

# Consumption Access and the Spatial Concentration of Economic Activity: Evidence from Smartphone Data\*

Yuhei Miyauchi<sup>†</sup>

*Boston University*

Kentaro Nakajima<sup>‡</sup>

*Hitotsubashi University*

Stephen J. Redding<sup>§</sup>

*Princeton University, NBER and CEPR*

June 3, 2021

## Abstract

Using smartphone data for Japan, we show that non-commuting trips are frequent, more localized than commuting trips, strongly related to the availability of nontraded services, and occur along trip chains. Guided by these empirical findings, we develop a quantitative urban model that incorporates travel to work and travel to consume non-traded services. We use the gravity equation predictions of the model to estimate theoretically-consistent measures of travel access. We show that consumption access makes a substantial contribution to the observed variation in residents and land prices and the observed impact of the opening of a new subway line.

Keywords: Cities, Consumption Access, Economic Geography, Transportation

JEL Classification: O18, R12, R40

---

\*Thanks to Gabriel Ahlfeldt, Daniel Sturm and Gabriel Kreindler and conference and seminar participants for helpful comments. We are grateful to Takeshi Fukasawa and Peter Deffebach for excellent research assistance. The usual disclaimer applies. “Konzatsu-Tokei (R)” Data refers to people flow data constructed from individual location information sent from mobile phones under users’ consent, through applications provided by NTT DO-COMO, INC (including mapping application *Docomo Chizu NAVI*). Those data are processed collectively and statistically in order to conceal private information. Original location data is GPS data (latitude, longitude) sent every five minutes, and it does not include information to specify individual. The copyrights of all tables and figures presented in this document belong to ZENRIN DataCom CO., LTD. We also acknowledge Yaichi Aoshima at Hitotsubashi University for coordinating the project with ZENRIN DataCom Co., LTD.; Heiwa Nakajima Foundation and The Kajima Foundation for their financial support; CSIS at the University of Tokyo for the joint research support (Project No. 954); the Ministry of Land, Infrastructure, Transport, and Tourism and the Miyagi Prefecture for the access to the travel survey data.

<sup>†</sup>Dept. Economics, 270 Bay State Road, Boston, MA 02215. Tel: 1-617-353-5682. Email: miyauchi@bu.edu.

<sup>‡</sup>Institute of Innovation Research, 2-1 Naka, Kunitachi, Tokyo 186-8603, Japan. Tel: 81-42-580-8417. E-mail: nakajima.kentaro@gmail.com

<sup>§</sup>Dept. Economics and SPIA, JRR Building, Princeton, NJ 08544. Tel: 1-609-258-4016. Email: red-dings@princeton.edu.

# 1 Introduction

Understanding the spatial concentration of economic activity is one of the most central challenges in economics. Traditional theories of cities emphasize production decisions and the costs of workers commuting between their workplace and residence. However, much of the travel that occurs within urban areas is related not to commuting but rather to the consumption of nontraded services, such as trips to restaurants, coffee shops and bars, shopping centers, cultural venues, and other services. Although several scholars have emphasized the “consumer city,” two major challenges in this area are a limited ability to measure non-commuting trips and the absence of a widely-accepted theoretical model of travel for consumption. In this paper, we provide new theory and evidence on the role of consumption and workplace access in understanding the spatial distribution of economic activity. We combine smartphone data including high-frequency location information with spatially-disaggregated census data to measure commuting and non-commuting trips within the Greater Tokyo metropolitan area. Guided by our empirical findings, we develop a quantitative urban model that incorporates both workplace and consumption access. We use the model to evaluate the role of consumption access in explaining the observed spatial variation in economic activity. We show that incorporating consumption access is quantitatively relevant for evaluating the observed impact of a new subway line.

We first use our smartphone data to provide fine resolution evidence on travel within the Greater Tokyo metropolitan area. Our data come from a major smartphone mapping application in Japan (*Docomo Chizu NAVI*), which records the Geographical Positioning System (GPS) location of each device every 5 minutes. In July of 2019, the data covers about 545,000 users, with 1.4 billion data points. We measure each location visited by a user using a “stay,” which corresponds to no movement within 100 meters for 15 minutes. We designate each anonymized user’s home location as her most frequent location (defined by groups of geographically contiguous stays) and her work location as her second most frequent location. We allocate non-commuting trips to other locations into different types using spatially-disaggregated census data on employment by sector. We validate our smartphone commuting measures by showing that they are highly correlated with the measures from the official census data.

Having validated our smartphone data, we show that focusing solely on these commuting trips provides a misleading picture of travel patterns. First, we show that non-commuting trips are more frequent than commuting trips, so that concentrating solely on commuting trips substantially underestimates the amount of travel within urban areas. Second, we show that these non-commuting trips are closely related to the availability of nontraded services, which is consistent with our modelling of them as travel to consume non-traded services. Third, we find that non-commuting trips have destinations closer to home than commuting trips, with semi-

elasticities of travel flows to travel times that are larger in absolute value than those for commuting trips. Therefore, the spatial patterns of non-commuting trips are not well approximated by those for commuting trips. Fourth, we show that trip chains are a relevant feature of the data, in which non-commuting trips occur along the journey between home and work, highlighting the relevance of jointly modelling commuting and non-commuting trips.

We next develop quantitative theory of internal city structure that incorporates both commuting and consumption trips. We consider a city that consists of a discrete set of blocks that differ in productivity, amenities, supply of floor space and transport connections. Consumer preferences are defined over consumption of a traded good, a number of different types of nontraded services, and residential floor space. The traded good and nontraded services are produced using labor and commercial floor space. We assume that workers' location decisions are nested. First, workers observe idiosyncratic preferences for amenities in each location and choose where to live. Second, workers observe idiosyncratic productivities in each workplace and sector, and choose where to work. Third, workers observe idiosyncratic qualities for the non-traded services supplied by each location, and choose where to consume these non-traded services. Fourth, workers observe idiosyncratic taste shocks for each route to consume these non-traded services, and choose which of these routes to take (e.g. home-work-consume-home versus home-consume-home). When making each of these choices, workers take into account their expected access to surrounding locations. Population mobility implies that workers must obtain the same expected utility from all populated locations.

We show that the model implies extended gravity equations for commuting and non-commuting trips, which provide good approximations to the observed data. We use these extended gravity equations to estimate a theoretically-consistent measure of travel access. Intuitively, we use the observed trips in the data and the structure of the model to reveal the relative attractiveness of locations for employment and consumption. From the model's population mobility condition, we derive a sufficient statistic for the relative attractiveness of locations, which incorporates both the residential population share and the price of floor space. We show that this sufficient statistic for the relative attractiveness of locations can be decomposed into our measure of travel access and a residual for residential amenities. Comparing our model incorporating both consumption and workplace access to a special case capturing only workplace access, we find a substantially larger contribution of travel access once we take into account consumption access (56 percent compared to 37 percent), and a correspondingly smaller contribution from the residual of residential amenities (44 percent compared to 63 percent). Taken together, this pattern of results is consistent with the idea that much economic activity in urban areas is concentrated in the service sector, and that access to surrounding locations to consume these services is an important determinant of workers' choice of residence and workplace.

We show how the model can be used to undertake a counterfactual for a transport infrastructure improvement, such as the construction of a new subway line. In addition to the initial shares of commuting trips, the predictions of these counterfactuals now also depend on the initial shares of non-commuting trips. As a result, frameworks that focus solely on commuting trips generally underestimate the welfare gains from transport infrastructure improvements, because they undercount the number of passenger journeys that benefit from the reduction in travel costs. Furthermore, these frameworks generate different predictions for the impact of the new transport infrastructure on the spatial distribution of economic activity, because of the different bilateral patterns of commuting and non-commuting trips. We compare the model’s counterfactual predictions for the opening of a new subway line to the estimated impact in the observed data. We show that the model has predictive power for the observed data. We show that undercounting of travel from focusing on commuting trips leads to a substantial underestimate of the welfare gains from the new subway line.

Our paper is related to a number of different strands of research. First, our findings relate to recent research on endogenous amenities and social and spatial frictions within urban areas. Evidence of endogenous amenities has been provided in the context of spatial sorting ([Diamond 2016](#), [Almagro and Domínguez-Iino 2019](#) and [Samuels, Hausman, Cohen, and Sasson 2016](#)), gentrification and neighborhood change within cities ([Glaeser, Kolko, and Saiz 2001](#), [Couture, Dingel, Green, and Handbury 2019](#), [Hoelzlein 2020](#) and [Allen, Fuchs, Ganapati, Graziano, Madera, and Montoriol-Garriga 2020](#)), and industry clustering ([Leonardi and Moretti 2019](#)). Evidence that both spatial and social frictions matter for agents’ location decisions has been provided using restaurant choice data ([Couture 2016](#), [Davis, Dingel, Monras, and Morales 2019](#)), credit card data ([Agarwal, Jensen, and Monte 2020](#) and [Dolfen, Einav, Klenow, Klopach, Levin, Levin, and Best 2019](#)), travel surveys and ride sharing data ([Gorback 2020](#) and [Zárate 2020](#)) and cellphone data ([Couture, Dingel, Green, and Handbury 2019](#), [Athey, Ferguson, Gentzkow, and Schmidt 2018](#), [Kreindler and Miyauchi 2019](#), [Gupta, Kontokosta, and Van Nieuwerburgh 2020](#), [Büchel, Ehrlich, Puga, and Viladecans 2020](#) and [Atkin, Chen, and Popov 2021](#)). Relative to these existing studies, we provide high-frequency and spatially-disaggregated data on non-commuting trips, and develop a quantitative urban model for estimating workplace and consumption access.

Second, our work contributes to research on transport infrastructure and the location of economic activity. One strand of empirical research has used quasi-experimental variation on the impact of transport infrastructure improvements, including [Baum-Snow \(2007\)](#), [Michaels \(2008\)](#), [Duranton and Turner \(2012\)](#), [Faber \(2014\)](#), and [Storeygard \(2016\)](#). A second line of work has used quantitative spatial models to evaluate general equilibrium impacts of transport infrastructure investments, including [Anas and Liu \(2007\)](#), [Donaldson \(2018\)](#), [Donaldson and](#)

Hornbeck (2016), Heblich, Redding, and Sturm (2020), Tsivanidis (2018), Severen (2019), Balboni (2019), and Zárate (2020). While existing research emphasizes the costs of transporting goods and commuting costs, a key feature of our work is to highlight the role of the transport network in providing access to consume nontraded services.

Third, our research is related to recent research on the internal structure of cities, including including Ahlfeldt, Redding, Sturm, and Wolf (2015), Allen, Arkolakis, and Li (2017), Monte, Redding, and Rossi-Hansberg (2018), Tsivanidis (2018), and Dingel and Tintelnot (2020). All of these studies emphasize commuting and the separation of workplace and residence. In contrast, one of our main contributions is to highlight the importance of travel to consume nontraded services in shaping agents’ location decisions.

The remainder of the paper is structured as follows. Section 2 introduces our data. Section 3 presents reduced-form evidence on travel patterns. Section 4 introduces our theoretical framework that we use to rationalize these findings. Section 5 uses the model to quantify the relative importance of consumption and workplace access for explaining the spatial concentration of economic activity. Section 6 shows that incorporating consumption access is quantitatively relevant for evaluating the counterfactual impact of transport infrastructure improvements, such as the construction of a new subway line. Section 7 concludes.

## 2 Data Description

In this section, we introduce our smartphone data and the other data used in the quantitative analysis of the model. In Subsection 2.1, we explain how we use our smartphone data to identify home location, work location, commuting trips and non-commuting trips. In Subsection 2.2, we discuss the spatially-disaggregated economic census data by sector and location that we use to distinguish between different types of non-commuting trips, and discuss our data on land values and other location characteristics. In Subsection 2.3, we report validation checks of the commuting measures from our smartphone data using official census data on employment by residence, employment by workplace and bilateral commuting flows.

### 2.1 Smartphone GPS Data

Our main data source is one of the leading smartphone mapping applications in Japan: *Do-omo Chizu NAVI*. Upon installing this application, individuals are asked to give permission to share location information in an anonymized form. Conditional on this permission being given, the application collects the Geographical Positioning System (GPS) coordinates of each smartphone device every 5 minutes whenever the device is turned on (regardless of whether the application is being used). These “big data” provide an immense volume of high-frequency and

spatially-disaggregated information on the geographical movements of users throughout each day. For example for the month of July 2019 alone, the data include 1.4 billion data points on 545,000 users (about 0.5 percent of the Japanese population).<sup>1</sup>

The raw unstructured geo-coordinates are pre-processed by the cell phone operator: NTT Docomo Inc. to construct measures of “stays,” which correspond to distinct geographical locations visited by a user during a day. In particular, a stay corresponds to the set of geo-coordinates of a given user that are contiguous in time, whose first and last data points are more than 15 minutes apart, and whose geo-coordinates are all within 100 meters from the centroid of these points.<sup>2</sup> We have data on the sequence of stays of anonymized users with the necessary level of spatial aggregation to deidentify individuals. Our data comprise a randomly selected sample of 80 percent of users in Japan, where the randomization is again to deidentify individuals.

This pre-processing also categorizes all stays in each month into three categories of home, work and other locations for each anonymized user. “Home” location and “work” locations are defined as the centroid of the first and second most frequent locations of geographically contiguous stays, respectively. To ensure that these two locations do not correspond to different parts of a single property, we also require that the “work” location is more than 600 meters away from the “home” location. In particular, if the second most frequent location is within 600 meters of the “home” locations, we define the “work” location as the third most frequent location. To abstract from noise in geo-coordinate assignment, all stays within 500 meters of the home location are aggregated with the home location. Similarly, all stays within 500 meters of the work location are aggregated with the work location. We assign “Work” location as missing if the user appears in that location for less than 5 days per month, which applies for about 30 percent of users in our baseline sample during April 2019. These users primarily include those with limited number of data observations due to infrequent smartphone use, and also include irregular workers with unstable job locations and those who work at home.<sup>3</sup> In Subsection 2.3 below, we report validation checks on our classification of home and work locations using commuting data from the population census. Stays which are neither assigned as home or work are classified as “other.” We distinguish between different types of these “other” stays, such as visits to restaurants and stores, using spatially-disaggregated data on economic activity by sector and location from the economic census, as discussed further in Section 2.2 below.

---

<sup>1</sup>The mapping application does not send location data points if the smartphone does not sense movement, in which case it is likely that the user has not moved from the last reported location. For this reason, the data points are less frequent than 5 minutes intervals in practice.

<sup>2</sup>See Patent Number “JP 2013-89173 A” and “JP 2013-210969 A 2013.10.10” for the detailed proprietary algorithm. This algorithm involves processes to offset the potential noise in measuring GPS coordinates.

<sup>3</sup>In Section A.4 of our online appendix, we show that the devices with missing “work” locations have significantly fewer number of active days (even at home locations), and that the probability of assigning missing “work” locations is uncorrelated with the observable characteristics of the municipality of residence.

For most of our subsequent analysis, we focus on the sample of users in the month of April 2019 who have home and work locations in the Tokyo Metropolitan Area (which includes the four prefectures of Tokyo, Chiba, Kanagawa, and Saitama). To abstract from overnight trips, we focus on the sample of user-day observations for which the first and last stay of the day is the user’s home location.

## 2.2 Other Data Sources

We combine our smartphone data with a number of complementary data sources.

**Spatial units:** Data are available for the Tokyo metropolitan area at three main levels of spatial aggregation: (i) The four prefectures of Tokyo, Chiba, Kanagawa and Saitama; (ii) The 242 municipalities (excluding islands); (iii) The 9,956 *Oaza*. Each *Oaza* has an area of around 1.30 squared kilometers and an average 2011 population of around 3,600.

**Population Census:** We measure residential population, employment by workplace and bilateral commuting flows using the 2015 population census, which is conducted by the Statistics Bureau, Ministry of Internal Affairs and Communications every five years. Residential population and total employment are available at the finest level of spatial disaggregation of 250-meter grid cells. Bilateral commuting flows are reported between pairs of municipalities.

**Economic Census:** We use data from the 2016 Economic Census on total employment and the number of establishments by one-digit industry for each 500-meter grid cell in the Tokyo metropolitan area, the finest level of disaggregation from publicly available data. We also use data on total revenue and factor inputs that are available at the municipality level.

**Building Data:** We measure floor space in each city block using the Zmap-TOWN II Digital Building Map Data for 2008. This data set contains polygons for all buildings in Japan, with their precise geo-coordinates and information on building use and characteristics. We measure floor space using the number of stories and land area for each building.

**Land Price Data:** We measure the residential land price for each city block using the evaluated land price that is used for the calculation of property tax. We take a simple average of these values to construct the average land prices per unit of land at the *Oaza* or Municipality level.

**Travel Time Data:** We measure travel time by public transportation using the web-based route choice service, *Eki-spert* API.<sup>4</sup> *Eki-spert* API provides the minimum travel times between any pairs of coordinates using public transport, including suburban rail, subway, and bus, and walking. We use the extracted travel time data from October 2, 2020 (weekday timetable). We also construct car travel time using the Open Source Routing Machine (OSRM).

**Municipality Income Tax Base Data:** We measure the average income of the residents in each

---

<sup>4</sup>See <https://roote.ekispert.net/en> for details.

municipality using official data on the tax base for that municipality.

### 2.3 Validation of Smartphone Data Using Census Commuting Data

We now report an external validation exercise, in which we compare our measures of “home” location, “work location” and “commuting trips” from the smartphone data to official census data that are available at the municipality level. In the left panel of Figure 1, we display the log density of residents in each municipality in our smartphone data against log population density in the census data. As our smartphone data cover only a fraction of the total population, the levels of the two variables necessarily differ from one another. Nevertheless, we find a tight and approximately log linear relationship between them, with a slope coefficient of 0.923 (standard error 0.011) and a R-squared of 0.968. The coefficient is slightly less than one, indicating that the smartphone data has higher coverage in less dense areas. In the right panel of Figure 1, we show the log density of workers in each Tokyo municipality in our smartphone data against log employment density by workplace in the census data. Again, we find a close and approximately log linear relationship between them, with a slope coefficient of 0.996 (standard error 0.008) and a R-squared of 0.985.

In Section A.1 of the online appendix, we provide further evidence on the representativeness of our smartphone data by comparing the coverage by residence characteristics (income, age and distance to city center) and workplace characteristics (employment by industry and distance to city center). In Section A.2, we show that we find the same pattern of decline of bilateral commuting with distance in smartphone data and official census commuting data. In Section A.3, we show that home stays tend to occur during nighttime (outside 6am-9pm) and both work and other stays rise during the daytime (from 6am-9pm), providing additional internal validation of our home and work classification from smartphone data.

## 3 Reduced-Form Evidence

In this section, we provide reduced-form evidence on commuting and non-commuting trips that guides our theoretical model below. First, we show that non-commuting trips are more frequent than commuting trips, so that concentrating solely on commuting trips underestimates the amount of travel within urban areas. Second, we demonstrate that non-commuting trips are closely-related to the availability of non-traded services, which is consistent with these trips playing an important role in determining consumption access. Third, we show that non-commuting trips exhibit different spatial patterns from commuting trips, so that abstracting from non-commuting trips yields a misleading picture of bilateral travel patterns. Fourth, we provide



Figure 1: Representativeness of Smartphone Users



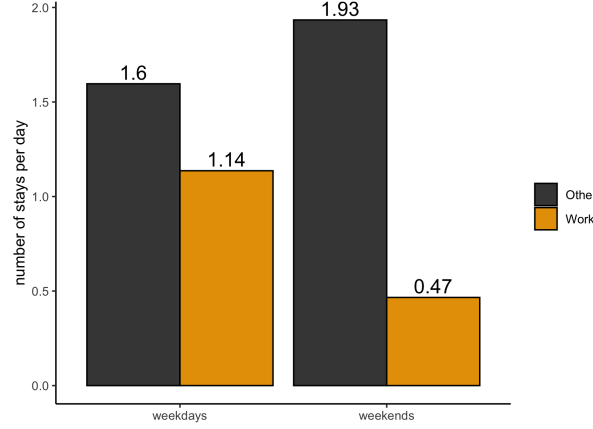
Note: Each dot is a municipality in the Tokyo metropolitan area. In the left panel, the vertical axis is the log of the number of smartphone users with a home location in the municipality divided by its geographic area, and the horizontal axis is the log of the number of residents in that municipality from the Population Census in 2011 divided by its geographic area. In the right panel, the vertical axis is the log of the number of smartphone users with a work location in the municipality divided by its geographic area, and the horizontal axis is the log of employment by workplace in that municipality from the Population Census in 2011 divided by its geographic area. The definitions of home and work in the smartphone data are discussed in the text of Subsection 2.1 above.

evidence of trip chains, in which non-commuting trips occur along the journey between home and work, highlighting the relevance of jointly modelling commuting and non-commuting trips.

**Fact 1. Non-commuting trips are pervasive.** In Figure 2, we display the average number of stays per day for work and non-work locations (excluding home locations) for our baseline sample of users with home and work locations in the Tokyo Metropolitan Area during April 2019. Note that the average number of work stays can be greater than one during weekdays, because workers can leave their workplace during the day and return there later the same day (e.g. after attending a lunch meeting outside their workplace). Similarly, the average number of work stays can be greater than zero at the weekend, because some workers can be employed during the weekend (e.g. in restaurants and stores). As apparent from the figure, even during weekdays, we find that non-commuting trips are more frequent than commuting trips, with an average of 1.6 non-work stays per day compared to 1.14 work stays per day. This pattern is magnified at weekends, with an average of 1.93 non-work stays per day compared to 0.47 work stays per day. These results are consistent with evidence from travel surveys, in which

commuting is only one of many reasons for travel.<sup>5</sup> A key advantage of our smartphone data is that they reveal bilateral patterns of travel at a fine level of spatial disaggregation within the urban area, and capture the sequence in which users travel between their home, work and consumption locations, as used to measure trip chains in our quantitative analysis of the model.

Figure 2: Frequency of Stays at Work and Other Locations (Excluding Home Locations)



Note: Average number of work and other stays per day for weekdays and weekends (excluding home stays) for our baseline sample users in the metropolitan area of Tokyo in April 2019. See Section 2 above for the definitions of home, work and other stays.

**Fact 2. Non-commuting trips are closely related to consumption.** We now show that non-commuting trips are closely related to consumption by combining our GPS smartphone data with spatially-disaggregated census data on employment by sector. In particular, we stochastically assign other stays (stays at neither home nor work locations) to different types based on the local economic activity undertaken at each geographical location, as captured by the share of service sectors in employment. For each  $500 \times 500$  meter grid cell in the Tokyo metropolitan area, we compute the employment share of each service sector in total service sector employment. We disaggregate service-sector employment into the following five categories: “Finance, Real Estate, Communication, and Professional”, “Wholesale and Retail”, “Accommodations, Eating, Drinking”, “Medical and Health Care”, and “Other Services”.<sup>6</sup> For each other stay in a given grid cell, we allocate that stay to these five categories probabilistically using their shares of service-sector employment. If no service-sector employment is observed in the grid cell, we allocate that other stay to the category “Z Others.”

<sup>5</sup>In Section A.5 of the online appendix, we show that this pattern of more frequent non-commuting stays than commuting stays holds in separate Japanese travel survey data, which are available for weekdays only.

<sup>6</sup>These sectors correspond to the one-digit classification of the Japan Standard Industrial Classification (JSIC), for which we have data available by  $500 \times 500$  meter grid cells. “Finance, Real Estate, Communication, and Professional” corresponds to sectors of G, J, K, L; “Wholesale and Retail” corresponds to I, “Accommodations, Eating, Drinking” corresponds to M, “Medical and Health Care” corresponds to P, and “Other Services” corresponds to Q.

Table 1: Frequency of Non-Commuting Trips and Service-Sector Employment Shares

Industry	Weekdays		Weekends		Employment Share in Service (%)	
	Stays / Day	Share (%)	Stays / Day	Share (%)	Total	Average (500m Grids)
GJKL finance realestate communication professional	0.23	14.3	0.21	10.7	11.9	23.2
I wholesale retail	0.69	43.4	0.91	46.1	32.0	28.7
M accomodations eating drinking	0.15	9.4	0.21	10.8	13.2	13.2
P medical welfare healthcare	0.23	14.2	0.27	13.7	18.7	15.2
Q other services	0.25	15.8	0.29	14.8	24.3	19.8
Z others	0.05	2.9	0.08	3.9		

Note: Average number of each type of other stay per day for weekdays and weekends (excluding home stays) for our baseline sample for the metropolitan area of Tokyo in April 2019. Other stays are allocated probabilistically to each category using the shares of these service sectors in total service-sector employment. The table also reports the share of each type of stay in the total number of other stays, the share of each service sector in total service-sector employment for the Tokyo metropolitan area, and the average share of each service sector in total service-sector employment across the  $500 \times 500$  meter grid cells. See Section 2 for the definitions of home, work and other stays.

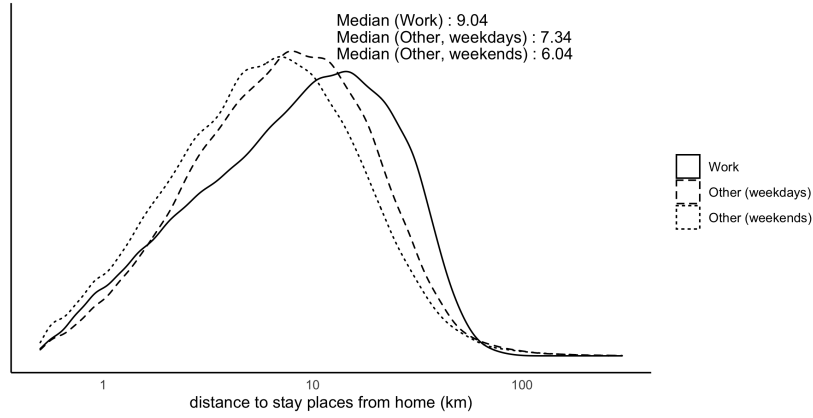
In Table 1, we report the average number of these different types of other stays per day during the working week and at weekends. We find that “Wholesale and Retail” stays are by far the most frequent, with an average of 0.69 per day on weekdays and 0.91 per day on weekends. To provide a point of comparison, we also report the share of each individual service sector in overall service-sector employment for the Tokyo metropolitan area as a whole (penultimate column) and the average share of each individual service sector in overall service-sector employment across the  $500 \times 500$  meter grid cells (final column). We find that “Wholesale and Retail” stays are substantially more frequent than would be implied by their shares of overall service-sector employment, accounting for 43.4 percent of weekday stays and 46.1 percent of weekend stays, compared to an aggregate employment share of 32.0 percent and an average employment share of 28.7 percent. This pattern of results implies that other stays are targeted towards locations with relatively high shares of the “Wholesale and Retail” sector in employment, which is consistent with these other stays capturing access to consumption opportunities. Although “Wholesale and Retail” stays are the most frequent, there is considerable variation in the composition of service-sector employment across the locations visited by users, with most sectors accounting for 10 percent or more of the total number of stays.<sup>7</sup>

As a check on our probabilistic assignment of other stays, Figure A.6.1 in Section A.6 of the online appendix displays the density of each type of other stay by hour and day, as a share of all stays for our baseline sample for the Tokyo metropolitan area in April 2019. We find that our probabilistic assignment captures the expected pattern of these different service-sector activities over the course of the week. First, we typically find a higher density of other stays during the middle of the day at weekends than during weekdays, which is in line with the fact that many of these services are consumed more intensively during leisure time. Second, we find that the peak

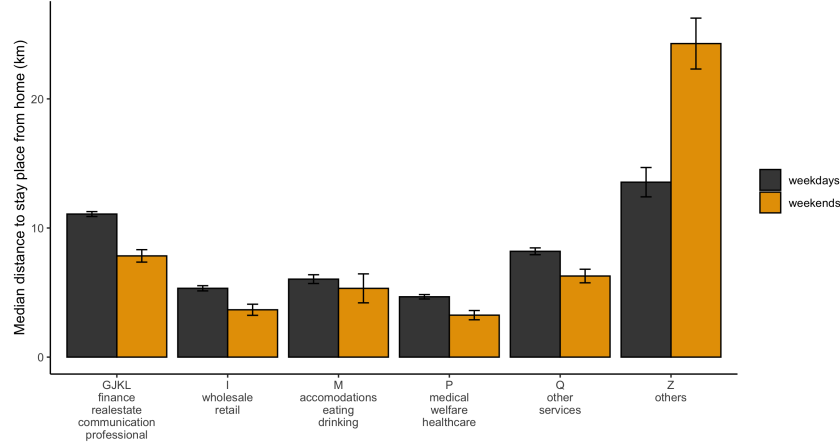
<sup>7</sup>Some non-commuting trips could be business-related (e.g., meetings). In Figure A.5.2 of the online appendix, we show that business-related trips are a minor fraction (20 percent) of all non-commuting weekday trips using separate travel survey data, where some of these business trips could involve consumption (e.g. lunches).

Figure 3: Distances of Commuting and Non-Communting Trips

(A) Distribution of Distances of Work and Other Stays from Home Locations



(B) Average Distances of Different Types of Other Stays from Home Locations



(0.75) (Note: Panel (A): Distributions of distance in kilometers of work locations from home location and of other stays from home locations during weekdays and weekends. Panel (B): Distributions of distance in kilometers for each type of other stay from home locations during weekdays and weekends. Results for our baseline sample of users in the Tokyo metropolitan area in April 2019.

densities of stays for “Wholesale and Retail” and “Accommodations, Eating, Drinking” occur at around 6pm on weekdays, corroborating the fact that these activities are typically concentrated after work during the week. Additionally, for “Accommodations, Eating, Drinking,” we find a smaller peak around noon on weekdays, capturing lunch time.

**Fact 3. Non-commuting trips are closer to home.** We now show that non-commuting trips exhibit different spatial patterns from commuting trips, such that observed bilateral commuting flows provide an incomplete picture of patterns of travel within urban areas. In Panel (A) of Figure 3, we display the distribution of distances from home locations to work locations and from home locations to other stays for our baseline sample of users in the Tokyo metropolitan area in the month of April 2019. We find that other stays are concentrated closer to home than

work stays, with average distances travelled of 7.34 and 9.04 kilometers respectively during weekdays, with an even larger difference in distances travelled at the weekend. In Panel (B) of Figure 3, we display the distribution of distances travelled for each type of other stay separately. We find that “Wholesale and Retail” and “Accommodations, Eating, Drinking” stays are concentrated closer to home than “Finance, Real Estate, Communication, and Professional” and “Other Services stays.” This clustering of other stays closer to home highlights the relevance of these non-commuting trips for residential location decisions. More generally, these differences in the geographical pattern of stays suggests that focusing on commuting trips yields an incomplete picture of bilateral patterns of travel.

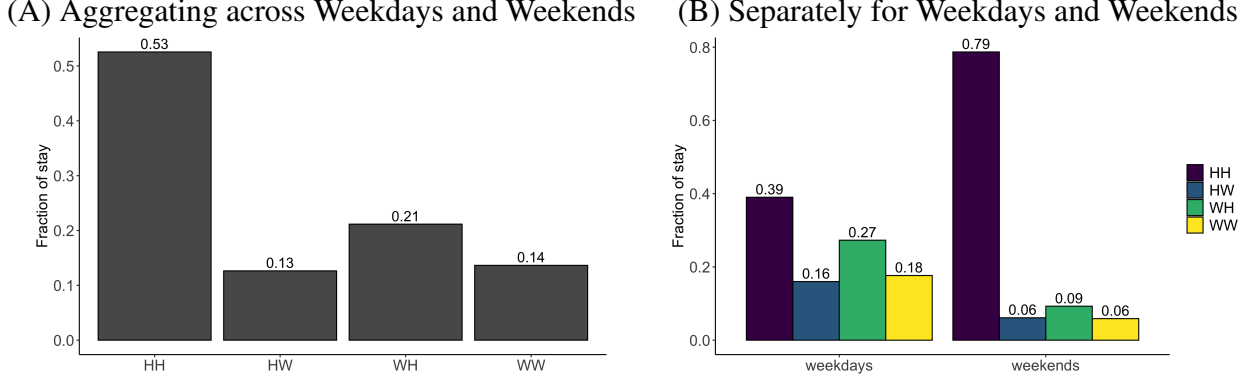
**Fact 4. Trip chains.** We now provide evidence of trip chains, in which non-commuting trips occur on the way from home and work. In Figure 4, we use the fact that in the smartphone data we observe the sequence of stays originating from a user’s home location and ending at a user’s home location (without going back home in between each stay), which we term “round trips.” Using this information, we divide all other stays that occur along such round trips into four mutually-exclusive categories: (i) HH stays, in which the other stay is part of a round trip that does not include the work location; (ii) HW stays, in which the other stay happens on the way from the home location to the work location; (iii) WH stays, in which the other stay happens on the way back from the work location to the home location; (iv) WW stays, in which the other stay happens in between two stays at the work location (e.g. a visit to a restaurant in the middle of the working day). Panel A shows the frequency of these four different types of other stays aggregating across weekdays and weekends, while Panel B shows their frequency for weekdays and weekends separately. We find that the majority of non-commuting trips occur separately from commuting trips (53 percent), which is driven primarily by weekends (79 percent) when users are significantly less likely to visit workplaces (Figure 2). Nevertheless, a substantial fraction of non-commuting trips (47 percent) occur as part of commuting trips (47 percent), highlighting the relevance of jointly modelling these two types of trips.

Taking the findings of this section as a whole, we have shown that non-commuting trips are frequent, are closely related to consumption, exhibit different spatial patterns from commuting trips, and can occur as part of trip chains. Each of these four features of our smartphone data guides our theoretical modelling of commuting and non-commuting trips in the next section.

## 4 Theoretical Framework

In this section, we develop our quantitative urban model of internal city structure that incorporates both commuting and non-commuting trips, where the derivations for all theoretical results in this section are reported in Section B of the online appendix.

Figure 4: Fractions of Different Types of Other Stays on Round Trips from Home



Note: Fractions of different types of other stays that occur as part of a round trip originating from a user's home location and ending at a user's home location (without going back home in between each stay); (i) on a round trip that does not include the work location (HH); (ii) on the way from home to work (HW); (iii) on the way back from work to home (WH); (iv) in between two stays at the work location (WW), such as a visit to restaurant in the middle of the working day; Panel A shows frequencies aggregating weekdays and weekends; Panel B shows frequencies separately for weekdays and weekends.

We consider a city (Tokyo) that is embedded in a larger economy (Japan). We consider both a closed-city specification (in which total city population is exogenous) and an open-city specification (in which total city population is endogenously determined by population mobility with the wider economy that offers a reservation level of utility  $\bar{U}$ ). The city consists of a discrete set of locations  $i, j, n \in N$  that differ in productivity, amenities, supply of floor space and transport connections. Utility is defined over consumption of a single traded good, a number of different types of non-traded services (e.g. restaurants, coffee shops, stores), and residential floor space use. Both the traded good and the non-traded services are produced with labor and commercial floor space according to constant returns to scale under conditions of perfect competition. Floor space is supplied by a competitive construction sector using land and capital according to a constant returns to scale construction technology.

A continuous measure of workers ( $\bar{L}$ ) choose a residence, a workplace and a set of locations to consume non-traded services in the city.<sup>8</sup> We assume the following timing or nesting structure for workers' location decisions. First, each worker observes her idiosyncratic preferences or amenities ( $b$ ) for each location within the city, and chooses her residence  $n$ . Second, given a choice of residence, each worker observes her idiosyncratic productivities ( $a$ ) for each workplace  $i$  and sector  $g$ , and chooses her sector and location of employment. Third, given a choice of residence and workplace, she observes idiosyncratic qualities ( $q$ ) for each type of non-traded service  $k$  available in each location  $j$ , and chooses her consumption loca-

<sup>8</sup>In the model, we assume a continuous measure of workers, which ensures that the expected values of variables equal their realized values. In our empirical analysis, we allow for granularity and a finite number of workers in our estimation (using the PPML estimator) and counterfactuals (using predicted shares from this estimation).

tion for each type of non-traded service. Fourth, given a choice of residence, workplace, and the set of consumption locations, she observes idiosyncratic shocks ( $\nu$ ) over different possible travel routes: home-consume-home, work-consume-work, home-consume-work-home, or home-work-consume-home. We choose this nesting structure because it permits a transparent decomposition of residents and land prices into the contribution of travel access and the residual of amenities, but the importance of consumption access is robust across other nesting structures. We also compare the predictions of our model with the special case abstracting from consumption trips, which corresponds to a conventional urban model, in which workers choose workplace and residence and consume only traded goods.

## 4.1 Preferences

The indirect utility for worker  $\omega$  who chooses residence  $n$ , works in location  $i$  and sector  $g \in K$ , and consumes non-traded service  $k \in K^S$  (where  $K^S \subset K$ ) in location  $j(k)$  using route  $r(k)$  is assumed to take the following Cobb-Douglas form:

$$\begin{aligned}
 U_{nig\{j(k)r(k)\}}(\omega) &= \left\{ B_n b_n(\omega) (P_n^T)^{-\alpha^T} Q_n^{-\alpha^H} \right\} \{ a_{i,g}(\omega) w_{i,g} \} \\
 &\times \left\{ \prod_{k \in K^S} [P_{j(k)}^S / (q_{j(k)}(\omega))]^{-\alpha_k^S} \right\} \left\{ d_{ni\{j(k)r(k)\}} \prod_{k \in K^S} \nu_{r(k)}(\omega) \right\} \\
 0 &< \alpha^T, \alpha^H, \alpha_k^S < 1, \quad \alpha^T + \alpha^H + \sum_{k \in K^S} \alpha_k^S = 1,
 \end{aligned} \tag{1}$$

where we use the notation  $j(k)$  to indicate that that non-traded service  $k$  is consumed in a single location  $j$  that is an implicit function of the type of non-traded service  $k$ ;  $r(k) \in \mathbb{R} \equiv \{HH, WW, HW, WH\}$  indicates the “route” choice of whether to visit consumption locations from home ( $HH$ ), from work ( $WW$ ), on the way from home to work ( $HW$ ), or on the way from work to home ( $WH$ ) for each non-traded service  $k$ ;  $K^S \subset K$  is the subset of sectors that are non-traded; the first term in brackets captures a residence component of utility; the second term in brackets corresponds to a workplace component; the third term in brackets reflects a non-traded services component; the fourth term in brackets reflects a travel cost component.

The first, residence component includes amenities ( $B_n$ ) that are common for all workers in residence  $n$ ; the idiosyncratic amenity draw for residence  $n$  for worker  $\omega$  ( $b_n(\omega)$ ); the price of the traded good ( $P_n^T$ ); and the price of residential floor space ( $Q_n$ ). We allow the common amenities ( $B_n$ ) to be either exogenous or endogenous to the surrounding concentration of economic activity in the presence of agglomeration forces, as discussed further below. The second, workplace component comprises the wage per efficiency unit in sector  $g$  in workplace  $i$  ( $w_{i,g}$ ) and the idiosyncratic draw for productivity or efficiency units of labor for worker  $\omega$  in sector

$g$  in workplace  $i$  ( $a_{i,g}(\omega)$ ).<sup>9</sup> The third, non-traded services component depends on the price of the non-traded service  $k$  in the location  $j(k)$  where it is supplied ( $P_{j(k)}^S$  for  $k \in K^S$ ) and the idiosyncratic draw for quality for that service in that location ( $q_{j(k)}(\omega)$  for  $k \in K^S$ ). The fourth component captures the iceberg travel cost for each combination of residence, workplace, consumption locations and routes ( $d_{ni\{j(k)r(k)\}}$ ) and the idiosyncratic draw for route preference for each non-traded sector ( $\nu_{r(k)}(\omega)$  for  $k \in K^S$ ).

To capture trip chains, we model the iceberg travel cost for each combination of residence  $n$ , workplace  $i$ , consumption location  $j(k)$  and route  $r(k)$  ( $d_{ni\{j(k)r(k)\}}$ ) as follows:

$$d_{ni\{j(k)r(k)\}} = \exp(-\kappa^W \tau_{ni}^W) \prod_{k \in K^S} \exp(-\kappa_k^S \tau_{nij(k)r(k)}^S). \quad (2)$$

The first term before the product sign captures the cost of commuting from residence  $n$  to workplace  $i$  without any detour to consume non-traded services, which depends on travel time ( $\tau_{ni}^W$ ) and the commuting cost parameter ( $\kappa^W$ ), where overall commuting travel time is the sum of that in each direction:

$$\tau_{ni}^W = \tau_{ni} + \tau_{in}. \quad (3)$$

The second term in equation (2) captures the additional travel costs involved in consuming each type of non-traded service  $k$  in location  $j(k)$  by the route  $r(k)$ , which depends on the additional travel time involved ( $\tau_{nij(k)r(k)}^S$ ) and the consumption travel cost parameter ( $\kappa_k^S$ ). This additional travel time depends on the route taken: whether the worker visits consumption location  $j(k)$  from home ( $r(k) = HH$ ), from work ( $WW$ ), on the way from home to work ( $HW$ ), or on the way from work to home ( $WH$ ):

$$\begin{aligned} \tau_{nij(k)HH}^S &= \tau_{nj} + \tau_{jn}, & \tau_{nij(k)WW}^S &= \tau_{ij} + \tau_{ji}, \\ \tau_{nij(k)HW}^S &= \tau_{nj} + \tau_{ji} - \tau_{ni}, & \tau_{nij(k)WH}^S &= \tau_{ij} + \tau_{jn} - \tau_{in}, \end{aligned} \quad (4)$$

where the negative terms on the second line above reflects the fact that the worker travels indirectly between residence  $n$  and workplace  $i$  via consumption location  $j$  on one leg of her journey between home and work, and hence does not incur the direct travel time between residence  $n$  and workplace  $i$  for that leg.<sup>10</sup>

We make the conventional assumption in the location choice literature following [McFadden \(1974\)](#) that the idiosyncratic shocks are drawn from an extreme value distribution. In particular,

<sup>9</sup>Although we model the workplace idiosyncratic draw as a productivity draw, there is a closely-related formulation in which it is instead modelled as an amenity draw.

<sup>10</sup>While we capture the relative importance of the consumption of non-traded services using the Cobb-Douglas expenditure shares ( $\alpha_k^S$ ), the frequency of trips can also differ across non-traded sectors, as shown in Figure 1. In Section C.1 of the online appendix, we explicitly incorporate this additional type of heterogeneity and show that the model is isomorphic up to a reinterpretation of the parameters  $\kappa_k^S$ . Therefore, all of our counterfactual results are unaffected by this extension of the model except for the interpretation of the estimated  $\kappa_k^S$ .



amenities ( $b$ ), productivity ( $a$ ), quality ( $q$ ), route preferences ( $\nu$ ) for worker  $\omega$ , residence  $n$ , workplace  $i$ , consumption location  $j(k)$  and route  $r(k)$  for non-traded service  $k$  are drawn from independent Fréchet distributions:

$$\begin{aligned} G_n^B(b) &= \exp\left(-T_n^B b^{-\theta^B}\right), & T_n^B > 0, \theta^B > 1, \\ G_{i,g}^W(a) &= \exp\left(-T_{i,g}^W a^{-\theta^W}\right), & T_{i,g}^W > 0, \theta^W > 1, \\ G_{j(k)}^S(q) &= \exp\left(-T_{j(k)}^S q^{-\theta_k^S}\right), & T_{j(k)}^S > 0, \theta_k^S > 1, k \in K^S, \\ G_{r(k)}^R(\nu) &= \exp\left(-T_{r(k)}^R \nu^{-\theta_k^R}\right), & T_{r(k)}^R > 0, \theta_k^R > 1, k \in K^S. \end{aligned} \quad (5)$$

where the scale parameters  $\{T_n^B, T_{i,g}^W, T_{j(k)}^S, T_{r(k)}^R\}$  control the average draws and the shape parameters  $\{\theta^B, \theta^W, \theta_k^S, \theta_k^R\}$  regulate the dispersions of amenities, productivity, quality and route preferences, respectively. The smaller these dispersion parameters, the greater the heterogeneity in idiosyncratic draws, and the less responsive worker decisions to economic variables.<sup>11</sup>

Using our assumption about the timing or nesting structure, the worker location choice problem is recursive and can be solved backwards. First, for given a choice of residence, workplace and sector, and consumption location for each non-traded service, we characterize the probability that a worker chooses each route for each non-traded sector (whether to visit consumption locations from home, from work, or in-between). Second, for given a choice of residence, workplace and sector, we characterize the probability that a worker chooses each consumption location in each non-traded sector, taking into account the expected travel cost for consumption trips. Third, for given a choice of residence, we characterize the probability that a worker chooses each workplace and sector, taking into account expected consumption access for that workplace and sector. Fourth, we characterize the probability that a worker chooses each residence, taking into account its expected travel access for both commuting and consumption.

## 4.2 Route Choices

We begin with the worker's choice of route for each non-traded service sector  $k$ . Conditional on her residence  $n$ , workplace  $i$ , and consumption location  $j(k)$ , she chooses whether to visit consumption location  $j(k)$  from home ( $r(k) = HH$ ), from work ( $WW$ ), on the way from home to work ( $HW$ ), or on the way from work to home ( $WH$ ). Given the indirect utility (1) and the specification of the travel cost (2), the component of the utility that depends on the route  $r(k)$

<sup>11</sup> Although we assume independent Fréchet distributions for amenities, productivity and quality, some locations can have high expected values for all these idiosyncratic shocks if they have high values for  $T_n^B$ ,  $T_{i,g}^W$ ,  $T_{j(k)}^S$  and  $T_{r(k)}^R$ . Additionally, correlations between the shocks can be introduced using a multivariate Fréchet distribution, as in [Hsieh, Hurst, Jones, and Klenow \(2019\)](#).

for non-traded service  $k$  is given by:

$$\delta_{nij(k)r(k)}(\omega) = \exp(-\kappa_k^S \tau_{nij(k)r(k)}^S) \nu_{r(k)}(\omega). \quad (6)$$

where the first component is the route-specific travel cost and the second component is the idiosyncratic route preference. Under our assumption of independent route draws  $\nu_{r(k)}(\omega)$  across each non-traded sector  $k$ , each worker chooses the route  $r(k)$  that maximizes  $\delta_{nij(k)r(k)}(\omega)$  independently for each sector  $k$ .

Using our independent extreme value assumption, the route choice probability is characterized by a logit form. In particular, the probability that a worker living in residence  $n$  and employed in workplace  $i$  consuming non-traded service  $k$  in location  $j(k)$  chooses the route  $r(k)$  ( $\lambda_{r(k)|nij(k)}^R$ ) is:

$$\lambda_{r(k)|nij(k)}^R = \frac{T_{r(k)}^R \exp(-\theta_k^R \kappa_k^S \tau_{nij(k)r(k)}^S)}{\sum_{r' \in \mathbb{R}} T_{r'(k)}^R \exp(-\theta_k^R \kappa_k^S \tau_{nij(k)r'(k)}^S)}. \quad (7)$$

Using the properties of the extreme value distribution, we can also compute the expected contribution to utility from the travel cost from consumption trips

$$d_{nij(k)}^S = \mathbb{E}_{nij(k)} [\delta_{nij(k)r(k)}(\omega)] = \vartheta_k^R \left[ \sum_{r' \in \mathbb{R}} T_{r'(k)}^R \exp(-\theta_k^R \kappa_k^S \tau_{nij(k)r'(k)}^S) \right]^{\frac{1}{\theta_k^R}} \quad (8)$$

where  $\vartheta_k^R \equiv \Gamma\left(\frac{\theta_k^R - 1}{\theta_k^R}\right)$  and  $\Gamma(\cdot)$  is the Gamma function.

### 4.3 Consumption Choices

We next describe the worker's decision of where to consume each type of non-traded service, given these expected travel costs. Conditional on living in residence  $n$  and being employed in workplace  $i$ , each worker chooses a consumption location  $j(k)$  for each non-traded service  $k$ , after observing her idiosyncratic draws for the quality of non-traded services ( $d$ ), but before observing her idiosyncratic route preferences ( $\nu$ ). Therefore, each worker chooses the consumption location  $j(k)$  that maximizes the contribution to indirect utility (1) from consuming that non-traded service  $k$ , taking into account the expected travel costs across alternative routes:

$$\gamma_{nij(k)}(\omega) = [P_{j(k)}^S / (q_{j(k)}(\omega))]^{-\alpha_k^S} d_{nij(k)}^S, \quad k \in K^S. \quad (9)$$

where  $d_{nij(k)}^S$  is the expected travel cost across these alternative routes from equation (8) above.<sup>12</sup>

---

<sup>12</sup>Although for simplicity we assume that workers choose a single consumption location for each non-traded service, it is straightforward to extend the model to incorporate multiple consumption locations, by allowing workers to make multiple discrete choices for each non-traded service.

Using our extreme value assumption, the probability that a worker living in residence  $n$  and employed in workplace  $i$  consumes non-traded service  $k$  in location  $j(k)$  ( $\lambda_{j(k)|ni}^S$ ) is:

$$\lambda_{j(k)|ni}^S = \frac{T_{j(k)}^S (P_{j(k)}^S)^{-\theta_k^S} (d_{nij(k)}^S)^{\frac{\theta_k^S}{\alpha_k^S}}}{\sum_{\ell \in N} T_{\ell(k)}^S (P_{\ell(k)}^S)^{-\theta_k^S} (d_{ni\ell(k)}^S)^{\frac{\theta_k^S}{\alpha_k^S}}}, \quad k \in K^S, \quad (10)$$

which we term the conditional consumption probability, since it is computed conditional on residence  $n$  and workplace  $i$ . This probability depends on destination characteristics (the price of non-traded services  $P_{j(k)}^S$  and their average quality  $T_{j(k)}^S$  in the numerator); expected travel costs (as determined by  $d_{nij(k)}^S$  in the numerator); and origin (residence and workplace) characteristics (as captured by the expected-travel-cost weighted average of destination characteristics in the denominator). Importantly, the frequency of consumption trips for each destination  $j(k)$  and non-traded service  $k$  depends on both the worker's residence  $n$  and her workplace  $i$ , because she can travel to consume non-traded services from either of these locations.

Using the properties of the extreme value distribution, we can also compute the expected contribution to utility from consuming non-traded service  $k$ , conditional on living in residence  $n$  and being employed in workplace  $i$ . This expectation for residence  $n$  and workplace  $i$  corresponds to a measure of consumption access for non-traded service  $k$ , and depends on the travel-time weighed average of destination characteristics:

$$\mathbb{S}_{nik} \equiv \mathbb{E}_{nik} [\gamma_{nij(k)}] = \vartheta_k^S \left[ \sum_{\ell \in N} T_{\ell(k)}^S (P_{\ell(k)}^S)^{-\theta_k^S} (d_{ni\ell(k)}^S)^{\frac{\theta_k^S}{\alpha_k^S}} \right]^{\frac{\alpha_k^S}{\theta_k^S}}, \quad k \in K^S. \quad (11)$$

where  $\vartheta_k^S \equiv \Gamma \left( \frac{(\theta_k^S/\alpha_k^S)-1}{(\theta_k^S/\alpha_k^S)} \right)$  and  $\Gamma(\cdot)$  is the Gamma function.

Noting that idiosyncratic quality is independently distributed across non-traded sectors, we can also compute the expected overall contribution to utility from non-traded services:

$$\mathbb{S}_{ni} \equiv \prod_{k \in K^S} \mathbb{S}_{nik} = \prod_{k \in K^S} \vartheta_k^S \left[ \sum_{\ell \in N} T_{\ell(k)}^S (P_{\ell(k)}^S)^{-\theta_k^S} (d_{ni\ell(k)}^S)^{\frac{\theta_k^S}{\alpha_k^S}} \right]^{\frac{\alpha_k^S}{\theta_k^S}}. \quad (12)$$

## 4.4 Workplace Choice

We next turn to the worker's choice of workplace, given consumption access. In particular, conditional on living in residence  $n$ , each worker chooses the workplace  $i$  and sector  $g \in K$  that offers the highest utility, taking into account the wage per efficiency unit ( $w_{i,g}$ ), the idiosyncratic draw for productivity ( $a_{i,g}(\omega)$ ), commuting costs ( $d_{ni}^W$ ), and expected consumption access ( $\mathbb{S}_{ni}$ ):

$$v_{ni,g}(\omega) = w_{i,g} a_{i,g}(\omega) d_{ni}^W \mathbb{S}_{ni}. \quad (13)$$

where  $d_{ni}^W \equiv \exp(-\kappa^W \tau_{ni}^W)$  is commuting travel cost from equation (2).

Using our independent extreme value assumption for idiosyncratic productivity, the model also implies a gravity equation for bilateral commuting, such that the probability that a worker in residence  $n$  commutes to workplace  $i$  in sector  $g$  ( $\lambda_{ig|n}^W$ ) is as follows:

$$\lambda_{ig|n}^W = \frac{T_{i,g}^W w_{i,g}^{\theta^W} (d_{ni}^W)^{\theta^W} (\mathbb{S}_{ni})^{\theta^W}}{\sum_{\ell \in N} \sum_{m \in K} T_{\ell,m}^W w_{\ell,m}^{\theta^W} (d_{n\ell}^W)^{\theta^W} (\mathbb{S}_{n\ell})^{\theta^W}}, \quad (14)$$

which we term the conditional commuting probability, since it is computed conditional on living in residence  $n$ . Bilateral commuting flows also depend on destination characteristics (the wage  $w_{i,g}$ , average efficiency units  $T_{i,g}^W$  and consumption access  $\mathbb{S}_{ni}$  in the numerator); bilateral travel costs (as captured by  $d_{ni}^W$  in the numerator); and origin characteristics (as captured by the travel-cost weighted average of destination characteristics across sectors in the denominator). Aggregating across the different sectors  $k \in K$ , we also obtain the overall commuting probability between residence  $n$  and workplace  $i$ :

$$\lambda_{i|n}^W = \sum_{g \in K} \lambda_{ig|n}^W. \quad (15)$$

Using the properties of the extreme value distribution, we can also compute an overall measure of travel access for residence  $n$  ( $\mathbb{A}_n$ ), which is a weighted average of the characteristics of each workplace  $i$ , including consumption access ( $\mathbb{S}_{ni}$ ):

$$\mathbb{A}_n = \mathbb{E}_n[v_{ni,g}] = \vartheta^W \left[ \sum_{\ell \in N} \sum_{m \in K} T_{\ell,m}^W w_{\ell,m}^{\theta^W} (d_{n\ell}^W)^{\theta^W} (\mathbb{S}_{n\ell})^{\theta^W} \right]^{\frac{1}{\theta^W}}, \quad (16)$$

where  $\vartheta^W \equiv \Gamma\left(\frac{\theta^W - 1}{\theta^W}\right)$  and  $\Gamma(\cdot)$  is the Gamma function.

## 4.5 Residence Choice

Having characterized a worker's consumption and workplace choices conditional on her residence, we now turn to her residence choice. Each worker chooses her residence after observing her idiosyncratic draws for amenities ( $b$ ), but before observing her idiosyncratic draws for productivity ( $a$ ), the quality of non-traded services ( $q$ ), and route preferences ( $\nu$ ). Therefore, each worker  $\omega$  chooses the residence  $n$  that offers her the highest utility given her idiosyncratic amenity draws ( $b_n(\omega)$ ), expected travel access ( $\mathbb{A}_n$ ), and other residence characteristics (the price of floor space ( $Q_n$ ), the price of the traded good ( $P_n^T$ ) and common amenities ( $B_n$ )):

$$U_n(\omega) = B_n b_n(\omega) (P_n^T)^{-\alpha^T} Q_n^{-\alpha^H} \mathbb{A}_n,$$

Using our extreme value assumption for idiosyncratic amenities, the probability that each worker chooses residence  $n$  ( $\lambda_n^B$ ) depends on its relative attractiveness in terms of travel access ( $\mathbb{A}_n$ ), and residential characteristics ( $B_n$ ,  $P_{T,n}$  and  $Q_n$ ):

$$\lambda_n^B = \frac{T_n^B B_n^{\theta^B} \mathbb{A}_n^{\theta^B} (P_n^T)^{-\alpha^T \theta^B} Q_n^{-\alpha^H \theta^B}}{\sum_{\ell \in N} T_\ell^B B_\ell^{\theta^B} \mathbb{A}_\ell^{\theta^B} (P_\ell^T)^{-\alpha^T \theta^B} Q_\ell^{-\alpha^H \theta^B}}. \quad (17)$$

Taking expectations over idiosyncratic amenities, expected utility from living in the city depends on the travel access and other residential characteristics of all locations within the city:

$$\mathbb{E}[u] = \vartheta^B \left[ \sum_{\ell \in N} T_\ell^B B_\ell^{\theta^B} \mathbb{A}_\ell^{\theta^B} (P_\ell^T)^{-\alpha^T \theta^B} Q_\ell^{-\alpha^H \theta^B} \right]^{\frac{1}{\theta^B}}. \quad (18)$$

where  $\vartheta^B \equiv \Gamma\left(\frac{\theta^B-1}{\theta^B}\right)$  and  $\Gamma(\cdot)$  is the Gamma function.

In Section 5.2, we use these residential choice probabilities to decompose the observed variation in economic activity into the contributions of travel access and a residual for amenities, without taking a stand on production technology and market structure in the traded and non-traded sectors. As a result, this quantitative analysis holds in an entire class of quantitative urban models, with different specifications for production technology and market structure.

Expected income in residence  $n$  ( $E_n$ ) in turn depends on the overall commuting probabilities ( $\lambda_{i|n}^W$ ) and expected income conditional on commuting from residence  $n$  to workplace  $i$  ( $E_{ni}$ ):

$$E_n = \sum_{i \in N} \lambda_{i|n}^W E_{ni}, \quad (19)$$

where  $E_{ni}$  depends on both wages and expected worker idiosyncratic productivity.

## 4.6 Production

When we undertake counterfactuals in Section 6, we do need to take a stand on a specific production technology and market structure. In particular, we assume that both the traded good and non-traded services are produced using labor and commercial floor space according a constant returns to scale technology. We assume for simplicity that this production technology is Cobb-Douglas and that production occurs under conditions of perfect competition.<sup>13</sup> Together these assumptions imply that profits are zero in each location with positive production:

$$\begin{aligned} P_i^T &= \frac{1}{A_{i,k}} w_{i,k}^{\beta^T} Q_i^{1-\beta^T}, & 0 < \beta^T < 1, & \quad k \in K/K^S, \\ P_{i(k)}^S &= \frac{1}{A_{i,k}} w_{i,k}^{\beta^S} Q_i^{1-\beta^S}, & 0 < \beta^S < 1, & \quad k \in K^S, \end{aligned} \quad (20)$$

<sup>13</sup>In Section C.2 of the online appendix, we show that our specification is isomorphic to a model of monopolistic competition under free entry, once we allow for agglomeration forces (equation (25) below).

where  $A_{i,k}$  is productivity in location  $i$  in sector  $k$ .

We allow productivity ( $A_{i,k}$ ) to be either exogenous or endogenous to the surrounding concentration of economic activity because of agglomeration forces, as discussed further below. We assume no-arbitrage between residential and commercial floor space, and across the different sectors in which commercial floor space is used, such that there is a single price for floor space within each location ( $Q_i$ ). In general, the wage per efficiency unit ( $w_{i,k}$ ) differs across both sectors and locations, because workers draw efficiency units for each sector and location pair, and hence each sector and location pair faces an upward-sloping supply function for effective units of labor. Finally, we assume that the traded good is costlessly traded within the city and wider economy and choose it as our numeraire, such that:

$$P_i^T = 1 \quad \forall i \in N. \quad (21)$$

## 4.7 Market Clearing

The price for non-traded service  $k$  in each location  $j$  ( $P_{j(k)}^S$  for  $k \in K^S$ ) is determined by market clearing, which equates revenue and expenditure for that sector  $k$  and location  $j$ :

$$P_{j(k)}^S A_{j,k} \left( \frac{\tilde{L}_{j,k}}{\beta^S} \right)^{\beta^S} \left( \frac{H_{j,k}}{1 - \beta^S} \right)^{1 - \beta^S} = \alpha_k^S \sum_{n \in N} R_n \sum_{i \in N} \lambda_{j(k)|ni}^S \lambda_{i|n}^W E_{ni}, \quad k \in K^S, \quad (22)$$

where expenditure on the right-hand side equals the sum across locations of workers travelling to consume non-traded service  $k$  in location  $j$ ;  $\tilde{L}_{j,k}$  is the labor input adjusted for expected idiosyncratic worker productivity in sector  $k$  in location  $j$ ;  $R_n$  is the measure of residents in location  $n$ ; and recall that  $\lambda_{j(k)|ni}^S$  is the conditional consumption probability and  $E_{ni}$  is expected worker income for residence  $n$  and workplace  $i$ .

Labor market clearing equates the measure of workers employed in workplace  $j$  in sector  $k$  to the measure of workers commuting from all residences  $n$  to that workplace  $j$  in sector  $k$ :

$$L_{j,k} = \sum_{n \in N} \lambda_{jk|n}^W R_n, \quad k \in K, \quad (23)$$

where we use  $L_{j,k}$  without a tilde to denote the measure of workers without adjusting for effective units of labor; and recall that  $\lambda_{jk|n}^W$  is the conditional commuting probability.

Land market clearing equates the demand for residential floor space ( $H_{i,U}$ ) plus commercial floor space in each sector ( $H_{i,k}$ ) to the total supply of floor space ( $H_i$ ):

$$H_i = H_{i,U} + \sum_{k \in K} H_{i,k}. \quad (24)$$

## 4.8 General Equilibrium

We begin by considering the case in which productivity ( $A_{i,k}$ ), amenities ( $B_i$ ) and the supply of floor space ( $H_i$ ) are exogenously determined. The general equilibrium of the model is referenced by the price for floor space in each location ( $Q_i$ ), the wage in each sector and location ( $w_{i,k}$ ), the price of the non-traded good in each service sector and location ( $P_{i(k)}^S$ ), the route choice probabilities ( $\lambda_{r(k)|nij(k)}^R$ ), the conditional consumption probabilities ( $\lambda_{j(k)|ni}^S$ ), the conditional commuting probabilities ( $\lambda_{ik|n}^W$ ), the residence probabilities ( $\lambda_n^B$ ), and the total measure of workers living in the city ( $\bar{L}$ ), where we focus on the open-city specification, in which the total measure of workers is endogenously determined by population mobility with the wider economy. These eight equilibrium variables are determined by the system of eight equations given by the land market clearing condition for each location (24), the labor market clearing condition for each location (23), the non-traded goods market clearing condition for each location and service sector (22), the route choice probabilities (7), the conditional consumption probabilities (10), the conditional commuting probabilities (14), the residence probabilities (17), and the population mobility condition that equates expected utility (18) to the reservation utility in the wider economy ( $\bar{U}$ ).

## 4.9 Agglomeration Forces and Endogenous Floor Space

We next extend the analysis to allow productivity and amenities to be endogenous to the surrounding concentration of economic activity through agglomeration forces and to allow for an endogenous supply of floor space. In both the traded and non-traded sector, we allow productivity ( $A_{i,k}$ ) to depend on production fundamentals and production externalities. Production fundamentals ( $a_{i,k}$ ) capture features of physical geography that make a location more or less productive independently of neighboring economic activity (e.g. access to natural water). Production externalities capture productivity benefits from the density of employment across all sectors ( $L_i/K_i$ ), where employment density is measured per unit of geographical land area:<sup>14</sup>

$$A_{i,k} = a_{i,k} \left( \frac{L_i}{K_i} \right)^{\eta^W} \quad (25)$$

where  $L_i = \sum_{k \in K} L_{i,k}$  is the total employment in location  $i$ , and  $\eta^W$  parameters the strength of production externalities, which we assume to be the same across all sectors.

In addition to the pecuniary externalities from consumption access, we allow residential amenities ( $B_n$ ) to depend on residential fundamentals and residential externalities. Residential fundamentals ( $b_n$ ) capture features of physical geography that make a location a more or less

---

<sup>14</sup>We assume for simplicity that production externalities depend solely on a location's own employment density, but we can also allow for the case in which there are spillovers of these production externalities across locations.

attractive place to live independently of neighboring economic activity (e.g. green areas). Residential externalities capture the effects of the surrounding density of residents ( $L_i/K_i$ ) and are modeled symmetrically to production externalities:<sup>15</sup>

$$B_n = b_n \left( \frac{R_n}{K_n} \right)^{\eta^B} \quad (26)$$

where  $\eta^B$  parameters the strength of residential externalities.

We follow the standard approach in the urban literature of assuming that floor space is supplied by a competitive construction sector that uses land  $K$  and capital  $M$  as inputs. In particular, we assume that floor space ( $H_i$ ) is produced using geographical land ( $K_i$ ) and building capital ( $M_i$ ) according to the following constant return scale technology:

$$H_i = M_i^\mu K_i^{1-\mu}, \quad 0 < \mu < 1. \quad (27)$$

Using cost minimization and zero profits, this construction technology implies a constant elasticity supply function for floor space as in [Saiz \(2010\)](#):

$$Q_i = \psi_i H_i^{\frac{1-\mu}{\mu}} \quad (28)$$

where  $\psi_i$  depends solely on geographical land area ( $K_i$ ) and parameters.

Given these agglomeration forces and endogenous floor space, the determination of general equilibrium remains the same as above, except that productivity ( $A_n$ ), amenities ( $B_n$ ) and the supply of floor space ( $H_n$ ) are now endogenously determined by equations (25), (26) and (28).

## 5 Quantitative Analysis

In this section, we use our theoretical model to quantify the contributions of workplace access and consumption access to location choices. The key insight underlying our approach is that the observed consumption and commuting probabilities in our smartphone data can be used to reveal the relative valuation placed by users on different locations as consumption and workplace locations, and hence can be used to estimate travel access in a theory-consistent way. In [Section 5.1](#), we develop a sequential procedure to estimate the model's parameters. In [Section 5.2](#), we use these estimated parameters and model's residential choice probabilities to quantify the relative importance of workplace access, consumption access and residential amenities in explaining the observed spatial concentration of economic activity.

---

<sup>15</sup>As for production externalities above, we assume that residential externalities depend solely on a location's own residents density, but we can allow spillovers of these residential externalities across locations.



## 5.1 Estimation Procedure

We begin by discussing the estimation and calibration of the model’s parameters. We proceed in a number of steps, where each step uses additional model structure. First, we calibrate the Fréchet dispersion parameters for commuting, consumption, and residence choices ( $\theta^W$ ,  $\theta_k^S$ ,  $\theta^B$ , respectively), and the shares of consumer expenditure on housing ( $\alpha^H$ ), traded goods ( $\alpha^T$ ), and each type of non-traded service ( $\alpha_k^S$ ) using central values from the existing empirical literature and the observed data. Second, we estimate the worker’s route choice problem for each non-traded service and obtain an estimate of the expected travel cost for consumption trips ( $d_{nij(k)}^S$ ). Third, we estimate her consumption choice problem conditional on her residence and workplace, and obtain an estimate of the travel time parameter for consumption trips ( $\phi_k^S = \theta_k^S \kappa_k^S / \alpha_k^S$ ) and consumption access ( $\mathbb{S}_{ni}$ ). Fourth, we estimate her commuting choice problem, and obtain an estimate of the travel time parameter for commuting trips ( $\phi^W = \theta^W \kappa^W$ ) and travel access ( $\mathbb{A}_n$ ). Fifth, we calibrate the remaining parameters using the observed data and central values from the existing empirical literature.

### 5.1.1 Preference Parameters ( $\theta^W$ , $\theta_k^S$ , $\theta^B$ , $\alpha^H$ , $\alpha^T$ and $\alpha_k^S$ ) (Step 1)

In our first step, we calibrate the preference dispersion parameters ( $\theta^W$ ,  $\theta_k^S$  and  $\theta^B$ ) and expenditure shares ( $\alpha^H$ ,  $\alpha^T$ ,  $\alpha_k^S$ ). We set the preference dispersion parameters for commuting, consumption and residence choices equal to  $\theta^W = \theta_k^S = \theta^B = 6$ , which is consistent with the range of estimated values for these parameters. In the existing literature on commuting, [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#) estimates a preference dispersion parameter for workplace-residence choices of 6.83 using the division of Berlin by the Berlin Wall; [Heblich, Redding, and Sturm \(2020\)](#) estimates a value for the same parameter of 5.25 using the construction of London’s 19th-century railway network; and [Kreindler and Miyauchi \(2019\)](#) estimates the same parameter of 8.3 using information on the spatial dispersion of income in Dhaka, Bangladesh. In Section D.2.1 of the online appendix, we provide an over-identification check on our model’s predictions, using the property that its predictions for residential income depend importantly on these parameter values. In particular, we compare the model’s predictions for residential income in each Tokyo municipality to separate data on residential income not used in its calibration. Although our model is necessarily an abstraction, we find a strong positive relationship between the model’s predictions and the observed data.

Fewer empirical estimates are available for the preference dispersion parameter for consumption trips ( $\theta_k^S$ ), which determines the elasticity of consumption trips and consumption expenditure with respect to changes in the cost of sourcing non-traded services. Our calibrated value for this parameter of  $\theta_k^S = 6$  is in line with the existing empirical literature that has esti-

mated elasticities of substitution across retail stores. In particular, [Atkin, Faber, and Gonzalez-Navarro \(2018\)](#) estimates an elasticity of substitution of 3.9 using Mexican data, while [Couture, Gaubert, Handbury, and Hurst \(2019\)](#) estimates an elasticity of substitution of 6.5 using US data. In Section D.2.2 of the online appendix, we provide another overidentification check on our model’s predictions, using the property that its predictions for non-traded service prices in each location are sensitive to this parameter value. Again we show that there is a strong positive relationship between the model’s predictions and the observed data.

Finally, we calibrate the Cobb-Douglas expenditure share parameters using aggregate data on observed expenditure shares in Japan. We set the share of expenditure on residential floor space equal to  $\alpha^H = 0.25$ , which also corresponds to the values in [Davis and Ortalo-Magné \(2011\)](#) and [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#). We set the expenditure share parameter for each type of non-traded service ( $\alpha_k^S$ ) equal to the observed expenditure share on that sector for the Tokyo metropolitan area. Lastly, we solve for the implied traded goods expenditure share:  $\alpha^T = 1 - \alpha^H - \sum_{k \in K^S} \alpha_k^S$ .

### 5.1.2 Estimating the Route-Choice Probabilities (Step 2)

In our second step, we estimate expected consumption travel costs ( $d_{nij(k)}^S$ ), using the model’s predictions for route choice ( $HH, WW, HW, WH$ ) and our smartphone data. From the route choice probability (7), the probability of choosing route  $r(k)$  for non-traded service  $k$  conditional on residence  $n$ , workplace  $i$ , and consumption location  $j(k)$  can be written as:

$$\lambda_{r(k)|nij(k)}^R = \frac{\exp(-\phi_k^R \tau_{nij(k)r(k)}^S) \xi_{r(k)}^R \exp(u_{nij(k)r(k)}^R)}{\zeta_{nij(k)}^R}, \quad (29)$$

where  $u_{nij(k)r(k)}^R$  is a stochastic error that captures idiosyncratic determinants of route choice, given residence, workplace, and consumption location.

We estimate this route choice probability using the Poisson Pseudo Maximum Likelihood (PPML) estimator of [Santos Silva and Tenreyro \(2006\)](#).<sup>16</sup> The estimated semi-elasticity of travel time ( $\phi_k^R$ ) in equation (29) is a composite of the response of consumption trips to travel costs ( $\theta_k^R$ ) and the response of travel costs to travel times ( $\kappa_k^S$ ), such that  $\phi_k^R = \theta_k^R \kappa_k^S$ . The estimated route fixed effect  $\xi_{r(k)}^R$  corresponds to the tendency that each route is chosen conditional on travel time, such that  $\xi_{r(k)}^R = T_{r(k)}^R$ . The estimated residence-workplace-consumption-location fixed effect  $\zeta_{nij(k)}^R$  captures the average tendency that routes are chosen for each residence, workplace, consumption location, such that  $\zeta_{nij(k)}^R = \sum_{\ell \in \mathbb{R}} T_{\ell(k)}^R \exp(-\theta_k^R \kappa_k^S \tau_{nij(k)\ell(k)}^S)$ .

Table 2 presents the estimation results for each of the different types of non-traded services: “Finance, Real Estate, Communication, and Professional”; “Wholesale and Retail”; “Accom-

<sup>16</sup>We find a similar pattern of results if we estimate this choice probability using the multinomial logit model.

modation, Eating and Drinking”; “Medical, Welfare and Health Care”; “Other Services”. In the first row, we report the coefficient on the travel time ( $\phi_k^R$ ). In the second to fourth row report, we report the coefficient on the dummy variables for each route choice, where  $r(k) = HH$  is the excluded category. Two features of Table 2 are noteworthy. First, we estimate a negative and statistically significant composite coefficient on travel time ( $-\phi_k^R = -\theta_k^R \kappa_k^S$ ), highlighting its relevance for route choice. Second, we estimate negative and statistically significant coefficients on the indicator variables for the included route choices ( $r(k) \in \{HW, WH, WW\}$ ) relative to the excluded category of  $r(k) = HH$ . These results imply a high average preference for consuming non-traded services from home, consistent with Figure 4 in Section 3.

Using these estimates of  $\phi_k^R$  and  $\xi_{r(k)}^R$ , we construct adjusted expected travel costs for consumption trips conditional on residence  $n$  and workplace  $i$  from equation (8) above as:

$$\tilde{d}_{nij(k)}^S \equiv (d_{nij(k)}^S)^{1/\kappa_k^S} = \vartheta_k^R \left[ \sum_{r' \in \mathbb{R}} \xi_{r'(k)}^R \exp(-\phi_k^R \tau_{nij(k)r'(k)}^S) \right]^{\frac{1}{\phi_k^R}}, \quad (30)$$

where  $\vartheta_k^R$  is again  $\vartheta_k^R \equiv \Gamma\left(\frac{\theta_k^R - 1}{\theta_k^R}\right)$  and recall  $\mathbb{R} = \{HH, HW, WH, WW\}$ .

In this second step of our estimation, the composite semi-elasticity of travel time ( $\phi_k^R = \theta_k^R \kappa_k^S / \alpha_k^S$ ) is a sufficient statistic for the impact of travel time on route choices, as estimated from the route choice probabilities (29). We do not need to separate out the contributions of  $\theta_k^R$  and  $\kappa_k^S$  to the overall value of this parameter. Similarly, our adjusted measure of expected travel costs ( $\tilde{d}_{nij(k)}^S \equiv (d_{nij(k)}^S)^{1/\kappa_k^S}$ ) from equation (30) is a sufficient statistic for the impact of expected travel costs on workers choice of consumption locations, workplace and residence in the subsequent steps of our estimation below. We do not need to separate out the contributions of  $1/\kappa_k^S$  and  $d_{nij(k)}^S$  to the overall value of adjusted expected travel costs ( $\tilde{d}_{nij(k)}^S$ ).

### 5.1.3 Estimating Consumption Access ( $\mathbb{S}_{ni}$ ) (Step 3)

In our third step, we estimate the consumption choice probability and consumption access ( $\mathbb{S}_{ni}$ ), using the observed frequencies of consumption trips to reveal the relative attractiveness of each location for each type of non-traded service. From the conditional consumption probabilities (10), the probability that a worker travels to consume non-traded service  $k$  in location  $j(k)$ , conditional on residence  $n$  and workplace  $i$  is:

$$\lambda_{j(k)|ni}^S = \frac{\xi_{j(k)}^S \left( \tilde{d}_{nij(k)}^S \right)^{-\phi_k^S} \exp\left(u_{nij(k)}^S\right)}{\zeta_{ni,k}^S}, \quad (31)$$

where  $\tilde{d}_{nij(k)}^S$  is our estimated adjusted expected travel costs from equation (30); and  $u_{nij(k)}^S$  is a stochastic error that captures idiosyncratic determinants of consumption travel costs.

Table 2: Estimation Results for Route Choice

Dependent Variable:	Route Choice Probability				
	Finance realestate communication professional (1)	Wholesale retail (2)	Accomoda- tions eating drinking (3)	Medical welfare healthcare (4)	Other services (5)
Model:					
<i>Variables</i>					
Travel Time (Hours)	-0.312 (0.004)	-0.269 (0.004)	-0.264 (0.004)	-0.297 (0.004)	-0.271 (0.004)
Dummy (HW)	-1.58 (0.010)	-1.66 (0.010)	-1.75 (0.009)	-1.67 (0.011)	-1.61 (0.012)
Dummy (WH)	-1.04 (0.009)	-1.16 (0.010)	-1.10 (0.009)	-1.23 (0.011)	-1.12 (0.010)
Dummy (WW)	-1.03 (0.009)	-1.13 (0.009)	-1.30 (0.008)	-1.09 (0.010)	-1.09 (0.011)
<i>Fixed-effects</i>					
Home-Work-Consumption Location	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
AIC	3,753,940.7	7,015,231.9	3,461,408.3	3,674,206.4	4,511,086.5
BIC	6,348,159.7	9,717,411.9	6,081,904.7	6,174,892.1	7,210,376.1
Observations	887,212	921,176	895,488	857,704	920,268

Note: Results of estimating the route choice probability (29) using the Poisson Pseudo Maximum Likelihood (PPML) estimator. Observations are triplets of municipalities in the Tokyo metropolitan area (residence  $n$ , workplace  $i$ , and consumption location  $j(k)$ ) for each type of non-traded service  $k$ . We construct the empirical frequencies of route choice ( $\lambda_{r(k)|nij(k)}^R$ ) using our smartphone data (aggregated across weekdays and weekends), as discussed in Figure 4 in Section 3 above. The dependent variable is these empirical frequencies ( $\lambda_{r(k)|nij(k)}^R$ ), where  $r \in \mathbb{R} \equiv \{HH, WW, HW, WH\}$  corresponds to the different route choices: consuming non-traded services from home ( $HH$ ), from work ( $WW$ ), on the way from home to work ( $HW$ ), and on the way from work to home ( $WH$ ). The independent variables are travel time and the dummy variables for the different route choices, where  $r(k) = HH$  is the excluded category. Regressions are weighted by the frequency of observations for each residence, workplace, consumption-location, and sector, where we stochastically assign each trip to each sector following the procedure described in Figure 4 in Section 3. Standard errors in parentheses are clustered at the level of the combination of residence, workplace, and consumption location.

In a conventional gravity equation, travel flows are determined for a bilateral pair of locations. In contrast, in our extended gravity equation (31), consumption trips are determined at the level of triplets of residence, workplace and consumption locations. Since workers can travel to consume non-traded services from either their residence or their workplace, the adjusted expected consumption travel cost ( $\tilde{d}_{nij(k)}^S$ ) from equation (30) to consume non-traded service  $k$  in location  $j(k)$  depends on both residence  $n$  and workplace  $i$ .

We estimate this extended gravity equation (31) separately for each type of non-traded service using the Poisson Pseudo Maximum Likelihood (PPML) estimator. This estimator yields theoretically-consistent estimates of the fixed effects (as shown in Fally 2015) and allows for granularity and zeros in travel flows (as discussed further in Dingel and Tintelnot 2020). We obtain three key sets of estimates from this extended gravity equation. First, the estimated elasticity of consumption trips with respect to travel costs ( $\phi_k^S$ ) is a composite of the elasticity of

consumption trips with respect to travel costs ( $\theta_k^S/\alpha_k^S$ ) and the elasticity of travel costs with respect to travel times ( $\kappa_k^S$ ) in equation (10), such that  $\phi_k^S = \theta_k^S \kappa_k^S / \alpha_k^S$ . Second, the estimated consumption destination fixed effect ( $\xi_{j(k)}^S$ ) in equation (31) captures the average attractiveness of consumption destination  $j(k)$  for service  $k$  in terms of its price for that non-traded service ( $P_{j(k)}^S$ ) and quality draws ( $T_{j(k)}^S$ ) in equation (10), such that:

$$\xi_{j(k)}^S = T_{j(k)}^S (P_{j(k)}^S)^{-\theta_k^S}. \quad (32)$$

Third, the estimated residence fixed effect in equation (31) corresponds to the denominator in the conditional consumption probability in equation (10) and captures the overall attractiveness of residence  $n$  in terms of its access to all consumption locations  $\ell(k)$  for service  $k$ :

$$\zeta_{ni,k}^S = \sum_{\ell \in N} T_{\ell(k)}^S (P_{\ell(k)}^S)^{-\theta_k^S} \left( \tilde{d}_{ni\ell(k)}^S \right)^{-\phi_k^S}. \quad (33)$$

From these estimated fixed effects, we recover a theoretically-consistent estimate of consumption access for each type of non-traded service ( $\mathbb{S}_{ni,k}$ ). Indeed, consumption access can be recovered from either the consumption destination fixed effects or the residence fixed effects. First, summing the estimated consumption destination fixed effects ( $\xi_{j(k)}^S$ ) weighted by the estimated bilateral travel cost ( $\left( \tilde{d}_{ni\ell(k)}^S \right)^{-\phi_k^S}$ ) across locations, and using our calibrated values of  $\theta_k^S$  and  $\alpha_k^S$ , we obtain our baseline estimate of consumption access:

$$\mathbb{S}_{ni} = \prod_{k \in K^S} \Gamma \left( \frac{\theta_k^S / \alpha_k^S - 1}{\theta_k^S / \alpha_k^S} \right) \left[ \sum_{\ell \in N} \xi_{\ell(k)}^S \left( \tilde{d}_{ni\ell(k)}^S \right)^{-\phi_k^S} \right]^{\frac{\alpha_k^S}{\theta_k^S}}. \quad (34)$$

Second, using the estimated residence fixed effects ( $\zeta_{ni,k}^S$ ), and our calibrated values of  $\theta_k^S$  and  $\alpha_k^S$ , we obtain another estimate of consumption access:  $\mathbb{S}_{ni} = \prod_{k \in K^S} \left[ \Gamma \left( \frac{\theta_k^S / \alpha_k^S - 1}{\theta_k^S / \alpha_k^S} \right) \right] (\zeta_{ni,k}^S)^{\frac{\alpha_k^S}{\theta_k^S}}$ . As sample size becomes sufficiently large, these two sets of estimates of consumption access converge asymptotically towards one another if the model is a correct specification of the true data generating process, as shown in an international trade context in Fally (2015). In practice, even in our finite sample, we find that these two estimates are extremely highly correlated with one another, as shown in Section D.4 of the online appendix.

In Table 3, we report the results of estimating the consumption extended gravity equation (31) for each type of non-traded service separately. In all cases, we estimate negative and statistically significant semi-elasticities of consumption trips with respect to travel costs ( $-\phi_k^S$ ). We find that these estimated semi-elasticities are relatively constant across the different types of consumption trips, ranging from -1.08 to -1.19, with the most localized consumption trips observed for “Finance, Real Estate, Communication, and Professional” and “Medical, Welfare

and Health Care”. In Section D.3 of the online appendix, we report a specification check in which model the relationship between consumption trips and travel costs non-parametrically and demonstrate a similar pattern of results.<sup>17</sup>

Table 3: Estimation Results for Consumption Location Choice

Dependent Variable:	Finance realestate communication professional (1)	Consumption Location Choice Wholesale retail (2)	Accommodations eating drinking (3)	Medical welfare healthcare (4)	Other services (5)
Model:					
<i>Variables</i>					
$\log \tilde{d}_{nij(k)}^S$	-1.15 (0.040)	-1.12 (0.036)	-1.09 (0.035)	-1.19 (0.038)	-1.08 (0.036)
<i>Fixed-effects</i>					
Home and Work Location Pairs	Yes	Yes	Yes	Yes	Yes
Consumption Location	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
AIC	129,480.5	130,876.3	131,837.4	128,831.5	132,020.7
BIC	291,657.6	293,164.9	294,084.1	290,841.4	294,295.3
Observations	2,981,924	2,983,860	2,983,134	2,979,020	2,983,618

Note: Results of estimating the consumption trip probability (31) using the Poisson Pseudo Maximum Likelihood (PPML) estimator. Observations are triplets of municipalities in the Tokyo metropolitan area (residence  $n$ , workplace  $i$ , and consumption location  $j(k)$ ). Each column regresses the consumption trip probability for each type of non-traded service on the adjusted expected travel cost ( $\tilde{d}_{nij(k)}^S$ ) from the previous step, consumption location fixed effects, and residence-workplace pair fixed effects. Standard errors in parentheses are clustered two-way on consumption location and residence-workplace pair.

#### 5.1.4 Estimating the Workplace Choice and Travel Access ( $\mathbb{A}_n$ ) (Step 4)

In our fourth step, we estimate the workplace choice probability and overall travel access ( $\mathbb{A}_n$ ), by using the observed frequencies of commuting trips to reveal the relative attractiveness of residences and workplaces. From our parameterization of commuting costs and equations (14) and (15), the probability that a worker commutes from residence  $n$  to workplace  $i$  can be written as the following extended gravity equation:

$$\lambda_{i|n}^W = \frac{\xi_i^W \exp(-\phi^W \tau_{ni}^W) (\mathbb{S}_{ni})^{\theta^W} \exp(u_{ni}^W)}{\zeta_n^W}, \quad (35)$$

where  $u_{ni}^W$  is a stochastic error that reflects idiosyncratic determinants of bilateral commuting costs not captured in bilateral travel times ( $\tau_{ni}$ ).

<sup>17</sup>As a specification check, we re-estimated the consumption gravity equation under the false assumption that all consumption trips originate from home. As shown in Table D.5.1 in Section D.5 of the online appendix, we find substantially smaller semi-elasticities in this robustness check (ranging from -0.8 to -0.6), highlighting the importance of endogenous route choice. Furthermore, we find a better model fit incorporating route choice than this alternative specification, as evident from the smaller Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC) than in Panel (B) of Table D.5.1.

In a conventional gravity equation for commuting, the key bilateral determinant of commuting flows is bilateral travel time  $\tau_{ni}^W$ . In contrast, in our extended gravity equation for commuting (35), a worker's choice of workplace depends on the extent to which it enhances the worker's access to consumption opportunities, which in turn depends on the worker's residence. Therefore, consumption access ( $\mathbb{S}_{ni}$ ) varies bilaterally with both workplace and residence, and enters as an additional determinant of bilateral commuting flows alongside bilateral travel time

We estimate this extended commuting gravity equation (35) using the Poisson Pseudo Maximum Likelihood (PPML) estimator, our measures of commuting travel times ( $\tau_{ni}^W$ ), and our estimates of bilateral consumption access ( $\mathbb{S}_{ni}$ ) from the previous step. We again obtain three key sets of estimates from this extended gravity equation. First, the estimated semi-elasticity of commuting flows with respect to travel times ( $\phi^W$ ) in equation (35) is again a composite of the response of commuting flows to commuting costs ( $\theta^W$ ) and the response of commuting costs to travel times ( $\kappa^W$ ) in equation (14), such that  $\phi^W = \theta^W \kappa^W$ . Second, the estimated workplace fixed effect ( $\xi_i^W$ ) in equation (35) captures the average attractiveness of workplace  $i$  across sectors in terms of its wage ( $w_{i,g}$ ) and productivity draws ( $T_{i,g}^W$ ):

$$\xi_i^W = \sum_{m \in K} T_{i,m}^W w_{i,m}^{\theta^W}. \quad (36)$$

Third, the estimated residence fixed effect ( $\zeta_n^W$ ) in equation (35) corresponds to the denominator in the conditional commuting probability in equation (14) and captures the overall attractiveness of residence  $n$  in terms of its travel-time weighted access to all workplaces:

$$\zeta_n^W = \sum_{\ell \in N} \sum_{m \in K} T_{\ell,m}^W w_{\ell,m}^{\theta^W} \exp(-\phi^W \tau_{n\ell}^W) (\mathbb{S}_{n\ell})^{\theta^W}. \quad (37)$$

From these estimated fixed effects, we recover a theoretically-consistent measure of overall travel access. Indeed, as for consumption access in the previous step, we can recover travel access in two different ways. First, summing the estimated workplace fixed effects ( $\xi_i^W$ ) weighted using the estimated bilateral travel costs ( $\exp(-\phi^W \tau_{ni}^W)$ ) across locations, and using  $\theta^W$ , we obtain our baseline estimate of travel access:

$$\mathbb{A}_n = \Gamma\left(\frac{\theta^W - 1}{\theta^W}\right) \left[ \sum_{\ell \in N} \xi_\ell^W \exp(-\phi^W \tau_{n\ell}^W) (\mathbb{S}_{n\ell})^{\theta^W} \right]^{\frac{1}{\theta^W}}. \quad (38)$$

Second, using the estimated residence fixed effects ( $\zeta_n^W$ ) and  $\theta^W$ , we obtain another estimate of workplace access:  $\mathbb{A}_n = \Gamma\left(\frac{\theta^W - 1}{\theta^W}\right) (\zeta_n^W)^{\frac{1}{\theta^W}}$ . As sample size becomes sufficiently large, these two sets of estimates of travel access again converge asymptotically towards one another if the model is a correct specification of the true data generating process. In practice, even in our finite sample, we find that these two estimates are extremely highly correlated with one another, as shown in Section D.4 of the online appendix.



In Table 4, we present the results of estimating our commuting extended gravity equation (31). We include the commuting travel time, consumption access ( $S_{nl}$ ) with a known exponent of  $\theta^W$ , workplace fixed effects and residence fixed effects. We estimate a negative and statistically significant semi-elasticity of commuting flows with respect to commuting time of  $-\phi^W = -0.617$ . This estimated value of  $\phi^W$  is significantly smaller than our estimates of  $\phi_k^S$  above, suggesting that consumption choices are more responsive to travel time than workplace choices. In Section D.3 of the online appendix, we report a specification check in which we model the relationship between commuting trips and travel time non-parametrically, and show that our semi-log specification provides a good approximation to the data.<sup>18</sup>

Table 4: Estimation Results for Workplace Choice

Dependent Variable: Model:	Commuting Choice Probability (1)
<i>Variables</i>	
Commuting Time (Hours)	-0.617 (0.037)
<i>Fixed-effects</i>	
Home Location	Yes
Work Location	Yes
<i>Fit statistics</i>	
AIC	1,212.0
BIC	5,557.3
Observations	58,564

Note: Results of estimating the commuting probability (35) using the Poisson Pseudo Maximum Likelihood (PPML) estimator. Observations are all pairs of municipalities in the Tokyo metropolitan area (residence  $n$  and workplace  $i$ ). Each column regresses the commuting probability on commuting time, workplace fixed effects, residence fixed effects, and consumption access ( $S_{nl}$ ) with a coefficient restricted to equal  $\theta^W$ . Standard errors in parentheses are clustered two-way on residence and workplaces.

### 5.1.5 Other Model Parameters (Step 5)

Together Steps 1-4 are sufficient to undertake our decomposition of the observed spatial variation in economic activity into the contributions of travel access and a residual for amenities, within an entire class of quantitative urban models with different specifications of production technology and market structure. However, when we undertake counterfactuals, such as for example for transport infrastructure improvements, we need to determine additional structural parameters related to the supply-side of the economy (land supply, tradable and non-tradable

<sup>18</sup>As a further specification check, we re-estimated the commuting gravity equation excluding consumption access ( $\log(S_{nl})^{\theta^W}$ ). As shown in Table D.5.2 in Section D.5 of the online appendix, we find a larger travel time semi-elasticity when omitting consumption access ( $-0.649$  instead of  $-0.617$ ), highlighting the relevance of controlling for this term. Furthermore, we find a better model fit for our specification incorporating consumption access, as measured by the Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC).



production, and production and amenity spillovers). In our fifth step, we calibrate these parameters directly from the data or using central values from the existing empirical literature.

We calibrate the Cobb-Douglas cost shares for labor in each sector  $(\beta^S, \beta^T)$  as 0.8, which are broadly consistent with the labor share on production costs for Tokyo metropolitan area. We assume a standard share of land in construction costs of  $\mu = 0.75$ . We explore a range of values for the production and residential agglomeration parameters ranging from zero to the values estimated in [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#):  $\eta^W \in [0, 0.08]$  and  $\eta^B \in [0, 0.15]$ , which spans most of the existing empirical estimates in the meta-analyses of [Melo, Graham, and Noland \(2009\)](#) and [Ahlfeldt and Pietrostefani \(2019\)](#).

## 5.2 Quantifying the Role of Workplace and Consumption Access

We now use our estimates from Steps 1-4 above to quantify the contributions of travel access and the residual of residential amenities to explaining the observed spatial concentration of economic activity and to examine the relative importance of workplace access and consumption access for overall travel access. Re-writing the residential choice probabilities (17), we have:

$$(\lambda_n^B)^{1/\theta^B} Q_n^{\alpha^H} = \mathbb{B}_n \mathbb{A}_n. \quad (39)$$

The left-hand side of this relationship corresponds to a summary measure of the relative attractiveness of locations. A larger share of residents  $(\lambda_n^B)$  and/or a higher price of floor space  $(Q_n)$  both imply that a location is a more attractive place to live. On the right-hand side,  $\mathbb{B}_n$  is a composite amenities parameter that includes common amenities  $(B_n)$ , the parameter determining average idiosyncratic amenities  $(T_n^B)$ , the common price of the traded good  $(P_n^T = P^T = 1)$ , and the common reservation level of utility  $(\bar{U})$ :

$$\mathbb{B}_n \equiv B_n (T_n^B)^{1/\theta^B} (P_n^T)^{-\alpha^T} (\bar{U}/\vartheta^B)^{-1} \quad (40)$$

In these residential choice probabilities (39), we observe the share of residents  $(\lambda_n^B)$  and the price of floor space  $(Q_n)$ , and we estimated travel access  $(\mathbb{A}_n)$  in equation (38). Therefore, we can use these residential choice probabilities (39) to recover the unobserved composite amenities  $(\mathbb{B}_n)$  as a structural residual that exactly rationalizes the observed data as an equilibrium of the model. This residential choice decomposition has an intuitive interpretation. If a location has a high share of residents  $(\lambda_n^B)$  and high price of floor space  $(Q_n)$  on the left-hand side, despite having relatively low values of composite access  $(\mathbb{A}_n)$  on the right-hand side, this is rationalized in the model by that location having relatively high residential amenities  $(\mathbb{B}_n)$ .

We now decompose the variance of our summary measure of the relative attractiveness of locations into the contributions of travel access  $(\mathbb{A}_n)$  and residential amenities  $(\mathbb{B}_n)$ . In particular, we use a regression-based variance decomposition from the international trade literature.

We estimate an ordinary least squares (OLS) regression of each of the components on the right-hand side of the residential choice probabilities (39) on our summary measure of the relative attractiveness of locations from the left-hand side:

$$\begin{aligned}\ln \mathbb{A}_n &= c_0^A + c_1^A \ln \left( (\lambda_n^B)^{1/\theta^B} Q_n^{\alpha^H} \right) + u_{nt}^A, \\ \ln \mathbb{B}_n &= c_0^B + c_1^B \ln \left( (\lambda_n^B)^{1/\theta^B} Q_n^{\alpha^H} \right) + u_{nt}^B,\end{aligned}\tag{41}$$

Noting that OLS is a linear estimator with mean zero residuals, and using the residential choice probabilities (39), we have  $c_0^A + c_0^B = 0$  and  $c_1^A + c_1^B = 1$ . Implicitly, this variance decomposition allocates the covariance terms equally across each of the two components. The relative values of the slope coefficients  $\{c_1^B, c_1^A\}$  provide measures of the relative importance of travel access ( $\mathbb{A}_n$ ) and residential amenities ( $\mathbb{B}_n$ ) in explaining the observed variation in our summary measure of the relative attractiveness of locations.

We next examine the relative importance of workplace access and consumption access for overall travel access, by considering a special case of our quantitative urban model without consumption trips ( $\alpha_k^S = 0$  for all  $k \in K^S$ ,  $\alpha^T = 1 - \alpha^H$ ,  $\lambda_{j(k)|ni}^S = 0$  and  $\mathbb{S}_{ni} = 1$ ). In this special case, we ignore the data on consumption trips, and estimate a standard quantitative urban model of workplace-residence choice using only the data on commuting trips. As a result, travel accessibility ( $\mathbb{A}_n^{\text{nocons}}$ ) depends on workplace access alone, and can be constructed using the estimates from our extended gravity equation estimation of equation (35), but omitting the consumption access term ( $\log(\mathbb{S}_{n\ell})^{\theta^W}$ ):

$$\mathbb{A}_n^{\text{nocons}} = \Gamma \left( \frac{\theta^W - 1}{\theta^W} \right) \left[ \sum_{\ell \in N} \xi_\ell^W \exp(-\phi^W \tau_{n\ell}^W) \right]^{\frac{1}{\theta^W}}, \tag{42}$$

where  $\phi^W$  and  $\xi_\ell^W$  are the estimated travel time coefficient and workplace fixed effects from the extended commuting gravity equation (35). Using this measure of travel access without consumption trips ( $\mathbb{A}_n^{\text{nocons}}$ ) in equation (39), we can recover a measure of amenities without consumption trips ( $\mathbb{B}_n^{\text{nocons}}$ ), and implement our variance decomposition in equation (41) above.<sup>19</sup>

Table 5 reports the results of these variance decompositions for our model including consumption trips (Panel A) and the special case excluding consumption trips (Panel B). Observations correspond to municipalities in the Tokyo metropolitan area for which we have land price data. We measure the price of floor space ( $Q_n$ ) using the observed land price data ( $\tilde{Q}_n$ )

<sup>19</sup>As a robustness check, Panel (B) of online appendix Table D.5.3 construct travel access without consumption trips ( $\mathbb{A}_n^{\text{nocons}}$ ) using the estimates of  $\phi^W$  and  $\xi_\ell^W$  from a conventional commuting gravity equation excluding consumption access. Although the estimated travel time coefficients differ between these two gravity equation specifications, we find a similar pattern of results for the relative importance of consumption access and residential amenities in this robustness test as in our baseline specification.

Table 5: Decomposition of our Summary Statistic for Relative Attractiveness ( $\log \left[ (\lambda_n^B)^{1/\theta^B} Q_n^{\alpha^H} \right]$ ) into Travel Access ( $\mathbb{A}_n$ ) and Residential Amenities ( $\mathbb{B}_n$ )

	$\log \mathbb{A}_n$	$\log \mathbb{B}_n$
	(1)	(2)
<b>Panel A: Baseline Model</b>		
$\log Q_n^{\alpha^H} (\lambda_n^B)^{1/\theta^B}$	0.566 (0.049)	0.434 (0.049)
Observations	201	201
R <sup>2</sup>	0.403	0.284
<b>Panel B: No Consumption Trips</b>		
$\log Q_n^{\alpha^H} (\lambda_n^B)^{1/\theta^B}$	0.373 (0.036)	0.627 (0.036)
Observations	201	201
R <sup>2</sup>	0.352	0.606

Note: Ordinary least squares (OLS) estimates of the regression-based variance decomposition in equation (41). Panel (A) corresponds to our baseline model, in which we compute travel access ( $\mathbb{A}_n$ ) incorporating consumption trips; Panel (B) corresponds to the special case of our model in which we abstract from consumption trips ( $\mathbb{A}_n^{\text{nocons}}$ ), such that  $\alpha_k^S = 0$  for all  $k \in K^S$ ,  $\alpha^T = 1 - \alpha^H$ ,  $\lambda_{j(k)|ni}^S = 0$  and  $\mathbb{S}_{ni} = 1$ . Observations are municipalities in the Tokyo metropolitan area. Heteroskedasticity robust standard errors in parentheses.

and our assumption of competitive construction sector (such that  $Q_n \propto \tilde{Q}_n^{1-\mu}$ ). In our model including consumption trips, we find that travel access ( $\mathbb{A}_n$ ) is about as important as the residual of residential amenities ( $\mathbb{B}_n$ ) in explaining variation in the relative attractiveness of locations ( $Q_n^{\alpha^H} (\lambda_n^B)^{1/\theta^B}$ ), with a contribution of 56 percent compared to 44 percent. In contrast, when we consider a conventional quantitative urban model excluding consumption trips, we find a substantially reduced contribution from travel access ( $\mathbb{A}_n^{\text{nocons}}$ ) of only 37 percent, with the residual of residential amenities making up the remaining 63 percent. These results suggest that a substantial component of the variation in conventional measures of residential amenities that do not control for consumption trips may reflect unobserved differences in consumption access. They also suggest that workplace access ( $\mathbb{A}_n^{\text{nocons}}$ ) is far from perfectly correlated with overall travel access incorporating consumption trips ( $\mathbb{A}_n$ ), because we find a much smaller contribution from travel access when we restrict attention to commuting information alone.

## 6 Counterfactuals

We next use our theoretical framework to undertake counterfactuals for changes in travel costs to provide further evidence on the role of consumption access in understanding the spatial concentration of economic activity. In particular, we examine the role of consumption trips in shaping

the welfare effects of transport infrastructure improvements. We undertake a counterfactual for the construction of a new subway (underground) line in the city of *Sendai* and compare the model’s predictions to the observed impact in the data.<sup>20</sup>

Before the opening of its new subway (underground) line, the city of *Sendai* had only one *Nanboku* (North-South) subway line, which had been in operation since 1987. In December 2015, the new *Tozai* (East-West) subway line opened, thereby providing a substantial expansion in the overall subway network. In Section 6.1, we report reduced-form evidence on the impact of the Tozai Subway line on floor space prices, residential population and travel access. We compare the results of differences-in-differences specifications estimated using the actual data and the counterfactual predictions of our model. In Section 6.2, we present the model’s counterfactual predictions for the welfare gains from the opening of the Tozai Subway Line and evaluate the contribution from consumption access towards these welfare gains.

To undertake the counterfactual simulation, we solve the system of equations for a general equilibrium of the model using an exact-hat algebra approach, in which we rewrite the counterfactual equilibrium conditions of the model in terms of the initial travel shares and endogenous variables of the model and the counterfactual changes in these endogenous variables, as shown in Section E of the online appendix. In our baseline specification, we use the fitted values for the initial travel shares from our gravity equation estimation to address potential concerns about granularity. In Section G.5 of the online appendix, we report a robustness test using the observed initial travel shares, and demonstrate similar results using both approaches. In our baseline specification, we consider the closed-city specification of the model, in which total population for the city as a whole ( $\bar{L}$ ) is exogenous, and hence the change in travel costs affects worker welfare.<sup>21</sup>

## 6.1 Difference-in-Difference Effects of Tozai Subway Line

We start by analyzing in the impact of the Tozai Subway Line in our observed smartphone and land price data. Our analysis is based on the following difference-in-difference regression:

$$\Delta \log Y_n = c_0 + c_1 T_n + u_n, \quad (43)$$

where  $n$  indexes Oaza;  $T_n$  is a dummy variable that equals one if the Oaza includes the new stations of the Tozai Subway Line (except for Sendai station which is also a station for the existing Nanboku Subway Line) and zero otherwise;  $\Delta \log Y_n$  is the log difference of an outcome

<sup>20</sup>In Section F of the online appendix, we provide further evidence on the relative importance of consumption and workplace access for location decisions by comparing the results of separate counterfactuals for changes in travel costs for commuting and consumption trips for the Tokyo metropolitan area.

<sup>21</sup>It is straightforward to instead consider the open-city specification, in which case total population is endogenous, and the welfare effects of the change in travel costs accrue only to landlords, as in the public finance literature following George (1879).

of interest before and after the opening of the Tozai Line; any fixed effect in the level of the outcome of interest is differenced out; the constant  $c_0$  captures any common change in the outcome of interest across all locations; and the coefficient  $c_1$  is an estimate of the treatment effect from the opening of a station on the new Tozai Subway Line. We consider the following outcomes: (i) the price of floor space ( $Q_n$ ); (ii) the residential probability or share of the city's residential population in each Oaza ( $\lambda_n^B$ ); (iii) travel access ( $\mathbb{A}_n$ ); and (iv) residential amenities ( $\mathbb{B}_n$ ).

We first estimate this regression using the observed data for the pre- and post-periods. We measure the price of floor space ( $Q_n$ ) using the observed land price data ( $\tilde{Q}_n$ ) and our assumption of competitive construction sector (such that  $Q_n \propto \tilde{Q}_n^{1-\mu}$ ). For the land price data, we use 2009 as the pre-period (the earliest available year to mitigate anticipation effects) and 2018 as the post-period. We construct the residential probability ( $\lambda_n^B$ ) using our smartphone data. We estimate travel access  $\mathbb{A}_n$  and residential amenities  $\mathbb{B}_n$  using our smartphone data for the pre- and post-period separately. For these variables constructed from our smartphone data, we use June 2015 as the pre-period (shortly before the opening of the new subway line), and we use June 2017 as our post-period (the same month two years after the pre-period). To better proxy the changes in travel time from the opening of the new subway line in this context where residents use different travel modes, we extend our baseline model to incorporate a mode choice between public transportation and cars, as discussed in Appendix G.1.<sup>22</sup>

In Panel (A) of Table 6, we present the results of estimating equation (43) using the observed data. As shown in Columns (1) and (2), we find larger increases in floor space prices and residential population in Oaza containing new stations than in other Oaza following the opening of the new subway line, which is consistent with these locations becoming relatively more attractive. As reported in Column (3), we also observe a larger increase in our estimate of travel access in locations with new stations, which is consistent with the idea that the increase in floor space prices and residential population in these location is driven by the model's mechanism of an improvement in travel access. In contrast, as shown in Column (4), there is no evidence of a larger increase in the structural residual of residential amenities in these locations. Therefore, we find that the model is quantitatively able to explain the observed increase in floor space prices and residential population through its mechanism of an improvement in travel access, without requiring increases in the residual of residential amenities in these locations. Notably, if we consider the special case of our model excluding consumption trips, we find a smaller increase in travel access (0.042 instead of 0.054) and a larger increase in the residual of residential amenities (0.017 instead of 0.004), as shown in Table G.3.1 in Section G.3 of the online appendix. Hence, we also find that incorporating consumption trips is important for

---

<sup>22</sup>We use the same parameters as above, except for  $\phi^W$  and  $\phi_k^S$ , which we re-estimate using our smartphone data for the city of Sendai, as discussed in Section G.2 of the online appendix.

the quantitative success of the model’s mechanism in explaining the observed data.

Table 6: Difference-in-Difference Estimates for the Opening of the Tozai Subway Line Using the Observed Data and our Model’s Counterfactual Predictions

	$\Delta \log Q_n$	$\Delta \log \lambda_n^B$	$\Delta \log \mathbb{A}_n$	$\Delta \log \mathbb{B}_n$
	(1)	(2)	(3)	(4)
<b>Panel A: Data</b>				
Dummy (Tozai Line Stations)	0.046 (0.014)	0.311 (0.210)	0.054 (0.008)	0.004 (0.036)
Observations	368	305	305	305
R <sup>2</sup>	0.030	0.007	0.123	0.0001
<b>Panel B: Model Prediction</b> ( $\eta^B = 0$ ; $\eta^W = 0.08$ )				
Dummy (Tozai Line Stations)	0.091 (0.010)	0.300 (0.032)	0.073 (0.008)	0.000 (0.000)
Observations	370	370	370	370
R <sup>2</sup>	0.197	0.191	0.199	

Note: Results of estimating the difference-in-difference regression (43) using the observed outcome variables (Panel A) and the counterfactual model predictions (Panel B). The treatment dummy is an indicator that takes the value one when the Oaza includes stations of the new Tozai Subway Line (except for Sendai station which is also a station for the existing Nanboku Subway Line) and zero otherwise. Observations are the 370 Oaza in the City of Sendai. In Panel (A), 2 observations are missing in Column (1) because land price data is not available, and 65 observations are missing in Columns (2)-(4), because we observe no residents in either the pre- or post-period in our smartphone data. Standard errors are clustered by Oaza.

To provide further evidence on the predictive power of our model, we next undertake counterfactuals for the impact of the reduction in travel time from the opening of the new subway line using only information from the pre-period, and estimate the same reduced-form regressions using the model’s counterfactual predictions. In our baseline specification, we assume the standard value for production agglomeration forces from the existing empirical literature ( $\eta^W = 0.08$ ), and assume that our mechanism of consumption access captures all agglomeration forces in residential decisions ( $\eta^B = 0$ ). In Panel (B) of Table 6, we present the results from estimating equation (43) using these counterfactual predictions for the change in each economic outcome of interest. We find that the model’s counterfactual predictions align closely with the observed patterns in the data. In Column (1), we estimate a positive and statistically significant treatment effect for the price of floor space, which is somewhat larger than that in the observed data, perhaps in part because the model may not fully capture the expansion in the supply of floor space following the opening of the new subway line. In Columns (2) and (3), we also estimate positive and statistically significant treatment effects for the residential probability and travel access, which lie within the 95 percent confidence intervals around the estimated treatments in the observed data. Finally, in Column (4), the model necessarily implies

zero treatment effect for residential amenities in the absence of residential agglomeration forces ( $\eta^B = 0$ ), which is consistent with our finding above using the observed data that the estimated treatment effect for residential amenities is close to zero and statistically insignificant.<sup>23</sup>

As an additional specification check, we estimate the same reduced-form regressions, but use a dummy variable that takes the value one for Oazas that contain stations on the existing Nanboku (North-South) Subway Line (which opened in 1987) rather than stations on the new Tozai (East-West) Subway Line (which opened in 2015). If there are positive or negative network effects from the new Tozai Subway Line on locations with stations on the existing Nanboku Subway Line, we would expect to again detect statistically significant treatment effects. In Section G.4 of the online appendix, we show that we find no evidence of statistically significant treatments effects on the price of floor space, residential population, travel access, and residential amenities for this existing Nanboku Subway Line. These results are consistent with a limited net impact of network effects on the existing subway line and suggest that our earlier estimates for the Tozai Subway Line are indeed capturing effects specific to this new subway line. Consistent with these findings using the observed data, we also find no evidence of statistically significant treatment effects for the existing Nanboku Subway Line using our counterfactual predictions of the model.

## 6.2 Welfare Gains from the Tozai Subway Line

We now use our baseline closed-city version of the model to evaluate the welfare impact of the opening of this new subway line. In Table 7, we present the results for the different model specifications shown in the left-most column. In the second column, we report the percentage point increase in expected utility for the residents of the city of Sendai. In our baseline specification in the first row, we again assume the standard value for production agglomeration forces from the existing empirical literature ( $\eta^W = 0.08$ ), and assume that our mechanism of consumption access captures all agglomeration forces in residential decisions ( $\eta^B = 0$ ). In the robustness checks in the subsequent rows, we report results for a number of alternative specifications. In the third column, we report the change in expected utility in each of these alternative specifications as a percentage of that in our baseline specification in the first row.

As reported in Row (1), we find an increase in the flow of expected utility from the opening of the new Tozai Subway Line of 2.74 percentage points in our baseline specification. Therefore, even though we take into account the existence of other modes of transport prior to the opening

---

<sup>23</sup>In a robustness test in Section G.3 of the online appendix, we estimate  $\eta^B$  using the identifying assumption that the log change in residential fundamentals ( $b_n$  in equation (26)) is uncorrelated with proximity to new subway stations. We find a small estimate of  $\eta^B = 0.01$ . In the special case of the model that abstracts from consumption trips, we obtain a somewhat larger estimate of  $\eta^B = 0.05$ , again highlighting the importance of incorporating consumption trips for the model's mechanism of travel access to explain the observed data.



of the new line (such as buses), we find substantial welfare gains from the reduction in bilateral travel times achieved by the opening of the new subway line. To provide a point of comparison, Row (2) reports results for the special case of our model excluding consumption trips ( $\alpha_k^S = 0$  for all  $k \in K^S$ ,  $\alpha^T = 1 - \alpha^H$ ,  $\lambda_{j(k)|ni}^S = 0$  and  $\mathbb{S}_{nt} = 1$ ). In this specification, we find a welfare gain from the new subway line of 1.44 percentage points, or 53 percent of that in our baseline specification. Therefore, we find that the undercounting of travel journeys from focusing solely on commuting trips is quantitatively important for the evaluation of the welfare effects of observed transport infrastructure improvements.

In Row (3), we consider another special case of the model, in which we falsely assume that all consumption trips originate from home locations, thereby ruling out travel to consume non-traded services from work or on the way between home and work.<sup>24</sup> In this special case, we find somewhat larger welfare gains from the new subway line of 2.99 percentage points, or 9 percent larger than our baseline specification. This pattern of results is intuitive, because excluding consumption travel from work or on the way between home and work increases average travel distances for consumption trips, and hence increases the magnitude of the welfare gain from the reduction in travel times achieved by the opening of the new subway line.

Table 7: Counterfactual Increase in Expected Utility in Sendai from the new Tozai Subway Line

	Percentage Point Increase in Residential Utility	Relative to Baseline (%)
(1) Baseline ( $\eta^B = 0; \eta^W = 0.08$ )	2.74	1.00
(2) No consumption trips	1.44	0.53
(3) No trip chains for consumption trips	2.99	1.09
(4) Include residential spillover ( $\eta^B = 0.15$ )	3.24	1.18
(5) Eliminate production spillover ( $\eta^W = 0$ )	2.61	0.95

Note: The second column reports model counterfactuals for the percentage point increase in expected utility as a result of the reduction in travel time from the opening of the new Tozai (East-West) subway line in the city of Sendai. The first row presents results for our baseline specification (residential agglomeration forces of  $\eta^B = 0$ , workplace agglomeration forces of  $\eta^W = 0.08$ ) and the subsequent rows present results for a number of alternative specifications. The third column reports the change in expected utility in each of these alternative specifications as a percentage of the change in our baseline specification in the first row.

In the remaining two rows, we examine the sensitivity of our results to alternative assumptions about the strength of residential and production agglomeration forces. In Row (4), we introduce residential agglomeration forces by assuming  $\eta^B = 0.15$  instead of  $\eta^B = 0$ . In this specification, we find welfare gains from the new subway line that are around 18 percent larger than those in our baseline specification. In Row (5), we exclude productivity spillovers by assuming  $\eta^W = 0$  instead of  $\eta^W = 0.08$ . In this case, we find welfare gains from the new subway line that are around 5 percent smaller than those in our baseline specification. Therefore, we find that agglomeration forces magnify the welfare gains from transport infrastructure improvements, consistent with the findings of existing studies, such as [Tsivanidis \(2018\)](#) and [Heblich,](#)

<sup>24</sup>More specifically, we consider the limiting case in which  $T_{r(k)}^R \rightarrow 0$  for  $r(k) \in \{WW, HW, WH\}$  and  $T_{HH}^R > 0$ , which ensures that workers always travel to consume non-traded services from home.



Redding, and Sturm (2020). Nevertheless, the impact of these agglomeration forces on the welfare gains from transport infrastructure improvements (comparing Rows (4) and (5) to Row (1)) is smaller than the impact of excluding consumption trips (comparing Row (2) to Row (1)), again highlighting the relevance of consumption access for the evaluation of the welfare effects of transport infrastructure improvements.

## 7 Conclusions

We provide new theory and evidence on the role of consumption access in understanding the spatial concentration of economic activity. We use smartphone data that records the global positioning system (GPS) location of users every 5 minutes to provide high-resolution evidence on patterns of travel by hour and day within the Tokyo metropolitan area. Guided by our empirical findings, we develop a quantitative model of internal city structure that captures the fact that much of the travel that occurs within urban areas is related not to commuting but rather to the consumption of non-traded services, such as trips to restaurants, coffee shops and bars, shopping expeditions, excursions to cinemas, theaters, music venues and museums, and visits to professional service providers.

We begin by establishing four key empirical properties of these non-commuting trips. First, we show that they are more frequent than commuting trips, so that concentrating solely on commuting substantially underestimates travel within urban areas. Second, we find that they are concentrated closer to home and are more responsive to travel time than commuting trips, which implies that focusing solely on commuting yields a misleading picture of bilateral patterns of travel within cities. Third, combining our smartphone data with highly spatially-disaggregated data on employment by sector, we show that these non-commuting trips are closely related to the availability of nontraded sectors, consistent with our modelling of them as travel to consume non-traded services. Fourth, we find evidence of trip chains, in which these consumption trips can occur along the journey between home and work, highlighting the relevance of jointly modelling both commuting and consumption trips.

We next develop our quantitative theoretical model of internal city structure that incorporates these consumption trips. Workers choose their preferred residence, workplace and consumption locations, taking into account the bilateral costs of travel and idiosyncratic draws for amenities for each residence, productivity for each workplace, service quality for each consumption location, and preferences for each route. We show that the observed travel data and model's gravity equations for commuting and consumption trips can be used to estimate theoretically-consistent measures of travel access. We use the model's residential choice probabilities to derive a summary measure of the relative attractiveness of locations based on the

observed share of residents and the price of floor space. We show that travel access is more important than the residual of residential amenities in explaining variation in this summary measure of relative attractiveness, with a contribution of 56 percent compared to 44 percent. In a special case of our model excluding consumption trips, we find a substantially smaller contribution from travel access of 37 percent, suggesting that conventional measures of amenities may in part capture consumption access, and highlighting the usefulness of smartphone data in measuring consumption trips that are otherwise hard to observe.

Finally, we show how the model can be used to undertake counterfactuals for changes in transport infrastructure. We compare the model's counterfactual predictions for the opening of a new subway line in the city of Sendai to the observed impact in the data. We show that our model incorporating consumption access generates a similar pattern of estimated treatment effects as in the observed data. We show that focusing solely on commuting trips leads to an underestimate of the welfare gains from the transport improvement by around one half, because of the substantial undercounting of trips that results from abstracting from the many other reasons besides commuting why individuals travel within urban areas.

Taken together, our findings suggest that access to consumption opportunities as well as access to employment opportunities plays a central role in understanding the spatial concentration of economic activity.

## References

- AGARWAL, S., J. B. JENSEN, AND F. MONTE (2020): “Consumer Mobility and the Local Structure of Consumption Industries,” *NBER Working Paper*, 23616.
- AHLFELDT, G., AND E. PIETROSTEFANI (2019): “The Economic Effects of Density: A Synthesis,” *Journal of Urban Economics*, 111, 97–107.
- AHLFELDT, G., S. REDDING, D. STURM, AND N. WOLF (2015): “The Economics of Density: Evidence from the Berlin Wall,” *Econometrica*, 83(6), 2127–2189.
- ALLEN, T., C. ARKOLAKIS, AND X. LI (2017): “Optimal City Structure,” Yale University, mimeograph.
- ALLEN, T., S. FUCHS, S. GANAPATI, A. GRAZIANO, R. MADERA, AND J. MONTORIOL-GARRIGA (2020): “Is Tourism Good for Locals? Evidence from Barcelona,” Dartmouth College, mimeograph.
- ALMAGRO, M., AND T. DOMÍNGUEZ-IINO (2019): “Location Sorting and Endogenous Amenities: Evidence from Amsterdam,” NYU, mimeograph.
- ANAS, A., AND Y. LIU (2007): “A Regional Economy, Land Use, And Transportation Model (RELU-TRAN): Formulation, Algorithm Design, and Testing,” *Journal of Regional Science*, 47(3), 415–455.
- ATHEY, S., B. FERGUSON, M. GENTZKOW, AND T. SCHMIDT (2018): “Experienced Segregation,” Stanford University, mimeograph.
- ATKIN, D., K. CHEN, AND A. POPOV (2021): “The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley,” MIT, mimeograph.
- ATKIN, D., B. FABER, AND M. GONZALEZ-NAVARRO (2018): “Retail globalization and household welfare: Evidence from Mexico,” *Journal of Political Economy*, 126(1), 1–73.
- BALBONI, C. (2019): “In Harm’s Way: Infrastructure Investments and the Persistence of Coastal Cities,” MIT, mimeograph.
- BAUM-SNOW, N. (2007): “Did Highways Cause Suburbanization?,” *Quarterly Journal of Economics*, 122(2), 775–780.
- BRODA, C., AND D. E. WEINSTEIN (2006): “Globalization and the Gains from Variety,” *The Quarterly Journal of Economics*, 121(2), 541–585.
- BÜCHEL, K., M. V. EHRLICH, D. PUGA, AND E. VILADECANS (2020): “Calling from the Outside: The Role of Networks in Residential Mobility,” *Journal of Urban Economics*, 119(103277).
- COUTURE, V. (2016): “Valuing the Consumption Benefits of Urban Density,” Vancouver School of Economics, mimeograph.
- COUTURE, V., J. DINGEL, A. GREEN, AND J. HANDBURY (2019): “Quantifying Social Interactions Using Smartphone Data,” Chicago Booth School of Business, mimeograph.
- COUTURE, V., C. GAUBERT, J. HANDBURY, AND E. HURST (2019): “Income Growth and the Distributional Effects of Urban Spatial Sorting,” University of California, Berkeley, mimeograph.
- DAVIS, D. R., J. I. DINGEL, J. MONRAS, AND E. MORALES (2019): “How Segregated Is Urban Consumption?,” *Journal of Political Economy*, 127(4), 1684–1738.
- DAVIS, M. A., AND F. ORTALO-MAGNÉ (2011): “Housing Expenditures, Wages, Rents,” *Review of Economic Dynamics*, 14(2), 248–261.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980–2000,” 106(3), 479–524.
- DINGEL, J., AND F. TINTELNOT (2020): “Spatial Economics for Granular Settings,” *NBER working Paper*, 27287.
- DOLFEN, P., L. EINAV, P. J. KLENOW, B. KLOPACK, J. D. LEVIN, L. LEVIN, AND W. BEST (2019): “Assessing the Gains from E-Commerce,” *NBER Working Paper*, 25610.
- DONALDSON, D. (2018): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, 108(4-5), 899–934.

- DONALDSON, D., AND R. HORNBECK (2016): "Railroads and American Economic Growth: A Market Access Approach," *Quarterly Journal of Economics*, 131(2), 799–858.
- DURANTON, G., AND M. A. TURNER (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79(4), 1407–1440.
- FABER, B. (2014): "Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country," *Review of Economic Studies*, 81(3), 1046–1070.
- FALLY, T. (2015): "Structural Gravity and Fixed Effects," *Journal of International Economics*, 97(1), 76–85.
- GEORGE, H. (1879): *Progress and Poverty*. Doubleday, Page & Company, New York.
- GLAESER, E. L., J. KOLKO, AND A. SAIZ (2001): "Consumer City," *Journal of Economic Geography*, 1(1), 27–50.
- GORBACK, C. (2020): "Ridesharing and the Redistribution of Economic Activity," National Bureau of Economic Research, mimeograph.
- GUPTA, A., C. KONTOKOSTA, AND S. VAN NIEUWERBURGH (2020): "Take the Q Train: Value Capture of Public Infrastructure Projects," *NBER Working Paper*, 26789.
- HEBLICH, S., S. REDDING, AND D. STURM (2020): "The Making of the Modern Metropolis: Evidence from London," *Quarterly Journal of Economics*, 135(4), 2059–2133.
- HOELZLEIN, M. (2020): "Two-sided Sorting and Spatial Inequality in Cities," Berkeley, mimeograph.
- HSIEH, C., E. HURST, C. I. JONES, AND P. J. KLENOW (2019): "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 87(5), 1439–1474.
- KREINDLER, G., AND Y. MIYAUCHI (2019): "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records," Boston University, mimeograph.
- LEONARDI, M., AND E. MORETTI (2019): "Agglomeration of Urban Amenities: Evidence from Milano Restaurants," .
- MCFADDEN, D. (1974): "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3(4), 303–328.
- MELO, P. C., D. J. GRAHAM, AND R. B. NOLAND (2009): "A Meta-analysis of Estimates of Urban Agglomeration Economies," *Regional Science and Urban Economics*, 39(3), 332–342.
- MICHAELS, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90(4), 683–701.
- MONTE, F., S. REDDING, AND E. ROSSI-HANSBERG (2018): "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 108(12), 3855–3890.
- SAIZ, A. (2010): "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125(3), 1253–1296.
- SAMUELS, P., N. HAUSMAN, M. COHEN, AND R. SASSON (2016): "Urban Pull: The Roles of Amenities and Employment," Harvard University, mimeograph.
- SANTOS SILVA, J., AND S. TENREYRO (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641–658.
- SEVEREN, C. (2019): "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification," Federal Reserve Bank of Philadelphia.
- STOREYGARD, A. (2016): "Farther on Down the Road: Transport costs, Trade and Urban Growth," *Review of Economic Studies*, 83(3), 1263–1295.
- TSIVANIDIS, N. (2018): "The Aggregate And Distributional Effects Of Urban Transit Infrastructure: Evidence From Bogotá's TransMilenio," University of California, Berkeley, mimeograph.
- ZÁRATE, R. D. (2020): "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City," University of California, Berkeley.