrade program at a distant state j has a strong positive effect on all three types of houses in district i. An additional month with rainfall levels less than 80% of the long-term average in j increased non-durable houses by 0.2%, durable houses by 2%, and vacant houses by 2.6% in i. The inclusion of state j in the GQ program led to 2.7% higher non-durable houses, 22% higher durable houses, and 30% higher vacant houses in district i. These effects are consistent with results from table 3 through table 7 which show that the distant shocks at j predict in-migration into i and that in-migration has a positive effect on urban population and households, thereby leading to higher housing demand. Out-migration has no such effect on housing demand at i. These results empirically confirm proposition 1 given in section 4.2. Second, consistent with table 6, higher consumption in i is associated with 0.2% fewer non-durable houses and vacant houses, and 0.06% higher durable houses. And finally, the urban surface area has a positive concave relationship with all three types of housing units.

The second-stage results are presented in table 9. The coefficients of log non-durable, durable, and vacant houses are the respective inverse elasticities for these individual housing unit categories. There are several things to be noted in these results. First, controlling for urban consumption, urban surface area, and the urban surface area squared at i, the inverse elasticity of housing for urban non-durable houses is -1.84 implying an urban non-durable housing supply elasticity of -0.545. Negative supply elasticity of housing is contrary to theoretical models and empirical estimates of urban housing supply seen elsewhere in the literature. The negative elasticity of supply for non-durable houses indicates that a process of urban gentrification is underway in Indian cities. Gentrification occurs in two ways. On the one hand, a simultaneous increase in rents paid by slum dwellers and land values around slums attract real estate developers. Slums are cleared for the construction of durable residential and commercial real estate space (Bhan, 2009). On the other hand, slums are upgraded through various government and non-government programs that convert non-durable units to durable ones (Rains *et al.*, 2019; Rains and Krishna, 2020). Second, the decadal elasticity of supply for durable housing units in urban India is 1.64. This estimate is substantially lower than the long-run housing supply elasticity estimates of 6-13 for metropolitan areas in the post-war United States obtained by Malpezzi and Maclennan (2001).[30] Saiz's more recent short-run housing supply elasticity estimate of 1.75 for the average metropolitan area in the United States is slightly higher than ours. Comparing supply elasticity estimates for urban India and metropolitan areas in the United States indicates that housing markets in Indian cities respond rather slowly to increases in prices even in the long run. Third, the elasticity of supply for vacant urban housing units is 2.63, which is substantially higher than that of urban durable housing units. This is perhaps due to the fact that developers are engaging in speculative building. Gandhi *et al.* (2021a) argues that a large number of buyers are investing in home purchases with the intention of selling them at a later stage when prices increase. Developers respond to such speculative demand-driven higher market prices by building more units. Hence, as prices go up, the number of vacant houses in markets with speculative buyers would increase. Fourth, 1% higher urban consumption is associated with 0.5% lower non-durable rents, but 0.5% higher durable rents and 0.4% higher overall rents. This is consistent with the fact that non-durable houses are an inferior good while durable houses are not. And finally, while non-durable rents increase with urban surface area, durable rents decrease. This indicates that while slum housing gets more expensive, formal housing is cheaper as cities grow.

The regression sample used for table 9 is restricted to those districts for which we have rent data for all types of housing. However, we have an unbalanced panel of 360 $\times$ 34 district-state pairs in 2001 and 437 $\times$ 34 district-state pairs in 2011, with durable and overall market rent data. In table 10 we show results from regressions given by equation (22) and equation (23) using the full sample of observations. The durable and vacant housing supply elasticities are 1.38 and 1.78, respectively, which are lower than those observed in table 9.

6.4 Robustness Checks

In section 6.1 through section 6.2, we discussed the effects of shock-induced inter-state migration on local urban population and household growth and the resulting local urban housing demand. We use a panel instrumental variable framework to identify these effects. In section 6.3, we discussed local urban housing supply estimates derived from using the distant

[30] Blackley (1999) obtained a much lower range of housing supply elasticity estimates between 1.6-3.7 for the post-war United States. However, we use the estimates given by Malpezzi and Maclennan (2001) for comparison because the latter estimates are robust to multiple specifications.

state exogenous events as demand shifters. In this section, we address two major identification concerns in the empirical estimation framework.

The first threat to the identification strategy, as mentioned in section 6.1, is that the negative rainfall shocks and the Golden Quadrilateral (GQ) highway upgrade program at the distant state might induce local labor supply effects in the construction industry, thereby affecting local construction wages. Hence, the distant state shocks would also have a local housing supply effect through changes in construction cost. This is particularly relevant in the Indian context since a large number of Indians migrate for construction work. However, such migrant workers are more likely to move seasonally for one to six months before moving back to their homes. The National Sample Survey on migration and employment conducted in 2007-08 indicates that while 36% of seasonal inter-state migrants move for construction work, only about 1.5% of long-term inter-state migrants do so. Hence, if we eliminate short-run migrants from our analysis, we alleviate the endogeneity concern arising from the housing supply effects of distant state shocks. We address this by redefining the migration variables to exclude short-run migrants who moved less than a year before the Census enumeration and re-estimating equations (17) to (20). The redefined m_{jit} and m_{ijt} represents the number of individuals who moved from j to i and i to j respectively over the 1-9 year period before Census year t.

The results with redefined long-term migration variables can be seen in table 11 and table 12. As seen in table 3, both the distant state negative rainfall shock and the distant state's inclusion in the GQ highway upgrade program have a positive significant effect on in-migration from j to i. The highway upgrade program also has a positive significant effect on out-migration from j to i. Controlling for consumption, urban area, and urban area squared at i, every additional month of rainfall levels below 80% of the long-term average at the distant state j leads to 0.07% additional in-migrants from j to i which is roughly half the effect of all migrants who moved in the decade leading up to the Census year (see table 3). Consistent with the literature on short-run migration in India, the reduced magnitude of rainfall shock effect on long-term in-migrants from j to i indicates that the negative rainfall shocks have a more pronounced effect on short-run migrants who move temporarily for a few months to supplement their farm incomes (Morten, 2019; Rosenzweig and Udry, 2014). Moreover, the distant state's inclusion in the GQ highway upgrade program increases in-migration from j to i by 23% and out-migration from i to j by 30%, which is very similar to the effects observed in table 3.

The estimated effect of instrumented long-term in-migration variables on urban population and household growth, as shown in table 12, are much larger than those in table 4. While all in-migrants from j to i leads to 0.96% growth in urban population and 1.13%

growth in urban households, long-term in-migrants lead to 1.9% growth in urban population and 2.22% growth in urban households. The relatively higher effect of long-term in-migrants on urban population and household growth is possibly due to two reasons. First, many short-term migrants who move for less than a year are circular-migrants who move back to their homes after a few months. And second, while short-term migrants do not settle down in their destination, many long-term migrants would likely do so and have children, thus contributing to both population and household growth. Consistent with results seen in table 4, long-term out-migrants have no effect on urban population and household growth. These results indicate that even though the magnitude of the effects change, the distant state negative rainfall shocks and inclusion of the distant state in the GQ highway upgrade program affect local urban population and household growth.

The second identification issue is with the spatial correlation of the shocks across state boundaries. Bhavnani and Lacina (2017) discusses the problem of spatial correlation of rainfall events across state boundaries and resolves the issue by controlling for rainfall at both the origin and destination of migrants. The GQ upgrade program across contiguous states might also potentially affect housing supply if better contiguous-state road networks lead to higher trading across neighboring states, and thus, reduced prices of construction material. To alleviate these endogeneity concerns, we implement the empirical estimation framework given by equations (17) to (28) using negative rainfall shocks and the GQ highway upgrade program implementation in distant non-contiguous states. We present the effect of these shocks on migration variables and the housing supply elasticity estimates in table 13 and table 14. The remaining estimation results can be seen in appendix tables A.1 to A.6. The coefficient estimates in table 13 are very similar in magnitude to those seen in table 3. However, negative rainfall events in distant non-contiguous states also reduce out-migration to those states. Every additional month of rainfall levels below 80% of long-term normal in distant non-contiguous states reduces out-migration to such states by 0.05%. This is consistent with the idea that the negative rainfall shocks reduce the attractiveness of distant states relative to neighboring states, and therefore, inhibit movement to the distant states. The elasticity magnitudes seen in table 14 are almost identical to those reported in table 9. These results indicate that the spatial correlation of distant state shocks does not pose a threat to the identification strategy discussed in equations (17) to (28).

6.5 State-level Elasticities

The previous section 6.3 discussed urban housing supply elasticity estimates at the national-level in India. However, in order to get a sense of spatial heterogeneity in these urban housing supply elasticity estimates, we also provide some state-level elasticity figures for durable and vacant housing units. We do this by exploiting district-level variation within a state for ten of the largest states in India.[31] We run these regressions using only durable and overall rent figures since data on non-durable rent is not available for enough districts to run regressions even in the largest states.[32] The durable and vacant housing supply elasticity for urban areas for the ten states is reported in table 15.

Among the reported states, Maharashtra has the highest urban housing supply elasticity with a value of 3.06. For reference, Maharashtra's decadal housing supply elasticity is very similar to the short-run housing supply elasticity of Austin-San Marcos and Charlotte–Gastonia–Rock Hill MSAs in the United States (Saiz, 2010). Bihar and West Bengal have the lowest urban housing supply elasticities with values of 0.49 and 0.39, respectively. These long-run elasticity values are lower than the United States' least short-run supply elastic MSAs: Miami and Los Angeles-Long Beach.

7 Conclusion

According to the United Nations, developing countries in Africa and South Asia are set to experience the next wave of urbanization (United-Nations, 2018). India will be one of the largest contributors to this growth. But, academic literature on housing in India is sparse. We fill this gap by estimating the supply elasticity of housing in urban India. We apply the Rosen-Roback spatial equilibrium framework to estimate the effect of distant state shock-induced migration on local urban population, households, and housing demand. We use a distant state's rainfall shocks and a highway upgrade program implementation status as demand shifters to estimate the supply elasticity of local urban housing markets.

We begin by presenting some stylized facts on housing and migration in India. Next, we provide a discussion of the underlying theoretical framework followed by a discussion of the empirical implementation of the model mechanisms. Then we discuss the empirical results and robustness checks. We find that both the negative rainfall shocks and the highway upgrade implementation at the distant state induce inter-state migration, thereby increasing local urban population, households, and the demand for local urban housing. We estimate national-level urban housing supply elasticity figures for durable, non-durable, and vacant residential housing units for India.

While national-level housing supply elasticity estimates do not paint an accurate picture

[31] We report the ten states in which at least 15 districts reported a durable rent figure for both years.

[32] For instance, Uttar Pradesh, the largest state in India, reports non-durable rents for both years in only 14 districts.

of metropolitan-level elasticities and the underlying heterogeneity across metropolitan areas of different sizes and regulations, it is a relevant parametric estimate in the context of a large, growing developing country. Further research with metropolitan-level price and new construction data would be required to provide granular estimates.

References

Bartik, T. J. (1993) Who benefits from local job growth: migrants or the original residents?, *Regional studies*, **27**, 297–311.

Baum-Snow, N. and Han, L. (2019) The microgeography of housing supply, *Work in progress, University of Toronto*.

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

Bhan, G. (2009) "This is no longer the city I once knew". Evictions, the urban poor and the right to the city in millennial Delhi, *Environment and urbanization*, **21**, 127–142.

Bhavnani, R. R. and Lacina, B. (2017) Fiscal federalism at work? Central responses to internal migration in India, *World Development*, **93**, 236–248.

Blackley, D. M. (1999) The long-run elasticity of new housing supply in the United States: Empirical evidence for 1950 to 1994, *The Journal of Real Estate Finance and Economics*, **18**, 25–42.

Brueckner, J. K. and Sridhar, K. S. (2012) Measuring welfare gains from relaxation of land-use restrictions: The case of India's building-height limits, *Regional Science and Urban Economics*, **42**, 1061–1067.

Brueckner, J. K. *et al.* (1987) The structure of urban equilibria: A unified treatment of the Muth-Mills model, *Handbook of regional and urban economics*, **2**, 821–845.

Census of India, P. (2011) *Census of India 2011 Provisional Population Totals*, Office of the Registrar General and Census Commissioner New Delhi.

Desa, U. *et al.* (2014) World urbanization prospects, the 2011 revision, *Population Division, Department of Economic and Social Affairs, United Nations Secretariat*.

Duranton, G. and Puga, D. (2015) Urban land use, in *Handbook of regional and urban economics*, Elsevier, vol. 5, pp. 467–560.

Elisseeff, V. (2000) *The silk roads: Highways of culture and commerce*, Berghahn Books.

Gandhi, S., Green, R. K. and Patranabis, S. (2021a) India's housing vacancy paradox: How rent control and weak contract enforcement produce unoccupied units and a housing shortage at the same time, *CSEP Working Paper*.

Gandhi, S., Tandel, V., Tabarrok, A. and Ravi, S. (2021b) Too slow for the urban march: Litigations and the real estate market in Mumbai, India, *Journal of Urban Economics*, **123**, 103330.

Ghani, E., Goswami, A. G. and Kerr, W. R. (2016) Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing, *The Economic Journal*, **126**, 317–357.

Ghose, D. (2019) Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?

Green, R. K., Malpezzi, S. and Mayo, S. K. (2005) Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources, *American Economic Review*, **95**, 334–339.

Greenwood, M. J. (1997) Internal migration in developed countries, *Handbook of Population and Family Economics*, **1**, 647–720.

Harari, M. (2020) Cities in bad shape: Urban geometry in India, *American Economic Review*, **110**, 2377–2421.

Harris, J. R. and Todaro, M. P. (1970) Migration, unemployment and development: a two-sector analysis, *The American Economic Review*, **60**, 126–142.

Imbert, C. and Papp, J. (2020) Short-term migration, rural public works, and urban labor markets: Evidence from India, *Journal of the European Economic Association*, **18**, 927–963.

India Meteorological Department, P. (1901-2017) *India Meteorological Department Rainfall Datasets*, http://www.imd.gov.in.

Jayachandran, S. (2006) Selling labor low: Wage responses to productivity shocks in developing countries, *Journal of Political Economy*, **114**, 538–575.

Joshi, H. *et al.* (2006) Identifying asset price bubbles in the housing market in India-preliminary evidence, *Reserve Bank of India Occasional Papers*, **27**, 73–88.

Kone, Z. L., Liu, M. Y., Mattoo, A., Ozden, C. and Sharma, S. (2018) Internal borders and migration in India, *Journal of Economic Geography*, **18**, 729–759.

Labor Bureau of India, D. (2001-2012) *Consumer Price Index Series*, http://www.labourbureau.gov.in/.

Malpezzi, S. and Maclennan, D. (2001) The long-run price elasticity of supply of new residential construction in the United States and the United Kingdom, *Journal of housing economics*, **10**, 278–306.

Mills, E. S. (1967) An aggregative model of resource allocation in a metropolitan area, *The American Economic Review*, **57**, 197–210.

Molloy, R., Smith, C. L. and Wozniak, A. (2011) Internal migration in the United States, *Journal of Economic perspectives*, **25**, 173–96.

Morten, M. (2019) Temporary migration and endogenous risk sharing in village India, *Journal of Political Economy*, **127**, 1–46.

Munshi, K. and Rosenzweig, M. (2016) Networks and misallocation: Insurance, migration, and the rural-urban wage gap, *American Economic Review*, **106**, 46–98.

National Sample Survey Organization, I. (1999-2012) *Indian National Sample Surveys, Rounds 55, 58, 66, and 69*, http://mospi.nic.in.

Niu, D., Sun, W. and Zheng, S. (2021) The role of informal housing in lowering China's urbanization costs, *Regional Science and Urban Economics*, p. 103638.

Olsen, E. O. (1969) A competitive theory of the housing market, *The American Economic Review*, **59**, 612–622.

Poterba, J. M. (1984) Tax subsidies to owner-occupied housing: an asset-market approach, *The Quarterly Journal of Economics*, **99**, 729–752.

Rains, E. and Krishna, A. (2020) Precarious gains: Social mobility and volatility in urban slums, *World Development*, **132**, 105001.

Rains, E., Krishna, A. and Wibbels, E. (2019) Combining satellite and survey data to study indian slums: Evidence on the range of conditions and implications for urban policy, *Environment and Urbanization*, **31**, 267–292.

Roback, J. (1982) Wages, rents, and the quality of life, *Journal of political Economy*, **90**, 1257–1278.

Rosen, S. (1979) Wage-based indexes of urban quality of life, *Current Issues in Urban Economics*, pp. 74–104.

Rosenzweig, M. R. and Stark, O. (1989) Consumption smoothing, migration, and marriage: Evidence from rural India, *Journal of political Economy*, **97**, 905–926.

Rosenzweig, M. R. and Udry, C. (2014) Rainfall forecasts, weather, and wages over the agricultural production cycle, *American Economic Review*, **104**, 278–83.

Saiz, A. (2010) The geographic determinants of housing supply, *The Quarterly Journal of Economics*, **125**, 1253–1296.

Saks, R. E. and Wozniak, A. (2011) Labor reallocation over the business cycle: New evidence from internal migration, *Journal of Labor Economics*, **29**, 697–739.

Sarsons, H. (2015) Rainfall and conflict: A cautionary tale, *Journal of Development Economics*, **115**, 62–72.

Sekhri, S. and Storeygard, A. (2014) Dowry deaths: Response to weather variability in India, *Journal of Development economics*, **111**, 212–223.

Sridhar, K. S. (2010) Impact of land use regulations: Evidence from India's cities, *Urban Studies*, **47**, 1541–1569.

Thapar, R. (2015) *The Penguin history of early India: from the origins to AD 1300*, Penguin UK.

Tiwari, P. and Parikh, J. (1998) Affordability, housing demand and housing policy in urban India, *Urban Studies*, **35**, 2111–2129.

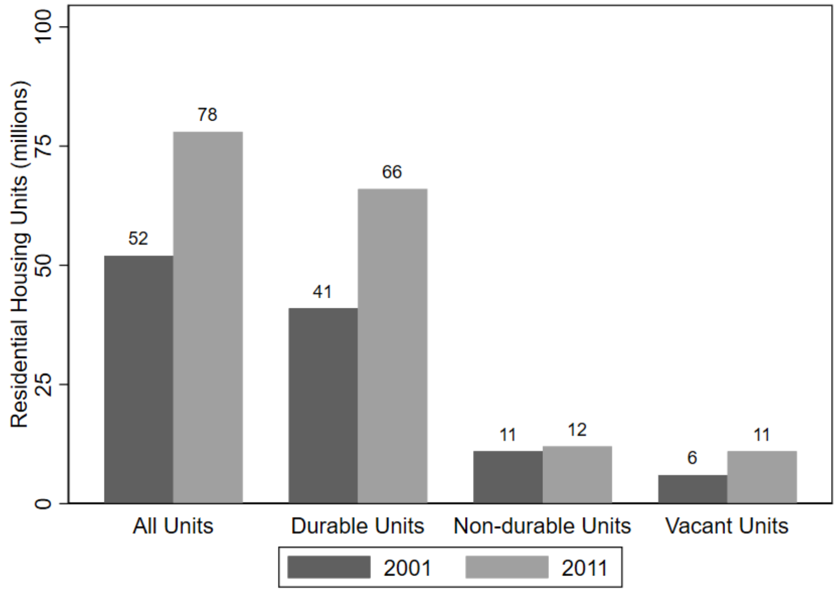
Tiwari, P., Parikh, K. and Parikh, J. (1999) Effective housing demand in Mumbai (Bombay) metropolitan region, *Urban Studies*, **36**, 1783–1809.

Todaro, M. P. (1969) A model of labor migration and urban unemployment in less developed countries, *The American Economic Review*, **59**, 138–148.

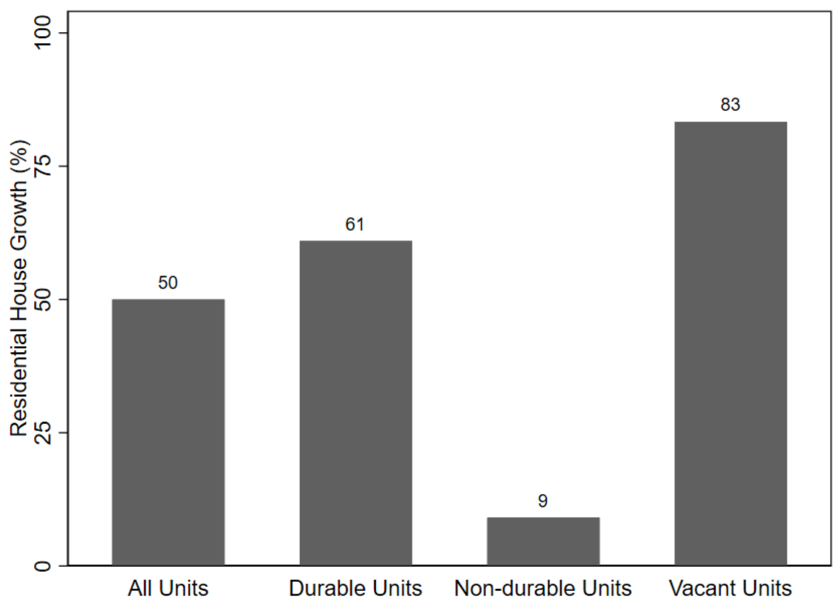
United-Nations (2018) World Urbanization Prospects: The 2018 Revision.

Zhang, K. H. and Shunfeng, S. (2003) Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses, *China Economic Review*, **14**, 386–400.

Figure 1: Urban Housing Units by Type and Structure



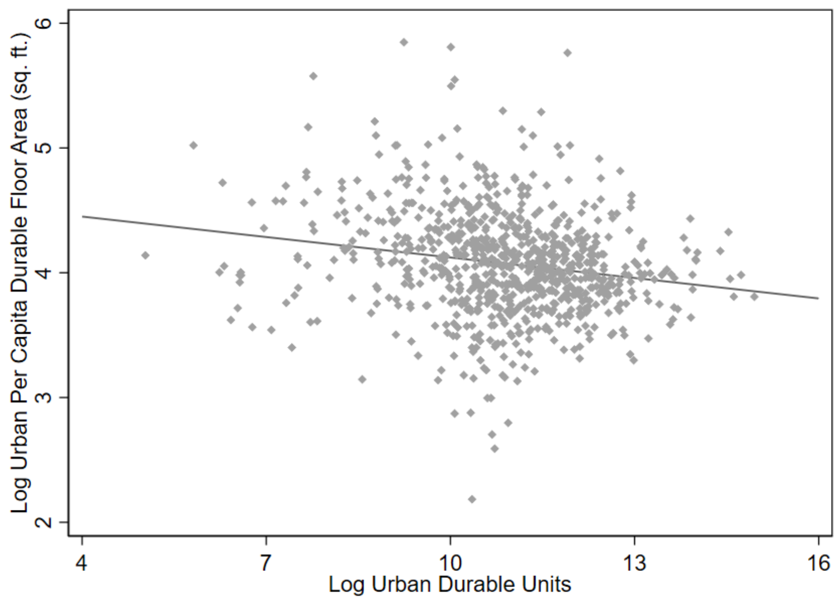
(a) Housing units by Census years



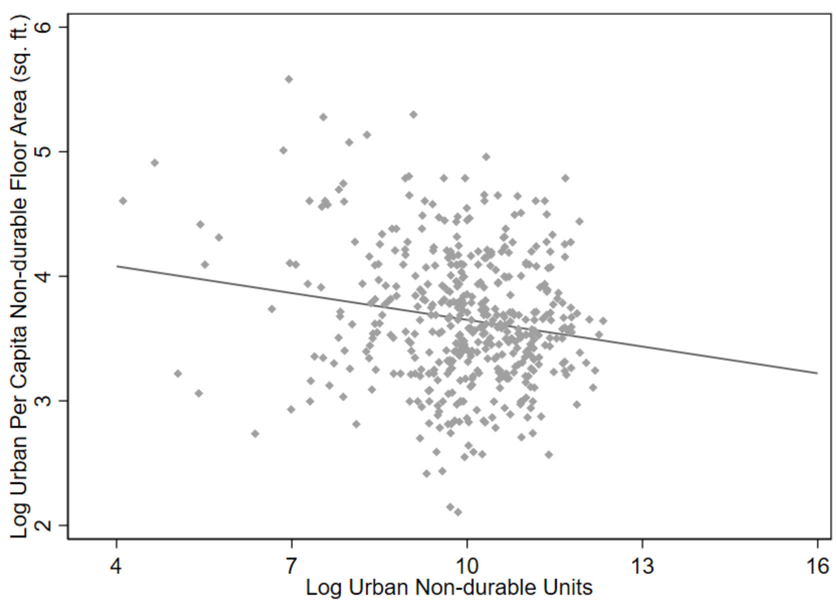
(b) Housing unit growth 2001-2011

Source: Author's calculations based on Census of India.
Note: Panel (a) presents the number of housing units (in millions) by type and structure. Panel (b) presents the percentage growth rate in urban housing units between 2001 and 2011. All bars are labeled by the corresponding values being represented.

Figure 2: Urban Housing Units and Per-Capita Floor Area



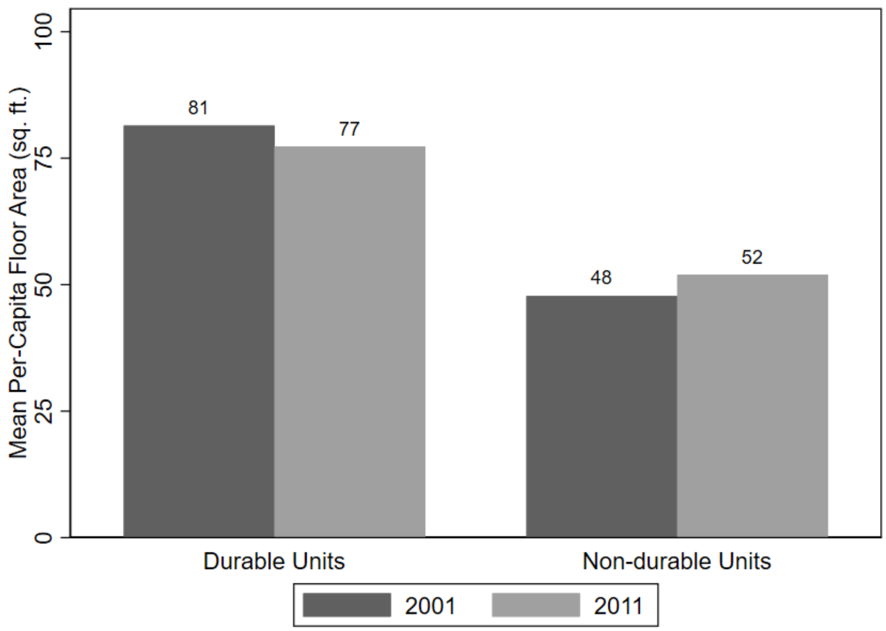
(a) Durable units and durable floor area per capita



(b) Non-durable units and Non-durable floor area per capita

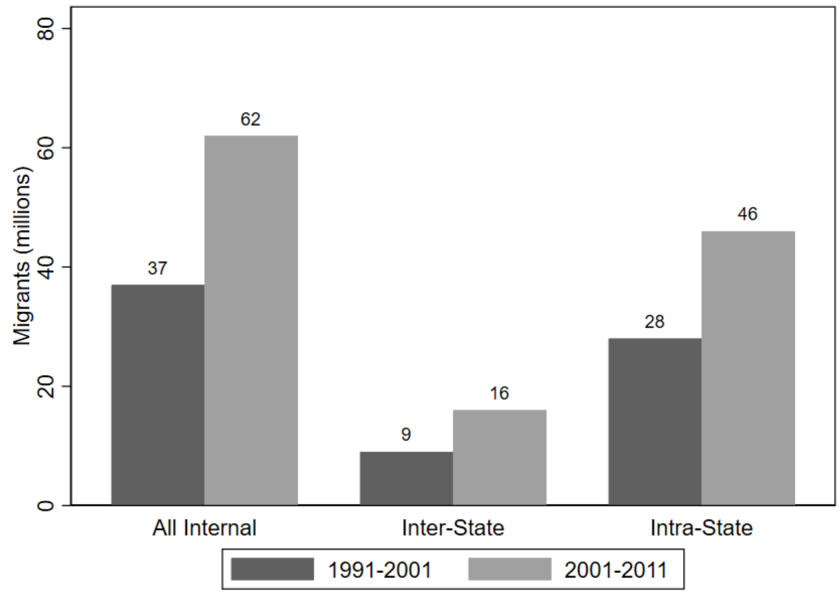
Source: Author's calculations based on Census and National Sample Survey of India.
Note: Panel (a) presents a scatter plot of the log of district-level urban durable housing units and the district-level mean per capita urban durable floor area. The regression line has a slope of -0.05 significant at the 99% interval. Panel (b) presents a scatter plot of the log of district-level urban non-durable housing units and the district-level mean per capita urban non-durable floor area. The regression line has a slope of -0.07 significant at the 99% interval.

Figure 3: Urban Per Capita Floor Area by Housing Structure

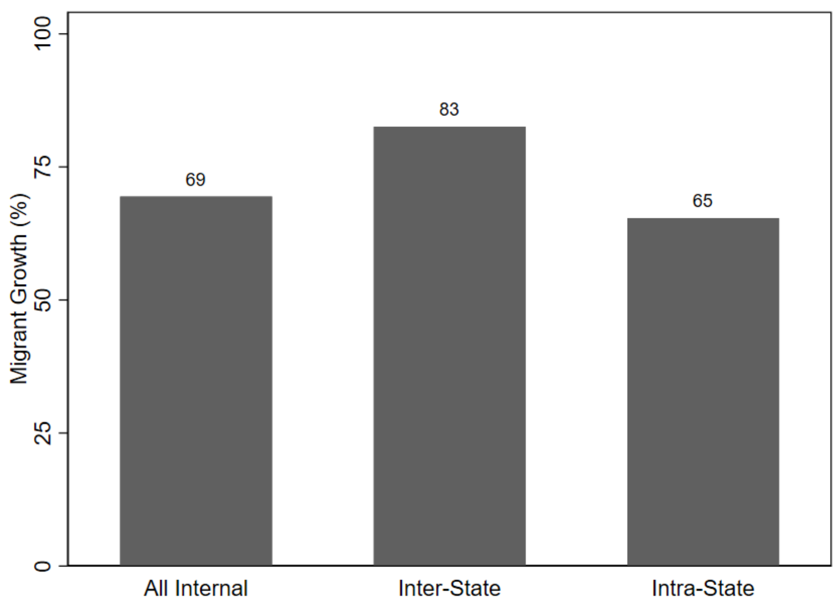


Source: Author's calculations based on National Sample Survey of India.
Note: Figure presents the district-level mean per capita floor area in square feet by structure of housing units and Census years. Bars are labeled by the corresponding values being represented.

Figure 4: Urban In-migration by Distance and Regions



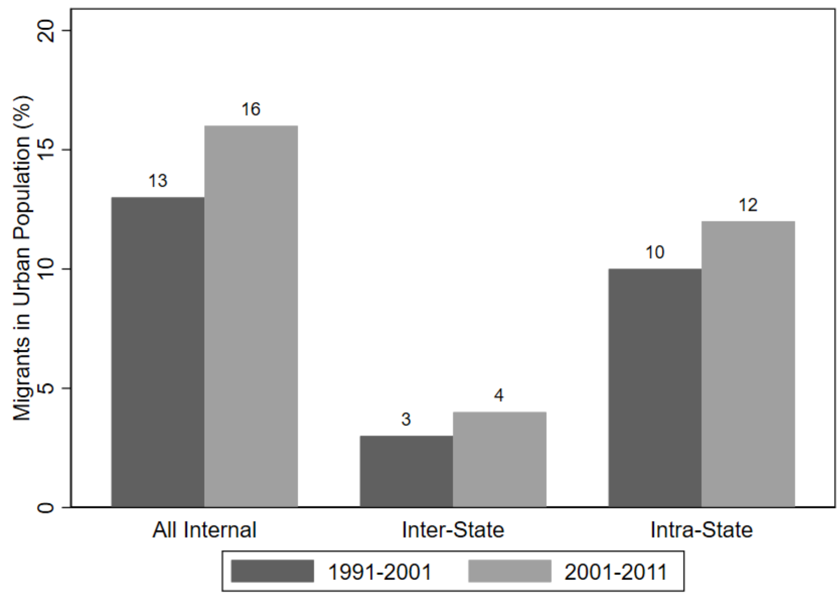
(a) Migration flows by decades



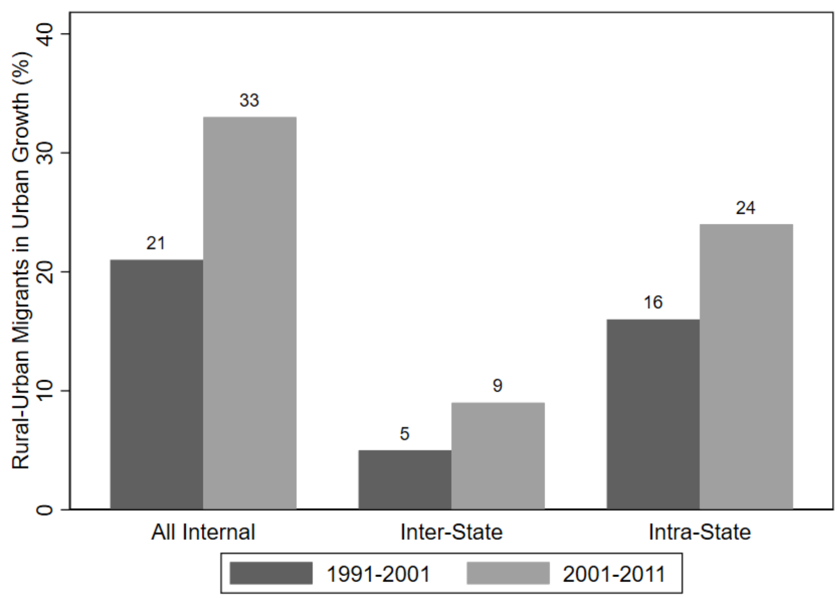
(b) Decadal growth in migration from 1991-2001 to 2001-2011

Source: Author's calculations based on Census of India.
Note: Panel (a) presents the number of urban in-migration flows (in millions) moving from rural and urban regions by distance. Panel (b) presents the percentage growth rate in urban in-migration flows from rural and urban regions during the decade of 1991-2000 through the decade of 2001-2010. All bars are labeled by the corresponding values being represented.

Figure 5: Migrants' Share in Urban Population



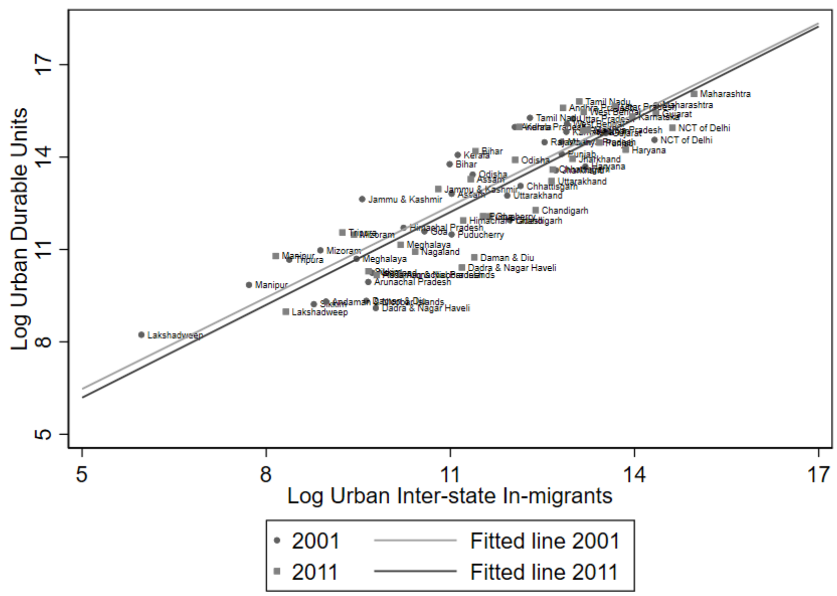
(a) All urban in-migrants' share in urban population



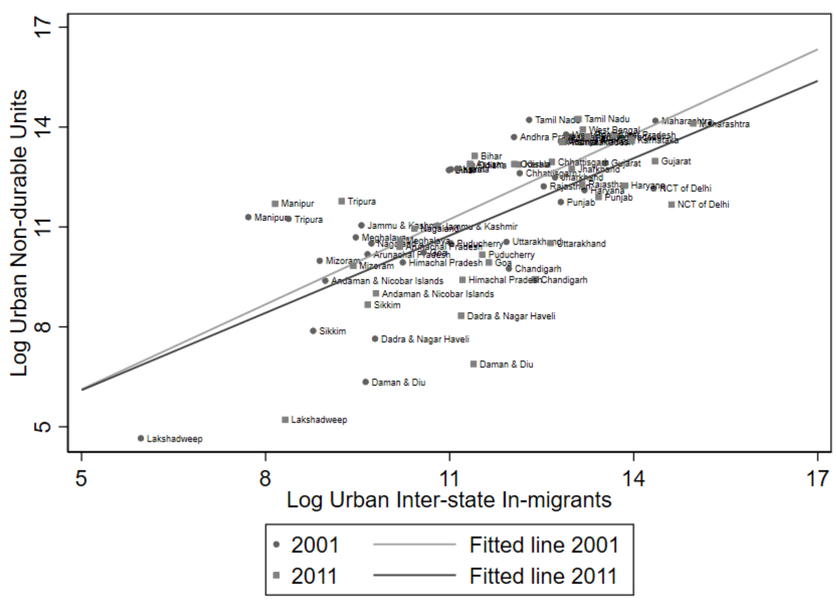
(b) Rural-urban migrants' contribution to urban growth

Source: Author's calculations based on Census of India.
Note: Panel (a) presents the percentage of urban population that are urban in-migrants moving from rural and urban regions by distance. Panel (b) presents the percentage share of rural-urban migrants in urban population growth between 2001 and 2011. All bars are labeled by the corresponding values being represented.

Figure 6: Urban Housing Units by Structure and In-migrants



(a) Urban durable units and in-migrants



(b) Urban non-durable units and in-migrants

Source: Author's calculations based on Census of India.
Note: Panel (a) presents a scatter plot of the log of state-level urban durable housing units and the log of inter-state in-migrants moving into urban areas. The regression lines have slopes of 0.99 and 1.00 respectively for 2001 and 2011, significant at the 99% interval. Panel (b) presents a scatter plot of the log of state-level urban non-durable housing units and the log of inter-state in-migrants moving into urban areas. The regression lines have slopes of 0.85 and 0.77 respectively for 2001 and 2011, significant at the 99% interval.

Table 1: **State-level Summary Statistics**

	Mean		Median		Std. dev.	
Variable	2001	2011	2001	2011	2001	2011
No. of urban inter-state in-migrants last decade ('000)	319	452	105	172	543	691
Urban population ('000)	8,175	10,774	3,439	4,399	10,670	13,629
No. of urban households ('000)	1,595	2,311	711	986	2,115	2,972
Urban surface area (sq. miles)	848	2,255	333	962	1,113	2,707
Urban mean monthly real per capita consumption last year (INR)	248	264	239	250	52	79
Months absolute rainfall <80% last decade	58	64	57	65	12	11
N	35	35	35	35	35	35

Source: National Sample Survey, India Meteorological Department, Census and Labor Bureau of India.
Note: Census state-level panel data used for years 2001 & 2011 for urban migrants, population, households, and urban surface area in a state. National Sample Survey (NSS) household-level data (rounds 55 & and 66) used to calculate state-level mean monthly per capita consumption expenditure for July 1999 - June 2000 and July 2009 - June 2010. Consumption expenditure values inflation-adjusted using the Industrial Worker Consumer Price Index (CPI) with base year 2001 (Labor Bureau). Except for rainfall deviation all values calculated for urban areas in a state. Migrants, total population, and households rounded off to the nearest thousandth integer. Urban surface area rounded off to the nearest square mile. The number of months when rainfall was less than 80% of the long term normal in the last decade rounded off to the nearest integer. Mean monthly real per capita consumption expenditure last year rounded off to the nearest INR.

Table 2: **District-level Summary Statistics**

	Panel (a): Census Variables							
	Mean		Median		Std. dev.		N	
Variable	2001	2011	2001	2011	2001	2011	2001	2011
Urban Population ('000)	597	787	302	411	1094	1372	479	479
Urban Households ('000)	117	169	53	76	227	307	479	479
No. of urban non-durable residential houses ('000)	23	24	12	15	30	29	479	479
No. of urban durable residential houses ('000)	86	138	36	56	187	276	479	479
No. of urban vacant residential houses ('000)	13	23	5	10	33	50	479	479
Mean urban household size	5.26	4.87	5.20	4.80	1.09	0.80	479	479
Median no. of rooms per urban household	2.03	2.02	2.00	2.00	0.53	0.32	479	479
Urban surface area (sq. miles)	62	82	36	48	84	102	479	479

	Panel (b): Rents							
	Mean		Median		Std. dev.		N	
Variable	2002	2012	2002	2012	2002	2012	2002	2012
Mean real rent of all urban residential houses (INR)	538	599	495	532	279	357	385	448
Mean real rent of urban non-durable residential houses (INR)	293	285	238	234	232	231	234	228
Mean real rent of urban durable residential houses (INR)	581	630	524	569	303	370	378	442

	Panel (c): Consumption Expenditure							
	Mean		Median		Std. dev.		N	
Variable	2000	2010	2000	2010	2000	2010	2000	2010
Urban mean monthly real per capita consumption (INR)	207	217	202	205	56	72	427	471

Source: National Sample Survey, Census and Labor Bureau of India.
Note: Census district-level panel data used for years 2001 & 2011 to calculate the urban population and households, the number of residential houses (by structure) and urban surface area presented in panel (a). Household-level sample survey data from the National Sample Survey (NSS) housing conditions rounds 58 and 69 used to calculate district-level average rent for years 2002 & 2012 presented in panel (b). Household-level sample survey data from the NSS consumption expenditure rounds 55 and 66 used to calculate district-level mean monthly per capita consumption expenditure for years July 1999 - June 2000 and July 2009 - June 2010 presented in panel (c). Average rent and consumption expenditure values inflation-adjusted using the Industrial Worker Consumer Price Index (CPI) with base year 2001 (Labor Bureau). All values calculated for urban areas in a district. Values for urban population, households, and residential houses rounded off to the nearest thousandth integer. Mean household size and the median no. of rooms in a district rounded off to two decimal places. Urban surface area rounded off to the nearest square mile. Mean values of real rents and consumption expenditure rounded off to the nearest INR.

Table 3: **First Stage: Effect of Distant State Shocks on Inter-state Urban Migration**

	Dep. var.			
	Log(Urban in-migration to i)		Log(Urban out-migration from i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.020***	0.015***	0.002	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Highway upgrade at j	0.323***	0.273***	0.391***	0.348***
	(0.038)	(0.041)	(0.033)	(0.036)
Log mean per capita consumption (urban) at i		0.111		-0.404**
		(0.178)		(0.188)
Urban surface area of state i		-0.018		0.131***
		(0.028)		(0.025)
Urban surface area of state i squared		0.010***		-0.009***
		(0.002)		(0.002)
Constant	4.99***	4.67***	5.91***	8.29***
	(0.182)	(0.981)	(0.156)	(1.072)
Angrist-Pischke F-stat	117***	44***	96***	48***
N	2,233	2,233	2,233	2,233
Adj. R-sq	0.144	0.182	0.126	0.145

Source: Authors' calculations.

Note: i is the local state and j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log of in-migration last decade into urban areas of local state i from rural and urban areas of distant state j. Dependent variable in columns (3) & (4) is log of out-migration last decade from urban areas of local state i to rural and urban areas of distant state j. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: **Second Stage: Effect of Inter-state Migration on Urban Population and Households**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Log urban in-migration last decade at $i^{\text{IV'd}}$	0.967***	0.958***	1.20***	1.13*
	(0.146)	(0.155)	(0.181)	(0.182)
Log urban out-migration last decade at $j^{\text{IV'd}}$	-0.136	-0.135	-0.189	-0.165
	(0.183)	(0.174)	(0.228)	(0.204)
Log mean per capita consumption (urban) at i		-0.084		-0.128
		(0.174)		(0.204)
Urban surface area of state i		0.131***		0.176***
		(0.035)		(0.041)
Urban surface area of state i squared		-0.017***		-0.021***
		(0.004)		(0.004)
Constant	9.62***	10.0***	6.88***	7.80***
	(0.448)	(1.16)	(0.559)	(1.366)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	2,233	2,233	2,233	2,233

Source: Authors' calculations.

Note: i is the local state j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is log of urban population and that for columns (3) & (4) is log of urban households. The endogenous (instrumented) independent variables in all regressions are the log of in-migrants into urban areas of local state i from rural and urban areas of distant state j and the log of out-migration to urban and rural areas of distant state j from urban areas of local state i. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: **First Stage: Effect of Distant State Shocks on Urban Population and Households**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.017***	0.009***	0.022***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)
Highway upgrade at j	0.183***	0.105***	0.243***	0.140***
	(0.005)	(0.005)	(0.006)	(0.006)
Log mean per capita consumption (urban) at i		-0.029**		-0.040***
		(0.013)		(0.014)
Urban surface area of district i		4.68***		5.54***
		(0.179)		(0.211)
Urban surface area of district i squared		-4.79***		-5.76***
		(0.370)		(0.442)
Mean urban household size at i		-0.082***		-0.137***
		(0.006)		(0.007)
Median no. of rooms (urban) at i		0.005		0.020**
		(0.008)		(0.009)
Constant	12.0***	12.7***	10.1***	11.2***
	(0.026)	(0.094)	(0.030)	(0.107)
Hansen J-stat	0.126	0.220	0.144	0.228
Angrist-Pischke F-stat	2734***	494***	4049***	703***
N	14,552	14,552	14,552	14,552
Adj. R-sq	0.429	0.631	0.489	0.693

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log urban population and that in columns (3) & (4) is log urban households. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Column (4) also includes mean urban household size and the median no. of rooms in a urban household in i as additional controls. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: **Second Stage: Effect of Urban Population on Urban Residential Houses**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban population at $i^{\text{IV'd}}$	0.217***	0.082**	1.74***	1.78***	2.12***	2.40***
	(0.015)	(0.036)	(0.014)	(0.035)	(0.020)	(0.045)
Log mean per capita consumption (urban) at i		-0.250***		0.027		-0.210***
		(0.014)		(0.018)		(0.021)
Urban surface area of district i		1.16***		-1.35***		-2.80***
		(0.235)		(0.232)		(0.330)
Urban surface area of district i squared		-1.78***		1.14***		1.24***
		(0.345)		(0.297)		(0.459)
Mean urban household size at i		-0.084***		-0.036***		-0.013
		(0.009)		(0.007)		(0.010)
Median no. of rooms (urban) at i		-0.147***		0.042***		-0.025**
		(0.007)		(0.006)		(0.011)
Constant	7.16***	10.9***	-11.7***	-12.2***	-18.5***	-20.7***
	(0.201)	(0.523)	(0.178)	(0.494)	(0.256)	(0.624)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban population at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: **Second Stage: Effect of Urban Households on Urban Residential Houses**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban households at $i^{\text{IV'd}}$	0.164***	0.061**	1.308***	1.32***	1.60***	1.78***
	(0.012)	(0.027)	(0.007)	(0.019)	(0.013)	(0.026)
Log mean per capita consumption (urban) at i		-0.250***		0.029**		-0.206***
		(0.014)		(0.013)		(0.016)
Urban surface area of district i		1.21***		-0.349**		-1.45***
		(0.221)		(0.157)		(0.231)
Urban surface area of district i squared		-1.82***		0.242		0.026
		(0.341)		(0.196)		(0.315)
Mean urban household size at i		-0.082***		-0.001		0.035***
		(0.010)		(0.006)		(0.008)
Median no. of rooms (urban) at i		-0.148***		0.024***		-0.048***
		(0.007)		(0.004)		(0.007)
Constant	8.12***	11.3***	-3.96***	-4.32***	-9.13***	-10.1***
	(0.135)	(0.370)	(0.086)	(0.257)	(0.147)	(0.331)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban households at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: **First Stage: Effect of Distant State Shocks on Urban Residential Houses**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock at j	0.004***	0.002***	0.029***	0.021***	0.036***	0.026***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Highway upgrade at j	0.042***	0.027***	0.315***	0.224***	0.391***	0.296***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Log mean per capita consumption (urban) at i		-0.198***		0.055***		-0.181***
		(0.015)		(0.018)		(0.026)
Urban surface area of district i		2.23***		7.72***		9.40***
		(0.168)		(0.300)		(0.369)
Urban surface area of district i squared		-3.31***		-8.59***		-11.9***
		(0.364)		(0.628)		(0.756)
Constant	9.78***	10.8***	9.26***	8.95***	7.06***	7.94***
	(0.028)	(0.083)	(0.037)	(0.094)	(0.052)	(0.136)
Hansen J-stat	0.052	0.002	0.063	0.064	0.291	0.240
Angrist-Pischke F-stat	85***	27***	4541***	1730***	3166***	1268***
N	14,552	14,552	14,552	14,552	14,552	14,552
Adj. R-sq	0.032	0.093	0.505	0.663	0.452	0.567

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: **Second Stage: Inverse Elasticity of Urban Housing Supply**

	Dep. var.					
	Log(Urban non-dur. rent i)		Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban non-durable houses at $i^{\text{IV'd}}$	0.180	-1.84***				
	(0.275)	(0.548)				
Log urban durable houses at $i^{\text{IV'd}}$			0.576***	0.611***		
			(0.024)	(0.037)		
Log urban vacant houses at $i^{\text{IV'd}}$					0.434***	0.381***
					(0.018)	(0.026)
Log mean per capita consumption (urban) at i		-0.475***		0.451***		0.399***
		(0.124)		(0.041)		(0.034)
Urban surface area of district i		5.72***		-1.96***		-0.182
		(1.53)		(0.481)		(0.410)
Urban surface area of district i squared		-3.29		2.23***		1.75***
		(2.22)		(0.720)		(0.666)
Constant	3.55	25.8***	-0.338	-3.01***	1.91***	0.227
	(2.75)	(5.99)	(0.263)	(0.452)	(0.165)	(0.314)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552
Elasticity	NA	-0.545	1.74	1.64	2.30	2.63

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean of log rent for urban non-durable residential houses, for columns (3) & (4) it's mean of log rent for urban durable residential houses, and for columns (5) & (6) it's mean of log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: **Second Stage: Urban Durable and Vacant Housing Supply Elasticity (Full Sample)**

	Dep. var.			
	Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)
Log urban durable houses at $i^{\text{IV'd}}$	0.527***	0.727***		
	(0.018)	(0.029)		
Log urban vacant houses at $i^{\text{IV'd}}$			0.441***	0.563***
			(0.014)	(0.021)
Log mean per capita consumption (urban) at i		0.179***		0.127***
		(0.024)		(0.023)
Urban surface area of district i		-6.43***		-5.60***
		(0.453)		(0.415)
Urban surface area of district i squared		7.25***		7.78***
		(0.717)		(0.689)
Constant	0.291	-2.43***	1.96***	0.505**
	(0.194)	(0.310)	(0.126)	(0.208)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	27,098	27,098	27,098	27,098
Elasticity	1.90	1.38	2.27	1.78

Source: Authors' calculations.
Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)360$ in 2001 and $i = 1(1)437$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean log rent for urban durable residential houses, and for columns (3) & (4) it's mean log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is log of urban durable houses, and for columns (3) & (4) it's log of urban vacant residential houses. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: **First Stage: Effect of Distant State Shocks on Inter-state Long-term Migration**

	Dep. var.			
	Log(Urban in-migration to i)		Log(Urban out-migration from i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.012***	0.007**	-0.002	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
Highway upgrade at j	0.269***	0.234***	0.317***	0.297***
	(0.038)	(0.040)	(0.035)	(0.037)
Log mean per capita consumption (urban) at i		0.262		-0.457**
		(0.169)		(0.182)
Urban surface area of state i		-0.056*		0.101***
		(0.029)		(0.025)
Urban surface area of state i squared		0.013***		-0.008***
		(0.002)		(0.002)
Constant	5.47***	4.27***	5.95***	8.53***
	(0.162)	(0.935)	(0.143)	(1.035)
Angrist-Pischke F-stat	63***	24***	55***	33***
N	2,210	2,210	2,210	2,210
Adj. R-sq	0.087	0.127	0.089	0.102

Source: Authors' calculations.
Note: i is the state of enumeration and j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log of in-migration in the last 1-9 years into urban areas of enumeration state i from rural and urban areas of distant state j. Dependent variable in columns (3) & (4) is log of out-migration in the last 1-9 years from urban areas of enumeration state i to rural and urban areas of distant state j. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the last 1-9 years. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: **Second Stage: Effect of Inter-state Long-term Urban Migration on Urban Population and Households**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Log urban in-migration 1-9 yr. ago at $i^{\text{IV'd}}$	1.77***	1.90***	2.19***	2.22***
	(0.394)	(0.558)	(0.490)	(0.655)
Log urban out-migration 1-9 yr. ago at $j^{\text{IV'd}}$	-0.726	-0.791	-0.920	-0.930
	(0.476)	(0.527)	(0.593)	(0.618)
Log mean per capita consumption (urban) at i		-0.786		-0.949
		(0.497)		(0.583)
Urban surface area of state i		0.268***		0.336***
		(0.088)		(0.103)
Urban surface area of state i squared		-0.036***		-0.043***
		(0.011)		(0.013)
Constant	8.32***	12.1***	5.29***	10.1***
	(1.17)	(3.17)	(1.46)	(3.72)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	2,210	2,210	2,210	2,210

Source: Authors' calculations.

Note: i is the state of enumeration and j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is log of urban population and that for columns (3) & (4) is log of urban households. The endogenous (instrumented) independent variables in all regressions are the log of in-migrants in the last 1-9 years into urban areas of enumeration state i from rural and urban areas of distant state j and the log of out-migration in the last 1-9 years to urban and rural areas of distant state j from urban areas of enumeration state i. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the last 1-9 years. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: **First Stage: Effect of Distant Non-contiguous State Shocks on Non-contiguous State Urban Migration**

	Dep. var.			
	Log(Urban in-migration to i)		Log(Urban out-migration from i)	
	(1)	(2)	(3)	(4)
Rainfall shock at non-contiguous j	0.020***	0.014***	-0.001	-0.005*
	(0.003)	(0.004)	(0.003)	(0.003)
Highway upgrade at non-contiguous j	0.325***	0.290***	0.385***	0.351***
	(0.043)	(0.045)	(0.037)	(0.039)
Log mean per capita consumption (urban) at i		0.195		-0.372*
		(0.191)		(0.198)
Urban surface area of state i		-0.030		0.128***
		(0.031)		(0.027)
Urban surface area of state i squared		0.011***		-0.009***
		(0.003)		(0.002)
Constant	4.55***	3.82***	5.63***	7.84***
	(0.199)	(1.05)	(0.168)	(1.13)
Angrist-Pischke F-stat	92***	39***	66***	40***
N	1,987	1,987	1,987	1,987
Adj. R-sq	0.133	0.168	0.105	0.120

Source: Authors' calculations.
Note: A panel of non-contiguous (i, j) state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)35$ for each of the 35 point of enumeration states in India and j is a set of non-contiguous states corresponding to each state i. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log of in-migration last decade into urban areas of enumeration state i from rural and urban areas of distant non-contiguous state j. Dependent variable in columns (3) & (4) is log of out-migration last decade from urban areas of enumeration state i to rural and urban areas of distant non-contiguous state j. Two exogenous variables — rainfall shocks and highway construction at the distant non-contiguous state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: **Second Stage: Inverse Elasticity of Urban Housing Supply with Distant Non-Contiguous State Shock IVs**

	Dep. var.					
	Log(Urban non-dur. rent i)		Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban non-durable houses at $i^{\text{IV'd}}$	0.190	-1.87***				
	(0.302)	(0.607)				
Log urban durable houses at $i^{\text{IV'd}}$			0.561***	0.596***		
			(0.025)	(0.039)		
Log urban vacant houses at $i^{\text{IV'd}}$					0.427***	0.378***
					(0.019)	(0.027)
Log mean per capita consumption (urban) at i		-0.448***		0.438***		0.401***
		(0.135)		(0.044)		(0.037)
Urban surface area of district i		5.91***		-1.94***		-0.255
		(1.70)		(0.509)		(0.437)
Urban surface area of district i squared		-3.73		2.28***		1.80**
		(2.48)		(0.767)		(0.709)
Constant	3.46	25.9***	-0.151	-2.75***	1.99***	0.263
	(3.017)	(6.62)	(0.280)	(0.480)	(0.177)	(0.340)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	12,505	12,505	12,505	12,505	12,505	12,505
Elasticity	NA	-0.535	1.78	1.68	2.34	2.65

Source: Authors' calculations.

Note: A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for point of enumeration districts. j is a set of non-contiguous states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean of log rent for urban non-durable residential houses, for columns (3) & (4) it's mean of log rent for urban durable residential houses, and for columns (5) & (6) it's mean of log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant non-contiguous state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: **State-level Elasticity of Housing Supply**

State	Durable	Vacant
Maharashtra	3.06	3.43
Odisha	2.05	1.77
Tamil Nadu	1.92	2.10
Andhra Pradesh	1.63	1.30
Madhya Pradesh	1.25	1.26
Uttar Pradesh	1.18	1.40
Rajasthan	1.06	1.57
Karnataka	0.75	0.60
Bihar	0.49	0.79
West Bengal	0.39	0.49

Source: Authors' calculations.
Note: All reported states have observations on durable housing rents and quantities for at least 15 districts in both 2001 and 2011. States arranged in decreasing order of overall elasticity values. All elasticity values rounded off to two decimal places.

Appendix

Table A.1: **Second Stage: Effect of Non-contiguous State Urban Migration on Urban Population and Households**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Log urban in-migration last decade at $i^{\text{IV'd}}$	0.966***	0.973***	1.20***	1.45*
	(0.138)	(0.162)	(0.172)	(0.820)
Log urban out-migration last decade at $j^{\text{IV'd}}$	-0.131	-0.159	-0.190	-0.251
	(0.187)	(0.187)	(0.233)	(0.343)
Log mean per capita consumption (urban) at i		-0.140		-0.320
		(0.196)		(0.446)
Urban surface area of state i		0.149***		0.296
		(0.038)		(0.225)
Urban surface area of state i squared		-0.019***		-0.032
		(0.004)		(0.021)
Mean urban household size at i				0.415
				(0.909)
Constant	9.86***	10.7***	7.25***	5.69
	(0.507)	(1.29)	(0.633)	(6.38)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	1,987	1,987	1,987	1,987

Source: Authors' calculations.

Note: A panel of non-contiguous (i, j) state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)35$ for each of the 35 local states in India and j is a set of non-contiguous states corresponding to each state i. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is log of urban population and that for columns (3) & (4) is log of urban households. The endogenous (instrumented) independent variables in all regressions are the log of in-migrants into urban areas of local state i from rural and urban areas of distant non-contiguous state j and the log of out-migration to urban and rural areas of distant non-contiguous state j from urban areas of local state i. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. IVs include rainfall shocks and highway construction at the distant non-contiguous state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: **First Stage: Effect of Distant Non-contiguous State Shocks on Urban Population and Households**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Rainfall shock at non-contiguous j	0.018***	0.010***	0.024***	0.013***
	(0.000)	(0.000)	(0.001)	(0.001)
Highway upgrade at non-contiguous j	0.178***	0.111***	0.231***	0.144***
	(0.006)	(0.006)	(0.007)	(0.006)
Log mean per capita consumption (urban) at i		-0.026*		-0.039**
		(0.014)		(0.016)
Urban surface area of district i		4.72***		5.58***
		(0.195)		(0.229)
Urban surface area of district i squared		-4.93***		-5.90***
		(0.401)		(0.476)
Mean urban household size at i		-0.079***		-0.133***
		(0.006)		(0.008)
Median no. of rooms (urban) at i		-0.001		0.013
		(0.009)		(0.010)
Constant	11.0***	12.6***	10.0***	11.1***
	(0.027)	(0.103)	(0.031)	(0.117)
Hansen J-stat	3.76*	3.82*	4.11**	3.92**
Angrist-Pischke F-stat	2297***	445***	3373***	633***
N	12,505	12,505	12,505	12,505
Adj. R-sq	0.427	0.626	0.488	0.688

Source: Authors' calculations.
Note: A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. j is a set of non-contiguous states outside the state in which district i is located. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log urban population and that in columns (3) & (4) is log urban households. Two exogenous variables — rainfall shocks and highway construction at the distant non-contiguous state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. Column (4) also includes mean urban household size and the median no. of rooms in an urban household in i as additional controls. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: **Second Stage: Effect of Urban Population on Urban Residential Houses Using Non-Contiguous State Shock IVs**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban population at $i^{\text{IV'd}}$	0.206***	0.074*	1.73***	1.76***	2.11***	2.36***
	(0.016)	(0.038)	(0.015)	(0.037)	(0.021)	(0.046)
Log mean per capita consumption (urban) at i		-0.246***		0.026		-0.226***
		(0.016)		(0.020)		(0.023)
Urban surface area of district i		1.23***		-1.31***		-2.64***
		(0.252)		(0.248)		(0.348)
Urban surface area of district i squared		-1.90***		1.15***		1.18**
		(0.375)		(0.320)		(0.486)
Mean urban household size at i		-0.085***		-0.041***		-0.017*
		(0.010)		(0.008)		(0.010)
Median no. of rooms (urban) at i		-0.145***		0.045***		-0.018
		(0.008)		(0.006)		(0.012)
Constant	7.28***	10.0***	-11.6***	-11.9***	-18.3***	-20.0***
	(0.216)	(0.553)	(0.195)	(0.524)	(0.275)	(0.650)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	12,505	12,505	12,505	12,505	12,505	12,505

Source: Authors' calculations.

Note: A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. j is a set of non-contiguous states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban population at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant non-contiguous state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: **Second Stage: Effect of Urban Households on Urban Residential Houses Using Non-Contiguous State Shock IVs**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban households at $i^{\text{IV'd}}$	0.155***	0.055*	1.31***	1.32***	1.59***	1.76***
	(0.013)	(0.029)	(0.008)	(0.021)	(0.014)	(0.028)
Log mean per capita consumption (urban) at i		-0.246***		0.031**		-0.221***
		(0.016)		(0.015)		(0.018)
Urban surface area of district i		1.28***		-0.341**		-1.33***
		(0.238)		(0.170)		(0.245)
Urban surface area of district i squared		-1.95***		0.242		-0.049
		(0.370)		(0.213)		(0.336)
Mean urban household size at i		-0.084***		-0.004		0.031***
		(0.010)		(0.006)		(0.009)
Median no. of rooms (urban) at i		-0.146***		0.025***		-0.044***
		(0.008)		(0.004)		(0.008)
Constant	8.19***	11.3***	-3.97***	-4.26***	-9.02***	-9.76***
	(0.146)	(0.394)	(0.095)	(0.277)	(0.159)	(0.349)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	12,505	12,505	12,505	12,505	12,505	12,505

Source: Authors' calculations.

Note: A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. j is a set of non-contiguous states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban households at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant non-contiguous state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: **First Stage: Effect of Distant Non-contiguous State Shocks on Urban Residential Houses**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock at non-contiguous j	0.003***	0.001***	0.031***	0.022***	0.036***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Highway upgrade at non-contiguous j	0.049***	0.034***	0.296***	0.217***	0.390***	0.306***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.012)	(0.012)
Log mean per capita consumption (urban) at i		-0.194***		0.055***		-0.196***
		(0.016)		(0.019)		(0.028)
Urban surface area of district i		2.29***		7.76***		9.49***
		(0.182)		(0.323)		(0.399)
Urban surface area of district i squared		-3.42***		-8.78***		-12.1***
		(0.395)		(0.673)		(0.813)
Constant	9.79***	10.8***	9.12***	8.84***	6.99***	7.99***
	(0.030)	(0.091)	(0.039)	(0.102)	(0.054)	(0.149)
Hansen J-stat	0.568	0.258	0.953	0.676	6.72***	5.16**
Angrist-Pischke F-stat	67***	21***	3816***	1522***	2614***	1089***
N	12,505	12,505	12,505	12,505	12,505	12,505
Adj. R-sq	0.030	0.089	0.507	0.665	0.448	0.564

Source: Authors' calculations.

Note: A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. j is a set of non-contiguous states outside the state in which district i is located. Results from six FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Two exogenous variables — rainfall shocks and highway construction at the distant non-contiguous state j — are independent variables. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: **Second Stage: Urban Durable and Vacant Housing Supply Elasticity Estimation with Non-Contiguous State Shock IVs**

	Dep. var.			
	Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)
Log urban durable houses at $i^{\text{IV'd}}$	0.511***	0.702***		
	(0.019)	(0.031)		
Log urban vacant houses at $i^{\text{IV'd}}$			0.430***	0.547***
			(0.015)	(0.023)
Log mean per capita consumption (urban) at i		0.174***		0.124***
		(0.026)		(0.025)
Urban surface area of district i		-6.22***		-5.44***
		(0.486)		(0.446)
Urban surface area of district i squared		7.09***		7.60***
		(0.766)		(0.735)
Constant	0.494**	-2.12***	2.07***	0.668***
	(0.208)	(0.333)	(0.135)	(0.226)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	23,054	23,054	23,054	23,054
Elasticity	1.96	1.43	2.33	1.83

Source: Authors' calculations.
Note: i is the district of enumeration and j is the distant non-contiguous state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)360$ in 2001 and $i = 1(1)437$ in 2011 for local districts. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean log rent for urban durable residential houses, and for columns (3) & (4) it's mean log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is log of urban durable houses, and for columns (3) & (4) it's log of urban vacant residential houses. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant non-contiguous state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: **First Stage: Effect of Distant State Shocks on Inter-state Urban Migration (Consumption at j Controlled)**

	Dep. var.			
	Log(Urban in-migration to i)		Log(Urban out-migration from i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.020***	0.015***	0.002	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Highway upgrade at j	0.323***	0.255***	0.391***	0.330***
	(0.038)	(0.042)	(0.033)	(0.037)
Log mean per capita consumption (urban) at i		0.115		-0.399**
		(0.178)		(0.189)
Log mean per capita consumption at j		0.416**		0.423***
		(0.174)		(0.147)
Urban surface area of state i		-0.041		0.109***
		(0.030)		(0.028)
Urban surface area of state i squared		0.011***		-0.008***
		(0.002)		(0.002)
Constant	4.99***	2.47*	5.91***	6.05***
	(0.182)	(1.340)	(0.156)	(1.429)
Angrist-Pischke F-stat	117***	40***	96***	41***
N	2,233	2,233	2,233	2,233
Adj. R-sq	0.144	0.186	0.126	0.151

Source: Authors' calculations.
Note: i is the local state j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log of in-migration last decade into urban areas of local state i from rural and urban areas of distant state j. Dependent variable in columns (3) & (4) is log of out-migration last decade from urban areas of local state i to rural and urban areas of distant state j. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: **Second Stage: Effect of Inter-state Urban Migration on Urban Population and Households (Consumption at j Controlled)**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Log urban in-migration last decade at $i^{\text{IV'd}}$	0.967***	0.936***	1.20***	1.10***
	(0.146)	(0.152)	(0.181)	(0.178)
Log urban out-migration last decade at $j^{\text{IV'd}}$	-0.136	-0.184	-0.189	-0.222
	(0.183)	(0.172)	(0.228)	(0.202)
Log mean per capita consumption (urban) at i		-0.095		-0.141
		(0.170)		(0.199)
Log mean per capita consumption at j		0.540***		0.638***
		(0.165)		(0.194)
Urban surface area of state i		0.108***		0.149***
		(0.034)		(0.040)
Urban surface area of state i squared		-0.016***		-0.019***
		(0.004)		(0.004)
Constant	9.62***	7.69***	6.88***	5.02***
	(0.448)	(1.20)	(0.559)	(1.42)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	2,233	2,233	2,233	2,233

Source: Authors' calculations.
Note: i is the local state j is the distant state. An unbalanced panel of 1,108 (i, j) state pairs for 2001 and 1,125 (i, j) state pairs for 2011 used for analysis. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is log of urban population and that for columns (3) & (4) is log of urban households. The endogenous (instrumented) independent variables in all regressions are the log of in-migrants into urban areas of local state i from rural and urban areas of distant state j and the log of out-migration to urban and rural areas of distant state j from urban areas of local state i. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in state i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: **First Stage: Effect of Distant State Shocks on Urban Population and Households (Consumption at j Controlled)**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.017***	0.009***	0.022***	0.012***
	(0.000)	(0.000)	(0.000)	(0.000)
Highway upgrade at j	0.183***	0.088***	0.243***	0.118***
	(0.005)	(0.005)	(0.006)	(0.005)
Log mean per capita consumption (urban) at i		-0.027**		-0.038***
		(0.012)		(0.013)
Log mean per capita consumption at j		0.466***		0.628***
		(0.020)		(0.022)
Urban surface area of district i		4.15***		4.83***
		(0.168)		(0.193)
Urban surface area of district i squared		-4.24***		-5.02***
		(0.342)		(0.397)
Mean urban household size at i		-0.053***		-0.098***
		(0.006)		(0.006)
Median no. of rooms (urban) at i		-0.004		0.008
		(0.007)		(0.008)
Constant	12.0***	10.1***	10.1***	7.70***
	(0.026)	(0.146)	(0.030)	(0.161)
Hansen J-stat	0.126	0.218	0.144	0.266
Angrist-Pischke F-stat	2734***	475***	4049***	704***
N	14,552	14,552	14,552	14,552
Adj. R-sq	0.429	0.669	0.489	0.737

Source: Authors' calculations.
Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log urban population and that in columns (3) & (4) is log urban households. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Column (4) also includes mean urban household size and the median no. of rooms in an urban household in i as additional controls. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: **Second Stage: Effect of Urban Population on Urban Residential Houses (Consumption at j Controlled)**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban population at $i^{\text{IV'd}}$	0.217***	0.081**	1.74***	1.78***	2.12***	2.40***
	(0.015)	(0.039)	(0.014)	(0.037)	(0.020)	(0.049)
Log mean per capita consumption (urban) at i		-0.250***		0.027		-0.210***
		(0.014)		(0.018)		(0.021)
Log mean per capita consumption at j		0.001		0.004		0.007
		(0.030)		(0.027)		(0.040)
Urban surface area of district i		1.17***		-1.36***		-2.80***
		(0.231)		(0.228)		(0.325)
Urban surface area of district i squared		-1.79***		1.15***		1.25***
		(0.342)		(0.294)		(0.454)
Mean urban household size at i		-0.084***		-0.036***		-0.012
		(0.009)		(0.007)		(0.009)
Median no. of rooms (urban) at i		-0.147***		0.042***		-0.025**
		(0.007)		(0.006)		(0.011)
Constant	7.16***	10.9***	-11.7***	-12.2***	-18.5***	-20.7***
	(0.201)	(0.486)	(0.178)	(0.461)	(0.256)	(0.568)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban population at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption at j and mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: **Second Stage: Effect of Urban Households on Urban Residential Houses (Consumption at j Controlled)**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban households at $i^{\text{IV'd}}$	0.164***	0.060**	1.31***	1.32***	1.60***	1.78***
	(0.012)	(0.029)	(0.007)	(0.021)	(0.013)	(0.029)
Log mean per capita consumption (urban) at i		-0.250***		0.029**		-0.206***
		(0.014)		(0.013)		(0.016)
Log mean per capita consumption at j		0.001		0.003		0.005
		(0.030)		(0.020)		(0.033)
Urban surface area of district i		1.21***		-0.355**		-1.45***
		(0.216)		(0.154)		(0.225)
Urban surface area of district i squared		-1.83***		0.247		0.025
		(0.337)		(0.192)		(0.310)
Mean urban household size at i		-0.082***		-0.001		0.035***
		(0.009)		(0.006)		(0.008)
Median no. of rooms (urban) at i		-0.148***		0.024***		-0.049***
		(0.007)		(0.004)		(0.007)
Constant	8.12***	11.3***	-3.60***	-4.34***	-9.13***	-10.1***
	(0.135)	(0.335)	(0.086)	(0.234)	(0.147)	(0.281)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban households at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption at j and mean urban household size, median no. of rooms (urban), log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: **First Stage: Effect of Distant State Shocks on Urban Residential Houses (Consumption at j Controlled)**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock at j	0.004***	0.002***	0.029***	0.019***	0.036***	0.024***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Highway upgrade at j	0.042***	0.022***	0.315***	0.177***	0.391***	0.235***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Log mean per capita consumption (urban) at i		-0.201***		0.032**		-0.210***
		(0.015)		(0.016)		(0.024)
Log mean per capita consumption at j		0.109***		0.967***		1.258***
		(0.024)		(0.027)		(0.040)
Urban surface area of district i		2.08***		6.39***		7.68***
		(0.168)		(0.264)		(0.322)
Urban surface area of district i squared		-3.14***		-7.07***		-9.88***
		(0.359)		(0.545)		(0.648)
Constant	9.78***	10.2***	9.26***	4.12***	7.06***	1.66***
	(0.028)	(0.151)	(0.037)	(0.149)	(0.052)	(0.218)
Hansen J-stat	0.052	0.002	0.063	0.066	0.291	0.243
Angrist-Pischke F-stat	85***	19***	4541***	1302***	3166***	951***
N	14,552	14,552	14,552	14,552	14,552	14,552
Adj. R-sq	0.032	0.097	0.505	0.731	0.452	0.636

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: **Second Stage: Inverse Elasticity of Urban Housing Supply (Consumption at j Controlled)**

	Dep. var.					
	Log(Urban non-dur. rent i)		Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban non-durable houses at $i^{\text{IV'd}}$	0.180	-1.83***				
	(0.275)	(0.655)				
Log urban durable houses at $i^{\text{IV'd}}$			0.576***	0.611***		
			(0.024)	(0.044)		
Log urban vacant houses at $i^{\text{IV'd}}$					0.434***	0.381***
					(0.018)	(0.031)
Log mean per capita consumption (urban) at i		-0.475***		0.451***		0.399***
		(0.144)		(0.040)		(0.035)
Log mean per capita consumption at j		-0.002		0.002		0.000
		(0.135)		(0.083)		(0.075)
Urban surface area of district i		5.72***		-1.96***		-0.191
		(1.64)		(0.471)		(0.399)
Urban surface area of district i squared		-3.29		2.24***		1.76***
		(2.42)		(0.709)		(0.658)
Constant	3.55	25.8***	-0.338	-3.03***	1.91***	0.220
	(2.75)	(6.69)	(0.263)	(0.432)	(0.165)	(0.350)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552
Elasticity	NA	-0.545	1.74	1.64	2.30	2.63

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean of log rent for urban non-durable residential houses, for columns (3) & (4) it's mean of log rent for urban durable residential houses, and for columns (5) & (6) it's mean of log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: **Second Stage: Urban Durable and Vacant Housing Supply Elasticity Estimation with Full Sample (Consumption at j Controlled)**

	Dep. var.			
	Log(Urban dur. rent i)		Log(Urban avg. rent i)	
	(1)	(2)	(3)	(4)
Log urban durable houses at $i^{\text{IV'd}}$	0.527***	0.727***		
	(0.018)	(0.035)		
Log urban vacant houses at $i^{\text{IV'd}}$			0.441***	0.563***
			(0.014)	(0.026)
Log mean per capita consumption (urban) at i		0.179***		0.127***
		(0.024)		(0.023)
Log mean per capita consumption at j		0.001		-0.001
		(0.061)		(0.060)
Urban surface area of district i		-6.43***		-5.60***
		(0.446)		(0.407)
Urban surface area of district i squared		7.26***		7.79***
		(0.706)		(0.680)
Constant	0.291	-2.44***	1.96***	0.504**
	(0.194)	(0.289)	(0.126)	(0.253)
Rainfall IV	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes
N	27,098	27,098	27,098	27,098
Elasticity	1.90	1.38	2.27	1.78

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)360$ in 2001 and $i = 1(1)437$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from four panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the mean log rent for urban durable residential houses, and for columns (3) & (4) it's mean log rent for all urban residential houses. The endogenous independent variables for columns (1) & (2) is log of urban durable houses, and for columns (3) & (4) it's log of urban vacant residential houses. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption at j and log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: **First Stage: Effect of Distant State Shocks on Urban Population and Households (without Median Rooms)**

	Dep. var.			
	Log(Urban population i)		Log(Urban households i)	
	(1)	(2)	(3)	(4)
Rainfall shock at j	0.017***	0.009***	0.022***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)
Highway upgrade at j	0.183***	0.105***	0.243***	0.141***
	(0.005)	(0.005)	(0.006)	(0.006)
Log mean per capita consumption (urban) at i		-0.029**		-0.041***
		(0.013)		(0.014)
Urban surface area of district i		4.67***		5.51***
		(0.178)		(0.209)
Urban surface area of district i squared		-4.77***		-5.71***
		(0.369)		(0.439)
Mean urban household size at i		-0.082***		-0.135***
		(0.006)		(0.007)
Constant	12.0***	12.7***	10.1***	11.2***
	(0.026)	(0.091)	(0.030)	(0.103)
Hansen J-stat	0.126	0.253	0.144	0.255
Angrist-Pischke F-stat	2734***	513***	4049***	722***
N	14,552	14,552	14,552	14,552
Adj. R-sq	0.429	0.631	0.489	0.692

Source: Authors' calculations.
Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from four FE regressions presented. Dependent variable in columns (1) & (2) is log urban population and that in columns (3) & (4) is log urban households. Two exogenous variables — rainfall shocks and highway construction at the distant state j — are independent variables. Columns (1) & (3) does not include controls while columns (2) & (4) does. Controls are log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. Column (4) also includes mean urban household size in i as additional controls. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: **Second Stage: Effect of Urban Population on Urban Residential Houses (without Median Rooms)**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban population at $i^{\text{IV'd}}$	0.217***	0.025	1.74***	1.80***	2.12***	2.39***
	(0.015)	(0.038)	(0.014)	(0.034)	(0.020)	(0.043)
Log mean per capita consumption (urban) at i		-0.244***		0.025		-0.208***
		(0.015)		(0.018)		(0.021)
Urban surface area of district i		1.66***		-1.49***		-2.72***
		(0.257)		(0.231)		(0.323)
Urban surface area of district i squared		-2.44***		1.33***		1.13**
		(0.385)		(0.296)		(0.453)
Mean urban household size at i		-0.099***		-0.032***		-0.015
		(0.009)		(0.007)		(0.009)
Constan	7.16***	11.4***	-11.7***	-12.3***	-18.5***	-20.6***
	(0.201)	(0.539)	(0.178)	(0.495)	(0.256)	(0.613)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.

Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban population at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 × highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.17: **Second Stage: Effect of Urban Households on Urban Residential Houses (without Median Rooms)**

	Dep. var.					
	Log(Urban non-dur. houses i)		Log(Urban dur. houses i)		Log(Urban vact. houses i)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log urban households at $i^{\text{IV'd}}$	0.164***	0.019	1.31***	1.33***	1.60***	1.77***
	(0.012)	(0.028)	(0.007)	(0.019)	(0.013)	(0.026)
Log mean per capita consumption (urban) at i		-0.244***		0.028**		-0.204***
		(0.015)		(0.013)		(0.016)
Urban surface area of district i		1.67***		-0.426***		-1.30***
		(0.241)		(0.156)		(0.227)
Urban surface area of district i squared		-2.45***		0.346*		-0.180
		(0.377)		(0.193)		(0.312)
Mean urban household size at i		-0.098***		0.001		0.030***
		(0.010)		(0.006)		(0.008)
Constant	8.12***	11.5***	-3.96***	-4.36***	-9.13***	-10.0***
	(0.135)	(0.377)	(0.086)	(0.257)	(0.147)	(0.327)
Rainfall IV	Yes	Yes	Yes	Yes	Yes	Yes
Highway IV	Yes	Yes	Yes	Yes	Yes	Yes
N	14,552	14,552	14,552	14,552	14,552	14,552

Source: Authors' calculations.
Note: i is the district of enumeration and j is the distant state. A panel of (i, j) district-state pairs for 2001 and 2011 Census years used for analysis. $i = 1(1)220$ in 2001 and $i = 1(1)208$ in 2011 for local districts. $j = 1(1)34$ for each of the 34 distant states outside the state in which district i is located. Results from six panel IV FE regressions reported. The dependent variables for columns (1) & (2) is the log of urban non-durable houses, for columns (3) & (4) it's log of urban durable houses, and for columns (5) & (6) it's log of urban vacant residential houses. The endogenous independent variable in all regressions is log of urban households at i. Columns (1), (3) & (5) does not include controls while columns (2), (4) & (6) does. Controls are mean urban household size, log mean monthly per capita consumption (urban), urban surface area and the urban surface area squared in district i. IVs include rainfall shocks and highway construction at the distant state j. Rainfall shocks are defined as the number of months when rainfall was less than 80% of long-term normal during the previous decade. Highway upgrade at j is defined as Post 2001 $\times$ highway upgrade dummy for j. Urban surface area unit in 1000 sq. miles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$