Cities in Bad Shape: Urban Geometry in India*

Mariaflavia Harari†

MIT

This version: 15 October 2014.

Preliminary and incomplete. Please do not cite or circulate.

Abstract

The spatial layout of cities is an important determinant of urban commuting efficiency. This paper investigates the economic implications of urban geometry in the context of India. A satellite-derived dataset of night-time lights is combined with historic maps to retrieve the geometric properties of urban footprints in India over time. My approach relies on a novel city-year level instrument for urban shape, which combines geography with a mechanical model for city expansion. I investigate how city shape affects the location choices of consumers, in a spatial equilibrium framework à la Roback-Rosen. Cities with more compact shapes are characterized by larger population, lower wages, and higher housing rents, consistent with compact shape being a consumption amenity. The implied welfare cost of deteriorating city shape is estimated to be sizeable. I also attempt to shed light on policy responses to deteriorating shape. The adverse effects of unfavorable topography appear to be exacerbated by building height restrictions, and mitigated by road infrastructure.

*I am grateful to my advisers Esther Duflo, Ben Olken and Dave Donaldson for their invaluable help throughout this project. I also thank Alex Bartik, Jie Bai, Nathaniel Baum-Snow, Alain Bertaud, Melissa Dell, John Firth, Ludovica Gazzè, Michael Greenstone, Gabriel Kreindler, Matthew Lowe, Rachael Meager, Yuhei Miyauchi, Hoai-Luu Nguyen, Paul Novosad, Arianna Ornaghi, Bimal Patel, Champaka Rajagopal, Otis Ried, Adam Sacarny, Albert Saiz, Ashish Shenoy, Chris Small, Kala Sridhar, William Wheaton and participants to the MIT Development Lunch, Applied Micro Lunch and Development Seminar for helpful comments and discussions at various stages of this work.

†harari@mit.edu

1 Introduction

Most urban economics assumes implicitly that cities are circular or star-shaped. Real-world cities, however, often depart significantly from this assumption: cities display considerable variation in shape. Urban geometries are jointly determined by a combination of geography and policy (Bertaud, 2004). On the one hand, existing natural and topographic constraints prevent cities from expanding radially in all directions. On the other, regulation and infrastructural investment can affect urban expansion both directly and indirectly. For instance, master plans can explicitly promote polycentricity, by planning satellite centers around existing ones. Land use regulations can encourage land consolidation, resulting in a more compact, as opposed to fragmented development pattern. Investments in road infrastructure can encourage urban growth along transport corridors.

While the economics literature has devoted very little attention to the geometry of cities, in fields such as urban planning and transportation engineering city shape is considered an important determinant of commuting efficiency, more compact shapes being characterized by shorter trips and more cost-effective transport networks. This, in turn, is claimed to affect productivity and welfare. Emphasizing the role of city shape in the functioning of public and private transport, Bertaud (2004) argues that contiguous and compact urban development can improve the welfare of city dwellers, by providing better and cheaper access to most of the jobs. This is thought to be particularly relevant in the megacities of the developing world, where most inhabitants cannot afford individual means of transportation or where the large size of the city precludes walking as a mean of getting to jobs. Cervero (2002) emphasizes how compact, accessible cities are potentially more productive, through a combination of labor market pooling, savings in transporting inputs and technological and information spillovers. This urban planning literature is mostly descriptive

In this paper I investigate empirically the economic implications of city shape in the context of India. More specifically, I examine two complementary sets of questions: first, how city geometry affects the location choices of consumers, in the framework of spatial equilibrium across cities, and what the welfare implications of deteriorating city shape are. Second, how policy interacts with geography in determining urban form.

India represents a promising setting to study urban spatial structures for a number of reasons. First, as most developing countries, India is experiencing fast urban growth. According to the

2011 Census, the urban population amounts to 377 million, increasing from 285 million in 2001 and 217 million in 1991, representing between 25 and 31 percent of the total. Although the pace of urbanization is slower than in other Asian countries, it is accelerating, and it is predicted that another 250 million will join the urban ranks by 2030 (Mc Kinsey, 2010). This growth in population has been accompanied by a significant physical expansion of urban footprints, typically beyond urban administrative boundaries (Indian Institute of Human Settlements, 2013; World Bank, 2013). This setting thus provides a unique opportunity to observe the shapes of cities as they evolve and expand over time. Such an exercise would not be feasible in a developed context: urban spatial structures in developed countries are typically very resilient and path dependent (Bertaud, 2004), which would prevent me from detecting significant variation over time.

Secondly, India is characterized by a diffused urbanization pattern. With an unusually large number of highly-populated cities, India contrasts from other developing countries characterized by the presence of a large capital city and very few other urban centers. This ensures a large enough sample of cities for the econometric analyses which I carry out.

Moreover, the challenges posed by urban expansion have been gaining increasing importance in India's policy discourse, which makes it particularly relevant to investigate these matters from an economics perspective. On the one hand, this debate has focused on the perceived harms of haphazardous urban expansion, most notably limited urban mobility and lengthy commutes. According to a recent Mc Kinsey report (2010), the average peak morning commute in Indian million-plus cities is in excess of one-and-a-half to two hours. Providing effective urban public transit systems has been consistently identified as a key policy recommendation for the near future (e.g. Mitric and Chatterton, 2005; World Bank, 2013; Mc Kinsey, 2013.). On the other hand, there is a growing concern that existing policies – in particular, urban land use regulations – might directly or indirectly contribute to distorting urban form (World Bank, 2013, Sridhar, 2010, Glaeser, 2011). The most paradigmatic of these measures are restrictions on building height, in the form of Floor Area Ratios (FARs). FARs in Indian cities are considered very restrictive compared to international standards, and have been indicated as one of the causes of sprawl in Indian cities (Brueckner and Sridhar, 2012; Bertaud and Brueckner, 2005). Another example is given by the Urban Land Ceiling and Regulation Act, which has been claimed to hinder intra-urban land consolidation and restrict the supply of land available for development within cities (Sridhar, 2010).

My broad research question concerns the economic implications of city shape. In particular, in this project I am interested in two complementary sets of issues. The first and main question relates to how city shape affects the location choices of consumers across cities. I investigate this in the framework of a model of spatial equilibrium across cities à la Roback-Rosen, modelling city shape as an amenity. My findings are broadly consistent with "good city shape" being a consumption amenity. I find robust evidence that cities with more compact shapes grow faster, all else being equal. There is also some evidence that consumers are paying a premium in terms of lower wages and higher rents in order to live in cities with better geometries. I estimate the implied welfare cost of deteriorating shape, finding that it is sizeable and considerably larger than the mere monetary cost associated to lengthier commutes.

A natural question which arises next concerns the role of policy: if city shape indeed has welfare implications, what is the optimal policy response to existing topographic constraints? Although a thorough investigation of this question might require more detailed data than those used in this study, in the concluding part of this paper I make a preliminary attempt to explore the interactions between topography, regulation and city shape. I find that restrictions on building height - in the form of low FARs - result in cities which are more spread out in space and less compact than their topography would predict. This is consistent with one of the most common arguments against restrictive FARs in India, namely that they cause sprawl by preventing cities from growing vertically (Glaeser, 2011).

My methodology can be outlined as follows. The bulk of my empirical analysis is conducted at the city-year level and is based on an instrumental variable approach. I assemble an unbalanced panel dataset of urban footprints for about 400 Indian cities, by combining historic maps (1950) with a satellite-derived dataset of night-time lights (1992-2010). I compute several geometric indicators for urban shape, used in urban planning as proxies for the patterns of within-city trips. I develop an original instrument for city shape by combining geography with a mechanical model for city expansion. The underlying idea is that, as cities expand in space, they face different geographic constraints - steep terrain or water bodies - leading to departures from an ideal circular expansion path. The construction of my instrument requires two steps. First, I use a mechanical model for city expansion to predict the area which a city should occupy in a given year. I consider two versions of this model: one based on each city's historic population growth rates, and one postulating a common growth rate for all cities. Second, I consider the largest contiguous patch of developable land within this predicted radius ("potential footprint") and compute its shape properties. I then proceed to instrument the shape properties of the *actual* city footprint in that given year with the shape properties of the *potential* footprint. The resulting instrument varies at the city-year level, allowing me to control for time-invariant city characteristics through city fixed effects. The identification of the impact of shape thus relies on comparing geography-driven *changes* in urban shape over time for each city. My main results are summarized below.

(i) I find a strong, robust first stage linking the *actual* shape that a city's footprint has at a given point in time and the *potential* shape which it can have given the layout of the geographic constraints surrounding it. This is not solely driven by extremely constrained topographies (e.g. coastal or mountainous cities) in my sample.

(ii) There is a strong and robust negative IV relationship between population and non-compactness, conditional on area. My most conservative estimate indicates that a one standard-deviation deterioration in city shape entails decrease in population density of 0.9 standard deviations.

(iii) There is suggestive evidence that households are paying for compact geometry both through higher housing prices and through lower wages. Using district level averages as proxies for city-level data, I find a marginally significant negative relationship between non-compactness and rents per square meter, and a positive relationship between non-compactness and wages. Although noisy, these results are consistent with "good shape" being valued a consumption amenity. Using these estimates in conjunction with a simple version of the Rosen-Roback model, I find that a one standard deviation deterioration in city shape entails a welfare loss equivalent to a 5% decrease in income.

(iv) Interacting city shape with different proxies for infrastructure, I find that the negative effects of poor geometry on population growth are mitigated by a denser road network and by the availability of motor vehicles, suggesting that commute times could be the main mechanism through which city shape affects consumers.

(v) Cities with poorer shapes are characterized by comparatively fewer slum dwellers and houses of better quality. This is consistent with the interpretation that low-income immigrants tend to sort into more compact cities, possibly because they are less able to cope with lengthy commutes.

The final part of my work attempts to look at the endogenous responses to unfavorable city shape. Due to data limitations, most of these results are based on a cross-section of cities only, and should therefore be interpreted as suggestive.

(vi) Restrictive FARs result in footprints which are both larger and less compact. Higher FARs, on the other hand, mitigate the negative effects of poor geometry on population growth.

(vii) Cities with worse geometries appear to have a less dense road network in absolute terms, but road density per capita is not significantly different from that observed in more compact cities.

(viii) Productive establishments appear less spatially concentrated in non-compact cities. This could reflect a tendency of cities to become less monocentric as their shape deteriorates.

This projects contributes to the existing literature in a number of directions. The first concerns the analysis of urban geometry from an economics standpoint. To my knowledge, this is the first work which studies city shape as an amenity. The second is methodological and relates to the measurement of the properties of urban footprints. I employ the OLS night-time lights dataset to track the shape of urban footprints over time, following an approach which has been used in urban remote sensing, but not in economics. The use of shape metrics borrowed from the urban planning literature to measure economically relevant properties of urban footprints is also novel. Furthermore, I develop an original instrument for city shape, which is potentially applicable to other contexts as well.

The remainder of the paper is organized as follows. Section 2 briefly reviews the existing literature. Section 3 outlines the conceptual framework. Section 4 documents my data sources and describes the geometric indicators which I employ. Section 5 presents my empirical strategy and describes in detail how my instrument is constructed. The empirical evidence is presented in the following two Sections. Section 6 discusses my main results, which pertain to the effects of city shape on the location choices of consumers. Section 7 provides some results on other endogenous responses to city shape, including interactions between topography and policy. Section 8 concludes with indications for future work.

2 Previous Literature

The economics literature on urban spatial structures has mostly focused on two aspects of urban form: city size and population density gradient. The classic monocentric city model (Alonso, 1964; Mills, 1967, 1972; Muth, 1969) postulates the existence of a single employment center and predicts the formation of a circular city with declining density gradient, and whose radius declines with transportation costs. Richer polycentric city models were subsequent developed, where the number, location and spatial extent of employment centers are determined endogenously in a linear or circular city (Fujita and Ogawa, 1982, Lucas and Rossi-Hansberg, 2002). Although these models can in principle be altered to generate non-circular cities, for instance, assuming the road network is not radial (Brueckner, 2011, p. 56), the focus of this literature is on the determinants of the internal structure of cities, and the implications of geometry are left unexplored. My approach is different since I take the shape of each city as given and I investigate its effects on the spatial equilibrium across, rather than within cities.

The conceptual framework I draw upon is that on spatial equilibrium across cities, after Rosen(1979) and Roback(1982). In the simplest version of this model, consumers and firms endogenously choose to locate in cities with different levels of amenities, and wages and rents adjust to equalize utility across locations. I model "compact city shape" as an amenity, which affects consumers and firms primarily by reducing the length of within-city trips. Although Glaeser (2008) acknowledges that amenities presumably include also commuting times, I am not aware of studies explicitly looking at the link between commuting, urban geometry and the spatial equilibrium across cities. A large empirical literature has employed the classic Rosen-Roback framework to investigate the value of amenities (e.g. Blomquist, 1988). This literature, however, has almost exclusively focused on the US, and has not always adequately addressed the endogeneity of amenities (Gyourko, Kahn and Tracy, 1999).

A large empirical literature investigates urban sprawl (Glaeser and Kahn, 2004; Burchfield et al, 2006), typically in the context of the US. Although there is not a single definition of sprawl, some studies identify sprawl with non-contiguous development (Burchfield et al, 2006), which is somewhat related to the notion of "compactness" which I investigate. In most analyses of sprawl, however, the focus is on decentralization and density. I focus on a different set of spatial properties of urban footprints: conditional on the overall amount of land used, I look at geometric properties aimed at proxying the pattern of within-city trips. My work is more closely related to that of Bento et al (2005), who incorporate a measure of city shape in their investigation of the link between urban form and travel demand. However, their analysis is based on a cross-section of US cities and does not address the endogeneity of city shape.

The geometric indicators of city shape which I employ are borrowed from the urban planning and landscape ecology literature (Angel and Civco, 2009; Angel et al., 2009), which is mostly descriptive. Urban planners emphasize the link between city shape, average intra-urban trip length and accessibility, claiming that contiguous, compact and predominantly monocentric urban morphologies are more favorable to transit. Bertaud (2004) argues that more compact cities are generally more favorable to the poor, since they provide better access to jobs, but at the same time have higher land prices, which would tend to reduce their housing floor space consumption. Cervero (2002) emphasizes how compact, accessible cities are potentially more productive, through a combination of labor market pooling, savings in transporting inputs and technological and information spillovers, and shows a correlation between labor productivity and accessibility in US cities[1].

In terms of methodology, my work is related to that of Burchfield et al. (2006), who also employ remotely sensed data to track urban areas over time. More specifically, they analyze changes in the extent of sprawl in US cities between 1992 and 1996. The data I employ comes mostly from night-time, as opposed to day-time imagery, and covers a longer time span (1992-2010).

Saiz (2011) also looks at geographic constraints to city expansion, by computing the amount of developable land within 50 km radii from US city centers and relating it to the elasticity of housing supply. I use the same notion of geographic constraints, but I employ them in a novel way to construct a time-varying instrument for city shape.

Finally, my work is also related to a growing literature on road infrastructure and urban growth in developing countries (Baum-Snow and Turner, 2012; Baum-Snow et al., 2013; Morten and Oliveira, 2014, Storeygard, 2014). In particular, Morten and Oliveira (2014) also estimate a model of spatial equilibrium across cities, and resort to an instrumental variables approach. Differently from these studies, I do not look at the impact of roads connecting cities, but instead focus on urban geometry as a proxy for the trips within cities.

[1] A subset of the urban planning and evironmental literature (reviewed by Martins et al, 2012) also looks at the environmental implications of city shape. Bertaud (2002, 2004) argues that although compact cities are characterized by shorter trip patterns, they might not necessarily be less polluted due to potential higher density of motor vehicles. Most empirical analyses (e.g. Clark et al. 2011) find that more compact and denser cities are less polluted, but do not separately identify the effect of shape and that of density.

3 Conceptual Framework

In this study I interpret city shape to first-order as a shifter of commuting costs: all else being equal, including city size, a city with a more compact geometry is characterized by shorter within-city trips[2]. It is plausible that consumers incorporate considerations on the relative ease of commutes when evaluating the trade-offs associated to different locations. This might be even more relevant in the context of India, in which most migrants to urban areas cannot afford individual means of transport. A natural starting point in thinking how city shape affects the location choices of consumers is to interpret "good geometry" as an amenity, in the framework of a model of spatial equilibrium à la the Rosen-Roback (Rosen 1979, Roback, 1982). The basic underlying idea of the Rosen-Roback model is that consumers and firms must be indifferent across cities, and wages and rents allocate people and firms to cities with different levels of amenities. The notion of spatial equilibrium across cities presumes that consumers are choosing across a number of different locations. The pattern of migration to urban areas observed in India is compatible with this element of choice: according to the 2001 Census, about 38 percent of internal migrants move to a location outside their district of origin, supporting the interpretation that they are effectively choosing a city rather than simply moving to the closest available urban location.

I draw upon the Roback model in its simplest form (Roback, 1982), following the exposition of the model by Glaeser (2008). Households consume a composite good C and housing H. They supply inelastically one unit of labor receiving a city-specific wage W. Their utility depends on net income, i.e. labor income minus housing costs, and on a city-specific bundle of consumption amenities θ. Their optimization problem reads:

$$\max_{C,H} U(C,H,\theta) \; s.t. \; C = W - p_h H \tag{1}$$

where p_h is the rental price of housing and

$$U(C,H,\theta) = \theta C^{1-\alpha} H^{\alpha}. \tag{2}$$

In equilibrium, indirect utility V must be equalized across cities, otherwise workers would move:

$$V(W - p_h H, \theta) = \overline{v} \tag{3}$$

[2]In Section 4.2 I illustrate the geometric indicators I employ to measure compactness, which are all based on the relative distances between points within a footprint. In Section 8 I discuss briefly other possible second-order channels through whih city shape might affect consumers.

which given the functional form assumptions yields the condition

$$\log(W) - \alpha \log(p_h) + \log(\theta) = \overline{V}. \tag{4}$$

The intuition for this condition is that consumers pay for amenities through lower wages (W) or through higher housing prices (p_h). The extent to which wages net of housing costs rise with an amenity is a measure of the extent to which that amenity decreases utility, relative to the marginal utility of income. Differentiating this expression with respect to some exogenous variable S - which could be (instrumented) city geometry - yields:

$$\frac{\partial \log(\theta)}{\partial S} = \alpha \frac{\partial \log(p_h)}{\partial S} - \frac{\partial \log(W)}{\partial S}. \tag{5}$$

This equation provides a way to evaluate the amenity value of S: the overall impact of S on utility can be found as the difference between the impact of S on housing prices, multiplied by the share of housing in consumption, and the impact of S on wages.

Firms in the production sector also choose optimally in which city to locate. Each city is a competitive economy that produces a single internationally traded good Y using labor N and a local production amenity A. Their technology also requires traded capital K and a fixed supply of non-traded capital $\overline{Z}$. Firms solve the following production maximization problem:

$$\max_{N,K} \left\{ Y(N, K, \overline{Z}, A) - WN - K \right\} \tag{6}$$

where

$$Y(N, K, \overline{Z}, A) = AN^{\beta} K^{\gamma} \overline{Z}^{1-\beta-\gamma}. \tag{7}$$

In equilibrium firms earn zero expected profits, which are equalized across cities. Under these functional form assumptions, the maximization problem for firms yields the following labor demand condition:

$$(1-\gamma)\log(W) = (1-\beta-\gamma)(\log(\overline{Z}) - \log(N)) + \log(A) + \kappa_1. \tag{8}$$

To close the model we need to specify the construction sector. Developers produce housing H using land l and "building height" h. In each location there is a fixed supply of land $\overline{L}$, as a result of land use regulations[3]. Denoting with p_l the price of land, their maximization problem

[3] In this framework, the amount of land to be developed is assumed to be given in the short run. It can be argued that in reality it is an endogenous outcome of factors such as quality of regulation, success of the city and geographic constraints. In my empirical analysis, city area will be specified as a mechanical function of predicted population growth, thus abstracting from these issues (see Section 5.2, Specification I).

reads:

$$\max_{H} \{p_h H - C(H)\} \tag{9}$$

where

$$\begin{aligned} H &= l \cdot h & (10)\\ C(H) &= c_0 h^{\delta} l - p_l l \ , \ \delta > 1. & (11) \end{aligned}$$

The construction sector operates optimally, with construction profits equalized across cities. By combining the housing supply equation resulting from the maximization problem of developers with the housing demand equation resulting from the consumers' problem, we obtain the following housing market equilibrium condition:

$$(\delta - 1)\log(H) = \log(p_h) - \log(c_0\delta) - (\delta - 1)\log(N) + (\delta - 1)\log(\overline{L}) \tag{12}$$

Using the three optimality conditions for consumers (4), firms (8) and (12) this model can be solved for the three unknowns N, W and p_h , representing respectively population, wages and housing prices, as functions of the model parameters and in particular as functions of the city specific productivity parameter and consumption amenities. Denoting all constants with K, this yields the following:

$$\log(N) = \frac{(\delta(1-\alpha)+\alpha)\log(A) + (1-\gamma)\left(\delta\log(\theta) + \alpha(\delta-1)\log(\overline{L})\right)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_N \tag{13}$$

$$\log(W) = \frac{(\delta-1)\alpha\log(A) - (1-\beta-\gamma)\left(\delta\log(\theta) + \alpha(\delta-1)\log(\overline{L})\right)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_W \tag{14}$$

$$\log(p_h) = \frac{(\delta-1)\left(\log(A) + \beta\log(\theta) - (1-\beta-\gamma)\log(\overline{L})\right)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_P. \tag{15}$$

These conditions translate into the following predictions:

$$\begin{aligned} \frac{d\log(N)}{d\log(A)} &> 0, \ \frac{d\log(N)}{d\log(\theta)} > 0, \ \frac{d\log(N)}{d\log(\overline{L})} > 0 & (16)\\ \frac{d\log(W)}{d\log(A)} &> 0, \ \frac{d\log(W)}{d\log(\theta)} < 0, \ \frac{d\log(W)}{d\log(\overline{L})} < 0 & (17)\\ \frac{d\log(p_h)}{d\log(A)} &> 0, \ \frac{d\log(p_h)}{d\log(\theta)} > 0, \ \frac{d\log(p_h)}{d\log(\overline{L})} < 0. & (18) \end{aligned}$$

Population, wages and rents are all increasing functions of the city-specific productivity parameter. Population and rents are increasing in the amenity parameter as well, whereas wages are decreasing in it. The intuition is that firms and consumers have potentially conflicting location preferences: firms prefer cities with higher production amenities, whereas consumers prefer cities with higher consumption amenities. Factor prices – W and p_H – are striking the balance between these conflicting preferences.

Consider now an indicator of urban geometry S, higher values of S denoting "worse" shapes, in the sense of shapes conducive to longer commute trips. Assume "non-compact shape" is purely a consumption disamenity, which decreases consumers' utility all else being equal but does not directly affect firms' productivity:

$$\frac{\partial \theta}{\partial S} < 0, \ \frac{\partial A}{\partial S} = 0 \tag{19}$$

In this case we should observe the following reduced-form relationships:

$$\frac{dN}{dS} < 0, \frac{dW}{dS} > 0, \frac{dp_h}{dS} < 0 \tag{20}$$

This framework predicts that a city with poorer shape should have *ceteris paribus* a smaller population, higher wages and lower house rents. The intuition is that consumers prefer to live in cities with good shapes, which drives rents up and bids wages down in these locations. Suppose instead that poor city geometry is both a consumption and a production disamenity, i.e. it depresses both the utility of consumers and the productivity of firms:

$$\frac{\partial \theta}{\partial S} < 0, \ \frac{\partial A}{\partial S} < 0 \tag{21}$$

This would imply the following:

$$\frac{dN}{dS} < 0, \frac{dW}{dS} \gtreqless 0, \frac{dp_h}{dS} < 0 \tag{22}$$

The model's predictions are similar, except that the effect on wages will be ambiguous. The reason for the ambiguous sign of $\frac{dW}{dS}$ is that now both firms and consumers want to locate in compact cities. With respect to the previous case, we now have an additional force which tends to bid wages up in compact cities: competition among firms for locating in low-S cities. The net effect on W depends on on whether firms or consumers value low S the most. If S is more a consumption than it is a production disamenity, then we should observe $\frac{dW}{dS} > 0$.

Assume now that

$$\begin{aligned} \log(A) &= \kappa_A + \lambda_A S & (23) \\ \log(\theta) &= \kappa_\theta + \lambda_\theta S. & (24) \end{aligned}$$

Plugging (23) and (24) into (13),(14) and (15) we obtain:

$$\log(N) = \frac{[(\delta(1-\alpha)+\alpha)\lambda_A + (1-\gamma)\delta\lambda_\theta]\cdot S + (1-\gamma)\alpha(\delta-1)\log(\overline{L})}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_N \tag{25}$$

$$\log(W) = \frac{[(\delta-1)\alpha\lambda_A - (1-\beta-\gamma)\delta\lambda_\theta]\cdot S - (1-\beta-\gamma)\alpha(\delta-1)\log(\overline{L})}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_W \tag{26}$$

$$\log(p_h) = \frac{[(\delta-1)\lambda_A + (\delta-1)\beta\lambda_\theta]\cdot S - (1-\beta-\gamma)(\delta-1)\log(\overline{L})}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} + K_P. \tag{27}$$

For ease of exposition, define:

$$B_N = \frac{(\delta(1-\alpha)+\alpha)\lambda_A + (1-\gamma)\delta\lambda_\theta}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)},\ L_N = \frac{(1-\gamma)\alpha(\delta-1)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} \tag{28}$$

$$B_W = \frac{(\delta-1)\alpha\lambda_A - (1-\beta-\gamma)\delta\lambda_\theta}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)},\ L_W = \frac{-(1-\beta-\gamma)\alpha(\delta-1)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)} \tag{29}$$

$$B_P = \frac{(\delta-1)\lambda_A + (\delta-1)\beta\lambda_\theta}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)},\ L_P = \frac{-(1-\beta-\gamma)(\delta-1)}{\delta(1-\beta-\gamma)+\alpha\beta(\delta-1)}. \tag{30}$$

This allows us to rewrite (25), (26), (27) in a more parsimonious form:

$$\log(N) = B_N S + L_N \log(\overline{L}) + K_N \tag{31}$$

$$\log(W) = B_W S + L_W \log(\overline{L}) + K_W \tag{32}$$

$$\log(p_h) = B_P S + L_P \log(\overline{L}) + K_P. \tag{33}$$

Note that (28), (29), (30) imply:

$$\lambda_A = (1-\beta-\gamma)B_N + (1-\gamma)B_W \tag{34}$$

$$\lambda_\theta = \alpha B_P - B_W. \tag{35}$$

The welfare impact of a marginal increase in S can be quantified using equation (5), which states that in equilibrium a marginal change in $\log(\theta)$ needs to be compensated 1 to 1 by a change in $\log(W)$:

$$\frac{\partial\log(\theta)}{\partial S} = \lambda_\theta = \alpha\frac{\partial\log(p_h)}{\partial S} - \frac{\partial\log(W)}{\partial S} = \alpha B_P - B_W. \tag{36}$$

Parameter λ_θ captures, in log points, the welfare loss form a marginal increase in S. Parameter λ_A captures the impact of a marginal increase in S on city-specific productivity. Denote with $\widehat{B_N}, \widehat{B_W}$ and $\widehat{B_P}$ the reduced-form estimates for the impact of S on respectively $\log(N), \log(W)$

and $\log(p_h)$.These estimates, in conjunction with plausible values for parameters β, γ, α, can be used to back out λ_A and λ_θ:

$$\widehat{\lambda_A} = (1-\beta-\gamma)\widehat{B_N} + (1-\gamma)\widehat{B_W} \quad (37)$$

$$\widehat{\lambda_\theta} = \alpha\widehat{B_P} - \widehat{B_W}. \quad (38)$$

Note that this approach captures the overall, net effect of S on the average city dweller, without explicitly modelling the mechanism through which S enters the decisions of consumers. In Section 6.2 I provide empirical evidence suggesting that the urban transit channel is indeed involved, and in the concluding section I discuss some alternative, second-order channels through which city shape might affect consumers. This simple model also does not explicitly address heterogeneity across consumers in tastes and skills. However, we expect that people will sort themselves into locations based on their preferences. The estimated differences in wages and rents across cities will thus be an underestimate of true equalizing differences for those with a strong taste for the amenity of interest, and an overestimate for those with weak preferences. The empirical evidence presented in Section 6.2 indicates that there might indeed be sorting across cities with different geometries.

Reduced-form estimates for B_N, B_W, B_P are presented in Sections 6.1 and 6.2, whereas Section 6.3 provides estimates for the structural parameters $\lambda_A, \lambda_\theta$.The next two Sections present the data sources and empirical strategy employed in the estimation.

4 Data Sources

I assemble an unbalanced panel of urban footprints-years, covering all Indian cities for which a footprint could be retrieved based on the methodology explained below.

4.1 Urban Footprints

I retrieve the boundaries of urban footprints from two sources. The first is the U.S. Army India and Pakistan Topographic Maps (U.S. Army Map Service, 195?), a series of detailed maps covering the entire Indian subcontinent at a 1:250,000 scale. I geo-referenced these maps and traced the reported perimeter of urban areas, which are clearly demarcated (Figure 1).

The second source is derived from the DMSP/OLS Night-time Lights dataset. This dataset is based on night-time imagery recorded by satellites from the U.S. Air Force Defense Meteorological Satellite Program (DMSP) and reports the recorded intensity of Earth-based lights, measured by a six-bit number (ranging from 0 to 63). These data are reported for every year between 1992 and 2010, with a resolution of 30 arc-seconds (approximately 1 square kilometers). Night-time lights have been employed in economics typically for purposes other than urban mapping (Henderson et al., 2012). However, the use of the DMSP/OLS dataset for delineating urban areas is quite common in urban remote sensing (Henderson et al., 2003; Small, 2005; Small et al. 2011). The basic methodology is the following: first, I overlap the night-time lights imagery with a point shapefile with the coordinates of Indian settlement points, taken from the Global Rural-Urban Mapping Project (GRUMP) Settlement Points dataset (CIESIN et al., 2011; Balk et al., 2006). I then set a luminosity threshold (e.g. 35) and consider spatially contiguous lighted areas surrounding the city coordinates with luminosity above that threshold. This approach, illustrated in Figure 2, can be replicated for every year covered by the DMSP/OLS dataset.

The choice of luminosity threshold results in a more or less restrictive definition of urban areas, which will appear larger for lower thresholds[4]. To choose luminosity thresholds appropriate for India, I overlap the 2010 night-time lights imagery with available Google Earth imagery. I find that a luminosity threshold of 35 generates the most plausible mapping for those cities covered by both sources[5]. In my full panel (including years 1950 and 1992-2010), the average city footprint occupies and area of approximately 63 squared kilometers[6].

Using night-time lights as opposed to alternative satellite-based products, in particular day-time imagery, is motivated by a number of advantages. Unlike products such as aerial photographs or high-resolution imagery, night-time lights cover systematically the entire Indian

[4]Determining where to place the boundary between urban and rural areas always entails some degree of arbitrariness, and in the urban remote sensing literature there is no clear consensus on how to set such threshold. It is nevertheless recommended to validate the chosen threshold by comparing the DMSP/OLS-based urban mapping with alternative sources, such as high-resolution day-time imagery, which in the case of India is available only for a small subset of city-years.

[5]For years covered by both sources (1990, 1995, 2000), my maps also appear consistent with those from the GRUMP - Urban Extents Grid dataset, which combines night-time lights with administrative and Census data to produce global urban maps (CIESIN et al., 2011; Balk et al., 2006).

[6]My results are robust to using alternative luminosity thresholds between 20 and 40. Results are available upon request.

subcontinent, and not only a selected number of cities. Moreover, they are one of the few sources which allow to detect changes in urban areas over time, due to their yearly temporal frequency. Finally, unlike multi-spectral satellite imagery such as Landsat- or MODIS- based products, which in principle would be available for different points in time, night-time lights do not require any sophisticated manual pre-processing[7]. An extensive portion of the urban remote sensing literature compares the accuracy of this approach in mapping urban areas with that attainable with alternative satellite-based products, in particular day-time imagery (inter alia, Henderson et al, 2003; Small et al., 2005). This cross-validation exercise has been carried out also specifically in the context of India by Joshi et al. (2011) and Roychowdhury (2009). The general takeout is that none of these sources is error-free, and that there is no strong case for preferring day-time over night-time satellite imagery if aerial photographs are not systematically available for the area to be mapped.

It is well known that urban maps based on night-time lights will tend to inflate urban boundaries, due to "blooming" effects (Small et al., 2005)[8]. This can only partially be limited by setting high luminosity thresholds. In my panel, urban footprints as reported for years 1992-2010 thus reflect a broad definition of urban agglomeration, which typically goes beyond the current administrative boundaries. This contrasts with urban boundaries reported in the US Army maps, which seem to reflect a more restrictive definition of urban areas (although no specific documentation is available). Throughout my analysis I include year fixed effects, which amongst other things control for these differences in data sources, as well as for different calibrations of the night-time lights satellites.

By combining the US Army maps (1950s) with yearly maps obtained from the night-time lights dataset (1992-2010) I thus assemble an unbalanced panel of urban footprints[9]. The criteria

[7] Using multi-spectral imagery to map urban areas requires a manual classification process, which relies extensively on alternative sources, mostly aerial photographs, to cross-validate the spectral recognition, and is subject to human bias.

[8] DMSP-OLS night-time imagery has a tendency to overestimate the actual extent of lit area on the ground, due to a combination of coarse spatial resolution, overlap between pixels, and minor geolocation errors (Small et al., 2005).

[9] The resulting panel dataset is unbalanced for two reasons: first, some settlements become large enough to be detectable only later in the panel; second, some settlements appear as individual cities for some years in the panel and then become part of larger urban agglomerations in later years.The number of cities in the panel ranges from 352 to 457, depending on the year considered.

for being included in the analysis is to appear as a contiguous lighted shape in the night-time lights dataset. This appears to leave out only very small settlements. Throughout my analysis, I instrument all the geometric properties of urban footprints, including both area and shape. This IV approach addresses problems of non-classical measurement error which could affect my data sources, for instance due to the well-known correlation between income and luminosity.

4.2 Shape Metrics

The indicators of city shape which I employ, based on those in Angel, Civco and Parent (2010)[10], are used in landscape ecology and urban studies to proxy for within-city trips and infer travel costs. They are all based on the distribution of points around the polygon's centroid[11] or within the polygon, and are measured in kilometers. Summary statistics for the indicators below are reported in Table 1.

(i) The *remoteness* index is the average distance between all interior points and the centroid. It can be considered a proxy for the average length of commutes to the urban center.

(ii) The *spin* index is computed as the average of the squared distances between interior points and centroid. This is similar to the remoteness index, but gives more weight to the polygon's extremities, corresponding to the periphery of the footprint. This index is more capable of identifying footprints that have "tendril-like" projections, often perceived as an indicator of sprawl.

(iii) The *disconnection* index captures the average distance between all pairs of interior points. It can be considered a proxy for commutes within the city, without restricting one's attention to those leading to the center.

(iv) The *range* index captures the maximum distance between two points on the shape perimeter, representing the longest possible of commute trips within the city.

All these measures are mechanically correlated with polygon area. In order to separate the effect of geometry *per se* from that of city size, in my preferred specifications I explicitly control for the area of the footprint - which is separately instrumented for. Conditional footprint area,

[10] I am thankful to Vit Paszto for help with the ArcGis shape metrics routines. I have renamed some of the shape metrics for ease of exposition.

[11] The centroid of a polygon, or center of gravity, is the point which minimizes the sum of squared Euclidean distances between itself and each vertex.

higher values of these indexes indicate longer within-city trips. Alternatively,it is possible to normalize each of these indexes, computing a version which is invariant to the area of the polygon. I do so by computing first the radius of the "Equivalent Area Circle" (EAC), namely a circle with an area equal to that of the polygon. I then normalize the index of interest dividing it by the EAC radius, obtaining what I define *normalized remoteness*, *normalized spin,* etc. One way to interpret these normalized metrics is as deviations of a polygon's shape from that of a circle, the shape which minimizes all the indexes above.

Figure 3 provides a visual example of how these metrics map to the shape of urban footprints. Among million-plus cities, I consider those with respectively the "best" and the "worst" geometry based on the indicators described above, namely Bengaluru and Kolkata (formerly known as Bangalore and Calcutta). The figure reports the footprints of the two cities as of year 2005, where Bengaluru's footprint has been rescaled so that they have the same area. The figure also reports the above shape metrics computed for these two footprints. The difference in the remoteness index between Kolkata and (rescaled) Bengaluru is 4.5 km; the difference in the disconnection index is 6.2 km. The interpretation is the following: if Kolkata had the same compact shape that Bengaluru has, the average trip to the center would be shorter by 4.5 km and the average trip within the city would be shorter by 6.2 km. The Indian Ministry of Urban Development (2008) estimates the average commute speed in million-plus cities to be of 12 km per hour in 2011, which is predicted to become 9 km by 2021. According to the 2011 estimated speed, the above differences in trip length translate to a difference in commute times of respectively 22.5 minutes (average trip to the center) and 31 minutes (average within-city trip). Although this is a very rough calculation, it is nevertheless revealing that city geometry indeed has potentially sizeable impacts on commute times.

4.3 Geography

Following Saiz (2011), I consider as "undevelopable" terrain which either covered by a water body or is characterized by a slope above 15%. I draw upon the highest resolution sources available: the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (NASA and METI, 2011), with a resolution of 30 meters, and the Global MODIS Raster Water Mask (Carroll et al. 2009), with a resolution of 250 meters. I combine these two raster datasets to classify pixels as "developable" or "undevelopable". Figure

4 illustrates this classification for the Mumbai area.

4.4 Population and Other Census Data

City-level data for India is notoriously hard to obtain (Greenstone and Hanna, forthcoming). The only systematic source which collects data explicitly at the city level is the Census of India, conducted every 10 years. In this project I employ population data from Census years 1871-2011. As explained in Section 5.1, historic population (1871-1941) is used to construct one of the two versions of my instrument, whereas population drawn from more recent waves (1951, 1991, 2001 and 2011) is used as an outcome variable[12].

Outcomes other than population are not consistently available for all Census years. I draw data on urban road length in 1991 from the 1991 Town Directory. In recent Census waves (1991, 2001, 2011) data on slum population and physical characteristics of houses are available for a subset of cities.

It is worth pointing out that "footprints" as retrieved from the night-time lights dataset do not always have an immediate Census counterpart in terms of town or urban agglomeration, as they sometimes stretch to include suburbs and towns treated as separate units by the Census. A paradigmatic example is the Delhi conurbation, which as seen from the satellite expands well beyond the administrative boundaries of the New Delhi National Capital Region. When assigning population totals to an urban footprint, I sum the population of all Census settlements which are located within the footprint, thus computing a "footprint" population total[13].

4.5 Wages and Rents

For outcomes other than those available in the Census I rely on the National Sample Survey and the Annual Survey of Industries, which provide at most district identifiers. I thus follow the approach of Greenstone and Hanna (forthcoming): I match cities to districts and use district urban averages as proxies for city-level averages. It should be noted that the matching is not

[12] Historic population totals were taken from Mitra (1980). Census data for years 1991 to 2001 were taken from the Census of India electronic format releases. 2011 Census data were retrieved from http://www.censusindia.gov.in/DigitalLibrary/Archive_home.aspx.

[13] In order to assemble a consistent panel of city population totals over the years I also take into account changes in the definitions of "urban agglomerations" and "outgrowths" across Census waves.

always perfect, for a number of reasons. First, it is not always possible to match districts as reported in these sources to Census districts, and through these to cities, due to redistricting and inconsistent numbering throughout this period. Second, there are a few cases of large cities which cut across districts (e.g. Hyderabad). Finally, there are a number of districts which contain more than one city from my sample. In these cases I follow several matching approaches: considering only the main city for that district, and dropping the district entirely. I show results following both approaches. The matching process introduces considerable noise and leads to results which are relatively less precise and less robust than those I obtain with city-level outcomes.

Data on wages are taken from the Annual Survey of Industries (ASI), waves 1990, 1994, 1995, 1997, 1998, 2009, 2010[14]. These are repeated cross-sections of plant-level data collected by the Ministry of Programme Planning and Implementation of the Government of India. The ASI covers all registered manufacturing plants in India with more than fifty workers (one hundred if without power) and a random one-third sample of registered plants with more than ten workers (twenty if without power) but less than fifty (or one hundred) workers. As mentioned by Fernandes and Sharma (2012) amongst others, the ASI data are extremely noisy in some years, which introduces a further source of measurement error. The average individual yearly wage in this panel amounts to 94 thousand Rs. at current prices.

Unfortunately there is no systematic source of data for property prices in India. I construct a rough proxy for the rental price of housing drawing upon the National Sample Survey (Household Consumer Expenditure schedule), which asks households about the amount spent on rent. In the case of owned houses, an imputed figure is provided. I focus on rounds 62 (2005-2006), 63 (2006-2007) and 64 (2007-2008), since they are the only ones for which the urban data is representative at the district level and which report total dwelling floor area as well. I use this information to construct a measure of rent per square meter. The average yearly total rent paid in this sample amounts to about 25 thousand Rs., whereas the average yearly rent per squared meter is 603 Rs., at current prices. These figures are likely to be underestimating the market rental rate, due to the presence of rent control provisions in most major cities of India (Dek, 2006). To cope with this problem, I also construct an alternative proxy for housing rents which focuses on the upper half of the distribution of rents per meter, which is less likely to include

[14] These are all the waves I could access as of June 2014.

observations from rent-controlled housing.

4.6 Other Data

Data on state-level infrastructure (motor vehicles density, state urban roads length) is taken from the Ministry of Road Transport and Highways, Govt. of India.

Data on the current road network is constructed from the maps available on Openstreetmap[15], a collaborative mapping project which provides crowdsourced maps of the world. Openstreetmap data has been favorably compared with proprietary data sources and is continuously updated. As such, it reflects the current state of the street network. I consider the most recent night-time lights-based map of urban boundaries in my sample - corresponding to 2010 - and overlap it with the street network map of India provided by Openstreetmap. Information on the type of road - whether trunk, residential, secondary etc. - is also provided. Given the collaborative nature of Openstreetmap, there is a concern that the level of detail of such maps might be higher in larger cities, or in neighborhoods with more economic activity. Upon visual inspection, it appears that smaller roads are reported only in relatively bigger cities. To avoid this source of non-classical measurement error, I exclude small roads and focus on those denoted as "trunk", "primary" or "secondary". I then compute the total road length by considering street segments contained within the urban boundary. The average road density in my sample as of year 2010, obtained by dividing total city road length by footprint area, is 2.4 km per squared km. Given the tendency of night-time lights to overestimate urban boundaries, this figure should be considered an underestimate of the actual road density.

Data on the maximum permitted Floor Area Ratios for a small cross-section of Indian cities (55 cities in my sample) is taken from Sridhar (2010), who collected them from individual urban local bodies as of the mid-2000s. FARs are expressed as ratios of the total floor area of a building over the area of the plot on which it sits. The average FAR in this sample is 2.3, a very restrictive figure compared to international standards. For a detailed discussion of FARs in India, see Sridhar (2010) and Bertaud and Brueckner (2005).

Data on the location of productive establishments in year 2005 is derived from the urban Directories of Establishments pertaining to the 5th Economic Census. The Economic Census

[15] http://www.openstreetmap.org/about

is a complete enumeration of all establishments, with the exception of those involved in crop production, conducted by the Indian Ministry of Statistics and Programme Implementation. Town or district identifiers are not provided to the general public. However in year 2005 establishments with more than 10 employees were required to provide an additional "address slip", containing a complete address of the establishment, year of initial operation and employment class. I geo-referenced all the addresses corresponding to cities in my sample through Google Maps API, retrieving consistent coordinates for about 240 thousand establishments in about 200 footprints. Although limited by their cross-sectional nature, these data provide an opportunity to study the spatial distribution of employment within cities. As a first, exploratory step in this direction I compute a simple measure of spatial concentration based on Moran's I statistic (Anselin, 1995), which has been used as an indicator of city "compactness" (Tsai, 2005).

5 Empirical Strategy

The objective of my empirical analysis is to estimate the effects of city shape on a number of city-year level outcomes, most notably population, wages and rents. My data has the structure of an unbalanced city-year panel. In every year, I observe the geometric properties of the footprint of each city, namely footprint area and the different shape metrics described above. The goal is to estimate the relationship between shape in a given city-year on a number of city-year level outcomes, conditional on city and year fixed effects. This strategy exploits variation in urban shape which is both cross-sectional and temporal. City and year fixed effects are included in all specifications, so as to account for time-invariant city characteristics and for country-level trends in population and other outcomes. The identification of the impacts of shape thus relies on comparing *changes* in urban shape over time for each city[16].

[16] As discussed below, a limited number of outcomes, analyzed in Section 7, are available only for a cross-section of cities. In these cases, I resort to a cross-sectional version of equation (39), which cannot include city fixed effects.

5.1 Instrument Construction

A major concern in estimating the above relationship is the endogeneity of city geometry. The observed urban footprint at a given point in time is the result of the interaction of local geographic conditions, city growth and policy. Cities which experience faster population growth might be expanding in a more chaotic and unplanned fashion, generating a "leapfrog" pattern of development which translates into less compact shapes. At the same time, cities which exhibit faster growth rates might be the object of more stringent regulations. For instance, restrictive Floor Area Ratios in India have been motivated by a perceived need to reduce urban densities (Sridhar, 2010). In order to address this endogeneity problem, I employ an IV approach, constructing an instrument for city shape which varies at the city-year level.

My instrument is constructed combining geography with a mechanical model for city expansion in time. The underlying idea is that as cities expand in space over time, they hit different geographic obstacles which constrain their shapes by preventing expansion in some of the possible directions. I instrument the *actual* shape of the observed footprint at a given point in time with the *potential* shape which the city can have, given the geographic constraints which it faces at that stage of its predicted growth. More specifically, I consider the largest contiguous patch of land which is developable, i.e. not occupied by a water body nor by steep terrain, within a given predicted radius around each city. I denote this contiguous patch of developable land as "potential footprint". I compute the shape properties of the *potential* footprint and use them as instruments for the corresponding shape properties of the *actual* urban footprint. What gives time variation to this instrument is the fact that the predicted radius is time-varying, and expands over time based on a mechanical model for city expansion. In its simplest form this mechanical model postulates a common growth rate for all cities; in my benchmark specifications, it is city-specific, and based on each city's projected historic population.

The procedure for constructing the instrument is illustrated in Figure 5 for the city of Mumbai. Recall that I observe the footprint of a city c in year 1951[17] (from the U.S. Army Maps) and then in every year t between 1992 and 2010 (from the night-time lights dataset). I take as starting point the minimum bounding circle of the 1951 city footprint (Figure 5a).

[17]The US Army Maps are from the mid-50s, but no specific year of publication is provided. The closest Census year is 1951. For the purposes of constructing the city-year panel, I am attributing to the footprints observed in these maps the year 1951, so that I can match them to population as of Census year 1951.

To construct the instrument for city shape in 1951, I consider the portion of land which lies within this bounding circle and is developable, i.e. not occupied by water bodies nor steep terrain. The largest contiguous patch of developable land within this radius is colored in green in Figure 5b and represents what I define as "potential footprint" of the city of Mumbai in 1951. In subsequent years $t \in \{1992, 1993..., 2010\}$ I consider concentrically larger radii $\widehat{r}_{ct}$ around the historic footprint, and construct corresponding potential footprints lying within these predicted radii (Figures 5c and 5d).

To complete the description of the instrument I need to specify how $\widehat{r}_{ct}$ is determined. The projected radius $\widehat{r}_{ct}$ is obtained by postulating a simple, mechanical model for city expansion in space. I consider two versions of this model: a "city-specific" version and a "common rate" one.

(A) City-specific: In this first version of the model, I make the rate of expansion of $\widehat{r}_{ct}$ vary across cities, depending on their historic (1871 - 1951) population growth rates. In particular, $\widehat{r}_{ct}$ answers the following question: if the city's population continued to grow as it did between 1871 and 1951 and population density remained constant at its 1951 level, what would be the area occupied by the city in year t? More formally, the steps involved are the following:

(i) I project log-linearly the 1871-1951 population of city c (from the Census) in all subsequent years, obtaining the projected population $\widehat{pop_{c,t}}$, for $t \in \{1992, 1993..., 2010\}$.

(ii) Denote the area of city c's actual footprint in year t as $area_{c,t}$ and the actual - not projected - population of city c in year t as $pop_{c,t}$. I pool together the 1951-2010 panel of cities and run the following regression:

$$\log(area_{c,t}) = \alpha \cdot \log(\widehat{pop_{c,t}}) + \beta \cdot \log\left(\frac{pop_{c,1950}}{area_{c,1950}}\right) + \gamma_t + \varepsilon_{c,t} \tag{39}$$

from which I obtain $\widehat{area_{c,t}}$, the *predicted* area of city c in year t.

(iii) I compute $\widehat{r_{c,t}}$ as the radius of a circle with area $\widehat{area_{c,t}}$:

$$\widehat{r_{c,t}} = \sqrt{\frac{\widehat{area_{c,t}}}{\pi}}. \tag{40}$$

The interpretation of the circle with radius $\widehat{r_{c,t}}$ from figures 5c and 5d is thus the following: this is the area which the city would occupy if it continued to grow as in 1871-1951, if its density remained the same as in 1951, and if the city could expand freely and symmetrically in all directions, in a fashion which optimizes the length of within-city trips.

(B) Common-rate: The second version of the mechanical model is even more parsimonious: the rate of expansion of the radii is the same for all cities, and equivalent to the average expansion rate across all cities in the sample. More formally, I run the following regression:

$$\log(area_{c,t}) = \theta_c + \gamma_t + \varepsilon_{c,t} \tag{41}$$

where θ_c and γ_t denote city and year fixed effects, from which I retrieve an alternative version of $\widehat{area_{c,t}}$ and corresponding $\widehat{r_{c,t}} = \sqrt{\frac{\widehat{area_{c,t}}}{\pi}}$.

This instrument seeks to isolate the variation in urban geometry which is induced by geography, excluding the variation which results from policy or other endogenous choices. Although resorting to geography arguably helps addressing issues of policy endogeneity, there is nevertheless a concern that geography affects location choices directly, for instance through the inherent amenity (or disamenity) value of water bodies, and not only through the constraints it posits on urban form. These concerns are mitigated by two features of my instrument. First, it is not purely cross-sectional but has time variation. This allows me to control for time-invariant effects of geography through city fixed effects. Second, it captures a very specific feature of geography: whether it allows for compact development or not. My instrument is not based on the generic presence of topographic constraints, nor on the share of constrained over developable terrain. Rather, it measures the *geometry* of available land. In one of my robustness checks, I show that my results are unchanged when I exclude mountain and coastal cities, which would be the two most obvious examples of cities where geography might have a specific (dis)amenity value.

5.2 Estimating Equations

Consider a generic shape metric S - which could be any of the indexes discussed in Section 4.2. Denote with $S_{c,t}$ the shape metric computed for the *actual* footprint observed for city c in year t, and with $\widetilde{S_{c,t}}$ the shape metric computed for the *potential* footprint of city c in year t, namely the largest contiguous patch of developable land within the predicted radius $\widehat{r_{c,t}}$.

Specification I

Consider outcome variable $Y \in (N, W, p_H)$ and let $area_{c,t}$ be the area of the urban footprint. My benchmark estimating equation (Specification I) corresponds to equation (31) in the model outlined in Section 3, augmented with city and year fixed effects:

$$\log(Y_{c,t}) = a \cdot S_{c,t} + b \cdot \log(area_{c,t}) + \mu_c + \rho_t + \eta_{c,t} \tag{42}$$

This equation contains two endogenous regressors: $S_{c,t}$ and $\log(area_{c,t})$. These are instrumented using respectively $\widetilde{S_{c,t}}$ and $\log(\widehat{pop_{c,t}})$ - the same projected historic population used in the city-specific model for urban expansion, step i.

This results in the following two first-stage equations:

$$S_{c,t} = \sigma \cdot \widetilde{S_{c,t}} + \delta \cdot \log(\widehat{pop_{c,t}}) + \omega_c + \varphi_t + \theta_{c,t} \tag{43}$$

and

$$\log(area_{c,t}) = \alpha \cdot \widetilde{S_{c,t}} + \beta \cdot \log(\widehat{pop_{c,t}}) + \lambda_c + \gamma_t + \varepsilon_{c,t}. \tag{44}$$

The counterpart of $\log(area_{c,t})$ in the conceptual framework is $log(\overline{L})$, where $\overline{L}$ is the amount of land which regulators allow to be developed in each period. It is plausible that regulators set this amount based on projections of past city growth, which rationalizes the use of projected historic population as an instrument.

One advantage of this approach is that it allows me to analyze the effects of shape and area considered separately - recall that the non-normalized shape metrics are mechanically correlated with footprint size. However, a drawback of this strategy is that it requires not only an instrument for shape, but also one for area.

Specification II

For robustness, I consider also an alternative approach which does not explicitly include city area in the regression, and therefore does not require including projected historic population among the instruments.

When focusing on population as an outcome variable, a natural way to do this is to normalize both right- and left-hand side by city area, considering respectively the normalized shape metric - see Section 4.2 - and population density. This results in the following, more parsimonious specification (Specification II): define population density[18] as

$$d_{c,t} = \frac{pop_{c,t}}{area_{c,t}}$$

[18]Note that this does not coincide with population density as defined by the Census, which reflects adiministrative boundaries.

and denote the normalized version of shape metric S with nS. The estimating equation becomes

$$d_{c,t} = a \cdot nS_{c,t} + \mu_c + \rho_t + \eta_{c,t} \tag{45}$$

which contains endogenous regressor $nS_{c,t}$. I instrument $nS_{c,t}$ with $\widetilde{nS_{c,t}}$, namely the normalized shape metric computed for the potential footprint. The corresponding first-stage equation is

$$nS_{c,t} = \beta \cdot \widetilde{nS_{c,t}} + \lambda_c + \gamma_t + \varepsilon_{c,t}. \tag{46}$$

The same approach can be followed for other outcome variables representing quantities - such as road length. Although it does not allow the effects of shape and area to be separately identified, this approach is less demanding. In particular, it does not require using projected historic population, and allows me to construct my shape instrument using both versions of $\widehat{r_{c,t}}$, the one obtained from the city-specific model as well as that obtained from the common rate model (see Section 5.1).

While population or road density are meaningful outcomes *per se*, it does not seem as natural to normalize factor prices - wages and rents - by city area. For these other outcome variables, the more parsimonious alternative to Specification I takes the following form:

$$\log(Y_{c,t}) = a \cdot S_{c,t} + \mu_c + \rho_t + \eta_{c,t} \tag{47}$$

where $Y \in (W, p_H)$.This equation does not explicitly control for city area other than through city and year fixed effects. Again, the endogenous regressor $S_{c,t}$ is instrumented using $\widetilde{S_{c,t}}$, resulting in the following first-stage equation:

$$S_{c,t} = \sigma \cdot \widetilde{S_{c,t}} + \omega_c + \varphi_t + \theta_{c,t}. \tag{48}$$

All of the specifications discussed above include year and city fixed effects, which capture time-invariant city characteristics - including the general effect of geography. The identification of the impacts of shape thus relies on comparing geography-driven *changes* in urban shape over time for each city[19]. Although the bulk of my analysis, presented in Section 6, relies on both

[19] The implicit underliying assumption of this approach is a "parallel trends" one, which would be violated if outcomes in cities with different geometries followed differential trends. To mitigate this concern, in one of my robustness checks (Appendix Tables 1 and 2) I augment the specifications above with year fixed effects interacted with the city's shape at the beginning of the panel.

cross-sectional and temporal variation, a limited number of outcomes, analyzed in Section 7, are available only for a cross-section of cities. In these cases, I resort to cross-sectional versions of equations (43) to (49),which cannot include city fixed effects.

In all specifications I employ robust standard errors clustered at the city level, to account for arbitrary serial correlation over time in cities.

6 Empirical Results: Amenity Value of City Shape

In this Section I address empirically the question of how city shape affects the spatial equilibrium across cities. The predictions of the conceptual framework suggest that, if city shape is valued as a consumption amenity by consumers, cities with longer trip patterns should be characterized by lower population, higher wages and lower rents.

6.1 First Stage

[Insert Table 2]

Table 2 presents results from estimating the first-stage relationship between city shape and the geography-based instrument described in Section 5.1. Each observation is a city-year. Panels, A, B, C and D each correspond to one of the four shape metrics discussed in Section 4.2: respectively, remoteness, spin, disconnection and range[20].Higher values of these indexes represent less compact shapes. Summary statistics are reported in Table 1. Columns 1 and 2 report the first stage for footprint shape (eq. (44)) and area (eq. (45)), which are relevant for Specification I. The dependent variables are city shape, measured in km and log city area, in km^2. The corresponding instruments are the shape of the potential footprint and log projected historic population, as described in Section 5.2. The construction of the potential footprint is based on the city-specific model for city expansion discussed in Section 5.1. Column 3 reports the first-stage for normalized shape (eq. (47)), which is the explanatory variable used in Specification II. Recall that normalized shape is an area-invariant measure of shape obtained normalizing a given shape metric by footprint radius. In this specification the construction of

[20]Recall that remoteness (panel A) is the average squared length of trips to the centroid; spin (panel B) is the average squared length of trips to the centroid; disconnection (panel C) is the average length of within-city trips; range (panel D) is the maximum length of within-city trips.

the potential footprint is based on the "common rate" model for city expansion outlined in Section 5.1.

Let us consider first Table 2A, which reports specification I estimated for the remoteness index. As discussed in Section 4.2, this index captures the length of the average trip to the footprint's centroid, and can be considered a proxy for the average commute to the central business district. Both first stages appear strong: the remoteness of the *potential* footprint is a highly significant predictor of the remoteness index computed for the *actual* footprint. Similarly, in column 2, projected historic population predicts footprint area. Column 2 reveals another interesting pattern: the area of the *actual* footprint is positively affected by the remoteness of the *potential* footprint. While this partly reflects the mechanical correlation between shape metric and footprint area, it also suggests that cities which are surrounded by topographic obstacles tend to expand more in space. An interpretation of this result is that the presence of topographic constraints induces a leapfrog development pattern, which is typically more land-consuming. It could also reflect an inherent difficulty in planning land-efficient development in constrained contexts, which could result in less parsimonious land use patterns. The first-stage of the more parsimonious Specification II, presented in column 3, is also strong: the normalized shape of the *potential* footprint is a significant predictor of the normalized shape of the *actual* footprint. The approach followed in column 3 is not based on projections of historic city population, and relies on the "common-rate" model for city expansion outlined in Section 5.1.[21] The results for the remaining shape indicators, reported in panels B, C, and D, are qualitatively similar.

6.2 Population

[Insert Table 3]

My main results on population and city shape are reported in Table 3. As in Table 2, each observation is a city-year and each panel corresponds to a different shape metric. Column 1 reports the IV results from estimating Specification I (equation (43)), which links population to city area and shape, separately instrumented for. The corresponding first stage is reported

[21] The normalized shape instrument can in principle be constructed also using the city-specific model for urban expansion. Results of the corresponding first-stage are not reported in the table for brevity, but are qualitatively similar to those in column 3 and are available upon request.

in column 1 of Table 2. Column 3 reports the corresponding OLS estimates. Column 2 reports the IV results from estimating Specification II (equation (46). The corresponding first stage is reported in column 3 of Table 2.

Interestingly, the OLS relationship between population and shape, conditional on area (column 3) appears to be positive due to an equilibrium correlation between city size and bad geometry: larger cities are typically also less compact. This arises from the fact that an expanding city has a tendency to deteriorate in shape. The intuition for this is the following: a new city typically arises in a relatively favorable geographic location; as it expands in space, however, it inevitably reaches areas with less favorable geography. Once shape is instrumented by geography (column 1), less compact cities are associated to a decrease in population, conditional on (instrumented) area, city and year fixed effects. To understand the magnitudes of this effect,consider the remoteness index (panel A), representing the length in km of the average trip to the footprint's centroid. A one-standard deviation in normalized remoteness (0.06) for the average-sized city (which has radius 4.5 km) corresponds to roughly 0.26 kilometers. Holding constant city area, a 0.26 kilometer increase in the average trip to the centroid is approximately associated to a 3% decline in population.

Column 2 reports results obtained from the more parsimonious Specification II, which links population density, measured in thousand inhabitants per km^2, to (instrumented) normalized shape. Recall that normalized shape metrics capture the departure of a city's shape from an ideal circular shape and are invariant to city area, higher values implying longer trips. The IV estimates of Specification II indicate that less compact cities are associated to a decline in population density. The magnitudes of this effect are best understood in terms of standardized coefficients. According to the estimates in panel A, a one standard deviation increase in normalized remoteness is associated to a decline in population density of 0.9 standard deviations. The results obtained with Specification I (Table 3, column 1) together with the first-stage estimates in Table 2 (column 2) indicate that this decline in density is driven both by a decrease in population and by an increase in footprint area.

The results for the remaining shape indicators, reported in panels B, C, and D, are qualitatively similar. The fact that these indexes are mechanically correlated with one another prevents me from including them all in the same specification. However, a comparison of the magnitudes of the IV coefficients of different shape metrics on population suggests that the most salient spatial properties are remoteness (Table 3A) and disconnection (Table 3C), which capture respectively the average trip length to the centroid and the average trip length within the footprint. This is plausible, since these two indexes are those which more closely proxy for urban commute patterns. Non-compactenss in the periphery, captured by the spin index (Table 3B), appears to have a precise zero effect on population. The effect of the range index (Table 3D), capturing the longest possible trip within the footprint, is significant but small in magnitude. For brevity, in the rest of my analysis I will mostly focus on the disconnection index, which measures the average within-city trip without restring one's attention to trips leading to the centroid. This index is the most general indicator for within-city commutes, and seems suitable to capture trip patters in polycentric as well as monocentric cities. Unless otherwise specified, in the rest of the tables "shape" will indicate the disconnection index.

[Insert Table 4]

As a robustness check, in Table 4 I re-estimate Specification I, excluding from the sample cities with severely constrained topographies, namely those located on the coast or in high-altitude areas. Such cities make about 9 % of cities in my sample. Out of 457 cities in the initial year of the panel (1951), those located on the coast and in mountainous areas are respectively 24 and 17 Both the first-stage (columns 1, 2, 4 and 5) and the IV estimates of the effect of shape on population (columns 3 and 6) are minimally affected by excluding these cities. This shows that my instrument has explanatory power also in cities without extreme topographic constraints[22], and that my IV results are not driven by a very specific subset of compliers.

Another robustness check is provided in Appendix Table 1. I re-estimate the IV impact of shape on population and density (columns 1 and 2 from Table 3C), including year fixed effects interacted with each city's shape at the beginning of the panel. This more conservative specification allows cities with different initial geometries to follow different time trends. Results are qualitatively similar to those obtained in Table 3. This mitigates the concern that diverging trends across cities with different geometries might be confounding the results.

[22] Recall that my instrument - the shape of the "potential" footprint - is not based on the severity of topographic constraints nor on the total share of land lost to such constraints, but is mostly driven by the relative *position* of constrained pixels.

6.3 Wages and Rents

The results presented thus far seem to indicate that consumers are affected by city shape in their location choices and that they dislike non-compact shapes. A natural question which arises next is whether we can put a price on "good shape". As discussed in Section 3, the Rosen-Roback model provides a framework for doing so by showing how urban amenities are capitalized in wages and rents. In particular, the model predicts that cities with better consumption amenities should be characterized by higher rents and lower wages.

Results on wages and rents are reported in Tables 5 and 6 respectively. As I discuss in Section 4.5, the measures of wages and rents which I employ are subject to significant measurement error. Both are urban district-level averages, derived respectively from the Annual Survey of Industries and the National Sample Survey Consumer Expenditure Schedule. The matching between cities and districts in not one-to-one. In particular, there are numerous instances of districts which include more than one city. I cope with these cases following three different approaches: (i) ignore the issue and keep in the sample all cities; (ii) drop from the sample districts with more than one city, thus restricting my sample to cities which have a one-to-one correspondence with districts; (iii) include in the sample only the largest city in each district. Tables 5 and 6 report estimation results obtained from all three approaches.

[Insert Table 5]

In Table 5 I report the OLS and IV relationship between average wages and city shape. The dependent variable is the log urban average of individual yearly wages in the city's district, in thousand 2014 Rupees. Columns 1, 4 and 7 report the IV results from estimating Specification I (equation (43)), which is conditional on instrumented city area. Columns 3, 6 and 9 report the same specification, estimated by OLS. Columns 2, 5 and 8 report the IV results from estimating Specification II (equation 48), which does not explicitly control for city area. The construction of the potential footprint is based on the city-specific model for city expansion in columns 1, 4 and 7, and on the common rate one in column 2, 5, 8 – see Section 5.1.

These estimates indicate that less compact shapes, as captured by higher values of the disconnection index, are associated to higher wages both in the OLS and in the IV. This pattern is consistent across different specifications and city-district matching approaches. Appendix Table 2, panel A, shows that these results are also robust to including year fixed effects interacted by initial shape. This positive estimated impact is compatible with the interpretation that consumers are paying a premium in terms of foregone wages in order to live in cities with better shapes. Moreover, it suggests that city shape is more a consumption than it is a production amenity. When city area is explicitly included in the regression, the reduced-form relationship between area and wages is negative, as predicted by the conceptual framework (condition 17 in Section 3).

[Insert Table 6]

Tables 6 reports the same set of specifications for house rents. In panel A the dependent variable considered is the log of yearly housing rent per square meter, in 2014 Rupees, averaged throughout all urban households in the district. In panel B the dependent variable is analogous, but constructed averaging only the upper half of the distribution of urban housing rents in each district. This addresses the concern that reported rents are a downward-biased estimate of market rents due to rent control policies. These estimates appear noisy or only borderline significant, with p values between 0.10 and 0.15. However, a consistent pattern emerges: the impact of disconnected shape on rents is negative in the IV and close to zero, or possibly positive, in the OLS. Appendix Table 2, panel B, shows that these results are qualitatively similar including year fixed effects interacted by initial shape. This is consistent with the interpretation that consumers are paying a premium in terms of higher housing rents in order to live in cities with better shapes. The reduced-form relationship between city area and rents is also negative, consistent with the conceptual framework (condition 18 in Section 3).

6.4 Interpreting estimates through the lens of the model

Tables 3, 5 and 6 provide estimates for the reduced-form relationship between city shape and respectively log population, wages and rents, conditional on city area. Although results for wages and rents should be interpreted with caution due to the data limitations discussed above, the signs of the estimated coefficients are consistent with the interpretation that consumers view compact shape as an amenity. In this sub-section I use these reduced-form estimates to back out the implied welfare loss associated to poor city geometry, according to the model outlined in Section 3. I focus here on the disconnection index, representing the average commute within the footprint. All monetary values are expressed in 2014 Rupees.

Recall that a one-unit increase in shape metric S has a welfare effect equivalent to a decrease in income of λ_θ log points, which, as derived in Section 3 (eq. 38) can be estimated as

$$\widehat{\lambda_\theta} = \alpha\widehat{B_P} - \widehat{B_W}$$

where α is the share of consumption spent on housing.

My most conservative point estimates for $\widehat{B_W}$ and $\widehat{B_P}$, from Specification II as estimated in Tables 5 and 6A respectively, amount to 0.038 and -0.518. To calibrate α, I compute the share of household expenditure devoted to housing for urban households, according to the NSS Household Consumer Expenditure Survey data in my sample - this figure amounts to 0.16. The implied $\widehat{\lambda_\theta}$ is -0.14. This implies that a one standard deviation increase in disconnection for the average-sized city (about 360 meters) entails a welfare loss equivalent to a 0.05 log points decrease in income.

It is interesting to compare this figure with the *actual* cost of an extra 360 meters in one's daily commute, under different commuting options. Postulating one round-trip per day, 5 days per week, this amounts to 225 extra kilometers per year. To compute the time-cost component of commuting, I estimate hourly wages by dividing the average yearly wage in my sample (93950 Rs.) by 312 yearly working days and 7 daily working hours, obtaining a figure of 43 Rs. per hour. Assuming that trips take place on foot, at a speed of 4.5 km per hour, a one-standard deviation deterioration in shape amounts to 50 extra commute hours per year, which is equivalent to 2.3% of the yearly wage. This figure is roughly 45% of the welfare cost I estimate. Assuming instead that trips occur by car, postulating a speed of 25 km per hour, a fuel efficiency of 5 liters per 100 km, and fuel prices of 77 Rs. per liter[23], the direct cost of increased commute length amounts to 1.3% of the yearly wage, or roughly one quarter of the welfare cost estimated above.

To complete the exercise, let us now consider the effect of city shape S on firms. The signs of the reduced-form estimates for B_W, B_N and B_P are in principle compatible with city shape being a production amenity or disamenity. The effect of S on productivity is pinned down by

[23]These figures are based respectively on: Ministry of Urban Development, 2008; U.S. Energy Information Administration, 2014; http://www.shell.com/ind/products-services/on-the-road/fuels/fuel-pricing.html accessed in August 2014.

equation (37) from Section 3:

$$\widehat{\lambda_A} = (1-\beta-\gamma)\widehat{B_N} + (1-\gamma)\widehat{B_W}$$

where parameters β and γ represent the shares of labor and tradeable capital in the production function postulated in equation (7). My most conservative point estimates for $\widehat{B_N}$ and $\widehat{B_W}$ are -0.0981 and 0.038, from Tables 3C and 5 respectively. Setting β to 0.4 and γ to 0.3 the implied $\widehat{\lambda_A}$ is -0.003. which implies a productivity loss of about 0.001 log points for a one standard deviation deterioration in city disconnection. These estimates appear very small, and sensitive to the choice of parameter values. Postulating a labor share of 0.5 the implied productivity effect becomes positive: 0.002 log points for a one-standard deviation deterioration in shape. This exercise thus yields inconclusive results. Although the currently available evidence seems to indicate that the cost of disconnected shape is mostly borne by consumers, more work will be required to understand the implications of city shape for firms, possibly disaggregating by sector.

6.5 Channels and Heterogeneous Effects

The results presented so far provide evidence that consumers have a preference for more compact cities. In this Section, I seek to shed light on the mechanisms through which city shape affects consumers, and on the categories of consumers who are affected the most by poor urban geometry.

Infrastructure

Discussing the role of city shape, the urban planning literature typically emphasizes commute times as the main argument in favor of more compact city shapes. If transit times are indeed the main channel through which urban shape matters, road infrastructure should mitigate the adverse effects of poor geometry. By the same argument, all else being equal, consumers with individual means of transport should be less affected by city shape.

[Insert Table 7]

In Table 7 I attempt to investigate these issues interacting city shape with a number of indicators for infrastructure. For ease of interpretation, I resort to the more parsimonious

Specification II, which links normalized shape to population density (eq. (46) and (47)). Population density is measured in thousand inhabitants per squared km. Results are reported for the disconnection and range index. Recall that these two indexes represent respectively the average and maximum length of within-city trips. While disconnection is a general indicator for city shape, the range index might be more suitable to capture longer within-city trips, which might be more likely to require motorized means of transportation.

This exercise is subject to a number of caveats. An obvious identification challenge lies in the fact that infrastructure is not exogenous, but rather jointly determined with urban shape. This issue is investigated explicitly in Section 7.2, in which I present some results on current road infrastructure and city shape. In Table 7 I partly address the endogeneity of city infrastructure by employing state-level proxies. In columns 1 and 2, instrumented normalized shape is interacted with state-level motor vehicles density measured in 1991 - the first year for which this statistic was available[24]. This figure is provided by the Ministry of Urban Transport and Highways. In columns 3 and 4, I interact shape with urban road density in 1991, as reported by the 1991 Census Town Directory; again, 1991 is the first year in which the Census provides this figure. To cope with the potential endogeneity of this variable, in columns 5 and 6 I employ instead a state-level, leave-out mean of 1991 urban road density in the state. The coefficients on all three interaction terms - with the exception of the one in column 5 - are positive and highly significant, which can be interpreted as evidence that infrastructure mitigates the negative effects of poor geometry. However, we cannot fully rule out potential confounding effects: for instance, these positive interaction coefficients might be capturing differential trends across cities with different incomes.

.**Housing quality**

[Insert Table 8]

A complementary question relates to who bears the cost of poor city geometry. Emphasizing the link between city shape, transit and poverty, Bertaud (2004) claims that compact cities are in principle more favorable to the poor because they reduce distance, particularly in countries where they cannot afford individual means of transportation or where the large size of the city

[24] Unfortunately, I cannot compute a leave-out mean because I don't have access to city-level data on motor vehicles density.

precludes walking as a means of getting to jobs. At the same time, however, if compact cities are also more expensive, this would tend to reduce the housing floor space which the poor can afford (Bertaud, 2004). In Table 8, I make an attempt to investigate the link between poverty and city shape by looking at two - admittedly very rough - proxies for city income: the total and the fraction of slum dwellers and a principal component wealth index based on physical characteristics of houses, including roof, wall and floor material as well as the availability of running water and a toilet. This information is provided in the Census for a limited number of cities and years. As in previous tables, I provide IV estimates obtained with both Specification I and II, as well as OLS estimates for Specification I. I find that cities with less compact shapes have overall fewer slum dwellers, both in absolute terms and relative to total population. Moreover, less compact cities are characterized by houses of marginally better quality. In principle, this could arise from higher equilibrium rents in compact cities, which could be forcing more people into substandard housing. A more interesting interpretation, however, is that poorer migrants sort into cities with more compact shapes, possibly because of their lack of individual means of transport and consequent higher sensitivity to commute lengths. It is plausible that the marginal migrant to urban India would start out as slum dweller, which substantiates the sorting hypothesis.

7 Empirical Results: Endogenous Responses to City Shape

The evidence presented so far indicates that city shape affects the spatial equilibrium across cities, and in particular that deteriorating urban geometry entails welfare losses for consumers. A natural question which arises next concerns the role of policy: given that most cities cannot expand radially due to their topographies, what kind of land use regulations, if any, and infrastructural investments best accommodate city growth? Although a thorough normative analysis requires more theoretical work and possibly more detailed data than those used in this study, this section provides some evidence on the interaction between topography, city shape and policy. Due to data availability constraints, part of the analyses carried out in this section are based on comparatively small samples and/or rely on cross-sectional variation only. These results should therefore be interpret cautiously.

FARs

I start by considering the most controversial among land use regulations currently in place in urban India: Floor Area Ratios (FAR). As explained in Section 4.6, FARs are restrictions on building height expressed in terms of the maximum allowed ratio of a building's floor area over the area of the plot on which it sits. Higher values allow for taller buildings. Previous work has linked the restrictive FARs in place in Indian cities to suburbanization and sprawl (Sridhar, 2010; Sridhar and Malpezzi, 2012), as measured by administrative data sources. Bertaud and Brueckner (2005) analyze the welfare impact of FARs in the context of a monocentric city model, estimating that restrictive FARs in Bengaluru carry a welfare cost ranging between 1.5 and 4.5%.

Information on FAR values across Indian cities is very hard to obtain. My data on FARs is drawn from Sridhar (2010), who collects a cross-section of the maximum allowed FAR levels as of the mid-2000s, for about 50 cities[25], disaggregating by residential and non-residential FAR. Based on discussions with urban local bodies and developers, it appears that FARs are very resilient, and have been rarely updated. While the data collected by Sridhar reflects FARs as they were in the mid-2000s, they are likely to be a reasonable proxy for FAR values in place throughout the sample.

[Insert Table 9]

In Table 9, panel A, I explore the interaction between topography and FARs in determining city shape and area. The three first-stage equations presented in Table 2 are reproposed here, augmented with an interaction between each instrument and FAR levels. Each observation is a city-year. In columns 1, 2 and 3 I consider the average of residential and non-residential FARs, whereas in columns 4, 5 and 6 I focus on residential FARs.

The main coefficients of interest are the interaction terms. The interaction between projected population and FARs appears to have a negative impact on city area (columns 2 and 4), indicating that laxer FARs cause cities to expand less in space than their projected growth would imply. This is in line with the results obtained by Sridhar (2010) who finds a cross-sectional correlation between restrictive FARs and sprawl using administrative, as opposed

[25] Sridhar (2010) collectes data for about 100 cities, but many of those cities are part of larger urban agglomerations, and do not have appear as individual footprints in my panel. Moreover, some are too small to be detected by night-time lights. This reduces the effective number of observations which I can use in my panel to 55. My analysis is thus subject to significant power limitations.

to remotely-sensed data. This interaction term has a negative impact on city shape as well (columns 1 and 5), suggesting that higher FARs also can slow down the deterioration in city shape which city growth entails. The interaction between potential shape and FARs in column 3 is also negative, indicating that cities with higher FARs have a more compact shape then their topography would predict. This seems to suggest that if growing and/or potentially constrained cities are allowed to build vertically, they will do so, rather than expand horizontally and face topographic obstacles.

In Table 9, panel B, I investigate the impact of FARs interacted with city shape on population and density. Again, each observation is a city-year. The specifications proposed here are equivalent to those in Table 3, augmented with interactions between the explanatory variables and FARs. The corresponding interacted first-stage equations are proposed in panel A. Results are mixed, possibly due to small sample size. However, the results from the interacted version of Specification II (columns 1 and 4) suggest that laxer FARs mitigate the negative impact of non-compactness on population: the interaction between instrumented shape and FARs is positive, and significant in the case of residential FARs. An interpretation for this result is that long potential commutes matter less in cites which allow taller buildings, since this allows more consumers to live in central locations. This result, however, is not robust to the alternative Specification I (columns 2 and 5).

My next question relates to the determinants of FARs, in particular whether urban form considerations appear to be incorporated by policy makers in setting FARs. Panel C of Table 9 provides an attempt to investigate this issue by regressing FAR values on urban form indicators - shape and area - as measured in year 2005. The estimating equations are a cross-sectional versions of eq. (43) and (48), with FARs as a dependent variable. Each observation is a city in year 2005. Subject to the limitations of cross-sectional inference and small sample size, these results indicate that larger cities have more restrictive FARs, both in the IV and in the OLS. This is consistent with one of the stated objectives of FARs: curbing densities in growing cities. At the same time, however, the coefficient on shape is positive, both in the IV and in the OLS. This could either indicate a deliberate willingness of policy makers to allow for taller buildings in areas with constrained topographies, or might reflect a historical legacy of taller architecture in more constrained cities[26].

Road infrastructure

Next, I turn to another type of policy response to city shape: road infrastructure. In Section 6.3 I provide some suggestive evidence that infrastructure might mitigate the adverse effects of poor shape. In this section I am interested in the complementary question of whether current road infrastructure responds to city shape, and whether denser road networks are compensating for longer potential trips in less compact cities. Although the Census provides urban road density for a subset of cities and years, these data appear to be based on administrative boundaries which are seldom updated. Instead, I compute the total length of the current road network as it appears on the maps available on Openstreetmap, overlapping the street network with my night-time lights-based urban footprint boundaries. More details can be found in Section 4.6. In Table 10 I investigate the cross-sectional relationship between current road network and city shape, measured in 2010 - the most recent year in my sample

[Insert Table 10]

Each observation in Table 10 is a city in year 2010. Results are reported for the disconnection (columns 1 to 3) and range index (column 4 to 6). Recall that these two indexes represent respectively the average and maximum length of within-city trips. While disconnection is a general indicator for city shape, the range index might be more suitable to capture cross-city trips, possibly requiring highways. Panel A considers total road length, whereas panel B considers road length per capita. Columns 1 and 4 report the IV results from estimating a cross-sectional version of Specification I (equation (43)). Columns 3 and 6 report the same specification, estimated by OLS. Columns 2 and 5 report the IV results from estimating a cross-sectional version of Specification II (equation (46)).

Results are somewhat mixed, and not always consistent across specifications. However, the general pattern discernible from the IV results seems to be the following: non-compact cities have a road network which is less dense in absolute terms, but road density per person which is not statistically different from that of compact cities[27] .

[26] Discussions with urban local bodies and developers suggest that FARs reflect to some extent the traditional architectural style of different cities.

[27] The OLS coefficients display the opposite pattern: more disconnected cities have a denser road network

Location of establishments

I conclude by providing some preliminary evidence on another type of endogenous response to deteriorating shape: a tendency to become less monocentric and more dispersed. It is conceivable that, as cities grow into larger and more disconnected footprints, resulting in lengthy commutes to the historic core, businesses might choose to locate further apart from each other, and/or possibly form new business districts elsewhere. I attempt to shed light on this hypothesis by analyzing the spatial distribution of productive establishments listed in the Urban Directories of the 2005 Economic Census This data source is described in greater detail in Section 4.6.

The literature has proposed a number of methodologies to detect employment sub-centers within cities, mostly based on relative employment densities (Anas et al., 1998). At this stage, I employ a very simple indicator of concentration: Moran's I statistic (Anselin, 1995), an index of spatial clustering which has been proposed as an indicator of monocentricity vis-a-vis dispersion (Tsai, 2005).

[Insert Table 11]

In Table 11 I estimate the relationship between spatial concentration of firms, area and shape, in a cross-section of footprints observed in 2005. The dependent variable is Moran's I statistic, computed for the locations of establishments in each footprint. Higher values indicate greater spatial clustering. Results are reported for the disconnection (columns 1 to 3) and range index (column 4 to 6). Again, I consider cross-sectional versions of Specifications I (eq. 43) and II (eq. 48) The IV estimates suggest that establishments might be more spread out in cities with more disconnected shapes. If confirmed by further analyses, this could indicate a more pronounced tendency towards dispersion or polycentricity in cities whose shape is deteriorating.

---

in absolute, but not in per capita terms. This could be generated by the correlation between city size and non-compactness discussed in Section 6.1. Growing cities are typically less compact, and also tend to have better infrastructure. The observed pattern can be due to the fact that the spurious correlation between non-compactness and population is stronger than that between non-compactness and infrastructure - possibly because in rapidly expanding cities infrastructure responds to population growth with a delay.

8 Conclusions

In this project I investigate empirically the economic implications of city shape in the context of India. My results indicate that city geometry does affect the location choices of consumers, in the framework of spatial equilibrium across cities. In particular, my findings are consistent with compact city shape being valued as a consumption amenity. I estimate the welfare cost of deteriorating city shape to be roughly twice as large as the direct implied increase in commute cost. I also investigate interactions between policy and geography in determining urban shape, finding that restrictions on building height - in the form of low FARs - result in cities which are more spread out in space and less compact.

This project leaves a number of open questions for future research. First, my approach captures the overall effects of urban geometry, measured at the city level, but does not explicitly address the implications of geometry for the spatial equilibrium *within* cities. More work will be required to develop a theoretical framework and deliver testable implications for how city shape affects the location choices and commute patterns of consumers *within* cities. Although my empirical analysis highlights that more compact cities are generally preferred by the average consumer, it does not answer the question of who bears the costs of disconnected geometry within each city, and through which mechanisms. More detailed, geo-referenced data at the sub-city level might help shed light on these issues.

It would be also important to pin down with more confidence the channels through which city shape affects consumers. Throughout this project I interpret city geometry mainly as a shifter of commuting costs, and I employ shape metrics specifically constructed to capture the implications of shape for transit. Moreover, the evidence provided in section 6.3 seems to support the interpretation that urban commutes are indeed the main channel through which city shape affects location choices. However, there could be other, second-order channels through which city shape might have economically-relevant effects. First, as noted by Bertaud (2002), geometry affects not only transportation but all kinds of urban utilities delivered through spatial networks, including those collecting and distributing electricity, water and sewerage. The optimal layout of such networks is founded on different engineering theories than those which found transport, and different shape indicators would be required to capture these aspects of shape. However, there might still be interactions with the shape metrics used in my analysis. Second, the same topographic obstacles which make cities "disconnected" from an urban transit perspective, might also act as boundaries, promoting the separation of a city in different, disconnected neighborhoods and/or administrative units. This could have implications both in terms of political economy and of segregation. Previous research, although in very different contexts, suggests this might be the case: for instance, Hoxby (2000) finds that the number of school districts in US cities is predicted by the number of natural barriers. Ananat (2011) finds that US cities which were subdivided by railroads into a larger number of physically defined neighborhoods became more segregated. Similarly, disconnected urban footprints might provide a "technology" for creating segregation. Again, more disaggregated data at the sub-city level will be required to investigate these ramifications.

Finally, more work is required to understand the response of firms to city shape. City shape has the potential to affect firms both directly, by increasing transportation costs,and indirectly, by distorting their location choices. In Section 8 I provide some preliminary evidence that firms tend to be more spread out in disconnected cities. This might reflect an attempt to counteract the spatial mismatch with potential employees, but could come at the expenses of agglomeration externalities. These effects are likely to be heterogeneous across sectors, which would raise the complementary question of whether firms in different sectors sort into cities with different geometries.

References

[1] Ananat, E.O. (2011), "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality", American Economic Journal: Applied Economics, 3(2), 34-66.

[2] Anas, A., R. Arnott and K. Small (1998), "Urban Spatial Structure", Journal of Economic Literature, 36 (3), 1426-1464.

[3] Alonso, W. (1964), *Location and Land Use*, Cambridge: Harvard University Press.

[4] Angel, S., S. C. Sheppard and D. L. Civco (2005), *The Dynamics of Global Urban Expansion*, Washington D.C.:The World Bank, Transport and Urban Development Department.

[5] Angel, S., J. Parent, and D.L. Civco (2009), "Ten Compactness Properties of Circles: A Unified Theoretical Foundation for the Practical Measurement of Compactness", The Canadian Geographer, 54 (4), 441–461.

[6] Angel, S., J. Parent, and D.L. Civco (2009), "Shape Metrics", ESRI working paper.

[7] Anselin, L. (1995), "Local Indicators of Spatial Association - LISA", Geographical Analysis, 27 (2), 93–115.

[8] Balk, D.L., U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. (2006), "Determining Global Population Distribution: Methods, Applications and Data", Advances in Parasitology, 62, 119-156.

[9] Baum-Snow, N., L. Brandt, V. Henderson, M.Turner, Q. Zhang (2013), "Roads, Railroads and Decentralization of Chinese Cities", working paper.

[10] Baum-Snow, N. and M. Turner (2012), "Transportation and the Decentralization of Chinese Cities", working paper.

[11] Bento, A., M.L. Cropper, A.M. Mobarak, and K.Vinha (2005), "The Effects of Urban Spatial Structure on Travel Demand in the United States", Review of Economics and Statistics, 87 (3), 466-478.

[12] Bertaud, A. (2002), "Note on Transportation and Urban Spatial Structure",Washington, ABCDE conference, working paper.

[13] Bertaud, A. (2004), "The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence?", working paper.

[14] Bertaud, A. and J.K. Brueckner (2005), "Analyzing Building-Height Restrictions: Predicted Impacts and Welfare Costs", Regional Science and Urban Economics, Elsevier, 35 (2), 109-125.

[15] Bertaud A. and S. Malpezzi (2001), "The Spatial Distribution of Population in 35 World Cities: The Role of Markets, Planning, and Topography", Wisconsin-Madison CULER working papers 01-03.

[16] Blomquist, G., M. Berger, and J. Hoehn (1988). "New Estimates of the Quality of Life in Urban Areas", American Economic Review, 78 (1), 89-107.

[17] Brueckner, J. and K. S. Sridhar (2012), "Measuring Welfare Gains from Relaxation of Land-use Restrictions: The Case of India's Building-Height Limits", Regional Science and Urban Economics, 42 (6),1061-67.

[18] Burchfield, M., H.G.Overman, D.Puga, and M.A.Turner (2006), "Causes of Sprawl: A Portrait from Space", The Quarterly Journal of Economics, 121 (2), 587-633.

[19] Carroll, M., J. Townshend, C. DiMiceli, P. Noojipady, R.Sohlberg (2009), "A New Global Raster Water Mask at 250 Meter Resolution", International Journal of Digital Earth, 2(4).

[20] Cervero, R. (2001), "Efficient Urbanisation: Economic Performance and the Shape of the Metropolis", Urban Studies, 38 (10), 1651–1671.

[21] CIESIN - Columbia University, IFPRI, The World Bank, and CIAT (2011), *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points,*Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

[22] Clark, L.P., D.B. Millet and J.Marshall (2011), "Air Quality and Urban Form in U.S. Urban Areas: Evidence from Regulatory Monitors" , Environ. Sci. Technol., 45 (16), 7028-7035.

[23] Dev, S (2006), “Rent Control Laws in India: A Critical Analysis”, CCS Working Paper No. 158, Centre for Civil Society, New Delhi.

[24] Fernandes, A. and G. Sharma (2012), "Determinants of Clusters in Indian Manufacturing: The Role of Infrastructure, Governance, Education, and Industrial Policy", IGC working paper.

[25] Glaeser, E. (2011), *Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier*, New York: Penguin Press.

[26] Glaeser, E. (2008), *Cities, Agglomeration and Spatial Equilibrium*, Oxford: Oxford University Press.

[27] Glaeser, E. and M. Kahn (2004), "Sprawl and Urban Growth" in V. Henderson and J. Thisse (Eds.), *The Handbook of Regional and Urban Economics*, v. 4, North Holland Press.

[28] Gyourko, J., M. Kahn and J. Tracy (1999), "Quality of Life and Environmental Comparisons", in E.S. Mills and P. Cheshire (Eds.), *The Handbook of Regional and Urban Economics*, v.3, North Holland Press.

[29] Greenstone, M., and R. Hanna (forthcoming), "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India", American Economic Review.

[30] Henderson, M., E. Yeh, P. Gong, and C. Elvidge (2003), "Validation of Urban Boundaries Derived from Global Night-time Satellite Imagery", International Journal of Remote Sensing, 24 (3), 595-609.

[31] Henderson, J. V., A. Storeygard, and D.N. Weil (2012). "Measuring Economic Growth from Outer Space", American Economic Review, 102 (2), 994-1028.

[32] Hoxby, C. (2000), "Does Competition among Public Schools Benefit Students and Taxpayers?", American Economic Review, 90 (5), 1209–1238.

[33] Indian Institute for Human Settlements (2013), "Urban India 2011: Evidence", working paper.

[34] Joshi, P. K., B.M. Bairwa, R. Sharma, V.S.P. Sinha (2011), "Assessing Urbanization Patterns over India Using Temporal DMSP–OLS Night-time Satellite Data", Current Science, 100 (10), 1479-1482.

[35] Mc Kinsey Global Institute (2010), "Globalisation and Urban Growth in the Developing World with Special Reference to Asian Countries".

[36] Mills, E.S. (1972), *Studies in the Structure of the Urban Economy*, Baltimore: Johns Hopkins Press.

[37] Ministry of Urban Development (2008), *Study on Traffic and Transportation Policies and Strategies in Urban Areas in Indi*a, Government of India: New Delhi.

[38] Mitra, A. (1980), *Population and Area of Cities, Towns, and Urban Agglomerations, 1872-1971*, Bombay: Allied.

[39] Mitric, S. and I.Chatterton (2005), "Towards a Discussion of Support to Urban Transport Development in India", Washington, DC: World Bank.

[40] Morten, M. and J. Oliveira (2014), "Migration, Roads and Labor Market Integration: Evidence from a Planned Capital City", working paper.

[41] Muth, R.F. (1969), *Cities and Housing*, Chicago: University of Chicago Press.

[42] NASA and Ministry of Economy, Trade and Industry of Japan (METI), Land Processes Distributed Active Archive Center (LP DAAC) (2011), ASTER Global Digital Elevation Model, Version 2, USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. 2001.

[43] Pandey B, Joshi P.K., Seto K.C. (2013), "Monitoring Urbanization Dynamics in India Using DMSP/OLS Night Time Lights and SPOT-VGT Data". International Journal of Applied Earth Observation and Geoinformation, 23, 49-61.

[44] Shetty, J.P., V. Shashikala, T.V. Ramachandra (2010), "Urban Footprint Dynamics in Indian Metropolis", Conference on Infrastructure, Sustainable Transportation and Urban Planning, IISc, Bangalore.

[45] Roback, J. (1982), "Wages, Rents and the Quality of Life", Journal of Political Economy, 90 (6), 1257-1278.

[46] Rosen, S.(1979), "Wage-Based Indexes of Urban Quality of Life", in Mieszkowski, P. and M. Strazheim (Eds.), *Current Issues in Urban Economics*, Baltimore, Johns Hopkins University Press.

[47] Roychowdhury, K., Jones, S. D., Arrowsmith, C. (2009), "Assessing the Utility of DMSP/OLS Night-time Images for Characterizing Indian Urbanization", 2009 IEEE Urban Remote Sensing Joint Event, Shanghai, China.

[48] Saiz, A. (2010), "The Geographic Determinants of Housing Supply", Quarterly Journal of Economics, 125 (3), 1253-1296.

[49] Small, C., F. Pozzi and C.D. Elvidge (2005), "Spatial Analysis of Global Urban Extent from DMSP-OLS Night Lights", Remote Sensing of Environment, 96 (3), 277-291.

[50] Small, C., & Elvidge, C. D. (2013), "Night on Earth: Mapping Decadal Changes of Anthropogenic Night Light in Asia", International Journal of Applied Earth Observation and Geoinformation, 22, 40-52.

[51] Sridhar, K.S. (2010), "Impact of Land Use Regulations: Evidence From India's Cities", Urban Studies 47 (7), 1541–1569.

[52] Storeygard, A. (2014), "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa", working paper.

[53] Tsai, Y. (2005), "Quantifying Urban Form: Compactness versus 'Sprawl'", Urban Studies 42 (1), 141–161.

[54] U.S. Army Map Service (1955), *India and Pakistan Topographic Maps, Series U502, 1:250,000,* U.S. Army Map Service Topographic Map Series.

[55] U.S. Energy Information Administration (2014), Report on India, *http://www.eia.gov/countries/analysisbriefs/India/.*

[56] World Bank (2013), "Urbanization Beyond Municipal Boundaries: Nurturing Metropolitan Economies and Connecting Peri-urban Areas in India", Washington D.C: The World Bank.

Table 1: Descriptive Statistics

	Obs.	Mean	St.Dev.	Min	Max
Area, km^2	6276	62.63	173.45	0.26	3986.02
Remoteness, km	6276	2.42	2.22	0.20	27.43
Spin, km^2	6276	12.83	39.79	0.05	930.23
Disconnection, km	6276	3.30	3.05	0.27	38.21
Range, km	6276	9.38	9.11	0.86	121.12
Norm. remoteness	6276	0.71	0.06	0.67	2.10
Norm. spin	6276	0.59	0.18	0.50	6.81
Norm. shape	6276	0.97	0.08	0.91	2.42
Norm. range	6276	2.74	0.35	2.16	7.17
City population	1440	422869	1434022	5822	22085130
City population density (per km^2)	1440	15011	19124	432	239179
Avg. yearly wage, thousand 2014 Rs.	2009	93.95	66.44	13.04	838.55
Avg. yearly rent per m^2, 2014 Rs.	895	603.27	324.81	104.52	3821.59

Table 2: First Stage

	(1)	(2)	(3)
	OLS	OLS	OLS
	Shape of actual footprint, km	Log area of actual footprint, km^2	Norm. shape of actual footprint
		A. Shape Metric: Remoteness	
Shape of potential footprint, km	0.352***	0.0562***	
	(0.0924)	(0.0191)	
Log projected historic population	0.390**	0.488***	
	(0.160)	(0.102)	
Norm. shape of potential footprint			0.0413***
			(0.0141)
Observations	6,276	6,276	6,276
		B. Shape Metric: Spin	
Shape of potential footprint, km	0.585***	-7.43e-05	
	(0.213)	(0.000760)	
Log projected historic population	2.470	0.571***	
	(2.456)	(0.101)	
Norm. shape of potential footprint			0.0249**
			(0.0103)
Observations	6,276	6,276	6,276
		C. Shape Metric: Disconnection	
Shape of potential footprint, km	1.392***	0.152***	
	(0.229)	(0.0457)	
Log projected historic population	-1.180***	0.307***	
	(0.271)	(0.117)	
Norm. shape of potential footprint			0.0663***
			(0.0241)
Observations	6,276	6,276	6,276
		D. Shape Metric: Range	
Shape of potential footprint, km	1.935***	0.0641***	
	(0.357)	(0.0212)	
Log projected historic population	-4.199***	0.310***	
	(0.977)	(0.117)	
Norm. shape of potential footprint			0.0754**
			(0.0344)
Observations	6,276	6,276	6,276
Model for $\hat{r}$	city-specific	city-specific	common rate
City FE	YES	YES	YES
Year FE	YES	YES	YES

Notes: this table reports estimates of the first-stage relationship between city shape and area, and the instruments discussed in Section 5.1. Each observation is a city-year. Columns (1), (2), (3) report the results from estimating respectively equations (44), (45), (47). The dependent variables are shape, in km, log area, in km^2, and normalized shape (dimensionless) of the actual city footprint. The corresponding instruments are shape of the potential footprint, in km, log projected historic population and normalized shape of the potential footprint. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1) and (2), and on a common rate one in column (3) - see Section 5.1. Shape is measured by different indexes in different panels. Remoteness (panel A) is the average length of trips to the centroid. Spin (panel B) is the average squared length of trips to the centroid. Disconnection (panel C) is the average length of within-city trips. Range (panel D) is the maximum length of within-city trips. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Estimation is by OLS. All specifications include city and year fixed effects. Standard errors are clustered at the city level.*** p<0.01,** p<0.05,* p<0.1.

Table 3: Impact of City Shape on Population

	(1)	(2)	(3)
	IV	IV	OLS
	Log population	Population density	Log population
		A. Shape Metric: Remoteness	
Shape of actual footprint, km	-0.137**		0.0338***
	(0.0550)		(0.0112)
Log area of actual footprint, km^2	0.785***		0.167***
	(0.182)		(0.0318)
Norm. shape of actual footprint		-311.8***	
		(92.18)	
Observations	1,440	1,440	1,440
		B. Shape Metric: Spin	
Shape of actual footprint, km	-0.00101		0.000887***
	(0.000657)		(0.000288)
Log area of actual footprint, km^2	0.547***		0.197***
	(0.101)		(0.0290)
Norm. shape of actual footprint		-110.9***	
		(37.62)	
Observations	1,440	1,440	1,440
		C. Shape Metric: Disconnection	
Shape of actual footprint, km	-0.0991**		0.0249***
	(0.0386)		(0.00817)
Log area of actual footprint, km^2	0.782***		0.167***
	(0.176)		(0.0318)
Norm. shape of actual footprint		-254.6***	
		(80.01)	
Observations	1,440	1,440	1,440
		D. Shape Metric: Range	
Shape of actual footprint, km	-0.0284***		0.00763***
	(0.0110)		(0.00236)
Log area of actual footprint, km^2	0.746***		0.171***
	(0.164)		(0.0305)
Norm. shape of actual footprint		-84.14***	
		(27.70)	
Observations	1,440	1,440	1,440
Model for $\hat{r}$	city-specific	common rate	
City FE	YES	YES	YES
Year FE	YES	YES	YES

Notes: this table reports estimates of the relationship between city shape and population. Each observation is a city-year. Column (1) reports the results from estimating equation (43) (Specification I). The dependent variable is log city population. The explanatory variables are log city area, in km^2, and city shape, in km. Column (2) reports the results from estimating equation (46) (Specification II). The dependent variable is population density, in thousands of inhabitants per km2. The explanatory variable is normalized shape (dimensionless). In columns (1) and (2) estimation is by IV. The instruments are discussed in Section 5.2, and the corresponding first-stage estimates are reported in Table 2. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1) and (2), and on a common rate one in column (3) - see Section 5.1. Column (3) reports the same specification as column (1), estimated by OLS. Shape is measured by different indexes in different panels. Remoteness (panel A) is the average length of trips to the centroid. Spin (panel B) is the average squared length of trips to the centroid. Disconnection (panel C) is the average length of within-city trips. Range (panel D) is the maximum length of within-city trips. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Population is drawn from the Census of India (1951, 1991, 2001, 2011). All specifications include city and year fixed effects. Standard errors are clustered at the city level.*** p<0.01,** p<0.05,* p<0.1.

Table 4: First Stage and Impact of City Shape on Population, Robustness to Excluding Cities with Extreme Topographies

	(1)	(2)	(3)	(4)	(5)	(6)
	FS(1)	FS(2)	IV	FS(1)	FS(2)	IV
	Shape of actual footprint, km	Log area of actual footprint, km^2	Log population	Shape of actual footprint, km	Log area of actual footprint, km^2	Log population
Shape of potential footprint, km	1.365***	0.155***		1.376***	0.158***	
	(0.228)	(0.0462)		(0.233)	(0.0488)	
Log projected historic population	-1.201***	0.278**		-1.213***	0.294**	
	(0.267)	(0.118)		(0.276)	(0.123)	
Shape of actual footprint, km			-0.100**			-0.109**
			(0.0414)			(0.0428)
Log area of actual footprint, km^2			0.776***			0.796***
			(0.190)			(0.185)
Observations	6,006	6,006	1,373	5,996	5,996	1,385
Cities	433	433	433	440	440	440
Sample excludes	coastal cities	coastal cities	coastal cities	mountainous cities	mountainous cities	mountainous cities
City FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Notes: this table presents a robustness check to Tables 2 and 3, excluding cities with “extreme” topographies from the sample. Each observation is a city-year. Columns (1), (2) and (4), (5) are equivalent to columns (1), (2) in Table 1. They report OLS estimates of the first-stage relationship between city shape and area and the instruments discussed in Section 5.1. Columns (3) and (6) are equivalent to column (1) in Table 3, and report IV estimates of the impact of shape on log city population. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. The construction of the potential footprint is based on a city-specific model for city expansion – see Section 5.1. Columns (1) to (3) exclude from the sample cities located within 5 km from the coast. Columns (4) to (6) exclude from the sample cities with an elevation above 600 m. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Population is drawn from the Census of India (1951, 1991, 2001, 2011). Elevation is from the ASTER dataset. All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 5: Impact of City Shape on Wages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	IV	IV	OLS	IV	IV	OLS	IV	IV	OLS
	All districts			*Only districts with one city*			*Only top city per district*		
	Dependent variable: log wage								
Shape of actual footprint, km	0.0381	0.109***	0.0538***	0.0626	0.0996***	0.0586***	0.0600*	0.112***	0.0512***
	(0.0386)	(0.0275)	(0.0172)	(0.0536)	(0.0336)	(0.0150)	(0.0360)	(0.0300)	(0.0181)
Log area of actual footprint, km^2	-0.409		-0.0478	-0.167		-0.00936	-0.337		-0.0308
	(0.390)		(0.0356)	(0.465)		(0.0516)	(0.392)		(0.0478)
Observations	2,009	2,009	2,009	1,075	1,075	1,075	1,517	1,517	1,517
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate		city-specific	common rate	
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: this table reports estimates of relationship between city shape and average wages. Each observation is a city-year. The dependent variable is the log urban average of individual yearly wages in the city's district, in thousand 2014 Rupees. The explanatory variables are log city area, in km^2, and city shape, in km. Columns (1), (4), (7) report the results from estimating equation (43) (Specification I) by IV. Columns (3), (6), (9) report the same specification, estimated by OLS. Columns (2), (5), (8) report the results from estimating equation (48) (Specification II) by IV. Instruments are described in Section 5.2. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1), (4) and (7), and on a common rate one in column (2), (5), (8) – see Section 5.1. In columns (4), (5), (6) the sample is restricted to districts containing only one city. In columns (7), (8), (9) the sample is restricted to cities which are either the only ones or the top ones in their respective districts. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010). Wages are from the Annual Survey of Industries, waves 1990, 1994, 1995, 1997, 1998, 2009, 2010. All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 6: Impact of City Shape on Housing Rents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	IV	IV	OLS	IV	IV	OLS	IV	IV	OLS
	All districts			*Only districts with one city*			*Only top city per district*		
	A. Dependent variable: log yearly rent per square meter								
Shape of actual footprint, km	-0.596	-0.303	0.00204	-0.518*	-0.636	-0.00857	-0.684	-0.198	0.0107
	(0.396)	(0.758)	(0.0505)	(0.285)	(1.661)	(0.0736)	(0.538)	(0.806)	(0.0516)
Log area of actual footprint, km^2	-1.529		-0.0176	-0.919		-0.0632	-1.407		-0.0355
	(1.409)		(0.0847)	(0.870)		(0.108)	(1.582)		(0.0962)
Observations	895	895	895	476	476	476	711	711	711
	B. Dependent variable: log yearly rent per square meter, upper 50%								
Shape of actual footprint, km	-0.675	-0.274	-0.00526	-0.594*	-0.595	-0.00736	-0.790	-0.143	0.00685
	(0.432)	(0.752)	(0.0577)	(0.324)	(1.626)	(0.0825)	(0.604)	(0.780)	(0.0590)
Log area of actual footprint, km^2	-1.649		-0.00333	-1.040		-0.0485	-1.586		-0.0141
	(1.542)		(0.0976)	(0.990)		(0.125)	(1.786)		(0.110)
Observations	895	895	895	476	476	476	711	711	711
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate		city-specific	common rate	
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: this table reports estimates of relationship between city shape and average housing rents. Each observation is a city-year. In panel A the dependent variable is the log urban average of housing rent per m^2 in the city's district, in 2014 Rupees. In panel B the dependent variable is analogous, but the average is calculated considering only the top 50% of the district's distribution of rents per m2. The explanatory variables are log city area, in km^2, and city shape, in km. Columns (1), (4), (7) report the results from estimating equation (43) (Specification I) by IV. Columns (3), (6), (9) report the same specification, estimated by OLS. Columns (2), (5), (8) report the results from estimating equation (48) (Specification II) by IV. Instruments are described in Section 5.2. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1), (4) and (7), and on a common rate one in column (2), (5), (8) – see Section 5.1. In columns (4), (5), (6) the sample is restricted to districts containing only one city. In columns (7), (8), (9) the sample is restricted to cities which are either the only ones or the top ones in their respective districts. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010). Housing rents are from the NSS Household Consumer Expenditure Survey, rounds 62 (2005-2006), 63 (2006-2007) and 64 (2007-2008). All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 7: Interactions of City Shape with Infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	Dependent variable: population density					
Norm. shape	-183.9***	-48.25***	-103.9***	-30.34***	-140.5***	-37.71***
	(39.53)	(9.047)	(18.10)	(5.485)	(44.85)	(9.934)
Norm. shape x state motor vehicles density, 1991	0.0462***	0.0246***				
	(0.0179)	(0.00441)				
Norm. shape x urban road density, 1991			0.204***	0.0692***		
			(0.0204)	(0.00776)		
Norm. shape x state urban road density, 1991					0.000462	0.000207*
					(0.000312)	(0.000107)
Observations	1,286	1,286	1,406	1,406	1,380	1,380
Shape metric	disconnection	range	disconnection	range	disconnection	range
City FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Notes: this table investigates the impact of shape, interacted with infrastructure, on population. Each observation is a city-year. All columns report IV estimates of a specification similar to equation (46) (Specification II), augmented with an interaction between normalized shape and different infrastructure proxies. The dependent variable is population density, in thousands of inhabitants per km^2. The construction of the potential footprint is based on a city-specific model for city expansion – see Section 5.1. The shape metrics considered are the normalized disconnection index (columns (1), (3), (5)), and normalized range index (columns (2), (4), (6)). Disconnection is the average length of within-city trips. Range is the maximum length of within-city trips. In columns (1), (2) normalized shape is interacted with state motor vehicles density in 1991 (source: Ministry of Road Transport and Highways). In columns (3), (4) normalized shape is interacted with urban road density in the city, as reported in the 1991 Census. In columns (4), (5) normalized shape is interacted with a leave-out mean of urban road density in the state calculated using 1991 Census data and data from the Ministry of Road Transport and Highways .City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Population is drawn from the Census of India (1951, 1991, 2001, 2011). All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 8: Impact of City Shape on Housing Quality

	(1) IV	(2) IV	(3) OLS	(4) IV	(5) IV	(6) OLS	(7) IV	(8) IV	(9) OLS
	Log slum population			***Log slum population share***			***Housing index***		
Shape of actual footprint, km	-0.134**	-0.0431	-0.0353**	-0.113**	-0.0501***	-0.0498***	0.195*	0.174*	-0.00698
	(0.0540)	(0.0437)	(0.0167)	(0.0561)	(0.0185)	(0.0169)	(0.118)	(0.0940)	(0.0312)
Log area of actual footprint, km^2	0.592		0.134	0.264		0.0501	0.438		0.141
	(0.492)		(0.0829)	(0.633)		(0.0938)	(1.483)		(0.120)
Observations	1,067	1,067	1,067	1,067	1,067	1,067	822	822	822
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate		city-specific	common rate	
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: this table reports estimates of relationship between city shape and proxies for housing quality. Each observation is a city-year. Columns (1), (4), (7) report the results from estimating equation (43) (Specification I) by IV. Columns (3), (6), (9) report the same specification, estimated by OLS. Columns (2), (5), (8) report the results from estimating equation (48) (Specification II) by IV. The explanatory variables are shape, in km, and log city area, in km^2. Instruments are described in Section 5.2. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1), (4) and (7), and on a common rate one in column (2), (5), (8) – see Section 5.1. The dependent variables are the following: log slum population in the city (columns (1), (2), (3)), log of the share of slum to total population in the city (columns (4), (5), (6)), and a housing quality principal component index (columns (7), (8), (9)). Data on slums are drawn from the 1991, 2001 and 2011 Census. The housing quality index is the city-year average of a principal component index based on the following characteristics of Census houses: roof, wall and floor material, availability of running water and toilet in-premise. The source is the 2001 and 2011 Census. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010). All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 9: Urban Form and Floor Area Ratios

A. First-stage impact of FARs on city shape

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
	Shape of actual footprint, km	Log area of actual footprint, km^2	Norm. shape of actual footprint	Shape of actual footprint, km	Log area of actual footprint, km^2	Norm. shape of actual footprint
Log projected historic population	2.998	1.984**		1.322	1.086*	
	(2.758)	(0.795)		(1.841)	(0.647)	
Log projected historic population x FAR	-1.958*	-0.686**		-1.315*	-0.342	
	(1.023)	(0.319)		(0.779)	(0.270)	
Shape of potential footprint, km	0.156	-0.186		0.235	0.0928	
	(1.182)	(0.233)		-0.671	(0.187)	
Shape of potential footprint, km x FAR	0.665	0.135		0.617**	0.0244	
	(0.487)	(0.106)		(0.279)	(0.0810)	
Norm. shape of potential footprint			0.435***			0.337***
			(0.123)			(0.117)
Norm. shape of potential footprint x FAR			-0.0908**			-0.0571
			(0.0452)			(0.0371)
Observations	1,183	1,183	1,183	1,183	1,183	1,183
Model for $\hat{r}$	city-specific	city-specific	common rate	city-specific	city-specific	common rate

B. Impact of city shape and FARs on population

	IV	IV	OLS	IV	IV	OLS
	Log population	Population Density	Log population	Log population	Population Density	Log population
Shape of actual footprint, km	-0.0979		0.0509	-0.271**		0.0283
	(0.124)		(0.0312)	(0.122)		(0.0285)
Shape of actual footprint, km x FAR	0.00290		-0.0160	0.0688*		-0.00767
	(0.0472)		(0.0129)	(0.0371)		(0.0103)
Log area of actual footprint, km^2	0.683**		0.0109	0.828***		-0.0443
	(0.338)		(0.128)	(0.308)		(0.103)
Log area of actual footprint, km^2 x FAR	0.0995		0.0665	0.00166		0.0981***
	(0.116)		(0.0471)	(0.0886)		(0.0368)
Norm. shape of actual footprint		-97.49			-85.46	
		(102.5)			(108.1)	
Norm. shape of actual footprint x FAR		-10.91			-16.89	
		(43.88)			(45.12)	
Observations	252	252	252	252	252	252
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate	
FAR	average	average	average	residential	residential	residential
City FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Notes: each observation is a city-year. Panel A of this table reports estimates of the first-stage relationship between Floor Area Ratios, city area and shape. Columns (1), (2) (3) and (4), (5), (6) report the same specifications reported in Table 2, with the addition of interactions between each instrument and FARs. Panel B reports estimates of the relationship between Floor Area Ratios, shape and population. Columns (1), (2) (3) and (4), (5), (6) report the same specifications reported in Table 3, with the addition of interactions between each explanatory variable and FARs. FARs are drawn from Sridhar (2010) and correspond to the maximum allowed Floor Area Ratios in each city as of the mid-2000s. FARs are expressed as ratios of the total floor area of a building over the area of the plot on which it sits. Columns (1), (2), (3) consider the average of residential and non-residential FARs, while columns (4), (5), (6) only consider residential FARs. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Population is from the Census (1951, 1991, 2001, 2011). Population density is measured in thousands of inhabitants per km^2. All specifications include city and year fixed effects. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 9 (continued): Urban Form and Floor Area Ratios

	C. FARs Determinants, 2005					
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	OLS	IV	IV	OLS
		Avg. FAR			***Residential FAR***	
Shape of actual footprint, km	0.0508	0.000290	0.021	0.0715*	0.00515	0.0439**
	(0.0409)	(0.00741)	(0.0140)	(0.0401)	(0.00914)	(0.0199)
Log area of actual footprint, km^2	-0.190		-0.105*	-0.280		-0.177**
	(0.179)		(0.0540)	(0.176)		(0.0876)
Observations	55	55	55	55	55	55
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate	

Notes: This table investigates the cross-sectional relationship between city area, shape and Floor Are Ratios as of year 2005. Each observation is a city in year 2005. Columns (1) and (4) estimate a cross-sectional version of equation (43) (Specification I), with log FARs as a dependent variable, estimated by IV. Columns (3) and (6) present the same specification, estimated by OLS. Columns (2) and (5) estimate a cross-sectional version of equation (48) (Specification II), with log FARs as a dependent variable, estimated by IV. FARs are drawn from Sridhar (2010) and correspond to the maximum allowed Floor Area Ratios in each city as of the mid-2000s. FARs are expressed as ratios of the total floor area of a building over the area of the plot on which it sits. Columns (1), (2), (3) consider the average of residential and non-residential FARs, while columns (4), (5), (6) only consider residential FARs. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset, in year 2005. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 10: Impact of City Shape on Road Network, 2010

	(1) IV	(2) IV	(3) OLS	(4) IV	(5) IV	(6) OLS
	A. Total Roads					
	Log road length, km	Road density, km/km^2	Log road length, km	Log road length, km	Road density, km/km^2	Log road length, km
Shape of actual footprint, km	-0.0769*		0.0231**	-0.0237**		0.00655**
	(0.0429)		(0.0111)	(0.0109)		(0.00320)
Log area of actual footprint, km^2	1.292***		0.861***	1.279***		0.871***
	(0.147)		(0.0415)	(0.115)		(0.0379)
Norm. shape of actual footprint		9.131			-3.563	
		(16.44)			(3.816)	
Observations	429	429	429	429	429	429
	B. Roads per capita					
	Log road length per capita	Road density per capita	Log road length per capita	Log road length per capita	Road density per capita	Log road length per capita
Shape of actual footprint, km	0.0403		-0.0677***	0.00753		-0.0190***
	(0.0467)		(0.0149)	(0.0111)		(0.00427)
Log area of actual footprint, km^2	-0.201		0.346***	-0.143		0.317***
	(0.178)		(0.0597)	(0.138)		(0.0547)
Norm. shape of actual footprint		-0.0377			-0.0152	
		(0.0785)			(0.0211)	
Observations	429	429	429	429	429	429
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate	
Shape metric	disconnection	disconnection	disconnection	range	range	range

Notes: This table estimates the cross-sectional relationship between city area, shape and road network as of year 2010. Panel A considers total road length and panel B considers road length per capita. Each observation is a city in year 2010. Columns (1) and (4) estimate a cross-sectional version of equation (43) (Specification I), estimated by IV. Columns (3) and (6) present the same specification, estimated by OLS. In columns (1), (3), (4), (6), panel A, the dependent variable is log road length in the footprint, in km. In columns (1), (3), (4), (6), panel B, the dependent variable is log road length per capita in the footprint, in km per thousand inhabitants. Columns (2) and (5) estimate a cross-sectional version of equation (46) (Specification II), estimated by IV. In columns (2), (5), panel A, the dependent variable is road density in the footprint, in km per km^2. In columns (2), (5) panel B, the dependent variable is road density per capita, in in km per km^2 per thousand inhabitants. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1) and (4), and on a common rate on in columns (2) and (5) – see Section 5.1. Road length is computed based on current (2014) *Openstreetmap* maps of Indian cities, focusing on roads denoted as "trunk", "primary" or "secondary". Disconnection is the average length of trips within the city footprint, in km. Range is the longest possible within-footprint trip, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset, in year 2010. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Table 11: Impact of City Shape on the Spatial Distribution of Productive Establishments, 2005

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	OLS	IV	IV	OLS
	Dependent variable: Moran's I index of spatial concentration					
Shape of actual footprint, km	-0.198**		-0.00633	-0.0410**		-0.000758
	(0.0978)		(0.0248)	(0.0182)		(0.00673)
Log area of actual footprint, km^2	0.395**		0.0519	0.244**		0.0449
	(0.181)		(0.0498)	(0.0958)		(0.0403)
Norm. shape of actual footprint		22.33			-1.117	
		(103.1)			(1.345)	
Observations	196	196	196	196	196	196
Model for $\hat{r}$	city-specific	common rate		city-specific	common rate	
Shape metric	disconnection	disconnection	disconnection	range	range	range

Notes: This table investigates the cross-sectional relationship between city area, shape and the spatial concentration of productive establishments in cities as of year 2005. Each observation is a city in year 2005. The dependent variable is the Moran's I index of spatial concentration, computed in each city for the location of productive establishments, as reported in the urban directories of the 2005 Economic Census. Higher values indicate more spatial clustering. Columns (1) and (4) estimate a cross-sectional version of equation (43) (Specification I), estimated by IV. Columns (3) and (6) present the same specification, estimated by OLS. Columns (2) and (5) estimate a cross-sectional version of equation (48) (Specification II), estimated by IV. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1) and (3), and on a common rate on in columns (2) and (4) – see Section 5.1. The shape metrics considered are the disconnection index (columns (1), (2), (3)), and range index (columns (4), (5), (6)). Disconnection is the average length of trips within the city footprint, in km. Range is the longest possible within-footprint trip, in km. City shape and area is calculated from maps constructed from the DMSP/OLS Night-time Lights dataset, in year 2005. Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1

Appendix Table 1: Impact of Shape on Population, Robustness to Initial Shape x Year Fixed Effects

	(1)	(2)
	IV	IV
	Log population	Population density
Shape of actual footprint, km	-0.194***	
	(0.0596)	
Log area of actual footprint, km^2	0.973***	
	(0.222)	
Norm. shape of actual footprint		-254.6***
		(80.01)
Observations	1,440	1,440
Model for $\hat{r}$	city-specific	common rate
City FE	YES	YES
Year FE	YES	YES
Initial shape x year FE	YES	YES

Notes: this table presents a robustness check to Table 3, augmenting the specification with year fixed effects interacted with initial shape. Each observation is a city-year. Columns (1) and (2) are equivalent to columns (1) and (2) in Table 3. They report IV estimates of the relationship between city shape and area and log city population (col. 1) and population density, in thousand inhabitants per km^2 (col.2). Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. The construction of the potential footprint is based on a city-specific model for city expansion in column (1) and on a common rate on in column (2) – see Section 5.1. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010) and U.S. Army maps (1951). Population is drawn from the Census of India (1951, 1991, 2001, 2011). All specifications include city and year fixed effects, as well as year fixed effects interacted with the city's disconnection index measured in the initial year of the panel (1951). Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

Appendix Table 2: Impact of Shape on Wages and Rents, Robustness to Initial Shape x Year Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	All districts		*Only districts with one city*		*Only top city per district*	
	A. Dependent variable: log wage					
Shape of actual footprint, km	0.0116	0.0989***	0.0393	0.108**	0.0381	0.102***
	(0.0420)	(0.0272)	(0.0768)	(0.0421)	(0.0386)	(0.0298)
Log area of actual footprint, km^2	-0.358		-0.164		-0.274	
	(0.395)		(0.465)		(0.379)	
Observations	2,009	2,009	1,075	1,075	1,517	1,517
	B. Dependent variable: log yearly rent per square meter					
Shape of actual footprint, km	-0.739	-0.393*	-0.670*	-0.465*	-0.839	-0.447*
	(0.487)	(0.221)	(0.359)	(0.265)	(0.664)	(0.251)
Log area of actual footprint, km^2	-1.301		-0.619		-1.092	
	(1.355)		(0.807)		(1.444)	
Observations	896	896	477	477	712	712
City FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Initial shape x year FE	YES	YES	YES	YES	YES	YES
Model for $\hat{r}$	city-specific	common rate	city-specific	common rate	city-specific	common rate

Notes: this table presents a robustness check to Tables 5 and 6, augmenting the specifications with year fixed effects interacted with initial shape. Each observation is a city-year. In panel A the dependent variable is the log urban average of individual yearly wages in the city's district, in thousand 2014 Rupees. In panel B the dependent variable is the log urban average of housing rent per m^2 in the city's district, in 2014 Rupees. Columns (1), (3), (5) report the results from estimating equation (43) (Specification I) by IV. Columns (2), (4), (6) report the results from estimating equation (48) (Specification II) by IV. Instruments are described in Section 5.2. The construction of the potential footprint is based on a city-specific model for city expansion in columns (1), (3), (5), and on a common rate one in column (2), (4), (6) – see Section 5.1. In columns (3), (4) the sample is restricted to districts containing only one city. In columns (5), (6) the sample is restricted to cities which are either the only ones or the top ones in their respective districts. Shape is captured by the disconnection index, which measures the average length of trips within the city footprint, in km. City shape and area are calculated from maps constructed from the DMSP/OLS Night-time Lights dataset (1992-2010). Wages are from the Annual Survey of Industries, waves 1990, 1994, 1995, 1997, 1998, 2009, 2010. Housing rents are from the NSS Household Consumer Expenditure Survey, rounds 62 (2005-2006), 63 (2006-2007) and 64 (2007-2008). All specifications include city and year fixed effects, as well as year fixed effects interacted with the city's disconnection index measured in the initial year of the panel (1951). Standard errors are clustered at the city level. *** p<0.01,** p<0.05,* p<0.1.

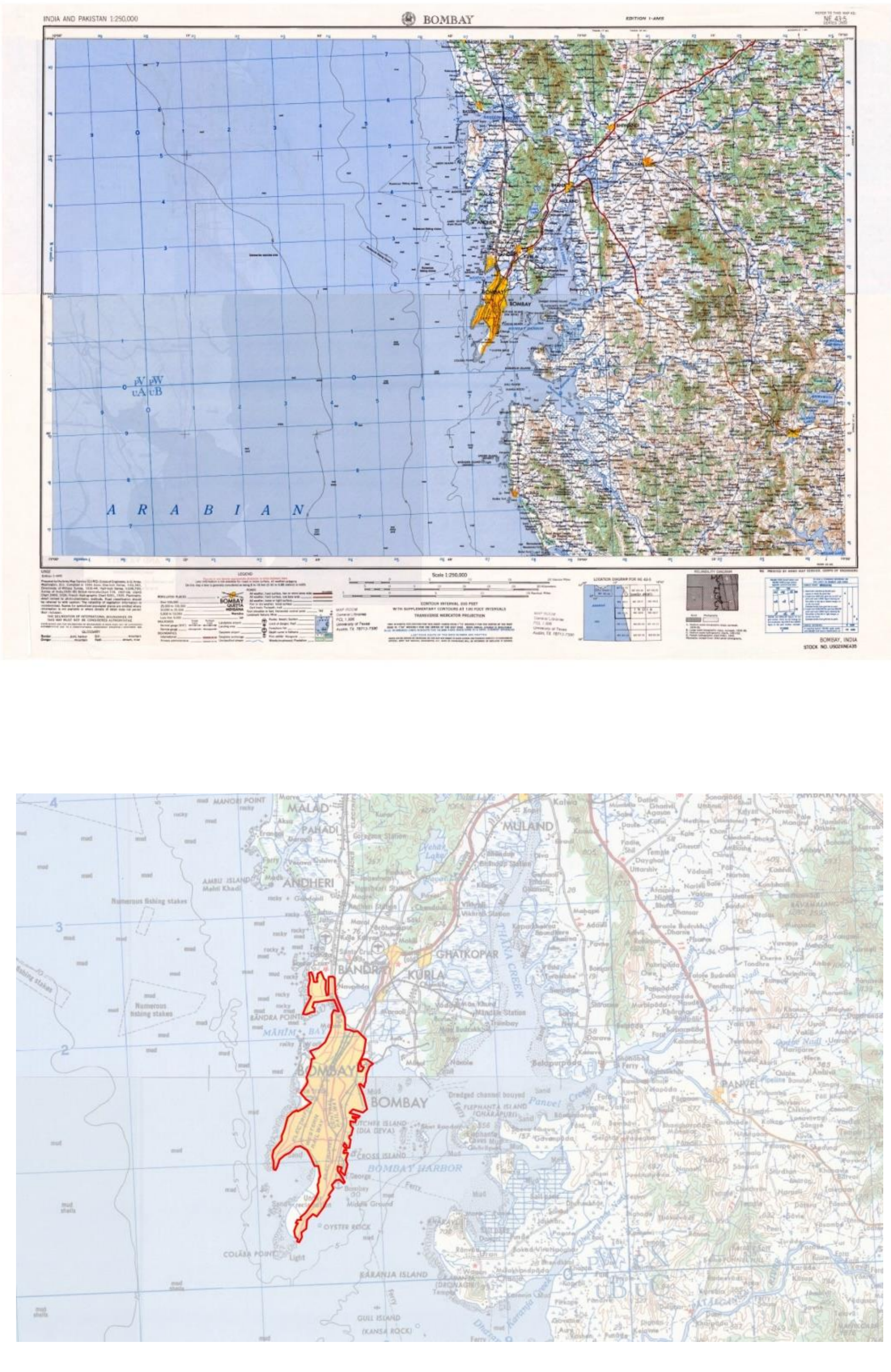


Figure 1:
U.S. Army India and Pakistan Topographic Maps

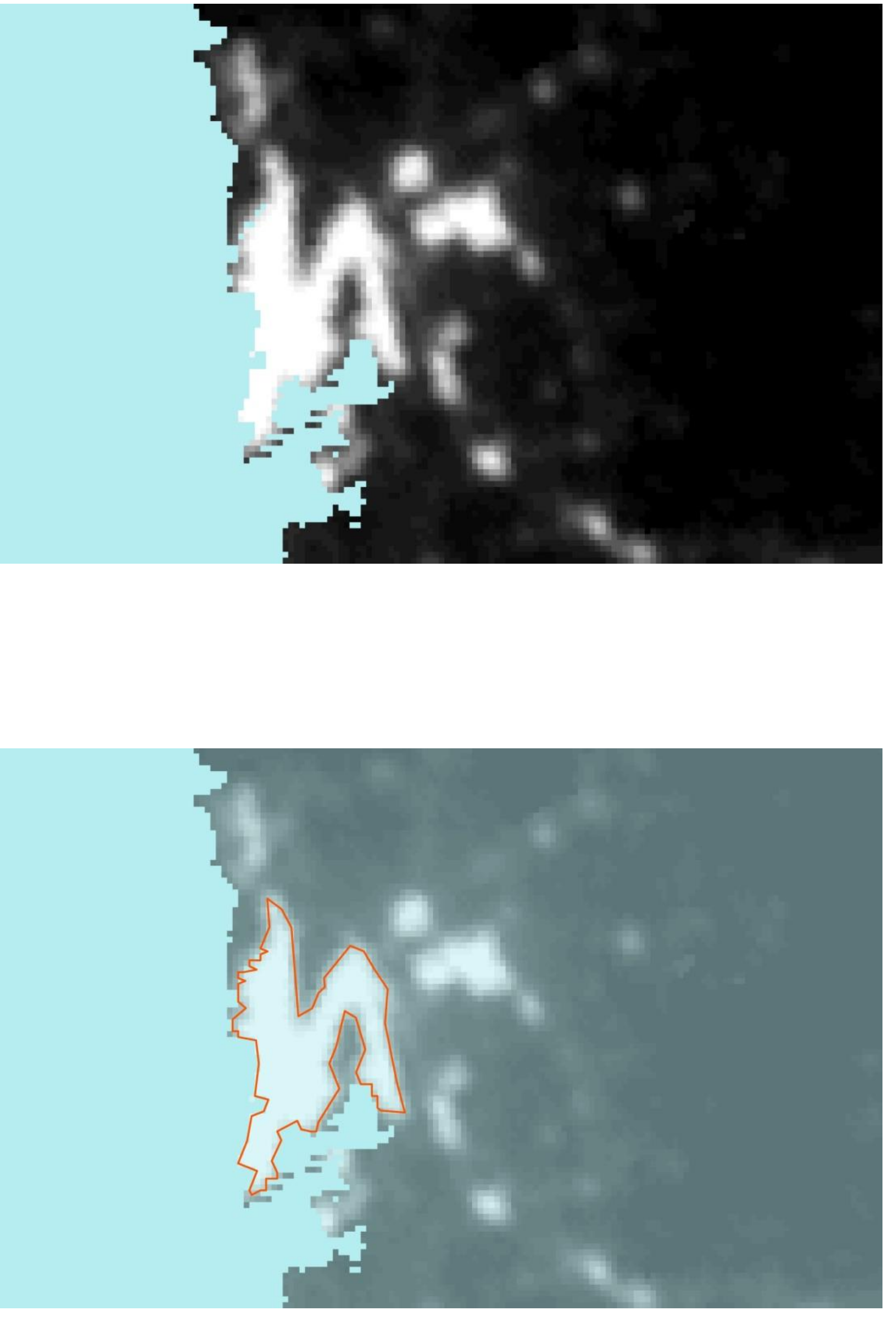


Figure 2
DMS/OLS nighttime lights, year 1992, luminosity threshold : 40.

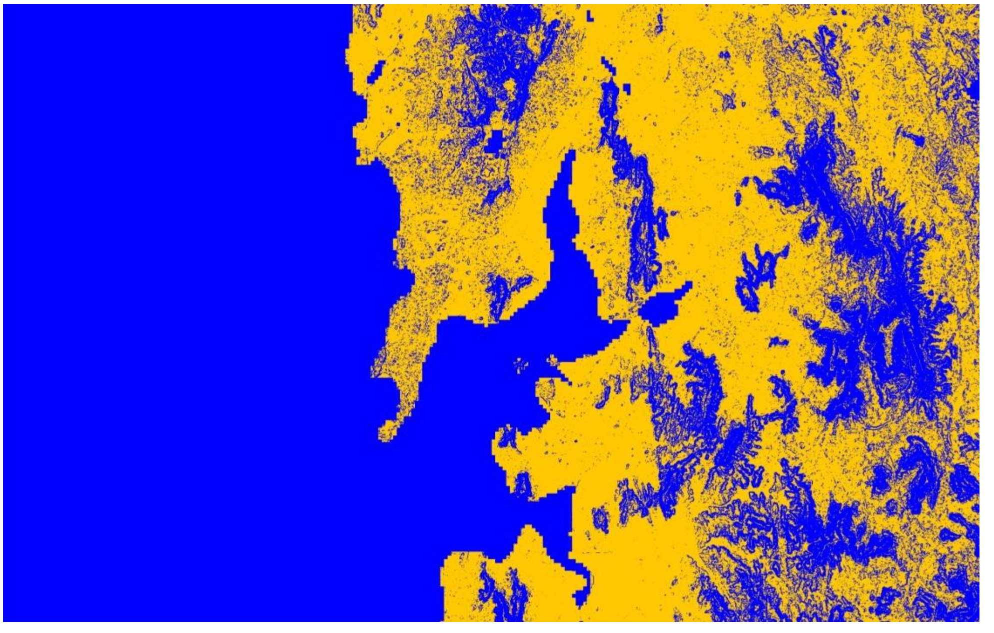
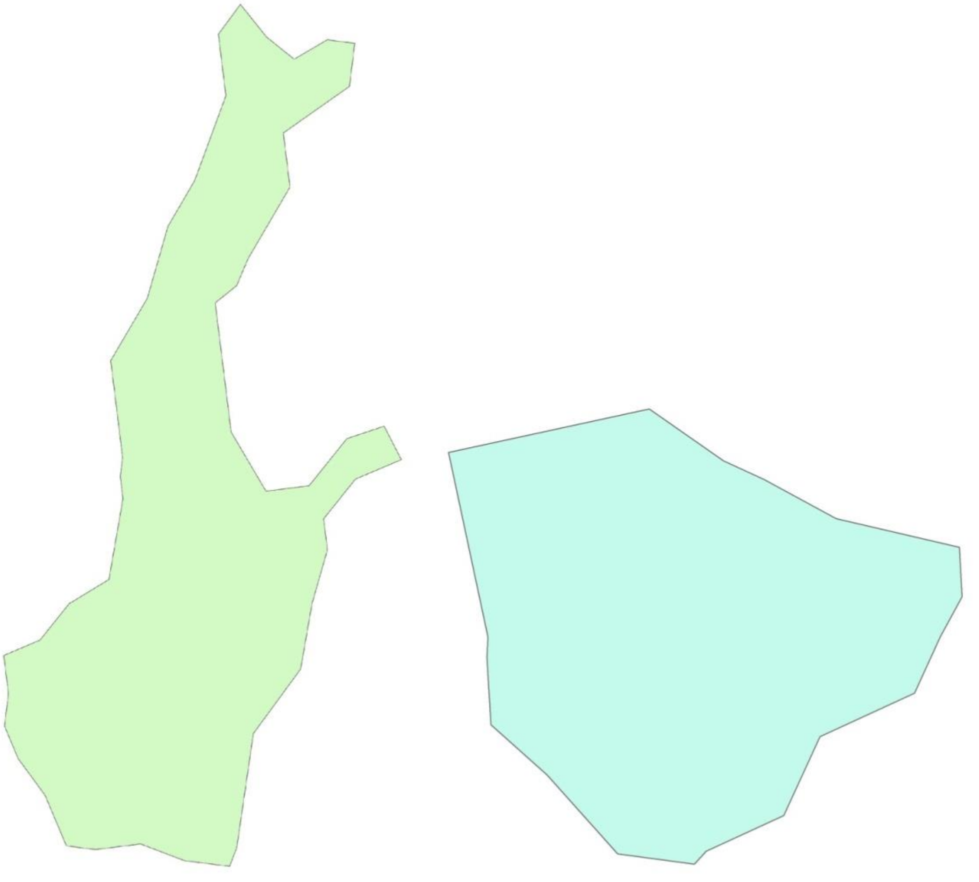


Figure 3
Developable vs. constrained land



	Kolkata		Bengaluru	
Shape metric		Normalized		Normalized
remoteness, km	14.8	0.99	10.3	0.69
spin, km^2	288.4	1.29	120.9	0.54
disconnection, km	20.2	1.35	14	0.94
range, km	62.5	4.18	36.6	2.45

Figure 4
Shape metrics: an example

Figure 5a

Figure 5b

Figure 5
Instrument construction

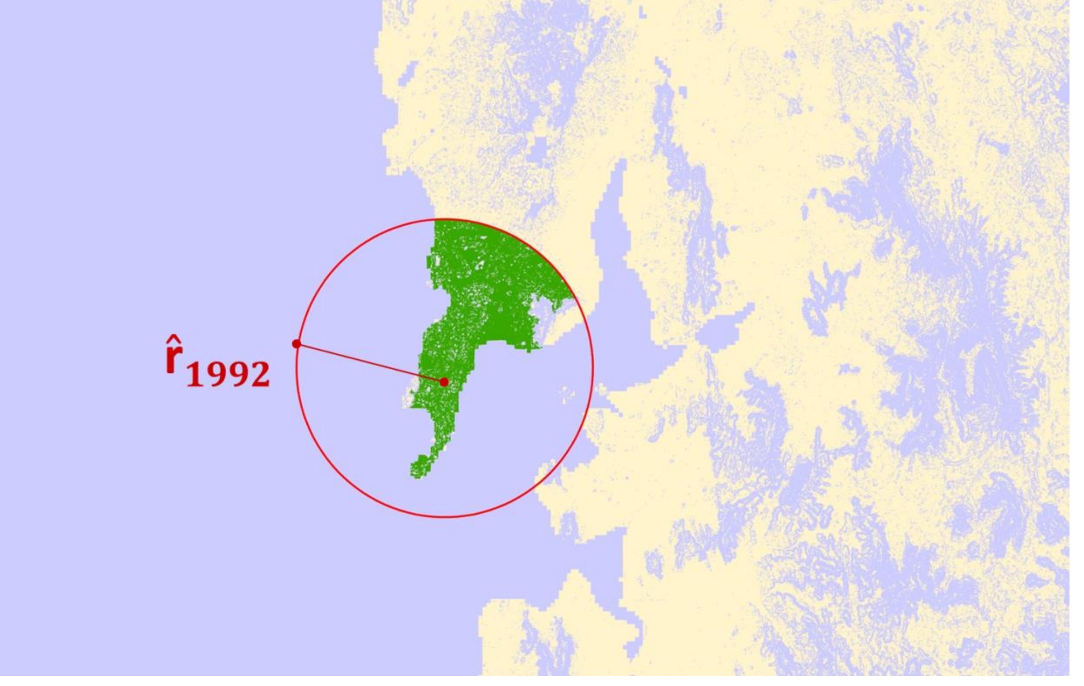


Figure 5c

Figure 5d

Figure 5
Instrument construction