Mobility and congestion in urban India

Prottoy A. Akbar[*] [†]
University of Pittsburgh

Victor Couture[*§]
University of California, Berkeley

Gilles Duranton[*‡]
University of Pennsylvania

Adam Storeygard[*¶]
Tufts University

June 2018

Abstract: Using a popular web mapping and transportation service, we generate information for more than 22 million counterfactual trip instances in 154 large Indian cities. We then develop a methodology to estimate robust indices of mobility for these cities. Our estimation allows for an exact decomposition of overall mobility into uncongested mobility and the congestion delays caused by traffic. We first document wide variation in mobility across Indian cities. We then show that this variation is driven primarily by uncongested mobility. Finally, we investigate correlates of mobility and congestion. Denser and more populated cities are slower, in part because of congestion, especially close to their centers. Urban economic development is correlated with better uncongested mobility, worse congestion, and overall with better mobility.

Key words: urban transportation, roads, traffic, determinants of travel speed, cities
JEL classification: R41

[*]This work is supported by the World Bank, the Zell Lurie Center for Real Estate at the Wharton School, the Fisher Center for Urban and Real Estate Economics at Berkeley-Haas, and we also gratefully acknowledge the support of the Global Research Program on Spatial Development of Cities at LSE and Oxford funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. We appreciate the comments from Leah Brooks, Ben Faber, Ejaz Ghani, Ed Glaeser, Vernon Henderson, Ki-Joon Kim, Emile Quinet, Christopher Severen, Kate Vyborny, and participants at conferences and seminars. Hero Ashman, Xinzhu Chen, Allison Green, Xinyu Ma, Gao Xian Peh, and Jungsoo Yoo provided us with excellent research assistance. We are immensely grateful to Sam Asher, Geoff Boeing, Arti Grover, Nina Harari, and Yue Li for their help with the data. The views expressed here are those of the authors and not of any institution they may be associated with.

[†]Department of Economics, University of Pittsburgh (email: `prottoyamanakbar@pitt.edu`); `https://pakbar.wordpress.com/`.

[§]Haas School of Business, University of California, Berkeley (email: `couture@haas.berkeley.edu`); `http://faculty.haas.berkeley.edu/couture/index.html`.

[‡]Wharton School, University of Pennsylvania (email: `duranton@wharton.upenn.edu`); `https://real-estate.wharton.upenn.edu/profile/21470/`.

[¶]Department of Economics, Tufts University (email: `Adam.Storeygard@tufts.edu`); `https://sites.google.com/site/adamstoreygard/`.

1. Introduction

Using a popular web mapping and transportation service, we generate information for more than 22 million counterfactual trip instances in 154 large Indian cities.[1] We then use this information to estimate a number of indices of mobility (speed) of motorized vehicle travel in these cities. We first assess the robustness of our indices to a wide variety of methodological choices. Second, we decompose overall mobility into uncongested mobility and the congestion delays caused by traffic. Third, we examine how indicators of urban economic development and other city characteristics correlate with mobility, uncongested mobility, and congestion delays. Finally, we provide additional mobility indices for walking and transit trips.

To the best of our knowledge, our paper provides the first systematic empirical investigation of mobility and congestion across cities in a developing country.[2] Our main substantive findings are the following. First, there are large differences in mobility across Indian cities. A factor of nearly two separates the fastest and slowest cities. Second, this variation is driven primarily by uncongested mobility, not congestion. An index of uncongested mobility explains 70% of the variance in overall mobility across cities. Traffic is generally slow in many Indian cities, even outside peak hours.[3] In the slowest decile, we find both small cities, which are slow even without congestion, and large congested cities. Congestion only really matters close to the center of the largest cities. Finally, we find that denser, more populated cities are slower, that there is a hill-shaped relationship between city per capita income and mobility, and that a city's mobility is related to characteristics of its road network.

This investigation is important for four reasons. First, there is an extreme paucity of useful knowledge about urban transportation, especially in developing countries. As a first building block towards a more serious knowledge base on urban transportation, some

[1]By counterfactual, we mean trip instances that have not been actually taken by a household. As we show below, these trips were selected to mimic some characteristics of trips that are taken by households in other contexts.

[2]Two new studies focusing on a single developing city complement our cross-city investigation: Kreindler (2018) studies the welfare impact of congestion pricing in Bangalore, and Akbar and Duranton (2018) measure the cost of congestion in Bogotá.

[3]We take a broad definition of 'congestion' and measure it as difference between travel time at a given time relative to travel time in the absence of traffic. Alternative natural measures of congestion with our data include, for instance, the ratio of the fastest to the slowest instance of trips.

stylized facts are needed.[4] For instance, we need to know how slow travel is in developing cities beyond the anecdotal evidence offered by disgruntled travelers. Equally important objects of interest are the differences between cities, between different parts of the same city, and across times of day within the same city.[5] We hope that our results, methodology, and data sources can help guide policy and future research on urban transportation in developing countries. We devote much of the last section of our paper to providing such guidance.

Second, there is a popular view that urbanization and economic development lead to ever larger cities and increased rates of motorization. According to this view, these two features will eventually lead to complete gridlock. We do find evidence of congestion in the largest Indian cities and a strong association between congestion and household access to motorized vehicles. However, economic development also brings about better travel infrastructure which facilitates uncongested mobility. In fact, indicators of urban economic development such as faster recent population growth, higher income levels, and higher motorization rates are generally associated with better overall mobility despite worse congestion.

Third, urban transportation in developing countries is prioritized for massive investments. For instance, transportation is the largest sector of lending by the World Bank and represents more than 20% of its net commitments as of 2016.[6] Among the many problems that these investments are trying to remedy, the lack of urban land devoted to the roadway is widely perceived to be a chief cause behind slow mobility and urban congestion. Providing an assessment of the determinants of mobility to guide policy is thus fundamental. For instance, we find suggestive evidence that better mobility is associated with a more regular grid network and more primary roads.

Fourth, the approach we develop here is an important stepping stone towards measuring

[4]In richer countries, much of our knowledge stems from representative surveys of household travel behavior. These surveys nonetheless have clear limitations, including a lack of precision in what travelers report. They are also prohibitively expensive to carry out broadly in developing countries. For the US, the Bureau of Transportation Statistics reports a cost per household of perhaps $300 to produce the National Household Transportation Survey or about $40 million in total (see `http://onlinepubs.trb.org/onlinepubs/reports/nhts.pdf`. Accessed, January 22, 2018.)

[5]Several software and data services such as Inrix and TomTom propose popular measures of congestion for a large sample of world cities. These services do not make the details of their methodology public. It seems that they monitor either specific roads or average traffic speed. We show below that measures of average speed are problematic and perform poorly.

[6]`http://pubdocs.worldbank.org/en/801011473440949738/WBAR16-FY16-Lending-Data.pdf`. Accessed, January 23, 2017.

accessibility, which is ultimately relevant to welfare.[7] In our companion paper (Akbar, Couture, Duranton, and Storeygard, 2018), we rely on the mobility (speed) index developed here as key component of an analogous accessibility (travel time) index. The other key component of accessibility is a proximity (distance to destinations) index, which also builds on the approach that we develop here.

Our investigation raises three challenges. The first is methodological. We propose a new approach to measure various forms of mobility from trip information, and to decompose them into uncongested mobility and delays caused by congestion. The second is a travel data challenge. There is no comprehensive source of data about urban transportation in Indian cities. Our approach is to collect data on predicted travel time from a popular website, Google Maps (GM).[8] For each city, we designed a sample of trips and sampled each trip at different times on different days. Our main worry is that these counterfactual trips may not be representative of the actual travel conditions faced by city residents. To address this worry, we use four different trip design strategies. These strategies aim to replicate some characteristics of actual trips taken by urban households in other countries. We show that our city mobility indices vary little by sampling strategies, type of trip destinations, origin and direction of travel, or time of day. Finally, we face the challenge of consistently defining and measuring the cities in which we measure counterfactual trips. To answer this challenge, we rely on a wide variety of sources including the census of India, OpenStreetMap, and satellite imagery.

2. Data collection

In this section we provide an overview of our data. Further details are available in Appendix A.

2.1 *City sample*

United Nations (2015) reports the names and locations of 166 cities in India that reached a population of 300,000 by 2014. Following Harari (2016) and Ch, Martin, and Vargas (2017),

[7]Formal welfare measures of accessibility were pioneered by Ben-Akiva and Lerman (1985) but their data requirements made it hard to implement them empirically. See Couture (2014) for recent developments and Duranton and Guerra (2016), Venter (2016), or Quinet (2017) for reviews on the topic.

[8]`https://en.wikipedia.org/wiki/Google_Maps`. Accessed, January 23, 2017. A number of new studies, which we discuss later in the paper, also use Google Maps to measure traffic in a developing city, notably Kreindler (2016), Hanna, Kreindler, and Olken (2017), and Akbar and Duranton (2018).

we initially define the spatial extent of these cities using nightlights. Within these light boundaries, we restrict attention to 40-meter pixels defined as built-up in 2014 according to the Global Human Settlements Layer (GHSL) of the European Commission's Joint Research Centre (JRC). After dropping cities for which no appropriate light exists, aggregating multiple cities within the same contiguous light, and dropping cities for which the relevant GHSL data are missing, we are left with an estimation sample of 154.

2.2 *Trips data*

We define a *trip* as a pair of points (origin and destination) within the same city as defined above. A *trip instance* is a trip taken at a specific time. Our target sample for city c is $15\sqrt{Pop_c}$ trips, where Pop_c is the projected 2015 population of city c from United Nations (2015), and 10 trip instances per trip, to ensure variation across times of day. For a city of population, say, one million, our sampling strategy thus targets 15,000 trips and 150,000 trip instances. Our sampling strategy is symmetrical, in the sense that each trip from origin o to destination d has a counterpart trip from origin d to destination o.[9] All trips are restricted to be at least one kilometer between origin and destination because Google results are less reliable for very short trips, few of which we expect to be motorized anyway. We sample across times of day to roughly match the weekday distribution of actual trips in Bogotá from Akbar and Duranton (2018). We oversample sparse overnight periods, and sample weekends at half the rate of weekdays.

We sample across four broad classes of trips, each designed to reflect key aspects of urban travel: radial, circumferential, gravity, and amenity trips.

Radial trips join a randomly located point within 1.5 kilometers of a city's center (as defined by United Nations, 2015) with another point in the city, either approximately 2, 5, 10, or 15 kilometers away, or at a distance percentile drawn from a uniform distribution. These trips are those predicted by the standard monocentric model of cities (Alonso, 1964, Mills, 1967, Muth, 1969). This models a reasonable first-order characterization of the distribution of population, density, and land and house prices in cities of many countries (see Duranton and Puga, 2015, for a survey).

Circumferential trips, orthogonal to radial trips, join a randomly located origin at least 2 kilometers from the city center with a destination at approximately the same radius but

[9] Unless otherwise indicated, random points are drawn with uniform probability from a support that is all valid 40-meter pixels within a city as defined above.

displaced approximately 30 degrees clockwise or counterclockwise.

Gravity trips join a random origin with a destination in a random direction, at a distance that is drawn from a truncated Pareto distribution with shape parameter 1 and support between one kilometers and 250 kilometers. Both commutes and city trips in general have been shown to reflect this distribution in many contexts (Ahlfeldt, Redding, Sturm, and Wolf, 2015, Akbar and Duranton, 2018).

Amenity trips join a random origin with an instance of one of 17 amenities (e.g. shopping malls, schools, train stations) as recorded in Google Places. The particular establishment selected is based on a combination of proximity and "prominence" assigned by Google. The weighting across these amenity types is based on a mapping of amenities to trip purposes whose share we draw from the 2008 US National Household Transportation Survey (NHTS) (Couture, Duranton, and Turner, 2018).

Using the sampling scheme above, we simulated 22,661,818 trip instances in Google Maps, covering 1,166,738 locations pairs and, hence, 2,333,476 trips across all cities and strategies, over 40 days between September and November of 2016.[10] For each trip, we record origin, destination, trip type, and length and estimated duration of Google's recommended route under current traffic conditions (which we sometimes refer to as real-time travel time), as well as the time required for the same route without traffic and with "typical" traffic.[11]

Google's route selection and speed estimates are based on the location and speed of mobile phones using the Android operating system, as well as other phones running Google software, especially Google Maps. Accurate measurement thus requires that drivers are providing information. It is therefore possible that estimates are worse in cities with lower mobile phone penetration. This is unlikely to affect our results. There were 300 million smartphone users in India as of the 4th quarter of 2016.[12] In December of 2015, 71% of

[10] A further 115,733 trip instances were collected for Bokaro Steel City in December 2017 as the UN database initially reported its location incorrectly. However, Bokaro is excluded from all results in section 6. We also describe the data we use for transit and walking trips below.

[11] While Google Maps does not report how it calculates travel time under regular traffic conditions, it generally provides the same answer for the same trip queried on different week days at the same time but not for the same trip queried at the different times.

[12] Source: http://www.counterpointresearch.com/press_release/indiahandset2016q4analysis/. While not all smartphones use Android, in the second quarter of 2016, 97% of smartphones shipped in India did. Source: http://indianexpress.com/article/technology/googles-android-captured-97-indian-smartphone-market-share-in-q2-2016-report-2957566/

mobile internet users were urban.[13] Given a 1.324 Billion population of India in 2016, and a 31% urbanization rate from the 2011 Census, a naive calculation implies that 52% of urban residents, including residents of smaller cities, and children, have smartphones. In setting up their phones, users may choose to opt out of sending information to Google. However, the opt-out rate, which Google does not publish, would have to be extremely high to affect our results. Crucially, to estimate slowed traffic on a block, Google only needs one vehicle with a phone, and by definition, time-varying congestion implies many vehicles. Put together, this suggests that all cities have enough phones to generate high-quality speed estimates. We discuss further evidence regarding the reliability of Google Maps information below.

2.3 *City-level data*

Several pieces of information were derived from administrative data. Daily labor earnings by district and gender are from the Employment and Unemployment Survey of the National Sample Survey (NSS-EUE) 2011–12. Population, and share of population with access to a car or motorcycle by "town" (fourth administrative level) are from the 2011 Census. We assign city populations as follows. The population of those towns falling completely within a city light are fully included. Towns falling partially within a city light contribute a share of their population defined by the share of the town's land area falling in the light. The other census variables (earnings, share of households with access to a car, motorcycle) are analogously aggregated using the resulting town population shares.

Weather data are from Weather Underground.[14] Data were available for 112 of 154 cities, for from one to 144 periods per day, with a median of eight. Population growth from 1990 to 2015 is from United Nations (2015). We also use variables that characterize 'urban shape' computed by Harari (2016). Data on characteristics of the road network within a (lights-based) city is from OpenStreetMap via GeoFabrik, and processed through OSMnx.[15]

---

[13] Source: `http://indianexpress.com/article/technology/tech-news-technology/mobile-internet-users-in-india-to-reach-371-mn-by-june-2016/`. While this is not just smartphones, presumably smartphone users are substantially more likely than other mobile phone users to be mobile internet users.

[14] `https://www.wunderground.com/`

[15] `http://download.geofabrik.de/asia/india.html` Accessed 2016/9/23.

3. A methodology for measuring mobility

3.1 A general conceptual framework

Consider the following general travel problem faced by a household. Its members work and conduct errands at several destinations, selected from a potentially large choice set. Potential destinations are costly to reach. To maximize utility, the household will choose to undertake some trips and not others. Some important decisions like household location and car purchases may also be made simultaneously with local mobility and accessibility. Fully modeling this presents overwhelming theoretical challenges and data requirements.

This travel problem is clearly not tractable unless we drastically simplify it. As a starting point, we note that the household travel problem is not unlike the standard consumption problem where consumers choose their basket from a large number of goods. We often simplify this consumption problem by considering a price index. We can do the same thing for the choice of destinations made by households. In each city, we can consider a number of residential locations and attempt to measure the cost of a 'typical' trip. The data requirements are still considerable but no longer overwhelming. The pitfalls of this approach are the same as those associated with typical price indices. Not knowing the preferences of households, it is unclear how travel costs (i.e., the prices) should be aggregated, keeping in mind that different households with different preferences face different price indices.

To minimize these pitfalls, we show that our mobility indices do not depend on how we weight different kinds of trips. In particular, our indices vary little by sampling strategies, type of trip destinations, origin and direction of travel, or time of day. This is because slower cities are slower at all times, for all types of trips, and throughout the city. As a result, we need not rely on a particular utility specification to tell us how to weight, say, a trip to the train station at peak hour on a weekday relative to a trip to a shopping destination on the weekend.[16]

[16] While generalized transportation costs involve money, time, and several dimensions of travel comfort and travel conditions (Small and Verhoef, 2007), here we can only focus on time. This generalization is not as extreme as it seems. First, if we think of travel time as home production and value it at half the wage as is customary in the literature, it represents a large share of the overall cost of travel. Second, many other components of travel costs such as gas consumption and vehicle depreciation are also correlated with travel distance and thus with travel time.

3.2 *Measuring mobility*

We want to measure the ease of going from an origin to a destination in cities. We focus on the speed of road travel using a motorized vehicle.[17] Measuring the speed of travel in a city raises a number of challenges since trips differ considerably in their length, location of origin and destination, time and day of departure, and mode.

The simplest approach is to compute a measure of mean speed for a given city:

$$S_c^m = \frac{\sum_{i \in c} D_i}{\sum_{i \in c} T_i}, \tag{1}$$

where c denotes a city and i is a trip instance. Because we sum the length D_i of all trip instances in city c and divide by the sum of trip durations T_i, the ratio S_c^m is a length-weighted measure of travel speed. It is straightforward to define the corresponding unweighted mean.

Means are attractive because of their simplicity and ease of computation. However, in our case means may not be comparable across cities for two reasons. First, although we sample a large number of trips, we may not observe trips in different cities taking place under exactly the same conditions such as time of departure. Second and most importantly, our trip generation strategy implies that trip length and distance to the center differ systematically across cities. As we show below, these characteristics are important determinants of trip speed. We can condition them out by estimating the following type of regression:

$$\log S_i = \alpha X_i' + s_{c(i)}^{fe} + \epsilon_i, \tag{2}$$

where the dependent variable is log trip speed ($S_i = D_i/T_i$), X_i is a vector of characteristics for trip instance i, $s_{c(i)}^{fe}$ is a fixed effect for city c, and ϵ_i is an error term.

If trip characteristics are appropriately centered and the errors are normally distributed, $\hat{S}_c^{fe} = \exp\left(\hat{s}_c^{fe} + \hat{\phi}^2/2\right)$ is a measure of predicted speed for a typical trip in city c where $\hat{\phi}$ is the estimator of the standard deviation of the error term ϵ. Note that for simplicity we can directly use $\hat{s}_c^{fe}$ as an index of mobility.

Equation (2) does not specify the exact content of the vector of characteristics X. In addition to the city within which a trip takes place, we expect the main variables that determine the speed of a motorized trip in our data to be its length, time of departure, distance to the center, and perhaps the type of the trip. We also expect trip speed to be

[17] Data from the 2011 Indian census suggests that 46% of urban commutes, and 55% of urban commutes longer than 1 kilometer, are by motorized road transport.

affected by weather conditions. We will test the robustness of our estimates of the city fixed effects with respect to which variables are included in the regression and how.

Travel conditions may also vary across cities in ways that may not be well captured by equation (2). For instance, we find below that peak hours are relatively slower and last longer in more congested cities. To capture this, we first estimate a more flexible version of equation (2) where we allow both the constant and the vector of coefficients to vary across cities:

$$\log S_i = \alpha_{c(i)} X_i' + s_{c(i)} + \epsilon_i \,. \tag{3}$$

Equation (3) includes many coefficients for each city. Comparing for instance the time of day effect for traffic between 9.30 and 10 p.m. across 154 cities will not be insightful. Rather than keep all these coefficients separate, we aggregate them into index measures of mobility for each city.

More specifically, we proceed as follows. We first estimate equation (3) for each city separately. Each of these 154 regressions can be used to generate a predicted speed for all trips in the data, telling us how fast trip i would be if it were taken in city c: $\hat{S}_{ci} = \exp\left(\hat{\alpha}_c X_i' + \hat{\phi}_c^2/2\right)$. We also predict speeds from an analogous 'national' regression using all trip instances by imposing common coefficients regardless of the city of travel: $\hat{S}_i = \exp\left(\hat{\alpha} X_i' + \hat{\phi}^2/2\right)$.

Then, we compute a predicted duration for each trip i if it were to take place in city c ($\hat{T}_{ci} = D_i/\hat{S}_{ci}$) or 'nationally' ($\hat{T}_i = D_i/\hat{S}_i$). Finally we can compute a relative speed index for each city:

$$L_c = \frac{\sum_i \hat{T}_i}{\sum_i \hat{T}_{ci}} \,. \tag{4}$$

The index L_c represents the time it would take to conduct all trip instances in the data at the estimated speed for city c relative to the predicted time it would take to conduct these trips at the average estimated 'national' speed. L_c is a unitless scalar, but we can multiply it by $\sum_i D_i / \sum_i \hat{T}_i$, the average national speed, to transform it into a predicted speed for city i.

We note that the index L_c defined in equation (4) resembles a Laspeyres price index in the sense that we compare the speed of trips across Indian cities for the same national bundle of trip instances. Like a standard Laspeyres index, L_c may be sensitive to sampling error or to out-of-sample predictions.

Alternatively, we can compute the predicted time it takes to undertake all city c trips in city c relative to the predicted time it needed to undertake all city c trips from a national regression. That is, we can compute:

$$P_c = \frac{\sum_{i \in c} \hat{T}_i}{\sum_{i \in c} \hat{T}_{ci}}. \tag{5}$$

This alternative speed index is analogous to a Paasche price index. Because we compare city trips at predicted city speed to city trips at predicted national speed, this Paasche index will be less sensitive to the problems of out-of-sample predictions that may afflict the Laspeyres index above. It is also straightforward to compute the corresponding Fisher index: $F_c = \sqrt{L_c \times P_c}$.

Finally, we can compute a broad class of mobility indices derived from logit or CES utility specifications. In the logit case of Ben-Akiva and Lerman (1985), the travel decision is a discrete choice over a set of trip destinations. In Appendix B, we derive the following mobility index, which resembles the (inverse of) the familiar CES price index:

$$G_c = \left(\frac{\sum_{i \in c} b_{ci} T_{ci}^{1-\sigma}}{\sum_{i \in c} b_{ci} \overline{T}_i^{1-\sigma}} \right)^{1/(\sigma-1)}, \tag{6}$$

where b_{ci} is a quality parameter for the destination of trip i in city c, and σ is an elasticity of substitution between trip destinations. In this standard utility maximization framework, cheaper (shorter) trips receive more weight, with the strength of that relationship governed by the elasticity of substitution σ. To construct the denominator of G_c, we use a non-parametric procedure to compute, from the national sample, the average duration $\overline{T}_i$ of trips with approximately the same length as trip i in city c. This procedure delivers a pure mobility index that depends only on speed differences across cities.[18]

Instead of tackling the difficult problem of estimating the parameters of G_c, we show that for a wide range of values of σ and b_{ci}, G_c is highly correlated with our benchmark index from equation (3). We also experiment with richer nesting structures, in which trips to similar destination types (e.g., work, shopping, medical/dental, etc) are more substitutable.[19]

It is important to keep in mind that the observations used to estimate equations (2) and (3) and to compute the indices in equations (4), (5), and (6) are counterfactual trips, not

[18] To see this, note that both the city-level numerator and the national-level denominator of G_c have the same number of trips, and the same distribution of trip lengths. The index in each city is therefore free of gains from variety and gains from closer proximity to travel destinations, and determined only by speed differences relative to a national sample.

[19] As another example, consider a utility function with limited scheduling flexibility, as in Kreindler (2018). Such a function would increase the weight of slow peak travel. Our approach is to show that mobility indices based on only peak time trips are highly correlated with those based on all trips.

actual trips. This presents both benefits and costs. The main advantage of our approach is that trips are exogenously chosen. Unlike Couture *et al.* (2018), we do not need to worry about the simultaneous determination of some variables such as trip length and speed, which could affect the estimates of city fixed effects in equations (2) and (3).[20] Conceptually, this approach is similar to measuring price indices from store price tags instead of from consumers' transactions.

This exogeneity is also a potential limitation of our method. The trip instances that we query do not correspond to actual trips and may not be representative of the travel conditions faced by urban travelers when they demand to travel. If our trips are far enough from representative, and if the speed of various types of trips varies across cities, then our mobility indices will be mismeasured.

To this criticism, we have four answers. The first is that some of the trips we created were designed to resemble what we know about actual trips in other cities, with respect to either their direction, the type of destination (and their frequency), or their length. Second, our four trip types (radial, circumferential, gravity, amenity) are designed to reflect reality in distinct ways. We show below that when we introduce a comprehensive sets of controls for other trip characteristics, the economic significance of the trip type indicators in equation (3) is small. Third and most important, our large sample allows us to estimate mobility indices for each trip type, destination, time of day, distance to city center, and various other subsamples. These indices are all highly correlated with our baseline index. As argued earlier, this result implies that our indices do not depend in an important way on the particular utility weight that each counterfactual trip could receive. Finally, Akbar and Duranton (2018) use Google Maps in Bogotá to measure the speed of actual trips reported in a transportation survey and counterfactual trips designed using the same strategy as here. Within short time intervals within days, the speeds of the two types of trips are virtually indistinguishable from each other, and from measures of speed reported by Uber for comparable trips.

3.3 Disentangling two sources of mobility: uncongested mobility and congestion.

Mobility can naturally be decomposed into two components: an uncongested or "free flow" speed, and a congestion factor. To separate the "intrinsic" slowness of a city from its

[20]For instance, as mobility gets better travelers may choose to travel to further destinations. In addition, the (counterfactual) trip instances that we query do not affect real traffic conditions.

congestion, we can adapt the approach proposed above. To measure mobility, we use as dependent variable in equation (2) the log of actual trip speed and estimate city fixed effects $\hat{s}_c^{fe}$ that we can interpret as an index of mobility. To measure mobility *in the absence of traffic*, we repeat the same estimation as with actual speed but use as dependent variable the log of speed in the absence of traffic returned by Google Maps for each query. The resulting city fixed effects $\hat{nt}_c^{fe}$ are our index of uncongested mobility.[21]

To measure congestion, we repeat the same estimation using the difference between log trip duration with traffic and log trip duration without traffic, $\log T_i - \log T_i^{nt} = \log(T_i/T_i^{nt})$, as the dependent variable. While strictly speaking, the city fixed effects, $\hat{f}_c^{fe}$, that we estimate are a measure of delay, we can interpret them as a broad index of congestion, which we refer to as the congestion factor.

The dependent variable when estimating mobility is $\log S_i = \log D_i - \log T_i$. The dependent variable when estimating mobility in the absence of traffic is $\log S_i^{nt} = \log D_i - \log T_i^{nt}$. It then follows that when estimating the congestion factor we have $\log T_i - \log T_i^{nt} = -(\log S_i^{nt} - \log S_i)$. Our third regression thus uses as dependent variable the difference between the dependent variables of the first two regressions. Because we estimate these three regressions for the same trip instances using the same set of covariates, it follows directly from simple econometrics that a city's congestion factor is the difference between its uncongested mobility factor and its overall mobility factor:

$$\hat{f}_c^{fe} = \hat{nt}_c^{fe} - \hat{s}_c^{fe}. \tag{7}$$

This result is useful on two counts. First, it provides us with an exact decomposition which we exploit below. Second, when we regress these three city fixed effects on the same set of city determinants below, the estimated coefficients will also conveniently add up. For instance, the estimated effect of city population on mobility will be equal to the estimated effect of city population on mobility in the absence of traffic minus the estimated effect of city population on the congestion factor.

[21] Alternatively, recall that we observe each trip an average of ten times and oversample times in the middle of the night when we expect very little traffic. We can treat the speed of the fastest trip instance as an estimate of uncongested speed. In practice, these two methods yield city congestion indices with a Spearman correlation coefficient of 0.96.

4. Trip-level results

4.1 Descriptive statistics

We queried 22,777,551 unique trip instances. After eliminating a small fraction of trips for which trip length is not well measured or larger than the haversine distance between origin and destination by more than 50 kilometers, we are left with 22,744,156 observations, 14.8% of which are weekend trips.[22]

Some basic trip statistics are reported in table 1. Average travel speed is 22 kilometers per hour. While the interquartile range is fairly small at only about 8 kilometers per hour, the tails of the distribution are quite long. Similar observations can be made for trip duration and length. The average trip under actual traffic conditions lasts about 13% more time than its counterpart without traffic. Keeping in mind that we oversampled trips taken at night, we return to this issue below. Finally, the average trip is about 50% longer than its "effective" (haversine) length.

Table 2 reports summary statistics for the 154 cities in our sample. They are on average large, with a mean population above 1.3 million, and fast growing, having doubled in population since 1990.[23] Variation across cities in rates of access to personal motorized transportation and road infrastructure stocks are substantial.

Table 3 reports descriptive statistics for various naive measures of mean city travel speed. Mean travel speed across cities is 24.4 kilometers per hour.[24] This is rather slow, especially given that faster night trips are somewhat oversampled. By comparison, Akbar and Duranton (2018) estimate a similar mean speed using a comparable methodology in Bogotá, Colombia, a highly congested city of nearly nine million, and Couture *et al.* (2018) report a mean trip speed by privately-owned vehicles of 38.5 kilometers per hour in US

[22]Google Maps often provides problematic routes for motorized travel on short trips. Furthermore, Google Maps rounds trip lengths, and moves our origin and destination points to the nearest road. In extreme cases, such as when a sampled origin is in the middle of a large park, this can lead to routes that are shorter than the haversine distance between the sampled origin and destination. To limit these problems we consider only trips longer than one kilometer. These problems still sometimes arise beyond one kilometer.

[23]The two sources of population differ both because of the target year and because they are based on slightly different boundaries. In most cases differences are small, but a few cities in Kerala are substantially smaller using our lights-based definition than in the UN database. These cities appear to have a particularly expansive urban agglomeration as defined by the Indian census.

[24]This cross-city mean is slightly larger than the overall population mean of 22.1 kilometers per hour reported in table 1 because travel speed is faster in smaller cities for which we have fewer observations.

Table 1: Trip statistics

	Mean	St. dev.	percentile: 1	10	25	50	75	90	99
Speed	22.1	7.1	11.5	14.7	17.1	20.6	25.4	31.6	45.8
Duration	20.0	17.6	4	7	9	14	23	40	93
Duration (no traffic)	17.2	14	4	6	9	13	20	33	76
Trip length	8.2	10	1.3	1.9	2.9	4.7	8.9	17.9	54.1
Effective length	5.4	7.0	1.0	1.2	1.8	2.9	5.5	11.9	39.6

Note: 22,744,156 observations. Durations are in minutes, lengths in kilometers; and speeds in kilometers/hour.

Table 2: Summary statistics for Indian cities

	Mean	St. dev.	Min.	Max.
Population ('000, Census/lights, 2011)	1,328	3,031	19	23,889
Population ('000; UN, 2015)	1,545	3,179	307	25,703
Population growth 1900-2015 (%)	106	65	31	399
Total area (km^2)	238	414	5.91	3,569
Total roads length (km)	1,393	3,451	10	32,513
Motorways (km)	43.9	64.5	0	437
Primary roads (km)	44.1	77.3	0	481
Share households with car access (%)	9.99	5.78	2.33	31.5
Share households with motorcycle access (%)	41.3	11.7	5.83	73.4
Mean daily earnings ($)	4.91	1.93	2.00	12.28

Notes: Cross-city averages not weighted by population. 153 cities except for vehicle registrations for which one city is missing.

Table 3: Summary statistics for travel speed in Indian cities

	Mean	St. dev.	Min.	Max.
All trips	24.4	3.79	16.2	34.9
Radial trips	22.2	3.79	14.8	32.8
Circumferential trips	20.6	3.23	14.3	29.5
Gravity trips	22.6	3.42	14.7	30.9
Amenity trips	26.9	6.08	16.6	42.0
All trips, unweighted by length	21.8	2.90	15.7	31.4
All trips, in absence of traffic	26.8	4.49	16.3	38.1
All trips, effective speed	16.4	2.77	11.6	24.0

Notes: 154 cities. Speed in kilometers per hour.

metropolitan areas.[25] This said, 24.4 kilometers per hour is much higher than the sometimes apocalyptic descriptions found in the popular press.

We note considerable differences in mean speed across cities. The standard deviation across cities is 3.8 kilometers per hour, more than half the standard deviation of 7.2 across trips in table 1. Mean speed for the slowest city is 16.2 kilometers per hour whereas it is more than twice as high for the fastest city at 34.9. We show below that these wide raw speed differences remain once we adequately control for features of our sampling strategy.

The second to the fifth rows of table 3 report mean speed for each type of trip separately. Circumferential trips are slower whereas amenity trips are faster. As we show below, these differences are mostly caused by differences in length and location.

The sixth row of table 3 reports a measure of mean speed by city, which, unlike the other rows, is not weighted by trip length. Because this increases the influence of shorter trips that are also slower, this unweighted mean of 21.8 kilometers per hour is slightly lower than the length-weighted mean of 24.4 reported in the first row.

The seventh row of table 3 exploits the information provided by Google Maps regarding trip duration in the absence of traffic. As expected, mean speed in the absence of traffic is higher but the difference is small. At 26.8 kilometers per hour, mean speed in the absence of traffic is only about 10% above the mean of actual speed reported in the first row. Interestingly, the variation across cities is not smaller for mean speed in the absence of traffic than for actual mean speed. If anything, it becomes slightly larger. We return to this intriguing finding below.

Finally, the last row of table 3 reports a measure of mean effective speed. Rather than trip length, we use the haversine distance between the origin and destination. Since the ratio between mean trip length and effective trip length is about 1.5 in table 1, we unsurprisingly find a roughly similar ratio between actual and effective trip speed.

4.2 Trip regressions

Before an in-depth analysis of mobility indices and their correlates, we first estimate a number of variants of the generic regression described by equation (2).

[25] If anything, 38.5 kilometers per hour understates true travel speed since it is measured from a travel survey where respondents view trip duration as much more than just the time spent driving in traffic.

Table 4: Determinants of log trip speed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log trip length	0.24^{a}	0.14^{a}	0.14^{a}	0.24^{a}	0.14^{a}	0.14^{a}	0.13^{a}
	(0.0036)	(0.012)	(0.012)	(0.0036)	(0.012)	(0.012)	(0.015)
log trip length^2		0.014^{a}	0.014^{a}		0.014^{a}	0.014^{a}	0.016^{a}
		(0.0034)	(0.0035)		(0.0034)	(0.0034)	(0.0045)
log distance to center		0.15^{a}	0.15^{a}		0.14^{a}	0.14^{a}	0.098
		(0.042)	(0.042)		(0.041)	(0.041)	(0.063)
log distance to center^2		0.025	0.025		0.031	0.031	0.041
		(0.023)	(0.023)		(0.022)	(0.022)	(0.034)
Type: circumferential	-0.015^{a}	-0.0039^{b}	-0.0040^{b}	-0.015^{a}	-0.0037^{b}	-0.0038^{b}	-0.0017
	(0.0020)	(0.0016)	(0.0016)	(0.0020)	(0.0016)	(0.0016)	(0.0019)
Type: gravity	0.077^{a}	-0.0032	-0.0032	0.079^{a}	-0.0027	-0.0027	0.00098
	(0.0065)	(0.0032)	(0.0032)	(0.0066)	(0.0032)	(0.0033)	(0.0043)
Type: amenity	0.082^{a}	0.0064^{c}	0.0063^{c}	0.083^{a}	0.0066^{c}	0.0065^{c}	0.0087
	(0.0058)	(0.0036)	(0.0036)	(0.0057)	(0.0036)	(0.0036)	(0.0054)
City effect	Y	Y	Y	Y	Y	Y	Y
Day effect	Y	Y	Y	weekd.	weekd.	weekd.	Y
Time effect	Y	Y	Y	Y	Y	Y	Y
Weather	N	N	Y	N	N	Y	only
Observations	22,744,156	-	-	19,385,656	-	-	10,319,939
R-squared	0.48	0.53	0.53	0.48	0.53	0.53	0.51
Cities	154	154	154	154	154	154	107

Notes: OLS regressions with city, day, and time of day (for each 30 minute period) indicators. Log speed is the dependent variable in all columns. Robust standard errors in parentheses. a, b, c: significant at 1%, 5%, 10%. All trip instances in columns 1-3. Only weekday trip instances in columns 4-6. Sample sizes for columns 1 and 4 apply to columns 1–3 and 4–6, respectively. Only weekday trip instances for which we have weather information in column 7. Weather in column 3 and 6 consists of indicators for rain (yes, no, missing), thunderstorms (yes, no, missing), wind speed (13 indicator variables), humidity (12 indicator variables), and temperature (8 indicator variables). These variables are introduced as continuous variables in column 7.

A first series of results is reported in table 4. Column 1 regresses log trip speed on city fixed effects controlling for log trip length, an indicator for each type of trip, each day of the week, and each thirty-minute period during the day. Column 2 introduces further controls: the square of the log trip length, log distance to the center (defining a trip's location as the midpoint between its origin and destination), and its square. Column 3 further adds weather variables (and indicators for missing weather data). Columns 4 to 6 repeat the specifications of columns 1 to 3 on a sample of only weekday trips. Column 7 is restricted to observations with non-missing weather data.

Table 4 reports selected coefficients. Longer trips are faster: the elasticity of trip speed with respect to trip length is 0.24 in columns 1 and 4, and larger for longer trips in the other columns where we introduce a quadratic term. This is a prominent feature of urban transportation data in other contexts.[26] Regressing log trip speed on log trip length without any further control yields an R^2 of 0.40.

Unsurprisingly, trips further from the center are also faster. The elasticity of trip speed with respect to distance from the center of 0.15 is a quite large, implying that a trip at 10 kilometers from the center of a city is about 40% faster than one a kilometer away.

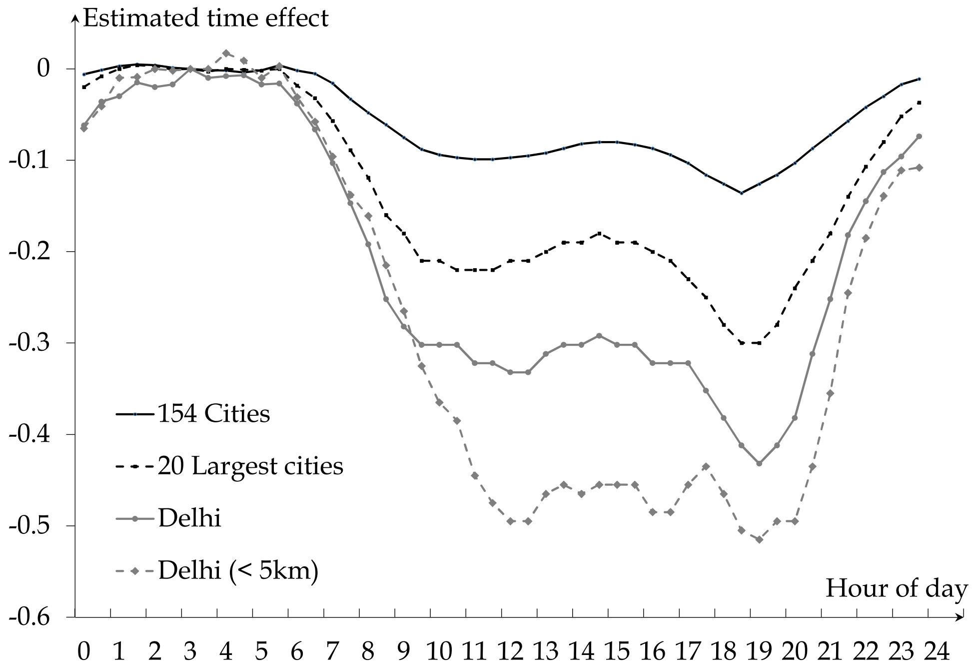
In column 1, we find fairly large differences of up to 10% in speed between different types of trips. These differences become mostly insignificant and economically small when controls for trip location are added in column 2. In the end, amenity trips are slightly faster while circumferential trips are slower but the speed difference between them is only about 1%. We also note that regressing log trip speed solely on trip type indicators yields an R^2 of only about 0.003. These two results are reassuring, and suggest that the design of our hypothetical trips is not driving our results. In Appendix C, we report versions of table 4 for each type of trip. While the non-linearities for the effect of trip length and distance to the center slightly differ, the results overall are similar to those in table 4, suggesting that the simple additive specification of table 4 is not obscuring deeper differences between trip types.

We now turn to the regression coefficients not reported in table 4. Starting with the weather, we find that characteristics associated with bad weather such as rain, high levels of humidity, high temperatures, and more windy conditions tend to be associated with slightly *higher* travel speeds. For instance, in columns 3 and 6, trips in rain are 2–3% faster.

To explain this contrast, we conjecture that roads in many Indian cities are 'multi-purpose' public goods used by various classes of motorized and non-motorized vehicles to travel and park as well as a wide variety of other users such as street-sellers, animals, or children playing. Non-transportation uses of the roadway arguably slow down motorized vehicles. Worse weather may reduce these activities and thus make travel faster. We provide further

[26] Couture *et al.* (2018) estimate a larger elasticity close to 0.40 using self-reported US data where the measure of trip duration also includes a fixed cost of getting into one's vehicle and getting into traffic. Using self-reported data, Akbar and Duranton (2018) find an even larger elasticity for Bogotá travelers, because their sample also includes transit trips, with even larger fixed costs. Using analogous Google Maps data for the same Bogotá trips, Akbar and Duranton (2018) find an elasticity of 0.21, very close to the elasticity estimated here.

Figure 1: Estimated time effects for weekday travel



The plain black line represents the time effects estimated in column 5 of table 4 for all 154 cities. The dashed black line represents the hour effects from the same estimation but restricts observations to the 20 largest cities. The plain gray line duplicates the same exercise for Delhi only. The dotted gray line only uses observations for which the distance to the center of the origin and destination is on average less than 5 kilometers in Delhi. All 3 - 3.30 a.m. effects are normalized to zero. All the plotted coefficients for 7am to midnight are significant at 1%.

indirect evidence for this conjecture below.[27]

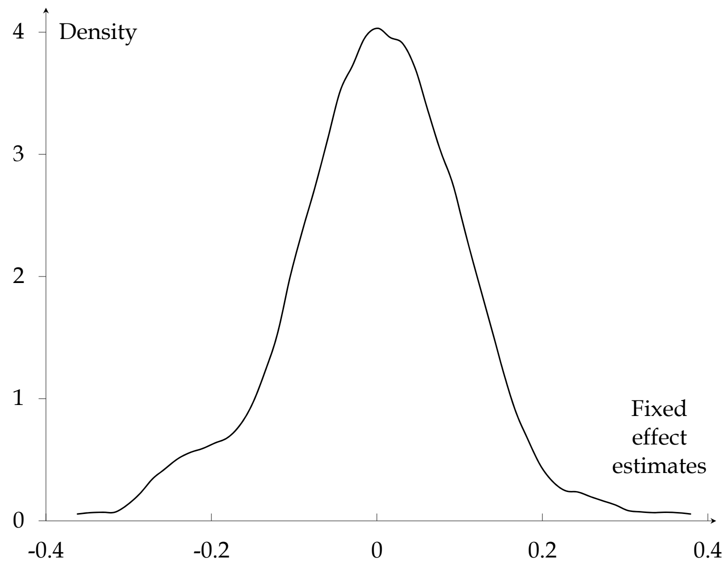
As expected, we also observe fluctuations in travel speed across times of day. In figure 1, the dark continuous line plots the fixed effect of each thirty-minute period estimated in column 5 of table 4. For all cities, the gap between the fastest time in the middle of the night and the slowest at 6.30 p.m. is just 13%. We also note that morning peak hours are more muted than the evening peak hours.[28] The figure also plots the same coefficients estimated only on the twenty largest cities. The patterns are much more marked. The slowest periods in the evening are now more than 25% slower than the fastest in middle of the night. In addition, travel speed starts declining earlier in the morning and recovers later in the evening.

While larger, this difference remains less important than that estimated by Akbar and

[27] However, it is important to note that our data collection period did not include monsoon season. Extreme weather conditions may affect mobility negatively, including for a period of time after they end.

[28] Although we do not report the results here, we can also estimate time of day effects more accurately using trip fixed effects. The resulting estimates for time of day effects are virtually indistinguishable.

Figure 2: Kernel density for estimated city effects



The city effects are as estimated in column 5 of table 4 for all 154 cities. Epanechnikov kernel with bandwidth of 0.031.

Duranton (2018) for Bogotá where the slowest period is about half as fast as the fastest. These mild within-day fluctuations may mask a lot of heterogeneity across Indian cities. To investigate this, we repeat the same exercise using only observations from the city of Delhi. Although Delhi is slow, we purposefully do not take the slowest city or a pathological case. The pattern is the same as for the 20 largest cities but more pronounced. The slowest time is now 35% slower than the fastest. Restricting attention further to trips taking place on average within five kilometers of the center of Delhi generates even more extreme patterns with the slowest time now being more than 40% slower than the fastest.[29]

If we take the difference between the fastest and slowest time as a summary measure of congestion, we can draw several lessons from figure 1. First, in many cities, there may not be that much congestion. Travel speed is slow and does not vary much throughout the day as the demand for travel changes. It is only in the largest cities and more particularly in their centers that travel speed experiences considerable variation during the day. We return to this below. Third, the evolution of travel speeds during the day reflects more than standard

[29]Since India is a vast country with a single time-zone, attenuated within-day fluctuations could be due to the timing of sunrise and sunset. Within our sample, there is range of up to a 98 minutes in sunrise and 126 minutes in sunset. To assess whether cities experience peak hours at different official hours, we produced a variant of figure 1 that defines the time of each trip as a fraction of the time between local sunrise and sunset (or between local sunset and sunrise). It is virtually indistinguishable from figure 1.

commuting patterns. Travel speed declines from roughly 5.30 a.m. to midday, the lowest speed are observed around 6.30 - 7 p.m., and only slowly recover late into the evening. This is consistent with the conjecture raised above that the roadway is used for multiple purposes from late in the morning until well into the evening.

We finally turn to city effects. As argued above, we can interpret them as mobility index values. They measure (log) trip speed in cities after conditioning out log trip length and its square, log trip distance to the center and its square, and day and time of day effects. Figure 2 represents a kernel density estimate of the distribution of city fixed effects from column 5 of table 4. The standard deviation is 0.106. The slowest city is 28% slower than the mean while the fastest city is 42% faster. This gap of a factor of two between the slowest and fastest city is extremely large. Using traveler-reported data and a different methodology, Couture *et al.* (2018) find a less than 30% difference in travel speed among the largest 50 US metropolitan areas. The analogous figure for the top 50 in India is 80%. These large differences are unlikely to be due to sampling bias. All cities have at least 70,000 observations, and the largest cities have more than half a million.

Tables 5 and 6 report the 20 slowest and 10 fastest cities, respectively. First, we note that seven of the 10 largest cities by population in 2015 are among the 20 slowest. The three exceptions are Ahmadabad and Surat in Gujarat and Jaipur in Rajasthan. The state of Gujarat stands out in India for its innovative and more efficient urban planning practices (Annez, Bertaud, Bertaud, Bhatt, Bhatt, Patel, and Phata, 2016). The list of the 20 slowest cities also contains 6 cities from the state of Bihar (among 8 in our data). Bihar is the poorest state in India. Most of the other slow cities are from the neighboring states of Jharkhand and Uttar Pradesh, which are also among the five poorest states in India.

The list of the fastest cities is more heterogeneous. Many are small and in more developed parts of India. Others are exceptional in different ways. The fastest, Ranipet, is an independent city based on our delineation procedure. However, it may be viewed more meaningfully for our purposes as a suburb of the city of Vellore, located about 20 kilometers away. Chandigarh hosts a population above a million, but unlike most Indian cities, it is a planned city characterized by a regular grid pattern laid out by the French architect Le Corbusier.[30] Both Srinagar and Jammu, which are in the disputed state of Jammu and

[30] Figure A.5 in appendix shows Chandigarh's road network, which has the most regular grid of all Indian cities in our sample.

Table 5: Ranking of the 20 slowest cities, slowest at the top

Rank	City	State	Index
1	Kolkata	West Bengal	-0.33
2	Bangalore	Karnataka	-0.25
3	Hyderabad	Andhra Pradesh	-0.25
4	Mumbai	Maharashtra	-0.24
5	Varanasi	Uttar Pradesh	-0.23
6	Patna	Bihar	-0.22
7	Delhi	Delhi	-0.22
8	Bhagalpur	Bihar	-0.22
9	Bihar Sharif	Bihar	-0.19
10	Chennai	Tamil Nadu	-0.17
11	Muzaffarpur	Bihar	-0.16
12	Aligarh	Uttar Pradesh	-0.15
13	Darbhanga	Bihar	-0.14
14	English Bazar	West Bengal	-0.14
15	Gaya	Bihar	-0.13
16	Allahabad	Uttar Pradesh	-0.13
17	Ranchi	Jharkhand	-0.12
18	Dhanbad	Jharkhand	-0.12
19	Akola	Maharashtra	-0.12
20	Pune	Maharashtra	-0.11

Notes: Mobility index is measured by the city effect estimated in column 5 of table 4.

Table 6: Ranking of the 10 fastest cities, fastest at the top

Rank	City	State	Index
1	Ranipet	Tamil Nadu	0.35
2	Srinagar	Jammu and Kashmir	0.26
3	Kayamkulam	Kerala	0.24
4	Jammu	Jammu and Kashmir	0.23
5	Thrissur	Kerala	0.19
6	Palakkad	Kerala	0.16
7	Chandigarh	Chandigarh	0.16
8	Alwar	Rajasthan	0.15
9	Thoothukkudi	Tamil Nadu	0.15
10	Panipat	Haryana	0.15

Notes: Mobility index is measured by the city effect estimated in column 5 of table 4.

Table 7: Determinants of log trip speed, variants

	(1) effective length	(2) typical traffic	(3) no traffic	(4) off peak	(5) peak	(6) high peak	(7) peak radial
log trip length	-0.18[a]	0.13[a]	0.16[a]	0.14[a]	0.13[a]	0.13[a]	0.040
	(0.012)	(0.012)	(0.012)	(0.011)	(0.013)	(0.012)	(0.030)
log trip length^2	0.085[a]	0.017[a]	0.019[a]	0.019[a]	0.013[a]	0.0098[a]	0.065[a]
	(0.0031)	(0.0039)	(0.0032)	(0.0031)	(0.0039)	(0.0034)	(0.010)
log distance to center	0.57[a]	0.16[a]	0.22[a]	0.23[a]	0.12[a]	0.087[b]	0.15[a]
	(0.036)	(0.046)	(0.046)	(0.048)	(0.042)	(0.036)	(0.051)
log distance to center^2	-0.13[a]	0.014	-0.037	-0.047[c]	0.054[b]	0.083[a]	-0.12[a]
	(0.015)	(0.026)	(0.025)	(0.027)	(0.023)	(0.019)	(0.044)
City effect	Y	Y	Y	Y	Y	Y	Y
Day effect	weekd.	weekd.	weekd.	weekd.	weekd.	weekd.	weekd.
Time effect	Y	Y	Y	Y	Y	Y	Y
Weather	N	N	N	N	N	N	N
Observations	19,385,656	19,385,656	19,385,656	4,910,731	10,469,622	2,375,960	826,539
R-squared	0.34	0.56	0.54	0.54	0.54	0.53	0.54
Cities	154	154	154	154	154	154	154

Notes: OLS regressions with city, day, and time of day (for each 30 minute period) indicators. Log effective speed is the dependent variable in column 1. Log speed under "typical" traffic conditions is the dependent variable in column 2. Log speed under 'no traffic' is the dependent variable in column 3. Log speed is the dependent variable in all subsequent columns. All columns only consider weekday observations. Column 4 considers observation from only off-peak hours (before 7.30 and after 22.30). Column 5 considers observation from only peak hours (from 8.30 a.m. to 5.30 p.m. and from 8 p.m. to 10 p.m.). Column 6 considers observations from only high peak hours (from 5.30 p.m. to 8 p.m.). Finally, column 7 considers only radial observation from peak and high peak hours (going towards the city center in the morning and back in the evening). Robust standard errors in parentheses. a, b, c: significant at 1%, 5%, 10%.

Kashmir, receive specific infrastructure funding from the federal government and have a strong police presence. These two features may lead to better mobility.

Table 7 reports a number of variants of our benchmark specification in table 4 column 5. Column 1 uses log effective speed (haversine length divided by time) instead of actual speed as dependent variable. The increase in effective speed with trip length and with trip distance to the center is even more pronounced than the increase in actual speed. This is consistent with shorter and more central trips being more tortuous. Column 2 uses speed under "typical" traffic conditions as dependent variable; results are very similar to those for the corresponding specification using actual speed in column 5 of table 4. Column 3 uses the same specification to predict speed with no traffic. Interestingly, trips taking place further from the center remain faster. While figure 1 above suggests that central parts of Delhi are more congestible, the bulk of the difference in speed between more central and more peripheral trips remains in the absence of traffic. This is plausibly caused by the expected greater density of intersections and narrower streets in more central parts of cities in India (and many other countries).

The second part of table 7 reports our preferred specification of table 4 for different times of day: off peak in column 4, peak in column 5, high peak in column 6, and radial trips at peak hours going towards the center in the morning and back towards the periphery in the afternoon in column 7. This last specification is meant to mimic archetypal commuting patterns. While again the curvature of the effect of trip length and distance to the center varies slightly, the results are generally very similar to those we obtained before.

4.3 Comparing mobility indices

We now turn to comparing mobility indices. Because many different variants of equations (2) and (3) are available and many different samples of trips can be selected, many mobility indices are possible. To explore these possibilities, we compute a wide variety of such indices. To avoid hard-to-digest matrices of pairwise correlations, we form our benchmark mobility index from the city fixed effects estimated from the specification reported in column 5 of table 4, and compare all our other indices to this one. We also report the standard deviation, maximum and minimum of each variant. Standard deviations vary very little, except for the mean speed indices, which are constructed on a different (linear) scale.

The results are reported in table 8. Panel A compares our benchmark mobility index to the analogous indices estimated in the other columns of table 4 that includes various trip level controls. All these correlations are above 0.98 when we include the square of trip length and distance to center and fall to about 0.92 when we do not.

Panel B compares our benchmark index to the analogous indices estimated using the same specification but considering different types of trips separately. The correlations are again high. The lowest at 0.90 is with perhaps our most artificial type of trips, circumferential trips, and the highest is with perhaps our most realistic, amenity trips. Even indices based on our 17 individual amenity classes, which represent less than 3% of a city's trips in nearly all cases, are highly correlated. Fifteen of them are correlated with the baseline index at

Table 8: Pairwise Spearman rank correlations with our benchmark mobility index

Index	Corr.	Std. Dev.	Min	Max
Panel A: Columns from table 4				
(1)	0.916	0.100	-0.232	0.332
(2)	>0.999	0.105	-0.321	0.347
(3)	0.992	0.108	-0.337	0.355
(4)	0.918	0.101	-0.240	0.332
(6)	0.991	0.109	-0.347	0.356
(7)	0.983	0.115	-0.329	0.374
Panel B: Trip subsamples				
Radial	0.926	0.117	-0.318	0.373
Circumferential	0.900	0.113	-0.286	0.330
Gravity	0.966	0.112	-0.375	0.308
Amenities	0.966	0.107	-0.345	0.373
Interact time/day with trip type	>0.999	0.105	-0.322	0.347
Panel C: Mean speeds				
Simple mean	0.476	3.790	16.212	34.903
Mean unweighted by length	0.619	2.899	15.7	31.4
Mean of "typical" traffic speed	0.452	3.814	16.2	35.1
Mean of uncongested speed	0.340	4.494	16.3	38.1
Mean effective speed	0.410	2.768	11.6	24.0
Panel D: Table 7 variants				
Effective speed	0.864	0.118	-0.430	0.392
"Typical" traffic	0.997	0.102	-0.301	0.345
No traffic	0.850	0.100	-0.242	0.339
Fastest trip instance	0.851	0.101	-0.261	0.298
Off peak	0.881	0.099	-0.255	0.316
Peak	0.991	0.113	-0.388	0.361
High peak	0.948	0.130	-0.430	0.367
Peak radial	0.915	0.133	-0.450	0.405
Panel E: Full indices				
Laspeyres	0.794	0.151	0.105	1.478
Paasche	0.941	0.107	0.767	1.478
Fisher	0.910	0.126	0.322	1.478
Logit/CES ($\sigma = 0$)	0.923	0.098	0.675	1.255
Logit/CES ($\sigma = 2$)	0.836	0.099	0.694	1.221
Logit/CES ($\sigma = 4$)	0.687	0.108	0.648	1.182
Panel F: Distance to center				
Trips within 5 km of center	0.970	0.108	-0.278	0.350
Trips within 3 km of center	0.918	0.111	-0.268	0.356
Trips within 2 km of center	0.827	0.116	-0.261	0.336
Weight by inverse dist. to center	0.959	0.106	-0.293	0.341
Panel G: Weight by powered congestion factor				
$\lambda = 0.2$	0.910	0.137	-0.533	0.378
$\lambda = 0.3$	0.927	0.123	-0.396	0.373

Notes: 154 cities in all rows except in the last row of panel A which uses 107. The first column reports the Spearman rank correlation between the index at hand and our preferred index from column 5 of table 4. The second column reports the standard deviation. The third and fourth column report the maximum and minimum respectively.

0.87 or higher. Finally, allowing time of day and weekend indicators to vary by trip type (radial inward, radial outward, circumferential, gravity, and 17 amenity types), so that, for example, trips to a temple on the weekend might be different than those on a weekday, also makes essentially no difference in rankings.

Next, panel C compares our benchmark index to various measures of mean speed computed above. The correlations are much lower than in the previous two panels. For instance, the correlation between our benchmark mobility index and mean speed computed as total travel length divided by total travel time is only 0.48. As noted in Couture *et al.* (2018) for US metropolitan areas, means of speed do not provide good descriptions of mobility in cities. This is because trip length, which varies systematically across locations, has a large explanatory power on trip speed. As a result, mean speeds are sensitive to sampling strategies, unlike our preferred mobility indices that control for trip length.

Panel D reports correlations between our benchmark mobility index and mobility indices computed from the estimations reported in table 7. The correlation of our benchmark mobility index with an index that measures speed using effective (haversine) rather than traveled trip length is 0.87. The 20 slowest cities reported in table 5 using our benchmark mobility index are all among the 30 slowest cities by effective speed. We can thus rule out the possibility that slow cities are more efficient at transporting travelers farther for the same number of straight line kilometers traveled. Slow cities are just slow.

Still in panel D, the correlation of our benchmark index with an uncongested mobility index, computed using travel times in the absence of traffic, is also relatively high at 0.85. This strongly suggests again that poor mobility is largely the outcome of generally slow travel. While congestion plays a role, it may not be the main driver of poor mobility in Indian cities. We return to this issue below. Interestingly, when ranking cities by uncongested mobility, we find that the five slowest cities in the absence of traffic are all in Bihar and 17 of the 20 slowest cities are in the poor northeastern part of India. Except for Kolkata which also ranks among the cities that are slow in the absence of traffic, most major Indian cities are in the middle of the distribution of uncongested mobility indices. For these cities, congestion is arguably an important determinant of why they are slow. Eight of the 10 fastest cities reported in table 6 are also among the 10 fastest cities in the absence of traffic.

The second part of panel D reports correlations between our benchmark index and mobility indices computed in the same manner as our benchmark but from observations taken at specific hours of the day. The correlation of our benchmark index with an index of peak-hour speed is extremely high. It is still high with an index computed only during the most extreme hours of the early evening, between 5.30 and 8 p.m., when traffic is generally at its slowest. The correlation is still 0.92 with an index computed using only the 5% of sample composed of radial trips at peak hours that go towards the center in the morning and away from the center in the evening.

Panel E reports correlations between our benchmark index and more sophisticated Laspeyres, Paasche, Fisher, and logit/CES indices computed as described by equations (4), (5), and (6). Row 1 uses a Laspeyres index computed from the same specification as for our benchmark index which allows all 58 regression coefficients to vary across cities. The correlation is still fair at 0.79. It jumps to 0.89 when we focus only on the 50 largest cities. The lower full-sample correlation is due to flawed out-of-sample predictions in small cities for long trips far from the center. Row 4 to 6 reports correlations with the logit/CES index for different values of the elasticity of substitution σ. The correlation for $\sigma = 0$, the perfect complement case for which all trips receive equal weight, is very high at 0.92, and only declines slightly to 0.84 for $\sigma = 2$. The correlation with our benchmark index remains relatively high at 0.69 even for an extreme value of $\sigma = 4$, which gives a two-kilometer trip about 400 times the weight of a longer 15-kilometer trip.[31] In Appendix B, we describe simulations showing that correlations remain invariably high across a wide range of random quality draws b_{ci}. In the same appendix, we describe mobility indices from models of travel demand with richer substitution patterns. These nested indices put less weight on destination types (e.g., shopping trips) that are relatively slower in a given city, because they allow travelers in each city to substitute away from costlier travel destination types. We find that such nested indices are highly correlated with our benchmark index. This finding further confirms that our benchmark index provides a robust characterization of travel cost differences across cities, because slow cities tend to be slow at all times, for all types of trip destinations, and across the city.

Panel F considers indices based on trips progressively closer to the center of the city. Correlations fall as expected, but even limiting to trips centered within 2 kilometers of the center, the correlation is still 0.83. Weighting trips close to the center more heavily, while including more peripheral trips, yields an index much more similar to the benchmark.

Finally, in panel G we try to weight each trip by how likely it is to be taken. Although this

[31] Atkin, Faber, and Gonzalez-Navarro (2018) estimate an elasticity of substitution across retail stores slightly smaller than 4 for poor Mexican households. This is almost certainly an upper bound: the index considered here covers a much broader set of destinations that are unlikely to be as substitutable as retail stores.

information is not directly available to us, we can use the implicit density of vehicles along the route as a proxy. To do so, we assume that (i) the speed of a trip instance is reduced from the maximum for that trip solely by congestion, (ii) the elasticity of trip speed with respect to the density of vehicles, λ, is constant, and (iii) the density of vehicles is constant along the route. Under these assumptions, we can weight each trip i by its length, D_i, times the implicit density of vehicles, $(T_i/T_i^{nt})^{1/\lambda}$. While these assumptions are unlikely to be strictly true, they manage to capture the fact that more vehicles slow down traffic and thus slower trip instances should receive a higher weight given that they represent more travelers. The question is of course which value to use for λ. We use $\lambda = 0.2$ and $\lambda = 0.3$. The value $\lambda = 0.2$ is a standard value in the traffic modelling literature (Small and Verhoef, 2007). The higher value $\lambda = 0.3$ reduces the weight put on slow trips since slower speeds in India may not be caused only by more traffic. With both values, the indices are highly correlated with our benchmark index.

We draw two important conclusions from this analysis. First, because trip length is such an important determinant of trip speed, and because trip length varies across cities of different sizes, appropriately estimating a city mobility index requires accounting for trip-length differences. Second, we find that once trip length is conditioned out, the mobility indices that we estimate for each city are not sensitive to the exact sample being used, and therefore to the weight that different kinds of trips receive. Although we use a variety of trips that reflect important differences in traveller behavior, these differences do not appear to matter when estimating city mobility.

5. Decomposition: uncongested mobility and congestion

We first decompose our indices of mobility into mobility in the absence of traffic (uncongested mobility) and the congestion factor following equation (7). This relationship allows us to perform an exact variance decomposition. The variance of the mobility index is equal to the sum of three terms: the variance of the index of uncongested mobility, the variance of the congestion factor, and minus twice the covariance between the index of uncongested mobility and the congestion factor.

As shown in the first row of Table 9 Panel A, the variance of the uncongested mobility index accounts for 88% of the variance of our benchmark mobility index while that of the congestion factor accounts for only 32%. This is a striking finding. Differences in mobility

Table 9: Variance decompositions of our baseline mobility index

Sample	Cities	All trips			Peak trips		
		Uncongested mobility	Congestion factor	Covariance	Uncongested mobility	Congestion factor	Covariance
Panel A: Full trip sample							
All	154	0.884	0.318	0.101	0.769	0.451	0.110
Largest 50%	77	0.646	0.346	-0.004	0.534	0.479	0.006
Smallest 50%	77	1.305	0.126	0.215	1.346	0.170	0.258
Largest 25%	38	0.526	0.287	-0.093	0.427	0.393	-0.090
Largest 10%	15	0.357	0.376	-0.134	0.270	0.474	-0.128
Panel B: Distance to city center less than 5 km							
All	154	0.963	0.366	0.164	0.807	0.552	0.179
Largest 50%	77	0.746	0.424	0.085	0.579	0.618	0.099
Smallest 50%	77	1.293	0.123	0.208	1.335	0.170	0.253
Largest 25%	38	0.580	0.434	0.007	0.422	0.604	0.013
Largest 10%	15	0.487	0.748	0.117	0.300	0.899	0.100
Panel C: Distance to city center less than 3 km							
All	154	1.042	0.384	0.213	0.887	0.593	0.240
Largest 50%	77	0.829	0.421	0.125	0.657	0.634	0.145
Smallest 50%	77	1.300	0.129	0.215	1.342	0.178	0.260
Largest 25%	38	0.607	0.484	0.045	0.434	0.672	0.053
Largest 10%	15	0.639	0.880	0.259	0.388	1.060	0.224
Panel D: By trip type							
Radial	154	0.960	0.369	0.164	0.821	0.534	0.178
Circumferential	154	1.034	0.397	0.216	0.898	0.577	0.238
Gravity	154	0.789	0.223	0.006	0.700	0.312	0.006
Amenities	154	0.841	0.302	0.071	0.733	0.418	0.075

between Indian cities are mostly driven by differences in their uncongested mobility, not by differences in how congested they are. As we show in the rest of this section, this finding is explained by both pervasive differences in uncongested mobility between cities and the fact that congestion remains modest in most cities. However, the finding is different when we focus on the largest cities. These cities face fairly similar uncongested mobility but are congested to different degrees.

This said, a possible caveat here is that our data collection oversamples trips at night and this may bias our mobility index towards uncongested mobility. Performing the same exercise with indices computed only from trips taken at peak hours, we find that the uncongested mobility index still represents 77% of the variance of the mobility index during peak hours whereas the congestion factor only represents only 45%.

We repeat the same exercise focusing on cities with population above the median. For these cities, the role of uncongested mobility falls, but remains larger than the congestion factor, and the covariance terms essentially goes to zero. For cities below the median population, the explanatory power of the congestion factor is very low. For cities in the top population quartile, the covariance term becomes negative, but the uncongested mobility still represents a larger share of the variance. Only in the top decile do the two factors have approximately even shares.

In the next two panels of Table 9, the role of congestion expands as we limit attention to city centers, especially at peak hours and in larger cities. Variance in uncongested mobility still however represents a substantial share of overall variance across cities in all samples. In the final panel, we repeat the same decomposition for each type of trip separately and find roughly similar results for the respective roles of uncongested mobility and congestion.

6. Correlation of mobility with city characteristics and urban development

We now explain mobility using city characteristics. We first consider basic characteristics like population and area. We then consider indicators of urban economic development, such as income levels, car ownership rates, and urban population growth. In addition, we consider road network measures that reflect urban development, such as the availability of primary roads and conformity to a regular grid pattern.

We report results for our benchmark mobility index in table 10. Table 11 panels A and B report the same specifications predicting the benchmark uncongested mobility and congestion indices, respectively. Because the mobility index is equal to the uncongested mobility index minus the congestion factor and we estimate the same specifications for all three dependent variables in each column, a given coefficient in table 10 is equal to the analogous coefficient in table 11 panel A minus the analogous coefficient in table 11 panel B.

In column 1 of table 10, we consider a simple specification with only log city population and log city area as explanatory variables. Because our dependent variable is a measure of log speed, we can interpret the coefficients as elasticities. For city population, we estimate an elasticity of -0.18. For city area, the elasticity is of opposite sign and equal to 0.15. These two variables explain more than half of the variation in mobility across Indian cities.

Further controls added in subsequent columns change these results little. The robustness of these results is further confirmed in appendix tables C.2 and C.3 where we use alternative measures of mobility as dependent variables.

These results suggest a large "gross density" effect since an increase in population keeping land area constant is, in effect, an increase in population density. This large increase in the cost of travel per unit distance can be contrasted with the usually much smaller estimates of analogous density elasticities for measures of urban productivity such as wages (Combes and Gobillon, 2015). By contrast, this increase in the cost of travel when population density increases is comparable but somewhat smaller than the elasticity of urban costs with respect to density estimated by Combes, Duranton, and Gobillon (2016) for French cities. This elasticity of urban costs, which is estimated indirectly using housing price at the centre of cities, may reflect more than just slower mobility when density increases.

On the other hand, the mostly offsetting nature of the coefficients on population and urban land area suggest that "net scale" effects are small, once we allow for land area to adjust to a larger population. Consistent with this, we estimate an elasticity of about -0.05 when regressing our preferred mobility index on log city population alone.[32]

In panels A and B of table 11, we estimate the same specifications as in table 10 using our preferred index of uncongested mobility and congestion factor as dependent variables. Consistent with our earlier decompositions of overall variance, we find that most of the effect of city population and city area on mobility works through uncongested mobility. For the congestion factor, we find an elasticity of city population of 0.02 in column 1. This coefficient remains between 0.02 and 0.03 in subsequent specifications. For the effect of city area on the congestion factor, we estimate small and insignificant elasticities in most specifications. Putting these results together, it appears that gross density mostly affects uncongested mobility while the negative net scale effects are mostly about congestion.

Column 2 of tables 10 and 11 adds the log of primary roads length. Here and in subsequent specifications, we estimate a small but robust elasticity of mobility with respect to primary road kilometers of about 0.01. We experimented with other measures of the

[32]We estimate a similar elasticity for US metropolitan areas using the preferred speed index computed by Couture *et al.* (2018). We nonetheless fail to replicate large gross density effects for US metropolitan areas when we also include log land area in the regression. This is perhaps because area is poorly measured by official definitions of metropolitan areas in the US. Couture *et al.* (2018) report a population elasticity of -0.12 when also conditioning out the roadway, perhaps because it more accurately reflects land area.

Table 10: Correlates of city mobility indices, benchmark mobility index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log population	-0.18^a	-0.18^a	-0.17^a	-0.17^a	-0.17^a	-0.17^a	-0.17^a	-0.16^a
	(0.016)	(0.015)	(0.016)	(0.017)	(0.016)	(0.018)	(0.016)	(0.017)
log area	0.15^a	0.14^a	0.13^a	0.12^a	0.12^a	0.12^a	0.12^a	0.11^a
	(0.016)	(0.017)	(0.017)	(0.018)	(0.017)	(0.019)	(0.017)	(0.019)
log roads		0.013^a	0.012^a	0.013^a	0.011^b	0.014^a	0.013^a	0.014^a
		(0.0043)	(0.0043)	(0.0040)	(0.0044)	(0.0045)	(0.0042)	(0.0042)
log income			0.22^b	0.23^b	0.20^c	0.23^b	0.22^b	
			(0.10)	(0.10)	(0.11)	(0.11)	(0.10)	
log^2 income			-0.064^b	-0.066^b	-0.055^c	-0.064^c	-0.065^b	
			(0.031)	(0.031)	(0.033)	(0.034)	(0.032)	
Network / shape				0.26^a	0.11^b	0.055^c		
				(0.096)	(0.049)	(0.029)		
Pop. growth 90-10							0.052^c	
							(0.030)	
share w. car								0.21
								(0.14)
share w. motorcycle								0.11^b
								(0.054)
Observations	153	153	153	153	153	142	153	152
R-squared	0.54	0.56	0.57	0.59	0.58	0.60	0.58	0.59

Notes: OLS regressions with a constant in all columns. The dependent variable is the city fixed effect estimated in the specification reported in column 5 of table 4. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Log population is constructed from town populations from the 2011 census. Log roads is log kilometers of primary roads within the city-light. Income is measured with male earnings from the 2011 census. The network / shape variable used in column 4 measures the share of edges in the road network that conform to the grid's main orientation, i.e., whose compass bearing are within 2 degrees of the modulo 90 modal bearing in the network. The network / shape variable in column 5 is a Gini index for the distribution of edge compass bearings in the road network. It also measures how grid-like the city is. The network / shape variable used in column 6 uses Harari's (2016) measure of the average distance between the centroid of the city and all the points that define its periphery. It measures the compactness of the city. The measure of population growth between 1990 and 2010 was constructed UN data. The share of households with access to a car or to a motorcycle is from the 2011 census.

roadway but failed to uncover other robust associations.[33] Interestingly, we find that the effect of primary roads on mobility mostly occurs through uncongested mobility while the effect of primary roads on the congestion factor is a precisely estimated zero. We think

[33]Surprisingly, more motorways - which are high capacity dual carriage roads equivalent to freeways in the United States - do not lead to a robust improvement in mobility. We note that many Indian cities do not have any motorways in our sample. Couture *et al.* (2018) estimate a much larger roads coefficient for US metropolitan areas but do not condition out land area.

Table 11: Correlates of city mobility indices, uncongested mobility and congestion factor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Uncongested mobility								
log population	-0.16[a]	-0.15[a]	-0.15[a]	-0.14[a]	-0.14[a]	-0.15[a]	-0.14[a]	-0.13[a]
	(0.017)	(0.016)	(0.015)	(0.016)	(0.015)	(0.017)	(0.014)	(0.017)
log area	0.17[a]	0.15[a]	0.13[a]	0.13[a]	0.13[a]	0.14[a]	0.13[a]	0.11[a]
	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)	(0.019)	(0.017)	(0.020)
log roads		0.013[a]	0.013[a]	0.014[a]	0.012[b]	0.015[a]	0.015[a]	0.017[a]
		(0.0050)	(0.0049)	(0.0045)	(0.0050)	(0.0053)	(0.0044)	(0.0045)
log income			0.15[c]	0.15[c]	0.13	0.14	0.14	
			(0.084)	(0.083)	(0.089)	(0.093)	(0.086)	
$\log^2$ income			-0.026	-0.029	-0.022	-0.023	-0.029	
			(0.026)	(0.025)	(0.027)	(0.028)	(0.026)	
Network / shape				0.21[c]	0.047	0.020		
				(0.11)	(0.063)	(0.032)		
Pop. growth 90-10							0.11[a]	
							(0.030)	
share w. car								0.55[a]
								(0.14)
share w. motorcycle								0.016
								(0.051)
R-squared	0.43	0.45	0.48	0.50	0.49	0.50	0.53	0.53
Panel B: Congestion factor								
log population	0.024[b]	0.024[b]	0.026[a]	0.025[a]	0.024[b]	0.023[b]	0.029[a]	0.030[a]
	(0.0096)	(0.0096)	(0.0095)	(0.0097)	(0.0099)	(0.010)	(0.0095)	(0.0098)
log area	0.018[b]	0.017[c]	0.0071	0.0084	0.0096	0.011	0.0044	0.0068
	(0.0086)	(0.0095)	(0.010)	(0.011)	(0.011)	(0.011)	(0.010)	(0.011)
log roads		0.00068	0.00094	0.00068	0.0017	0.00081	0.0017	0.0028
		(0.0028)	(0.0029)	(0.0029)	(0.0029)	(0.0031)	(0.0027)	(0.0028)
log income			-0.079	-0.080	-0.064	-0.091	-0.081	
			(0.060)	(0.060)	(0.061)	(0.065)	(0.057)	
$\log^2$ income			0.037[b]	0.038[b]	0.032[c]	0.041[b]	0.036[b]	
			(0.019)	(0.019)	(0.019)	(0.020)	(0.018)	
Network / shape				-0.046	-0.061[c]	-0.035[b]		
				(0.045)	(0.035)	(0.016)		
Pop. growth 90-10							0.054[a]	
							(0.018)	
share w. car								0.34[a]
								(0.091)
share w. motorcycle								-0.096[b]
								(0.042)
R-squared	0.54	0.54	0.60	0.60	0.61	0.61	0.63	0.63
Observations	153	153	153	153	153	142	153	152

Notes: OLS regressions with a constant in all columns. The dependent variable is the city fixed effect estimated for uncongested mobility in panel A, and the congestion factor in panel B. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. See the footnote of table 10 for further details about the explanatory variables.

these findings reflect two facts. First, primary roads are intrinsically faster than secondary or tertiary roads. Second, the absence of an effect on the congestion factor is consistent with the fundamental law of congestion: more primary roads attract new traffic and eventually leave congestion unchanged (Duranton and Turner, 2011). We return to this issue below.

Column 3 of table 10 further includes log city income and its square.[34] We find evidence of a hill shape where mobility first increases with income and then declines. The turning point corresponds to a city slightly below the top quartile of income. This finding is consistent with our rankings of the fastest and slowest cities in tables 5 and 6. Many of the fastest are middle-income cities, while the slowest are either among the poorest or richest cities in the country. When we examine the separate effects of income on uncongested mobility and the congestion factor in table 11 we find that the overall shape of the income-mobility relationship reflects two opposing forces. Uncongested mobility improves with income, perhaps because of better roads. The congestion factor also increases with income, perhaps as residents have more vehicles and travel more. This second force appears to kick in at higher levels of income as evidenced by the fact that it is captured by the squared log income term in the regression. This is also consistent with our earlier findings that congestion is important in only a small number of cities.

In columns 4 and 5 of table 10, we consider two different measures of how well the road network of a city conforms to a regular grid.[35] Both measures suggest a positive association between a more grid-like pattern and better mobility in cities. The magnitude of the coefficients reported in the table for these measures is hard to interpret directly. A normalization indicates that a standard deviation in our grid variable is associated with 0.16 (in column 4) or 0.11 (in column 5) standard deviation in log mobility. This finding provides preliminary evidence in support of calls for more regular grid patterns for the roadway of emerging cities (Angel, 2008, Fuller and Romer, 2014).

We also experimented with the measures of urban form constructed by Harari (2016) and found a robust association between mobility and her measure of urban sprawl. The results are reported in column 6. That more sprawl is positively correlated with mobility is

[34]Our income measure is log daily earnings for men. Since it is measured at the district level, it is subject to substantial measurement error. We exclude women due to lower labor force participation.

[35]The first measure captures the share of edges in the network that conform to the grid's main orientation i.e., whose compass bearing are within 2 degrees of the modulo 90 modal bearing in the network. The second measure is a Gini index for the distribution of edge compass bearings. Appendix A provides details. We also experimented with measures of the density of intersections and the length and circuitry of road segments but failed to uncover any robust association with our measures of mobility.

consistent with earlier results by Glaeser and Kahn (2004) for the US.

In column 7 of tables 10 and 11, we introduce a measure of past population growth. Cities that experienced faster population growth between 1990 and 2010 enjoy both faster uncongested mobility and more congestion. Overall the positive effect happening through uncongested mobility appears to dominate. While we leave a deeper investigation of these results for future research, we emphasize that they are inconsistent with typical claims that rapid urban population growth in developing countries is necessarily associated with worse mobility. Congestion may worsen with population growth but this negative effect is more than offset by faster roads.

Finally, in column 8 we no longer consider income but instead introduce two measures for the share of population with access to car (or equivalent) and (separately) a motorcycle. The insignificant positive coefficient for cars in explaining mobility in table 10 results from two offsetting effect where more cars are strongly and positively associated with both uncongested mobility and congestion in table 11. Motorcycles are associated with faster travel via less congestion, consistent with them taking up less room than cars, but inconsistent with them being a response to congestion. Again, causal identification is beyond our scope here but we would like to highlight that standard indicators of urban economic development such as higher incomes, faster population growth, and more cars are generally associated with better mobility outcomes despite higher congestion.

Although our findings above are generally stable across a wide variety of specifications, they may be subject to bias due to omitted city-level variables. In results reported in Appendix D, we control for city fixed effects, using within-city variation in population, area, and roads, at the level of concentric rings (0 to 2 kilometers from the center, 2 to 5, 5 to 10, 10 to 15, and 15 and beyond) to gain further insights about variation in mobility. Within cities, rings with more population and less urban area are slower, just as in the across-city results above.

7. Transit and walking

While roughly half the households in the average city in our data have access to a private vehicle – sometimes a car but more often a motorcycle – we recognize that city dwellers in India also often walk and use transit. To investigate these two alternative modes of travel, we also collected travel time data for walking and transit for all our trip instances.

For walking trips, speeds typically do not vary much across our trips and remain constant within trip. Mean walking speed is 4.8 kilometers/hour with a standard deviation of 0.1 kilometers/hour. We first estimate a city effect for walking trips in the same spirit as our baseline mobility index above. The standard deviation for the city effects is unsurprisingly tiny at 0.006. When we try to explain city effects for walking trip using the same approach as in table 10, the only robust correlate of our walking mobility index is a measure of average slope in the city. As Google Maps' algorithm reflects, steeper slopes slow down walking.

As described in Appendix A, we also collected transit data. These data have two important limitations. Google Maps only appears to return transit information for formal transit, and it bases its information on official timetables. This ignores informal transit and delays or missed services in formal transit. With these caveats in mind, we first note that only about 20% of our trip instances have a transit alternative that we define as 'viable': it requires less than an hour wait, and is strictly faster than walking. Despite this selection, viable transit trips take on average 2.3 times as long as trips with private vehicles. In regressions not reported here, we additionally find that unsurprisingly, the transit time penalty is higher for shorter trips, trips further from the centre, and nighttime trips.

Next, for 141 cities we can estimate an index analogous to our baseline mobility index for transit. Unlike with walking, there is a lot of cross-city variation for transit. The standard deviation for our transit mobility index is about twice that of our baseline mobility index for private vehicles. This variation does not seem to be due to sampling problems as these indices are precisely estimated and alternative transit indices are all highly correlated.

The correlation between our mobility index for transit and our baseline mobility index (for private vehicles) is extremely low at 0.02. This correlation even becomes negative when we focus on the largest cities. However, when we re-estimate our mobility index for private vehicles on a sample limited to trip instances for which a viable transit alternative is possible, this correlation increases from 0.02 to 0.17. This difference suggests a fair amount of selection regarding which trip instances have a viable transit alternative known to Google. To confirm the low correlation between transit and vehicle travel times we regress log transit travel time on log private vehicle travel time and log walking time. In this regression, the coefficient on log vehicle travel time is only 0.19 while the coefficient on log walking time, which is essentially a measure of trip length, is 0.52.

Finally, we also replicated the regressions of table 10 for our transit mobility index. We did not find any robust correlates of transit mobility at the city level. Given the sizable variation across cities in transit mobility, this may seem surprising. Nonetheless, this result is consistent with the weak correlation between (private vehicle) mobility and transit mobility. Although we must remain cautious given the caveats that apply to our transit data, taken together these results suggest to us to that transit mobility depends much more on the coverage and frequency of transit than on driving speeds.

8. Conclusions

We propose a novel approach to measuring vehicular mobility within cities, and decomposing it into uncongested mobility and a congestion factor. We apply it using novel large scale data on counterfactual trips in 154 Indian cities collected from Google Maps. After showing that various sampling and estimation strategies yield similar estimates of mobility, we document a number of important facts about mobility in Indian cities. Among the most important, we first highlight large mobility differences across cities. Second, slow mobility is primarily due to cities being slow all the time rather than congested at peak hours. We do nonetheless find an important role for congestion in the largest cities, especially close to their centers. Third, several city attributes are consistently correlated with mobility and its components. We find that population and land area are key correlates of city mobility. A larger population leads to slower uncongested mobility as well as more congestion. We also find that both recent population growth and a measure of cars per capita are positively associated with uncongested mobility but also with congestion. More primary roads and a more regular grid-patterns are associated with moderately faster mobility. Higher income cities have higher uncongested mobility, but also higher congestion, leading to a hill-shaped relationship between income and overall mobility. Overall, these indicators of urban economic development are associated with better mobility despite worse congestion, contrary to a conventional wisdom that urban growth and development condemns developing cities to complete gridlock. While in principle, variation in uncongested mobility could be due to many city attributes beyond those we consider here in our regressions, such as the state of the vehicle stock or driving culture, we interpret it as being primarily due to the quality of the road network. Most old cars can be driven 45 kilometers per hour (the 99th percentile of our trip speed distribution), and Google Maps' algorithm is likely to pick out a high moment of the block speed distribution in order to distinguish motorized from non-motorized vehicles.

We hope that this first set of cross-city evidence on urban mobility and congestion in a developing country can help guide policy and future research. We now review three of our findings that have research and policy implications. First, we document that congestion in India is not a nationwide problem, but rather is highly concentrated near the center of the largest Indian cities. Given their importance to the Indian economy, these areas with the highest levels of congestion, such as the center of Kolkota and Bangalore, should be the focus of policy effort to alleviate congestion, and of future research to identify the most effective policies, as in Kreindler (2018).

Second, we compared travel patterns in India with those from more developed cities, and we uncovered important differences. In particular, Indian cities do not experience the familiar twin peak congestion patterns due to morning and evening commutes. There is almost no distinct morning peak, and instead a slow build up of congestion that often persists until late into the evening. Light rainfall appears to speed up traffic slightly. These unique patterns are consistent with Indian roads being multi-purpose public goods serving a wide variety of uses other than motorized transport that slow down travel. If this conjecture is correct, then further research on technologies and policies for separating roadway uses appears especially promising, with appropriate consideration for the costs of restricting non-vehicle uses. More generally, our findings of unique Indian travel patterns imply that country-specific policies are necessary, and that using our data sources and methodology to study other countries individually may uncover distinctive patterns.

Third, our most surprising and perhaps controversial finding is that in most Indian cities travel is slow at all times, not just peak times. As a result, standard policy recommendations like congestion pricing, HOV lanes, or other types of travel restrictions may do little to improve mobility. Instead, potentially costly travel infrastructure may be the only way to improve uncongested mobility. Our paper provides a first set of results suggesting a modest positive role for the design of a regular network grid and the presence of more primary roads. We hope that future research and engineering studies can identify cost-effective ways to build faster urban networks. On an optimistic note we find that better uncongested mobility generally correlates with the process of economic development. Unfortunately, this relationship is neither perfect nor linear.

We believe a lot more can be learned from the data we use here. In an extension of this paper (Akbar *et al.*, 2018), we provide complementary measures of urban accessibility in Indian cities, decompose accessibility into proximity and mobility, and provide an analysis of the urban correlates of accessibility and proximity. This sort of data can thus be used to learn about the fundamentals of urban travel beyond mobility and congestion. It can also potentially play an important role in our understanding of patterns of land use and property prices in cities in relation to transportation. Relative to more traditional travel surveys, the information used here is less complete but can be gathered at a small fraction of the cost, hundreds of dollars instead of tens of millions for full travel survey. The type of data we used here is also much more versatile and can thus be targeted at narrower issues or areas without fear of losing statistical power. It can also be collected at much higher frequency than the typical 5 to 8 year gap between consecutive traditional travel surveys.

This type of data is also particularly interesting to evaluate the effects of policy changes in the short-run. For instance, Kreindler (2016) uses a data collection comparable to ours for Delhi to examine the short-term congestion benefits of a new driving restriction based on vehicle plate numbers. Hanna *et al.* (2017) use a similar strategy to assess the effects of the relaxation of a high-occupancy vehicle constraints on certain major arteries in Jakarta. We believe these studies and future studies of this type will shed useful light on many aspects of transportation policy in cities. Many other possible applications are possible. They include, for instance, the monitoring of city recovery after major natural disasters.

We also hope that more data underlying the production of real-time travel information will be made available for research. The data that we use allow us to learn about mobility, and the price (time cost) of travel for all possible trips at all times. The analogous quantities (i.e., number of travelers) are potentially knowable from the same underlying data. With both prices and quantities, the detailed study of congestion, both on particular road segments and in larger areas, will be possible. Repeated observations of the same travelers would also enable a much better analysis of individual travel behavior. For instance, Kreindler (2018) uses a panel of trip-level data for 2,000 commuters from a smartphone app to learn about individual response to peak travel congestion, and to measure the welfare impact of various pricing policies to alleviate congestion in Bangalore. With appropriate regard for privacy, the availability for larger trip-level samples across cities would allow for a comprehensive analysis of the welfare consequences of better urban mobility and accessibility.

References

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. The economics of density: Evidence from the Berlin Wall. *Econometrica* 83(6):2127–2189.

Akbar, Prottoy A., Victor Couture, Gilles Duranton, and Adam Storeygard. 2018. Accessibility in urban India. Work in progress, University of Pennsylvania.

Akbar, Prottoy A. and Gilles Duranton. 2018. Measuring congestion in a highly congested city: Bogotá. Processed, University of Pennsylvania.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Anderson, Simon P., André de Palma, and Jacques-Francois Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge, ma: MIT Press.

Angel, Shlomo. 2008. An arterial grid of dirt roads. *Cities* 25(3):146–162.

Annez, Patricia Clarke, Alain Bertaud, Marie-Agnes Bertaud, Bijal Bhatt, Chirayu Bhatt, Bimal Patel, and Vidyadhar Phata. 2016. Ahmedabad. More but different government for "slum free" and livable cities. World Bank Policy Research Working Paper 6267.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2018. Retail globalization and household welfare: Evidence from Mexico. *Journal of Political Economy* 126(1):1–73.

Ben-Akiva, Moshe E. and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Ch, Rafael, Diego Martin, and Juan F. Vargas. 2017. Measuring the size and growth of cities using nighttime light. Processed, CAF Latin America Development Bank.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2016. The costs of agglomeration: House and land prices in French cities. Processed, Wharton School, University of Pennsylvania.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: Elsevier, 247–348.

Couture, Victor. 2014. Valuing the consumption benefits of urban density. Processed, University of California Berkeley.

Couture, Victor, Gilles Duranton, and Matthew A. Turner. 2018. Speed. *Review of Economics and Statistics* Forthcoming.

Duranton, Gilles and Erick Guerra. 2016. Developing a common narrative on urban accessibility: An urban planning perspective. Brookings Institution, Moving to Access.

Duranton, Gilles and Diego Puga. 2015. Urban land use. In Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: North-Holland, 467–560.

Duranton, Gilles and Matthew A. Turner. 2011. The fundamental law of road congestion: Evidence from US cities. *American Economic Review* 101(6):2616–2652.

Fuller, Brandon and Paul Romer. 2014. Urbanization as opportunity. World Bank Policy Research Paper 6874.

Glaeser, Edward L. and Matthew E. Kahn. 2004. Sprawl and urban growth. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2481–2527.

Hanna, Rema, Gabriel Kreindler, and Benjamin A. Olken. 2017. Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357(6346):89–93.

Harari, Mariaflavia. 2016. Cities in bad shape: Urban geometry in India. Processed, Wharton School of the University of Pennsylvania.

Kreindler, Gabriel. 2016. Driving Delhi? Behavioural responses to driving restrictions. Processed, MIT.

Kreindler, Gabriel. 2018. The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications. Processed, MIT.

Mills, Edwin S. 1967. An aggregative model of resource allocation in a metropolitan area. *American Economic Review (Papers and Proceedings)* 57(2):197–210.

Muth, Richard F. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Quinet, Emile. 2017. Accessibility in practice. Processed, Paris School of Economics.

Sheu, Gloria. 2014. Price, quality, and variety: Measuring the gains from trade in differentiated products. *American Economic Journal: Applied Economics* 6(4):66–89.

Small, Kenneth A. and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. New York (NY): Routledge.

United Nations. 2015. *World Urbanization Prospects: The 2014 Revision*. New York (NY).

Venter, Christo. 2016. Developing a common narrative on urban accessibility: A transportation perspective. Brookings Institution, Moving to Access.

Appendix A. Further data description

City sample and extent

United Nations (2015) reports the population and location of 166 cities in India that reached a population of 300,000 by 2014. Following (Harari, 2016) and Ch *et al.* (2017), we define the spatial extent of these cities as sets of contiguous 30 arc-second pixels with a lights-at-night digital number (DN) of at least 35 whose boundaries reach within 3 kilometers of the UN's reported latitude and longitude. The lights data are the stable lights product from the F-18 satellite.[36] The UN database initially reported an incorrect location for one city (Bokaro Steel City); it has since been corrected. We resampled Bokaro in December 2017 once we discovered this problem.

We drop two cities (Cherthala and Malappuram) that are not within 3 kilometers of a DN>35 light, one (Santipur) that belongs to a light with exactly one DN>35 pixel, and thus an implausibly small extent, and five cities that are too far east to be in the land use dataset described below (Agartala, Aizawl, Guwahati, Imphal, and Shillong). Four city-lights contain two cities each: Raipur and Durg-Bhilainagar, Mumbai and Bhiwandi, Asansol and Durgapur, and Bangalore and Hosur. We treat each of these four pairs as an individual city, with the center of the larger member of each pair kept as the center of the combined city. Our primary sample thus includes 154 cities.

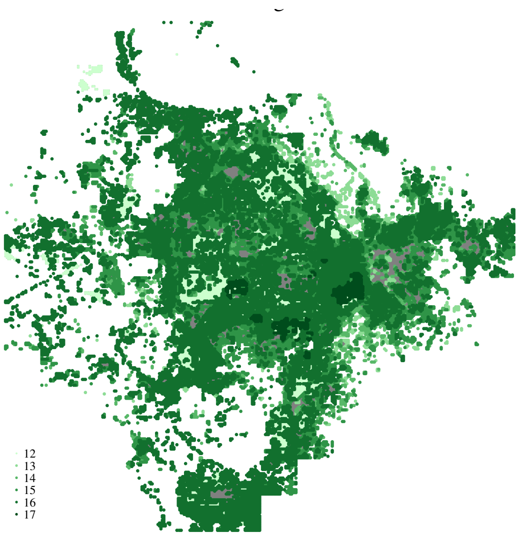
We further restrict city boundaries for the purpose of defining trip origins and destinations by excluding water bodies and non-urban land using 40-meter resolution land cover classifications from the Global Human Settlement Layer (GHSL) of the European Commission's Joint Research Centre (JRC). Cells identified as at least partially built up or roads within a city light are retained. Panel A of figure A.1 shows the lit and built-up portions on a median-sized city, Jamnagar in Gujarat, which we use for illustrative purpose throughout this appendix.

Trip sample

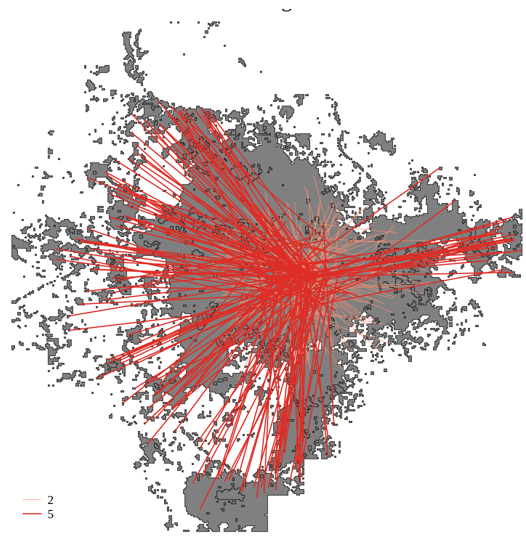
This section describes how we determine the within-city trips to query on Google Maps. We define a *trip* as a pair of points (origin and destination) within the same city as defined above. A *trip instance* is a trip taken at a specific time on a specific day. A *location*/point refers to a pair of longitude-latitude coordinates identifying the centroid of a roughly 40-meter GHSL pixel. We require that trip location pairs are at least one kilometer apart in haversine length, for three reasons. First, the rounding of travel times and lengths introduce potentially

[36] Available at `https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`.

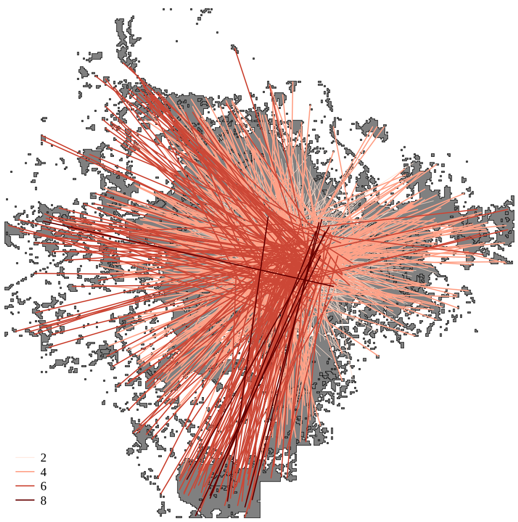
Figure A.1: Illustrations for the city of Jamnagar



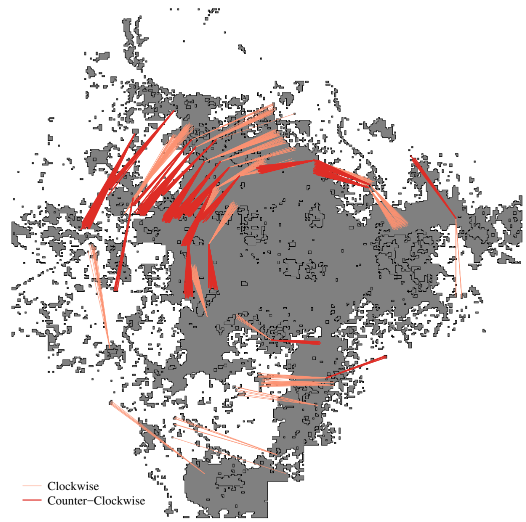
Panel A: Built-up categories within lit area
(higher numbers reflect greater built-up intensity)



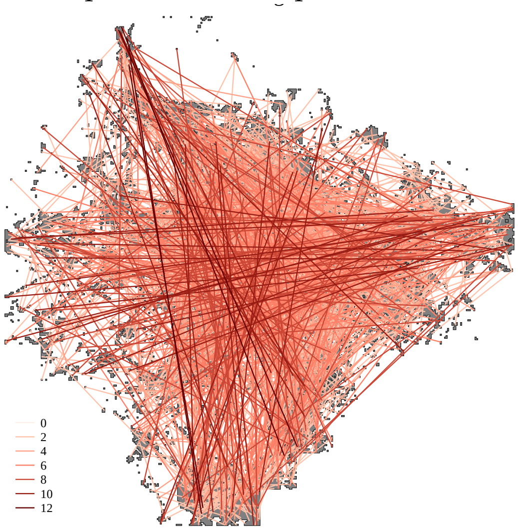
Panel B: Radial trips of absolute lengths
2 km, 5 km, 10 km, and 15 km from the center



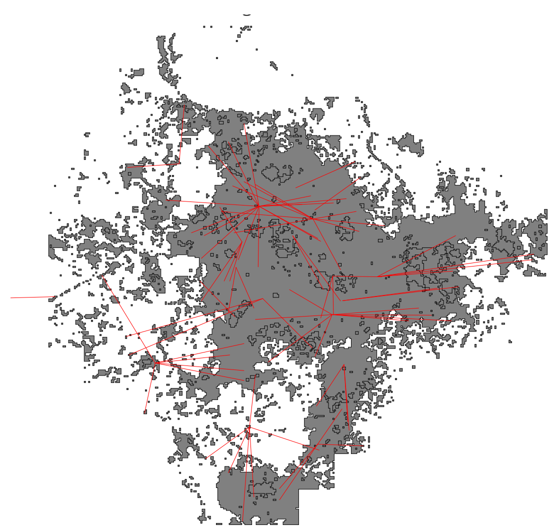
PanelC: Radial trips over uniformly
picked distance percentiles



Panel D: Circumferential trips around the center

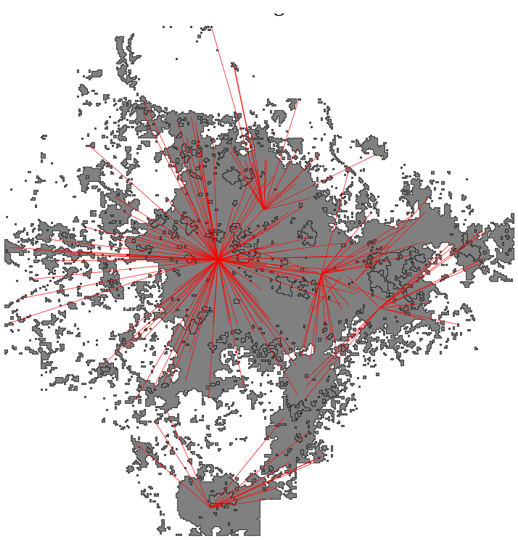


Panel E: Gravity trips

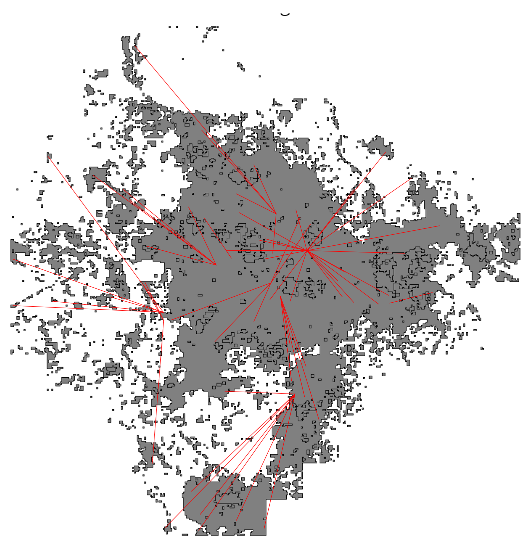


Panel F: Trips to school

Figure A.1 (continued): Illustrations for the city of Jamnagar



Panel G: Trips to shopping malls



Panel H: Trips to hospitals

non-classical measurement error in our computations of travel speed. Second, Google does not always return a driving time under traffic conditions for very short trips. Even when it does, the travel times can sometimes be very inconsistent or require taking unnecessary detours. Third, walking is an easy alternative to driving for short trips, and sources of error such as the unobserved time cost of finding parking, etc. will be a more significant component of the trip.

Our target sample for city c is 15 $\sqrt{Pop_c}$ trips, where Pop_c is the projected 2015 population of city c from United Nations (2015), and 10 trip instances per trip, to ensure variation across times of day. That is approximately 82,000 trip instances for the smallest of our cities, 116,000 instances for a median-sized city, and 760,000 instances for the largest city (Delhi).[37]

We define four types of trips: radial (2/9 of all trips), circumferential (1/9), gravity (1/3), and amenity (1/3).

Radial trips

Radial trips are defined in a polar coordinate system with respect to a city center. They have one end at a randomly located point within 1.5 kilometers of the city centroid as defined by United Nations (2015). Distance from the centroid is drawn from a truncated normal distribution with mean 0, standard deviation 0.75 kilometer and support [0,1.5] kilometers. For convenience, we call this the destination, but in practice trips in both directions are

[37]By comparison, in the 2008 US National Household Transportation Survey (NHTS), the 187th, 100th, 50th, 10th and 1st most sampled US metro areas have about 200, 800, 2,200, 12,000, and 29,000 trips, respectively.

sampled. For each destination, the point of origin is determined using two methods with equal probability:

1. *Absolute distances* of $AbsDist \in \{2,5,10,15\}$ kilometers (equally weighted) are drawn. For each of these four distances, we (uniform) randomly pick a point of origin within the lit-up area of the city that is between $(AbsDist - 0.2)$ kilometers and $(AbsDist + 0.2)$ kilometers from the given destination. See panel B of figure A.1 for illustration with the city of Jamnagar. Darker shades of red distinguish longer trips.

2. *Distance percentiles* relative to the largest possible distance for any trip from a lit-up area of the city to that destination are drawn from a uniform distribution from the 1st to 99th percentile (excluding distances less than 1 kilometer). See panel C of figure A.1 for illustration with the city of Jamnagar.

If a city has no valid trips for a given absolute distance +/-0.2 kilometer, the trips assigned to that distance are reallocated to the distance percentiles sample.[38] Similarly, if there are not enough unique 40 m pixel centroids $AbsDist$ +/-0.2 kilometer from the center destination to fill a given absolute distance's quota, the remainder of the quota is filled with randomly drawn distance percentiles instead.

Circumferential trips

Like radial trips, circumferential trips are also defined in a polar coordinate system with respect to a city center. Circumferential trips originate at a random origin at least 2 kilometers away from the city centroid. The analogous destination is at the same distance (+/-0.2 kilometer) from the centroid, 30 (+/-3) degrees clockwise or counter-clockwise from the origin. For three small cities, the city centroid according to United Nations (2015) is far from the geographic center of the city-light, so it was not possible to fill the circumferential trip quota. See panel C of figure A.1 for illustration with the city of Jamnagar.

Gravity trips

Gravity trips are designed to match the length profile of trips sampled in the US NHTS and the Bogotá Travel Survey. We identified each location-pair using the following algorithm:

1. Consider a uniformly randomly picked initial point ($GravityPoint$) and a length ($GravityLength$ kilometers) drawn from a truncated pareto distribution with shape

[38] Only 43 cities have a maximum distance to centroid of 15 kilometers or more. 78, or roughly half, of the cities have a maximum distance of 10 kilometers or more. 132 cities have a maximum distance of 5 kilometers or more, and all cities have a maximum distance greater than 2 kilometers (with the smallest maximum distance being 2.8 kilometers).

parameter 1 and with support between 1 kilometer and 250 kilometers (corresponding to a mean of roughly 5.52 kilometers).[39]

2. Choose a point randomly from among all points at a straight-line length between $(GravityLength - 0.2)$ kilometers and $(GravityLength + 0.2)$ kilometer from the point $GravityPoint$. If there are no such points, start over from (1) with a new pair of $(GravityPoint, GravityLength)$.

See panel E of figure A.1 for illustration with the city of Jamnagar. Darker shades of red distinguish longer trips.

Amenity trips

Amenity trips join a random origin with an instance of one of 17 amenities (e.g. shopping malls, schools, train stations) as recorded in Google Places. The particular instance we used is based on a combination of proximity and "prominence" assigned by Google. The weighting across these amenity types is based on a mapping of amenities to trip purposes for the 100 largest MSA in the US from the 2008 US National Household Transportation Survey (NHTS) (Couture *et al.*, 2018). NHTS has nine categories of trip purpose (trip share in parentheses): Work (23.6%), Work-related business (3.3%), Shopping(21.8%), School & Religious practice (4.6%), Medical/dental (2.2%), Vacation & visiting friends/relatives (6.0%), Other social/recreational (13.8%), Other family/personal business (24.3%), and Other (0.5%).

The Google Places API classifies points of interest using one or more of roughly 100 Google-defined place "types". We match each NHTS trip purpose to the most relevant Google Places types, using city hall for Work, under the assumption that employment is relatively concentrated near the city center. Since we cannot identify types associated with Other family/personal business, we reallocated its 24.3% share among the rest of the categories except Work using the following formula. If place type v gets $TripTypeShare_v$% of the trips otherwise, then they get an additional $\frac{24.3(23.6 - TripTypeShare_v)}{\sum_w (23.6 - TripTypeShare_w)}$. Less popular place types get a larger share of Other family/personal business as we do not want too few absolute trips in any category. The final allocation is shown below. The first number in each category is its initial allocation, and the second is its share of Other family/personal business.

- Work: city hall (23.6%+0%)
- Work-related business: gas station (3.3%+1.5%)
- Shopping: shopping mall (7.3%+1.2%), convenience store (7.3%+1.2%), grocery/ supermarket (7.2%+1.2%)

[39]This mean of 5.52 kilometers is slightly smaller than the mean of 6.51 kilometers for Bogotá from Akbar and Duranton (2018).

- Social/recreational: movie theater (5.7%+1.3%), park (5.7%+1.3%), stadium (2.4%+1.5%)
- School & religious practice: school (2.3%+1.6%), place of worship (2.3%+1.6%)
- Medical/dental: hospital (1.1%+1.7%), doctor (1.1%+1.7%)
- Vacation & visits: train station (3.0%+1.5%), airport (1.0%+1.7%), bus station (2.0%+1.6%)
- Other: police (0.25%+1.75%), post office (0.25%+1.75%)

We set a different maximum radius of the search around any initial point based on the place type:

- 50 kilometers radius: city hall, airport, stadium
- 20 kilometers radius: train station, bus station, hospital, doctor
- 10 kilometers radius: movie theater, school, police
- 5 kilometers radius: shopping mall, convenience store, grocery/supermarket, park, place of worship, gas station, post office

A query request to Google Places API specifies a search location and a 'type'. For each query, we randomly draw (without replacement) a new location within our city's lit-up boundary. We call a query to the API successful if it returns at least one place. For a given city, if a query by 'type' is unsuccessful more often than not after at least 50 unsuccessful queries, we switch to querying by 'keyword', which is more likely to return results but also more likely to include badly matched returns, e.g. return coordinates for some segment of a road named "Airport Road" instead of coordinates for the airport. If queries by keyword also continue to be unsuccessful more often than not, after 50 unsuccessful queries we reallocate the remaining share of the location pairs evenly among the rest of the place types under the same trip purpose category. For example, suppose we require 100 location pairs for 'convenience stores' and the first 50 queries by type return zero results. So we switch to querying by keyword. Suppose, the 80th query by keyword is the 50th unsuccessful one. Then we stop there, get 30 location pairs from the successful queries for 'convenience stores' and reallocate the remaining 70 required location pairs to 'shopping mall' and 'grocery/supermarket' (35 each). If all place types in the same trip purpose category yield zero place returns more often than not and we have yet to fulfil our quota of location pairs in the category, then we re-distribute the count of unqueried location pairs evenly across all the rest of the place types.

From each successful query, we collect only the first twenty places returned by Google in order of "prominence", as determined "by a place's ranking in Google's index, global popularity, and other factors". For each place, Google's Places API returns us: geographical coordinates, "name", "vicinity" (this might be either an address or nearby landmarks), and the "types" it is classified under. We only keep places that are at least one kilometer in straight-line distance from the random initial point. Then we use the "name", "vicinity" and "types" of the place to score the relevance/quality of each place return. We drops places below a minimum threshold (i.e. more likely to be a bad match), and use the highest scoring place, breaking ties first with length differentials over one kilometer (i..e keeping the closest), and then by "prominence" (i.e., the order in which they are reported by Google). This ensures that small differences in length are ignored in favor of Google's recommendation.

Since not all successful queries return good quality places, we make 50% more queries than needed. When choosing the final set of trips to query for traffic, we prioritise trips to places that scored highly on relevance. If we need to break ties here, we pick randomly. Panels F, G, and H of figure A.1 illustrate for the city of Jamnagar our selection of trips to schools, shopping malls, and hospitals, respectively.

Querying trips on Google Maps

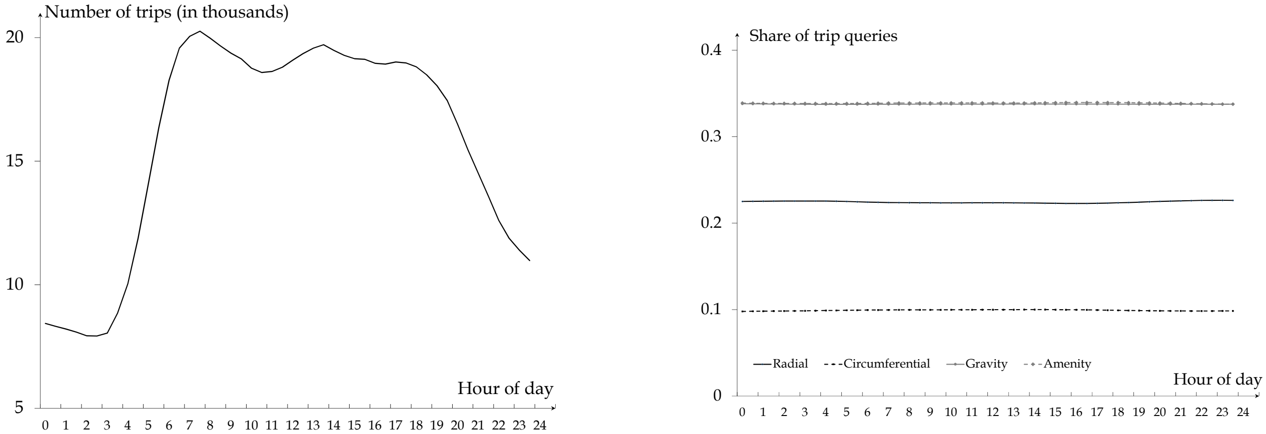
Our target sample was 2,373,764 trips across all cities and strategies, corresponding to 1,186,882 locations pairs. Because of some overlaps between trips and because Google Maps did not return any route for few hundred trips, we ended up with 2,333,762 queried trips, or 98.3% of our target. Across cities, the mean is 98.7% with a coefficient of variation of 1.34%.

We simulated 22,766,881 trip instances across 40 days between September and November of 2016. This corresponds to 92.5% of our target on average in Indian cities with a coefficient of variation of 4.06%. The median (as well as the mean) trip was queried 10 times (with a standard deviation of 1.9) and 99% of the trips were queried at least 8 times. Missing trip instances are due mostly to empty returns from Google Maps or minor technical glitches such as early computer disconnections, formatting problems in the returns, etc.

We wanted the distribution of trip departure/query times to roughly resemble the distribution of departure times on a typical weekday.[40] However, we also wanted enough trip queries from each time period of the day for the fixed effects to be credible, so we oversampled the early morning. At any hour of the day, we had the following number of machines querying trips on Google: 12 a.m. - 4 a.m.: 15, 4 a.m. - 5 a.m.: 20, 5 a.m. - 6 a.m.: 35, 6 a.m. - 8 a.m.: 40, 8 a.m. - 12 p.m.: 35, 12 p.m. - 1 p.m.: 40, 1 p.m. - 5 p.m.: 35, 5 p.m. - 7 p.m.: 40, 7 p.m. - 9 p.m.: 30, 9 p.m. - 10 p.m.: 25, 10 p.m. - 12 a.m.: 20. All the machines had identical processing power, so the number of machines also reflects the distribution of our trip queries across hours of the day. Panel A of figure A.2 shows the realized distribution of query times across hours of the day.

[40]We rely on a household transportation survey from Bogota, Colombia as a reference for this.

Figure A.2: Queries by time of the day



Panel A: Number of trip instances across times of the day

Panel B: Proportion of types by time of day

We wanted to have an even spread of days and times across cities and trip types/strategies. So the order in which the trips were queried was randomized to alternate between strategies and cities (based on the size of the city, e.g. city A - with twice as many trips as city B - is queried twice between every city B query). Once we have run through the ordered list of trips, we start over at the beginning of the list. Panel B of figure A.2 shows the stable realized proportion of trip types across hours of the day.

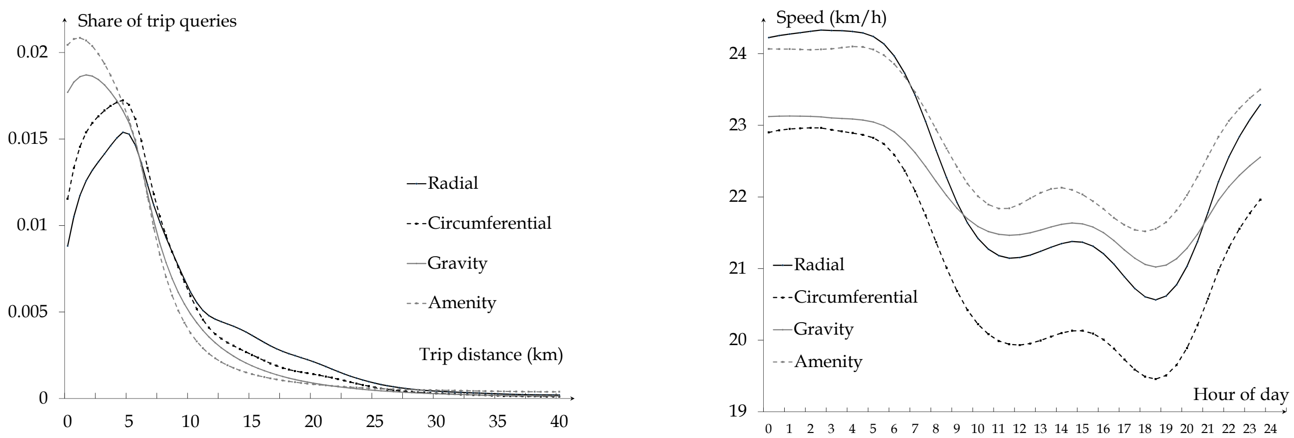
As the ordering of trips stays the same, one may worry that if the time it takes to cycle through the list is roughly a multiple of 24 hours, there will be too little variation in time of day across instances of the same trip. So we split the day into four 6-hour time slots (12 a.m. - 6 a.m., 6 a.m. - 12 p.m., 12 p.m. - 6 p.m., 6 p.m. - 12 a.m) and forced randomization within each of them by maintaining a separate trip query list for each slot. That means, at the end of each 6 hour slot we bookmarked our location on the query list and came back to it in 18 hours. This makes sure that no trip is randomly over- or under-queried at any given 6-hour slot of day. We managed to make sure that 95% of the trips were queried at all four 6-hour time slots, and every trip was queried at, at least, three of the four slots.

We sampled weekends at 50% of our weekday rate, using the same method. While we might prefer to oversample "Other family/personal business" trips on weekends, as discussed above we cannot narrow down the set of destinations for this category.

Travel lengths and speeds

The median Google-reported travel length across all our trips is 5 kilometers (with a standard deviation of 10.5 kilometers). However, there are noticeable differences across our four trip selection strategies. Figure A.3 shows the distribution of travel lengths for the portfolio

Figure A.3: Travel length and speed



Panel A: Distribution of travel lengths by trip class

Panel B: Travel speeds across time of day by trip class

of trips under each strategy. Amenity trips are relatively shorter in length, with a median of 4.2 kilometers. This is understandable as our algorithm weakly prefers closer destinations for any given amenity. Radial trips are the longest, with a median of 6.6 kilometers. This is probably because we force a large share of the trips to be of fixed haversine lengths of 5 , 10 and 15 kilometers, which translate to even larger actual travel lengths.[41] Recall that the gravity trips are designed to mirror the distribution of travel lengths that have been observed in other cities.

Panel B of figure A.3 shows how travel speeds through the day vary across our trip selection strategies. As we would expect, speeds are highest in the early hours of the morning and late at night and lowest during the day, in particular around the 6 - 7 p.m. evening rush hour. Some of the differences in speeds across strategies may be explained by the differences in trip lengths, as longer trips also tend to be faster. But, clearly there is more to it: circumferential trips experience the lowest speeds, and speeds for the radial and circumferential trips seem relatively more sensitive to daytime increases in traffic.

Walking and transit trips

We do not expect walking times for a given trip to vary by either the day or the hour of day. However, walking speeds do vary based on slope and the density of the network of streets and pedestrian paths. So, unlike for driving times, we query each location pair only once, in one direction, for walking times.

[41] In fact, the ratio of total travel length to total haversine length is 1.53.

Table A.1: Ranking of cities by transit network coverage

Rank	City	State	Coverage
1	Chennai	Tamil Nadu	0.74
2	Bangalore	Karnataka	0.73
3	Pune	Maharashtra	0.73
4	Mysore	Karnataka	0.69
5	Mumbai	Maharashtra	0.67
6	Ahmedabad	Gujarat	0.65
7	Chandigarh	Chandigarh	0.63
8	Rajkot	Gujarat	0.62
9	Kolkata	West Bengal	0.61
10	Jaipur	Rajasthan	0.61

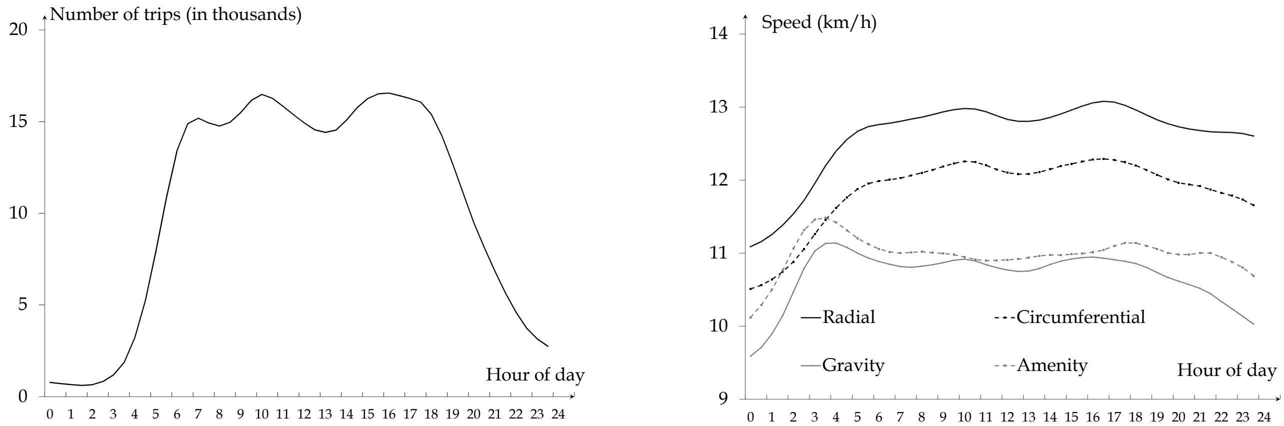
Notes: Coverage refers to the share of trip instances with viable transit routes returned by Google Maps.

Google does not generally track transit in real time, but instead relies on public transportation schedules made available by transit authorities and open General Transit Feed Specification data. Thus, for any given trip, we do not expect any meaningful variation across weekdays in our travel times by transit. Scheduled transit frequency does however vary by time of day. We thus re-queried each weekday trip instance in our driving data as a transit trip, at its original time of day, but on 10 January 2018. This was a Wednesday that did not coincide with any public holidays in India to our knowledge.

There are several important caveats to these data. First, 22% of queries, including all queries in 14 cities, returned no routes. Second, we do not expect the schedules to include informal transit providers, which own the large majority of India's bus fleet.[42] Third, some returned routes are implausible. Specifically, we exclude routes that (1) require walking all the way, (2) require waiting over an hour to start the trip, or (3) are slower than their walking counterpart, which happens when Google uses inter-city rail, presumably because it is the only nearby transit alternative, to create highly convoluted itineraries. Following these exclusions, only 20% of our driving trip instances offer viable transit alternatives, and they are highly concentrated in the largest cities. In 133 of our 154 cities, less than 8% of trips are viable by transit. We cannot distinguish whether the absence of a viable transit route is due to limitations in the city's transit network or limitations in Google Maps' coverage of the transit network. With that in mind, we report the 10 cities with the largest share of our trip instances covered by Google Maps in Table A.1.

[42]See https://data.gov.in/catalog/number-buses-owned-public-and-private-sectors-india, consulted 26 April 2018. Note also that Google Maps only officially lists transit authorities spanning 12 Indian cities, corresponding to 10 of our cities, and four multi-region services that share their transit schedules (http://maps.google.com/landing/transit/cities/), but queries in an additional 130 cities returned transit components.

Figure A.4: Transit data



Panel A: Distribution of query times across times of the day

Panel B: Travel speeds across time of day by trip class

Road network data

Our measure of road network characteristics come from OpenStreetMap (OSM), a collaborative worldwide mapping project. We downloaded OSM data within the light-based boundary of each city through Geofabrik in September 2016.[43] We then used OSMNX, a python pacakge created by Geoff Boeing, to process the OpenStreetMap network as a directed graph of edges and nodes.

Road length

Each edge in the OSM network receives a tag which characterizes its road type. We measure total road length in kilometers for three types of roads:

1. Motorways: The highest capacity roads in a country, equivalent to freeways in the United States. Motorways generally consist of restricted access dual carriage ways with 2 or more lanes in each direction plus emergency hard shoulder.[44]
2. Primary Roads: The next most important road in a country's transportation system, after motorways and trunks. Generally not dual carriage ways.
3. Total Road Length: aggregation of all road types driveable by motor vehicles and public for everyone to use.[45]

[43] `http://download.geofabrik.de/asia/india.html`

[44] We also include the less frequent OSM type "trunks" in the motorways category. Trunk are the next most important types of roads after motorway, and often but not always consist of dual carriage ways.

[45] In the OSM network, both carriage ways of a motorway count as separate edges (in each direction). We experimented with counting dual carriage ways only once when measuring length, and also with measuring lane-kilometers, instead of just edge kilometers. These adjustments generates measures of length by road type that are very highly correlated with that without adjustments that we show in the paper.

We note that certain cities have incomplete street networks on OSM. Using satellite data, we visually identified a set of cities for which the road network appear incomplete (Jhansi, on the left-hand panel of Figure A.5, is one such cities.) The results are robust to limiting the sample to the subset of cities for which we have a more complete road network.

Characterizing the road network

OSMNX calculates the compass bearing ("bearing" for short) from each directed edge's origin node to its destination node. The bearing captures the orientation of the edge with respect to true north. We use the distribution of edge bearings in a city to characterize how 'grid-like' its road network is. We measure how grid-like a network is in two separate ways: 'orientation' which captures the share of edges conforming to the network's main grid orientation, and 'Gini' which captures the dispersion in the distribution of edge bearings. We now describe both measures of how grid-like a road network is in more detail.

Orientation. A grid is a series of roads intersecting at perpendicular angles. If a city were a perfect grid network, then all bearings for would be either be perpendicular or parallel to each other. The orientation grid metric measures the proportion of edges in a city's road network that conform to the dominant grid orientation in that they are perpendicular or parallel to the modal edge bearing.

Let g index each edge in the road network of city c, and let x_{cg} be the edge bearing rounded to the nearest degree, and x_c^{modal} be the modal edge bearing modulo 90 of city c. For example, if a city's grid were oriented N-E-S-W, then x_c^{modal} would equal 0. Let $\delta_{gc,x_c^{modal},\nu}$ be an indicator for whether edge g in city c conforms to grid orientation x_c^{modal} within a bandwidth error of ν:

$$\delta_{ic,x_c^{modal},\nu} = \begin{cases} 1 & \text{if } (x_g - x_{0ct}) \mod 90 <= \nu \\ 1 & \text{if } (x_g - x_{0ct}) \mod 90 >= (90-\nu) \\ 0 & \text{else.} \end{cases} \tag{A1}$$

We then compute our grid-like measure as:

$$Orientation_c = \frac{\sum_{g \in I_c} \delta_{gc,x_c^{modal},\nu}}{Q_c}, \tag{A2}$$

where I_c is the set of all edges in city c, and Q_c is the number of edges in I_c.

In the paper, we report results using a narrow error bandwidth of $\nu = 2^\circ$. We experimented with a wider bandwidth of 5°. We also experimented with allowing for more than one dominant grid orientation, because for instance larger cities can have smaller sub-grids whose orientation differs from that of the main grid.[46] These variations produce highly

[46] We also experimented with weighting edges by length, but visual inspection suggests that such measures overestimate how grid-like small cities with few very long roads are.

Figure A.5: Most and least grid-like city road network using orientation grid metric

Panel A: Chandigarh - Grid Score = 0.54

Panel B: Jhansi - Grid Score = 0.08

correlated rankings of cities, and we therefore prefer the simplest version above. Visual inspection suggests that our methodology performs well at ranking road networks by how grid-like they are. Figure A.5 shows the most and least grid-like cities according to the orientation metric, side-by-side.[47]

Gini. We modify the definition of the Gini index for income inequality to measure the normalized dispersion of edge bearings. For each city c, we define 360 different possible bearings, indexed by k, and ranked by their frequency such that $k = 1$ is the least frequent bearing and $k = 360$ is the most frequent bearing. In a perfectly gridded city, the four most frequent bearings, spaced 90 degrees apart, would account for 100% of edge bearings. Therefore, we can interpret high values of the following Gini index as corresponding to cities with a more grid-like network:

$$Gini_c = \frac{Q_c \times 360 - 2\sum_{k=1}^{360}\sum_{l=1}^{k}\theta_{cl}}{Q_c \times 360}, \tag{A3}$$

where θ_{cl} is the number of edges in city c with bearing l. The Gini and orientation metric have a correlation of 0.53.

The assumption of 360 possible distinct bearings is arbitrary, and we also computed Gini indices after rounding up each bearing to the nearest even degree (i.e., by assuming 180 possible bearings.) We also experimented with defining modulo 90 bearings (instead of modulo 360 as above).[48] These variations produce Gini indices that are highly correlated

[47] It is also possible to compute measures of how grid-like the road network is separately for different types of road defined above, instead of only for the total road network. However, visual inspection suggest that these measures do not perform well at capturing overall how grid-like cities are, and for instance motorways are often curved and outside of the main grid.

[48] For some smaller cities with sparser road networks, the number of distinct edge bearings is less than 360. In these cases, we adjust the calculation to consider only the total set of bearings present in that city, which may be less than 360.

with the index defined above that we use in the paper.

Weather data

Hourly and daily historical weather data (rain, thunderstorm, temperature, humidity, and wind speed) are from the Weather Underground website.[49] Weather Underground (WU) links each city to a station nearby (if there is one) and reads the weather reported by the station at the time it was reported.

We recovered weather data for 112 cities during the trips collection period. The median city-day has 8 weather readings, with a range from 1 to 144. On an average day, 25 of the cities report weather at least once every hour and 13 of them (mostly cities with international airports) report every half hour or more. The number of readings per day for a given city varies little across days.

The remaining 42 cities are missing data for one or more of the following three reasons First, WU does not recognize the city name (4 cities). Second, WU recognizes the city name, but has no data on it (i.e., not linked to any weather stations – 31 cities). Third, WU re-directs to a different city name, either because: (a) WU recognizes our entry as an alternative name to the returned city, or (b) WU treats the city as a suburb or extension of a larger city nearby (20 cities). In this case, we accepted the returned city as a proxy as long as it was within 50 kilometers of the queried city (8 of 20 cities). Over the two months when we collected weather data, it rained 4.5% of the time and there were thunderstorms 2% of the time.

Appendix B. Derivation and computation of the logit/CES mobility index.

We define the utility from visiting the destination of trip i in city c as:

$$u_{ci} = \log(b_{ci}) + (1-\sigma)\log(t_{ci}) + \epsilon_{ci}, \tag{B1}$$

where $t_{ci} = \gamma T_{ci}$ is the time cost of a trip to destination i in city c that takes T_{ci} units of time at value of time γ per unit, and ϵ_{ci}, the random component of utility, has a Type I extreme value distribution.[50] The parameter $\sigma > 1$ is an elasticity of substitution across destinations, and b_{ci} is a trip-specific quality parameter capturing all factors other than time costs making some destinations more desirable than others.[51]

[49] https://www.wunderground.com/history

[50] Ben-Akiva and Lerman (1985) are the first to show how to derive a travel accessibility index from a logit model of travel demand. Anderson, de Palma, and Thisse (1992) are the first to show the correspondence between the logit and CES models.

[51] In Table 8, we present an index computed at $\sigma = 0$. Technically, values of $\sigma < 1$ are inconsistent with utility maximization. In practice, the index at $\sigma = 0$ simply weights all trips equally and intuitively corresponds to a perfect complement case.

The expected utility of a traveler in city c is equal to the expected value of u_{ci}'s maximum across the N_c travel destinations available in city c:[52]

$$\mathbf{E}\left(\max_{i\in N_c}\{u_{ci}\}\right) = \log\left(\sum_{i=1}^{N_c} \exp\left[\log(b_{ci}) + (1-\sigma)\log(t_{ci})\right]\right) = \log\left(\sum_{i=1}^{N_c} b_{ci} t_{ci}^{1-\sigma}\right) . \tag{B2}$$

Now consider two cities, c and c'. Define a relative price index $G_{c,c'}$ as the factor by which travel costs in city c would have to change in order to equalize expected utility in the two cities:

$$\log\left(\sum_{i=1}^{N_c} b_{ci}(G_{c,c'} t_{ci})^{1-\sigma}\right) = \log\left(\sum_{i=1}^{N_c} b_{c'i} t_{c'i}^{1-\sigma}\right) . \tag{B3}$$

It is easy to show that

$$G_{c,c'} = \left(\frac{\sum_i^{N_{c'}} b_{c'i} t_{c'i}^{1-\sigma}}{\sum_i^{N_c} b_{ci} t_{ci}^{1-\sigma}}\right)^{1/(1-\sigma)} = \left(\frac{\sum_i^{N_{c'}} b_{c'i} T_{c'i}^{1-\sigma}}{\sum_i^{N_c} b_{ci} T_{ci}^{1-\sigma}}\right)^{1/(1-\sigma)} , \tag{B4}$$

where the second equality uses $t_{ci} = \gamma T_{ci}$. The relative price index $G_{c,c'}$ is best characterized as a relative travel accessibility index. It is low when comparing cities that have many destinations to those with few (gains from variety), and when comparing cities where travel to those destinations is short-distance and fast to those where it is long-distance and slow.

We now develop a simple non-parametric procedure to isolate a pure mobility index determined only by speed differences across cities. To do this, we replace the denominator of $G_{c,c'}$ with a 'national index' that has exactly the same distribution of trip length as in city c, and the same number of trips. This leads to equation (6) in the main text. Note that we inverted the index to ensure that G_c increases with faster speed (the index derived above is a price index increasing with time costs.) We compute $\overline{T}_{ci}$ as the average travel time of all trips in the national sample with length within 1% of that of trip i in city c. We drop any trip with fewer than 10 corresponding trips within 1% of its length in the national sample (less than 0.01% of trips).

We investigate robustness to the parametrization of the quality parameters b_{ci}. For this investigation, we restrict the sample to amenity trips. We do not observe the quality of destinations, but we sampled amenity trips to match the trip shares in the US NHTS, so assuming that $b_{ci} = 1$ for all amenity trips is a reasonable starting point to compute G_c. We then compute variations of this index using random draws of $b_{ci} \in U[1{,}100]$, thus randomly allowing certain destinations to be more desirable and to carry a higher weight in the index. Indices obtained from these draws are highly correlated with one another and with our benchmark index. This exercise is not a particularly demanding robustness test, but it corroborates other findings from Table 8, showing that slow cities are slow for all types of

[52] See Anderson *et al.* (1992), pp. 60–61, for a proof of the equality in equation (B2).

trips, and that weighting certain trips more than others has little impact on our mobility indices.

Finally, we divide trips into M groups and compute the following nested CES/logit mobility index:

$$G_c^{nest} = \frac{\left(\sum_{m=1}^{M} G_{mc}^{1-\mu}\right)^{\frac{1}{1-\mu}}}{\left(\sum_{m=1}^{M} \overline{G}_{mc}^{1-\mu}\right)^{\frac{1}{1-\mu}}}, \tag{B5}$$

and

$$G_{mc} = \left(\sum_{i=1}^{N_{mc}} b_{ci} T_{ci}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}, \quad \overline{G}_{mc} = \left(\sum_{i=1}^{N_{mc}} b_{ci} \overline{T}_{i}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}, \tag{B6}$$

where $\mu > 1$ is the elasticity of substitution across groups, $\sigma > \mu$ is the elasticity of substitution within groups, and N_{mc} is the number of trips in group m in city c.[53] As an example, we can define eight groups, one for each amenity type recorded in Appendix A. In this case, the nested index G_c^{nest} puts less weight on destination types that are relatively slower in city c; travelers substitute away from them because they are costlier. We compute these indices using exactly the same methodology as before. Setting $\mu = 1.5$ and $\sigma = 2.5$, we experiment with various nesting structures defined by time (e.g., non-peak, peak, high-peak), area (e.g., rings), types of destinations (e.g., amenity types), and find high correlation with our benchmark index in all cases.

Appendix C. Further results

The four panels of table C.1 duplicate the results of table 4 for each type of trip separately. Table C.2 duplicates table 10 but uses as dependent variable a fixed effect from a trip regression where trips are weighted by how slow they are relative to their speed in absence of traffic ($\lambda = 0.2$). Finally, table C.3 duplicates the specification of column 6 in table 10 but uses as dependent variables further alternative mobility indices.

Appendix D. A ring analysis of mobility in Indian cities

Although our main findings of city-level correlations in Section 6 are generally stable across a wide variety of specifications, they may be subject to bias due to omitted city-level

[53] Sheu (2014) extends the equivalence result in Anderson *et al.* (1992) to show that the nested-CES price index below can be also derived from modifications of a standard discrete choice nested logit model.

Table C.1: Correlates of log trip speed for specific trip classes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Radial trips							
log trip length	0.28[a]	0.073[b]	0.074[b]	0.28[a]	0.069[b]	0.069[b]	0.049
	(0.0063)	(0.033)	(0.033)	(0.0063)	(0.033)	(0.033)	(0.059)
log trip length^2		0.055[a]	0.055[a]		0.057[a]	0.057[a]	0.063[a]
		(0.011)	(0.011)		(0.011)	(0.011)	(0.018)
log distance to center		0.15[a]	0.15[a]		0.16[a]	0.16[a]	0.15[b]
		(0.048)	(0.048)		(0.049)	(0.049)	(0.076)
log distance to center^2		-0.11[b]	-0.11[b]		-0.11[b]	-0.11[b]	-0.13[b]
		(0.042)	(0.042)		(0.042)	(0.042)	(0.061)
Observations	5,102,925	-	-	4,347,207	-	-	2,313,862
R-squared	0.53	0.54	0.54	0.53	0.54	0.54	0.57
Panel B. Circumferential trips							
log trip length	0.26[a]	0.056[c]	0.056[c]	0.26[a]	0.053[c]	0.053[c]	0.037
	(0.0083)	(0.030)	(0.030)	(0.0082)	(0.030)	(0.030)	(0.058)
log trip length^2		0.060[a]	0.059[a]		0.060[a]	0.060[a]	0.066[a]
		(0.010)	(0.010)		(0.010)	(0.010)	(0.018)
log distance to center		0.15[b]	0.15[b]		0.15[b]	0.15[b]	0.14
		(0.060)	(0.060)		(0.062)	(0.062)	(0.11)
log distance to center^2		-0.13[b]	-0.13[b]		-0.13[b]	-0.13[b]	-0.14[c]
		(0.051)	(0.051)		(0.052)	(0.052)	(0.082)
Observations	2,261,556	-	-	1,934,692	-	-	1,018,394
R-squared	0.45	0.46	0.47	0.45	0.46	0.47	0.51
Panel C. Gravity trips							
log trip length	0.21[a]	0.13[a]	0.13[a]	0.21[a]	0.13[a]	0.13[a]	0.14[a]
	(0.0032)	(0.012)	(0.012)	(0.0032)	(0.012)	(0.012)	(0.014)
log trip length^2		0.016[a]	0.016[a]		0.015[a]	0.015[a]	0.013[a]
		(0.0031)	(0.0031)		(0.0032)	(0.0032)	(0.0035)
log distance to center		0.18[a]	0.18[a]		0.18[a]	0.18[a]	0.13[c]
		(0.051)	(0.051)		(0.050)	(0.050)	(0.077)
log distance to center^2		0.031	0.031		0.036	0.036	0.047
		(0.026)	(0.026)		(0.025)	(0.025)	(0.038)
Observations	7,672,821	-	-	6,539,528	-	-	3,495,291
R-squared	0.38	0.45	0.45	0.37	0.45	0.45	0.46
Panel D. Amenity trips							
log trip length	0.25[a]	0.17[a]	0.17[a]	0.25[a]	0.17[a]	0.17[a]	0.16[a]
	(0.0045)	(0.011)	(0.011)	(0.0045)	(0.010)	(0.010)	(0.014)
log trip length^2		0.0064[c]	0.0064[c]		0.0059[c]	0.0059[c]	0.0081[c]
		(0.0034)	(0.0034)		(0.0033)	(0.0033)	(0.0044)
log distance to center		0.21[a]	0.21[a]		0.21[a]	0.21[a]	0.17[a]
		(0.037)	(0.037)		(0.036)	(0.036)	(0.052)
log distance to center^2		0.0052	0.0052		0.0097	0.0097	0.019
		(0.019)	(0.019)		(0.019)	(0.018)	(0.027)
Observations	7,706,854	-	-	6,564,229	-	-	3,492,392
R-squared	0.55	0.60	0.60	0.55	0.60	0.60	0.54
City effect	Y	Y	Y	Y	Y	Y	Y
Day effect	Y	Y	Y	weekd.	weekd.	weekd.	Y
Time effect	Y	Y	Y	Y	Y	Y	Y
Weather	N	N	Y	N	N	Y	only

Notes: OLS regressions with city, day, and time of day (for each 30-minute period) indicators. Log speed is the dependent variable in all columns. Robust standard errors in parentheses. a, b, c: significant at 1%, 5%, 10%. 154 cities in columns 1-7 and 107 in column 8. All trip instances in columns 1-3. Only weekday trip instances in columns 4-6. Only weekday trip instances for which we have weather information in column 7. Weather in column 3 and 6 consists of indicators for rain (yes, no, missing), thunderstorms (yes, no, missing), wind speed (13 indicator variables), humidity (12 indicator variables), and temperature (8 indicator variables). These variables are introduced as continuous indicator variables in column 7. Sample sizes for columns 1 and 4 apply to columns 1–3 and 4–6, respectively.

Table C.2: Correlates of city mobility indices, mobility index for which trips are weighted by powered congestion factor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log population	-0.19^a	-0.19^a	-0.18^a	-0.18^a	-0.18^a	-0.17^a	-0.18^a	-0.17^a
	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.023)	(0.021)	(0.021)
log area	0.13^a	0.11^a	0.11^a	0.10^a	0.11^a	0.10^a	0.11^a	0.085^a
	(0.020)	(0.022)	(0.022)	(0.022)	(0.021)	(0.025)	(0.022)	(0.024)
log roads		0.015^a	0.013^b	0.016^a	0.011^c	0.014^b	0.014^b	0.014^b
		(0.0057)	(0.0060)	(0.0052)	(0.0058)	(0.0060)	(0.0060)	(0.0056)
log income			0.46^a	0.48^a	0.42^b	0.50^b	0.46^a	
			(0.17)	(0.18)	(0.18)	(0.21)	(0.18)	
log^2 income			-0.15^a	-0.15^a	-0.14^b	-0.16^b	-0.15^a	
			(0.055)	(0.055)	(0.055)	(0.062)	(0.055)	
Network / shape				0.36^a	0.18^a	0.097^b		
				(0.11)	(0.062)	(0.037)		
Pop. growth 90-10							0.020	
							(0.037)	
share w. car								-0.18
								(0.19)
share w. motorcycle								0.31^a
								(0.081)
Observations	153	153	153	153	153	142	153	152
R-squared	0.51	0.52	0.56	0.58	0.58	0.59	0.56	0.58

Notes: OLS regressions with a constant in all columns. The dependent variable is the city fixed effect estimated in the specification reported in column 5 of table 4 where trips are weighted by how slow they are relative to their speed in absence of traffic ($\lambda = 0.2$). Robust standard errors in parentheses. a, b, c: significant at 1%, 5%, 10%. See the footnote of table 10 for further details about the explanatory variables.

variables. We now use within-city variation in population, area, and roads to avoid this problem and gain further insights about variation in mobility.

Specifically, we divide each city in our sample into concentric rings. Among other advantages, nearly all radial trips will pass through the same rings, regardless of route. We apply the following transformation of equation (2) which uses the location of trips within cities to estimate a mobility index for each ring within each city:

$$\log S_i = \alpha X_i' + \sum_r R_{rc(i)}\, share_{rc(i)}(i) + \epsilon_i\,, \tag{D1}$$

where $share_{rc(i)}(i)$ is the share of trip i which takes places within ring r of city c and R_{rc} is a mobility index for ring r of city c. We consider (up to) 5 rings around each city center: 0 to 2 kilometers, 2 to 5, 5 to 10, 10 to 15, and 15 and beyond. We compute each trip's share

Table C.3: Correlates of city mobility indices with alternative mobility indices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var.	Effect. sp.	Peak hrs.	Mean	Simp. FE	Amenity	cent.<5 km	Laspeyres	Paasche
log population	-0.15[a]	-0.18[a]	-0.17[a]	-0.17[a]	-0.17[a]	-0.17[a]	-0.17[a]	-0.17[a]
	(0.019)	(0.017)	(0.029)	(0.015)	(0.017)	(0.016)	(0.021)	(0.018)
log area	0.072[a]	0.12[a]	0.21[a]	0.15[a]	0.12[a]	0.13[a]	0.12[a]	0.15[a]
	(0.020)	(0.018)	(0.032)	(0.016)	(0.019)	(0.017)	(0.024)	(0.021)
log roads	0.014[a]	0.011[b]	-0.0066	0.012[a]	0.013[a]	0.013[a]	0.027[a]	0.011[b]
	(0.0048)	(0.0044)	(0.0097)	(0.0044)	(0.0047)	(0.0043)	(0.0062)	(0.0051)
log income	0.25[b]	0.26[b]	-0.032	0.20[b]	0.17	0.19[c]	0.25	0.15
	(0.10)	(0.11)	(0.23)	(0.094)	(0.12)	(0.10)	(0.18)	(0.12)
$\log^2$ income	-0.073[b]	-0.080[b]	0.0083	-0.054[c]	-0.049	-0.048	-0.088[c]	-0.046
	(0.032)	(0.034)	(0.064)	(0.029)	(0.036)	(0.032)	(0.052)	(0.037)
Observations	153	153	153	153	153	153	153	153
R-squared	0.56	0.59	0.28	0.55	0.56	0.54	0.42	0.47

Notes: OLS regressions with a constant in all columns. The dependent variable is the city fixed effect estimated using effective speed in column 1, only peak hour observations in column 2, a simpler speed regression in column 4, only amenity trips in column 5, only trips taking place within 5 kilometres from the center in column 6, our benchmark Laspeyres index in column 7, and a benchmark Paasche index in column 8. The dependent variable in column 3 is the log of a simple mean speed (length-weighted). Robust standard errors in parentheses. a, b, c: significant at 1%, 5%, 10%. Log population is constructed from the town population from the 2011 census. Log roads is log kilometers of primary roads within the city-light.

in each ring using information about the origin and destination. For instance, a radial trip that starts 9 kilometers from the center and finishes one kilometer from the center on the same side will receive a share of 12.5% (=1/8) for the first ring of 0 to 2 kilometers, 37.5% for the second ring, 50% for the third ring and 0% for the fourth and fifth ring. We estimate equation (D1) using as controls log trip length, time of day and day of week indicators in manner that is consistent with our baseline index.

In a second step, we can estimate the following regression

$$\hat{R}_{rc} = \kappa_r + \beta_c + \alpha X'_{rc} + \epsilon_i \,, \tag{D2}$$

where κ_r is a ring fixed effect, β_c is a city fixed effect, and X_{rc} is a vector of explanatory variables at the level of the city-ring. In our dataset, only land area, population, and roads are available separately by city-ring. Two caveats must be kept in mind. First, we winsorize the top and bottom 5% of city-ring effects before estimating equation (D2). This is because some cities barely enter an outer ring and therefore these city-rings have a tiny number of trips. Second, we also expect some equilibrium effects across rings as, for instance, population in nearby rings may affect mobility locally. Given the limited precision of our

Table D.1: Correlates of city mobility indices, rings analysis

	(1) Base	(2) No Step 1 Control	(3) < 5 km	(4) < 3 km	(5) Base	(6) <5m	(7) Peak	(8) Peak <5 km
log ring population	-0.084^a	-0.13^a	-0.086^a	-0.089^a	-0.085^a	-0.088^a	-0.084^a	-0.086^a
	(0.013)	(0.018)	(0.010)	(0.010)	(0.013)	(0.011)	(0.013)	(0.010)
log ring area	0.038^b	0.039	0.053^a	0.058^a	0.028	0.050^a	0.038^b	0.058^a
	(0.018)	(0.024)	(0.014)	(0.014)	(0.019)	(0.015)	(0.018)	(0.014)
log roads	-0.010	-0.0017	-0.017^b	-0.018^b			-0.010	-0.018^b
	(0.0095)	(0.013)	(0.0076)	(0.0076)			(0.0095)	(0.0076)
ring 2	0.14^a	0.30^a	0.084^a	0.062^a	0.10^b	0.049	0.14^a	0.077^a
	(0.019)	(0.026)	(0.015)	(0.015)	(0.043)	(0.035)	(0.019)	(0.015)
ring 3	0.20^a	0.44^a	0.095^a	0.058^a	0.19^a	0.082^a	0.20^a	0.086^a
	(0.025)	(0.033)	(0.020)	(0.020)	(0.036)	(0.029)	(0.025)	(0.020)
ring 4	0.19^a	0.46^a	0.041^c	0.00043	0.15^a	-0.010	0.19^a	0.032
	(0.031)	(0.042)	(0.024)	(0.025)	(0.045)	(0.036)	(0.031)	(0.025)
ring 5	0.20^a	0.53^a	0.040	-0.0067	0.15^a	0.051	0.20^a	0.026
	(0.042)	(0.057)	(0.034)	(0.034)	(0.055)	(0.044)	(0.042)	(0.034)
roads per ring	N	N	N	N	Y	Y	N	N
Observations	467	467	466	465	467	466	467	466
R-squared	0.56	0.72	0.47	0.42	0.57	0.48	0.56	0.45

Notes: OLS regressions with a city fixed effect and a ring fixed effect in all columns (145 cities in all regressions). The dependent variable is the city-ring fixed effect estimated as per equation (D2). Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Column 1 is our baseline estimation for which city-ring effects are estimated as described in the text. Column 2 considers city ring effects estimated with out trip controls in the first step. Columns 3 and 4 only consider trips with a length of less than 5 and 3 kilometers respectively. Columns 5 and 6 estimate separate roads effects for each ring. Columns 7 and 8 duplicate columns 1 and 3 but only consider peak-hour trips.

population data, detecting such effects may be out of reach here. This said, this rings approach may better capture rerouting within city as drivers substitute across routes.

We report results in table D.1. The coefficient on population is -0.084 in our baseline specification, and similar in the rest of the table.[54] We note that the population coefficients estimated in table D.1 are only about half those estimated in table 10. This may be because our measures of ring population are less precise. We also expect mobility within ring to be determined by population in neighboring rings.[55] Consistent with table 10, table D.1 also

[54] It is only when we do not control for trip characteristics in the first step in column 2, that we estimate a slightly larger coefficient in absolute value. This is likely because longer trips are faster and predominantly take place in outer rings where population is less dense.

[55] We experimented with specifications that also included population in neighboring rings. Estimated coefficients are generally small and insignificant.

reports small positive coefficients for area. On the other hand, the coefficient on roads is generally negative, though it is only significantly different from zero when we focus on the city centers. Although we do not report the details here, this negative coefficient is driven mainly by the central ring when roads effects are allowed to vary by ring in columns 5 and 6. Finally, table D.1 also reports that mobility is generally faster in outer rings, which confirms earlier results from section 4.