Appendix

A Data Construction

A.1 Household Travel Survey Data

Here we describe the procedures we used to clean the 2010 and 2014 Beijing Household Travel Surveys (BHTS), geocode home and work addresses, and construct commuting routes for trips in the BHTS data and hypothetical trips in the mortgage data. The notation in this appendix follows as closely to that in the main text as possible.

BHTS is designed to be representative using a multistage cluster sampling of households in Beijing. In the first stage, BTRC randomly selects a subset of Traffic Analysis Zones (TAZs) from the entire city. TAZs are one to two square km on average and their size is inversely proportional to the density of trip origins and destinations: the TAZs are smaller closer to the city center. For the first stage of sampling, BTRC selected 642 out of 1,191 TAZs in 2010 and 667 out of 2050 TAZs in 2014, respectively. In the second stage, about 75 and 60 households were randomly selected for in-person interviews for each TAZ.

BHTS includes detailed individual and household demographic (e.g., income, household size, vehicle ownership, home ownership, age, gender) and occupations, availability of transportation options (vehicles, bikes, etc.), and a travel diary on all trips taken during the preceding 24 hours. Household income is reported in bins and we use bin midpoints to measure income.

We focus on six commuting modes: *Walk*, *Bike*, *Bus*, *Subway*, *Car*, and *Taxi*. In principle, a traveler could take arbitrary combinations of different travel modes. In our data, single-mode trips account for over 95% of all trips. We therefore eschew multi-mode commuting trips, except for subway and bus trips where we allow commuters to walk to and from subway stations and bus stops.

We use Baidu's API to geocode addresses because its quality of matching Chinese character strings is higher than alterantive APIs such as Google Maps. We found that Baidu's Geocoding API performed best for home addresses and its Place API performed best for work addresses. About 36% of 2010 respondents and 44% of 2014 respondents are dropped because their home or work addresses cannot be geocoded.

The 2010 survey contains 46,900 households, 116,142 individuals, and 253,648 trips, while the 2014 survey contains 40,005 households, 101,827 individuals, 205,148 trips. We dropped trips with the origin or destination that could not be geocoded (40%), trips on weekends and holidays (10%), trips of non-working aged respondents (age> 65 or age <16, 12%), trips using mixed travel modes among subway, bus, and driving (3%), and trips with implausible trip distance and travel time (3%). The remaining sample includes 78,246 trips by 29,770 individuals in the year 2010 and 98,730 trips by 38,829 individuals in the year 2014. The analysis in the main text focuses on work commuting trips, with a total of 73,154 observations.

The size of the choice set varies across commuters and trips. Driving is available for households with personal vehicles on non-restricted days. Car rental is uncommon in Beijing and the mode share of rental cars is nearly zero in the travel survey. Walk, bike, taxi, and subway modes are available for all trips. We assume that households walk to/from the nearest subway stations if they take subways. Bus availability is determined by the home and work locations. We remove bus from the choice set if Gaode Maps API fails to provide any bus route, indicating a lack of the public bus service in the vicinity.

The monetary cost for walking is zero. For biking, the cost is zero if a household has a bike and the rental price (free for the first hour and then ¥1 per hour with ¥10 as the maximum payment for 24 hours) otherwise. The bus fare is set by the municipality at 0 for senior citizens, ¥0.2 for students, ¥0.4 for people with public transportation cards, and ¥1 for people without public transportation cards. The subway cost per trip is set by the public transport authority at ¥2 and adjusted by the type of public transportation card the traveler holds. Fuel cost is a major component of the monetary cost associated with driving. Based on the average fuel economy reported by vehicle owners in BHTS, we use 0.094 liter/km (10.6 km/liter) for 2010, and 0.118 liter/km (8.5 km/liter) for 2014. Gasoline prices are ¥6.87/liter in 2010 and ¥7.54/liter in 2014. We also assume a tear-and-wear cost which is 0.3 yuan per km. In 2010, the taxi charge is ¥10 for the first 3 km, ¥2 for each additional km, and ¥1 for the gasoline fee. In 2014, the charge increases to ¥13 for the first 3 km, ¥2.3 for each additional km plus ¥1 for the gasoline fee.

The construction of the travel time and distance via API and GIS is illustrated in Appendix Figure A4. To take into consideration differences between peak and off-peak traffic, we queried Baidu and Gaode API at the same departure time on the same weekday as that reported in the survey to obtain driving time predictions. Then we use historic levels of Beijing's Travel Congestion Index (TCI) to adjust the travel time for driving, taxi, and bus to the relevant historical years. We use ring-road specific TCIs for the travel time adjustment.

For subway commuting, we identified the nearest subway stations to home and to work using ArcGIS maps of the historical subway stations and used Baidu's API to calculate walking distances and time from home and work to the nearest subway station. For the BHTS travel survey data, the subway commuting time is calculated using the historical subway system at the time of the survey, including additional time when transferring lines. For hypothetical trips considered by home buyers in the mortgage data, we assume buyers are forward-looking and use the subway network two years after the home purchase date. This is because the subway construction goes through a lengthy process and it takes a few months to a few years from the the public announcement of subway station locations to the actual operation. Households are likely to be aware of new subway stations in the near future and we allow households to consider this in their purchase decisions. We also conduct a robustness check using a one-year projection window in constructing subway time and obtain similar results. Appendix Figure A5 shows travel time and cost of six routes for a particular trip based on the procedure.

The constructed travel distances and reported travel distances of chosen modes in the final dataset are highly correlated (correlation= 0.81). Correlation is highest among walking trips (0.99), followed by bicycle trips (0.98), subway trips (0.94), bus trips (0.88), car trips (0.61), and taxi trips (0.49).

A.2 Mortgage Housing Transaction Data

As part of the social safety net, the mortgage program aims to encourage home ownership by offering prospective homeowners mortgages with a subsidized interest rate. Similar to the retirement benefit, employees and employers are required to contribute a specific percentage of the employee's monthly wage to a mortgage account under this program. The savings contributed to this account can only be used for housing purchases and rental. Workers with formal employment were eligible for this government-backed mortgage program, upon which our data are based.

Although the mortgage data have a good representation of Beijing's middle-class, it under-represents low-income households without employment and high-income households who do not take loans for home purchases. To increase the representativeness of the mortgage data, we re-weight them based on two larger datasets that are more representative of home buyers in Beijing.

The first dataset include sales of new properties that are compiled from home registration records from Beijing Municipal Commission of Housing and Urban-Rural Development, accounting for 90% of all new home sales. It does not include employer-provided/subsidized housing. The second dataset includes 40% of all transactions in Beijing's second-hand market during our data period and is sourced from China's largest real estate brokerage company, Lianjia, that is present at all neighborhoods and across housing segments (Jerch et al., 2021). Different from the mortgage data, these datasets do not include information on the work location of the owners, therefore preventing us from using it for the main empirical analysis. To ease explanation below, we call these two larger transaction datasets "population dataset".

To improve the representativeness of the mortgage data, we match the distributions of housing price, size, age, and distance to the city center in the mortgage data to those in the population dataset using entropy balancing following Hainmueller (2012). Specifically, we solve the following constrained optimization problem to match sample moments between the mortgage data and the population dataset:[36]

$$\min_{w_i} H(w) = \sum_i h(w_i) = \sum_i w_i \log(w_i) \tag{A1}$$

subject to balance and normalizing constraints:

$$\frac{1}{N}\sum_{i \in \text{new homes}} w_i (X_{ij}^{mortgage} - \mu_j^{mortgage})^r = E_{\text{new homes}}[(X_j - \mu_j)^r]$$
$$\frac{1}{N}\sum_{i \in \text{resales}} w_i (X_{ij}^{mortgage} - \mu_j^{mortgage})^r = E_{\text{resales}}[(X_j - \mu_j)^r]$$
$$\sum_i w_i = N = \text{total number of new homes + resales in the mortgage data}$$
$$w_i \geq 0 \text{ for all } i.$$
$$\frac{\sum_{i \in \text{new homes}} w_i}{\sum_{i \in \text{resales}} w_i} = E\left[\frac{\text{new homes}}{\text{resales}}\right]$$

[36]The objective function, h, is a special case of a Kubelock divergence function, where the base weight, which w_i within the logarithm is divided by, is set to 1.

Here w_i is property i's weight, $\sum_{i \in \text{new homes}} w_i (X_{ij} - \mu_j)^r$ is the rth order (weighted) moment of matching covariate X_j among new home transactions in the mortgage data, and $E_{\text{new homes}}[(X_j - \mu_j)^r]$ is the rth order moment of covariate X_j among new home transactions in the population dataset. Similarly, $\sum_{i \in \text{resales}} w_i (X_{ij} - \mu_j)^r$ is the rth order moment of covariate X_j among second-hand housing transactions in the mortgage data, and $E(X_j - \mu_j)^r_{\text{resales}}$ is the rth order moments among used properties in the population data.

Matching covariates include housing prices, sizes, building ages, and distances to city center. We match both the mean and variance ($r = 1$ and $r = 2$). The third constraint normalizes the sum of weights to the total number of homes in the mortgage dataset (which is N). The fourth constraint requires weights to be positive. The last constraint requires the ratio of new homes to resales to be the same as the official statistics provided by Beijing Municipal Commission of Housing and Urban-Rural Development. We solve for optimal weights, w_i^* using the `entropy` package in STATA. The reweighted mortgage data match the larger representative datasets quite well. In all the empirical analysis, we use the reweighted the mortgage data to better represent home buyers in Beijing.

B Theoretical Model

This appendix presents the monocentric city model discussed in Section 3. The city is linear with a fixed population of N_R rich and N_P poor households ($N_R + N_P = N$). All households work at the urban center (CBD) at location 0, where wage income for the rich is larger $y_R > y_P$. The rest of urban space is occupied by homes with lot size normalized to 1 (roads take up no space in this model with a linear city). Land rents are remitted to absentee landlords. Housing consumption (in square meters) is provided by perfectly competitive developers facing constant returns to scale. Beyond the residential area is agricultural land with rental value p_a. The model is a closed-city model with intracity but not intercity migration. Both assumptions could be relaxed without affecting the key predictions of the model.[37] Since both the population and land use per household are fixed, the location of the urban boundary $\bar{x}$ which reflects the overall city size is fixed and equal to the lot size multiplied by population.

We begin by assuming that the subway network covers the entire urban area and then relax this assumption when considering the role of public transportation infrastructure. While there could be many urban configurations, we focus on those where the city is segmented by household income and commuting modes, such as the one illustrated in Figure A11.

B.1 A Theoretical Sorting Model

Household Utility Maximization and Housing Demand Households consume two goods: a numeraire good c with a unitary price, and housing $q(x)$ which varies in quantity depending on the distance x from CBD. There are two commuting modes $m \in M = \{C, S\}$: personal vehicles C and subway S. Households maximize

[37] Brueckner (1987) provides an analysis of a monocentric city model with a perfectly competitive supply side for both a closed and open city.

utility given income y_d, fixed (θ_m) and variable commuting cost ($w_{d,m}$) that vary by mode and household income via differences in the value of time:

$$u_d = \max_{c_{d,m}, q_{d,m}} \quad u\left(c_{d,m}, q_{d,m}\right) \quad \text{s.t.} \quad c_d + p(x) \cdot q_d = y_d - \theta_m - w_{d,m}(x), \quad d = R, P; m \in M. \tag{A2}$$

The solution to (A2), $\{c^*_{d,m}, q^*_{d,m}\}_{d=R,P;m \in M}$ determines housing demand conditional on mode choice m. We rewrite the housing demand as the bid-rent equations for each income type $d = R, P$ conditioning on the mode choice $m \in M$:

$$p^*_{d,m}(x; w) \equiv \max_{c_{d,m}, q_{d,m}} \left\{ \frac{y_d - \theta_m - w_{d,m} - c_{d,m}}{q_{d,m}} \right\}. \tag{A3}$$

Travel Mode Choice Commuting is costly: the fixed cost and variable (per-kilometer) cost are θ_C and $w_{d,C}$ for personal vehicles and θ_S and variable cost $w_{d,S}$ for subway. Car commuting has higher fixed costs but lower variable costs (when there is no congestion) than subway commuting. Variable costs include time (monetized by the value of time, VOT) and pecuniary costs. Rich households have a higher VOT: $\nu_R > \nu_P$.

We ignore the potential congestion in subway transportation and assume subway commuting's variable cost is linear in distance:

$$w_{d,S}(x) = \nu_d / \xi \cdot x$$

where ν_d is the value of time for income group d and ξ is the average subway speed. Car commuting's variable cost depends on congestion. Congestion at a given location x in turn depends on the flow of vehicles $n_C(x)$ by rich and poor households from the urban boundary to that location:

$$n_C(x) = \int_x^{\bar{x}} \mathbf{1}_R\{m = C\}(s) ds + \int_x^{\bar{x}} \mathbf{1}_P\{m = C\}(s) ds. \tag{A4}$$

The indicator function $\mathbf{1}_d\{m = C\}(x)$ is equal to one if a household of type $d = R, P$ commutes via car (*i.e.*, $m = C$) from location x. The commute cost per unit of travel distance is given by:

$$t_{d,C}(x) = \nu_d \mathscr{C}\left(n_C(x)\right), \quad d = R, P \tag{A5}$$

The congestion function $\mathscr{C}(\cdot)$, which depends on the flow of car commuters $n_C(x)$, is measured in travel time per unit of travel distance with positive first and second derivative. Total variable car commuting costs for a household living at x is:

$$w_{d,C}(x) = \int_0^x t_{d,C}(s) ds. \tag{A6}$$

Market Clearing Conditions and Spatial Equilibrium For simplification, we assume consumption of the numeraire good is fixed, and that housing quantity varies by income group d and travel mode m but is invariant of x conditioning on d, m. This allows us to focus on the effect of congestion on the bid-rent curve across location x.

A spatial equilibrium is determined by a bid-rent function that is the envelope of willingness to pay across households, keeping the utility for each income type fixed at $\bar{u}_d$, $d = R, P$:

$$p^*(x) = \max_{d,m} \left\{ p\left(y_d - \theta_m - w_{d,m}(x), \bar{u}_d\right) \right\}. \tag{A7}$$

Equations (A2) - (A6) make clear the simultaneous determination of housing locations and traffic congestion across the city. Solving for congestion at any location x requires knowing the distribution of car users at all points in the city. The slope of the bid-rent function for car commuters is equal to the derivative of $\frac{w_{d,C}(x)}{q_{d,m}}$ with respect to x:

$$p'_{d,C}(x) = \frac{t_{d,C}(x)}{q_{d,m}}.$$

Subway commuters do not experience congestion. The slope of their bid-rent function is constant. In residential regions with car commuting, moving toward the city center means adding additional car commuters, further increasing per kilometer commuting time costs and steepening the bid-rent function.

The market clearing conditions for a spatial equilibrium are:

1. The equilibrium bid-rent at the urban boundary $p^*(\bar{x})$ must adjust to equate to the agricultural rent p_a.

2. Given identical preferences, all households in the same income group attain the same level of utility, $\bar{u}_d$:

$$u\left(c_d^*, q_d^*\right) = \bar{u}_d \quad \text{for all } x \in [0, \bar{x}], \quad d = R, P.$$

3. Congestion from car commuting at each location $x \in [0, \bar{x}]$ is determined by equation (A4).

4. Bid-rent curves across income and commuting types intersect at some $x_i \in (0, \bar{x}), i = 1, ..., 6$:

$$p_{R,S}(x_1) = p_{P,S}(x_1), p_{R,S}(x_2) = p_{R,C}(x_2), p_{R,S}(x_3) = p_{P,C}(x_3)$$
$$p_{P,S}(x_4) = p_{R,C}(x_4), p_{P,S}(x_5) = p_{P,C}(x_5), p_{R,C}(x_6) = p_{P,C}(x_6)$$

5. The market clearing bid-rent function is the envelope of conditional bid-rent functions across all income and commuting groups:

$$p^*(x) = \max_{d,m}\{p_{d,m}(x)\}, \quad d = R, P, \quad m = C, S.$$

   This envelope determines the pattern of residential locations as well as commuting choices that clear the housing market at each location $x \in (0, \bar{x})$. The commuting boundaries between each group are defined by the intersections of bid-rent curves across income and commuting types that determine the equilibrium bid-rent function. There must be at least 1 and at most 3 intersections for a spatial equilibrium to exist.

In equilibrium, the bid-rent function $p^*(x)$ and utility level $\bar{u}$ are determined endogenously based on the level of commuting cost t, incomes y_d, and agricultural rent p_a. Many urban configurations are possible, though we focus on a specific one. Given sufficiently high fixed costs for driving relative to subway, high variable costs for subway relative to driving, and large enough differences in the value of time between the rich and poor, a spatial configuration as in Figure A11 may emerge where a mass of rich households live closest to the CBD and commute by subway. Beyond this group, a mass of poor households also commute by subway, followed by a mass of rich households commuting by car who consume more housing than their subway commuting counterparts ($q_{R,S} < q_{R,C}$) to compensate for longer commuting. Finally a mass of car commuting poor households live at the urban boundary given their lower value time, but also consume more housing than their subway commuting counterparts ($q_{P,S} < q_{P,C}$). Bid-rents are steeper for the rich than for the poorer for each respective commuting mode because the rich have a higher value of time.

Effect of Congestion on Bid-Rent Functions The bid-rent functions for each transportation technology evaluated at the CBD ($x = 0$) are:

$$p^0_{d,m} = \frac{1}{q_{d,m}}\left[y_d - \theta_m - c_d\right], \quad d = R,P; m = C,S,$$

where $\frac{y_R}{q_{R,S}}$ is sufficient large compared to $\frac{y_P}{q_{P,S}}$ that $p^0_{R,S} > p^0_{P,S}$. Similarly, the fixed costs of driving are sufficiently high that $\frac{y_R-\theta_C}{q_{R,C}} > \frac{y_P-\theta_C}{q_{P,C}}$ so $p^0_{R,C} > p^0_{P,C}$, and both are smaller than those for subway.

The bid-rent function for subway riders is:

$$p_{d,S}(x) = \frac{1}{q_{d,S}}\left[y_d - \theta_S - c_d - v_d/\xi \cdot x\right], \quad d = R,P$$

where ξ is the average subway speed. The bid-rent function for drivers is:

$$p_{d,C}(x) = \frac{1}{q_{d,C}}\left[y_d - \theta_C - c_d - v_d\mathscr{C}\left(\int_x^{\bar{x}} \mathbf{1}_R\{m = C\}(s)ds + \int_x^{\bar{x}} \mathbf{1}_P\{m = C\}(s)ds\right)\right], \quad d = R,P,$$

where $\mathscr{C}()$ is an increasing, convex congestion function of car commuting.

Consider the effect of an exogenous increase in vehicle traffic congestion. Increases in congestion make the slope and curvature of the car commuting bid-rent functions adjust based on the extent of car commuting across the city. To make this clear, consider the changes in the slope of $p_{R,C}$ evaluated at x_B, the boundary between poor subway commuting and rich car commuting, when congestion increases from $n_C(x_B)$ to $n_C(x_B) + \varepsilon$:

$$\widetilde{p'_{R,C}}(x_B) = \frac{t_{R,C}(x_B + \varepsilon)}{q_{R,C}} = \frac{v_R}{q_{R,C}}\left[\mathscr{C}(n_C(x_B + \varepsilon))\right] > \frac{v_R}{q_{R,C}}\left[\mathscr{C}(n_C(x_B))\right] = p'_{R,C}(x_B),$$

where ε also corresponds to the mass of commuters living along that distance given the assumption of fixed lot size at 1 and $q_{R,C}$ does not vary with x as mentioned above. This demonstrates that a change in endogenous congestion in the model has two effects: it steepens the bid-rent curve overall, and the curve itself gets steeper after passing through residential areas with car commuters as the flow of vehicles onto the roadway builds up.

B.2 Effect of Transportation Policies on Urban Structure and Welfare

Now we consider the effect of transportation policies on the urban structure to motivate our empirical analysis in the main text. While there are many potential equilibrium spatial configurations depending on model parameters, we calibrate the parameters of the model to produce the patterns depicted in Panel (a) of Figure A11.[38] LeRoy and Sonstelie (1983) demonstrate that the urban configuration can be explained, in part, by the relation between the income elasticity of marginal commuting costs relative to the income elasticity of housing demand. The spatial pattern in the figure is consistent with the case where the income elasticity of marginal commuting costs is larger than the income elasticity of housing demand. therefore the rich outbid the poor to live close to the city center to save commuting costs.

Congestion Pricing. First we consider a typical first-best approach: a per-kilometer congestion charge. The optimal level would be equal to the marginal external cost of congestion and can be derived from differentiating (A6) with respect to $n_C(x)$ giving the increased cost from one additional car commuter. Multiplied by the number of car commuters, this yields:

$$\tau_C(x) = \left(v_R \frac{n_{R,C}(x)}{n_C(x)} + v_P \frac{n_{P,C}(x)}{n_C(x)} \right) \mathscr{C}' \left(n_C(x) \right). \tag{A8}$$

The revenue from congestion pricing can then be recycled lump sum to each resident of the city. Panel (b) of Figure A11 shows the effect of the congestion pricing on spatial equilibrium in this hypothetical city. Due to the recycling of the revenue, the bid-rent curves shift up for subway users. Under the uniform recycling of the revenue, the shift is larger for the poor than for the rich due to the smaller home size among the poor (the denominator of the intercept). The intercept of the bid-rent curves for car users moves up as well. For poor drivers, the slope of the bid-rent curve steepens as the congestion toll defined above would be larger than the savings from improved speed (due to their low VOT), hence leading to a higher travel cost per unit of distance (the numerator of the slope). For richer drivers, the bid recent curve would be flatter as the congestion toll would be smaller than the savings from improved speed (due to high VOT). However, the curve to the right of x'_B becomes steeper as it moves to x'_B from the right. This is due to the fact that congestion worsens as it is closer to x'_B from the right, leading to an increasing unit travel cost.

There are two competing forces at work that affect the spatial pattern of residential locations. First, congestion pricing increases the unit travel cost and incentivizes residents to move closer, hence bidding up home prices near the city center. Second, the reduction in congestion leads to time savings and reduces the

[38] The bid-rent curve is convex in the standard monocentric city model where the travel cost is assumed to linear in distance: prices do not need to fall as fast as the increase in travel cost to keep residents indifferent since they are compensated by living in a larger home the further they live away from the city center. The bid-rent curve becomes concave only if the travel cost is sufficiently convex in distance. To ease exposition, we draw linear bid-rent curves in the figures.

travel cost. However, due to the differences in VOT, the time saving is more valuable to the rich than to the poor. That is, the second force is relatively stronger compared to the first force for the rich than for the poor. Given the initial spatial configuration, congestion pricing results in some poor residents shifting away from driving to subway while moving their residence from the outer ring to the inner ring. At the same time, some rich residents move out of the inner ring and switch to driving while living in larger homes. If congestion pricing or the cost of driving increases enough, the poor may occupy the city center as in the case when driving is prohibitively expensive for the poor (LeRoy and Sonstelie, 1983).

Subway Expansion. Panel (a) of Figure A12 examines subway expansion by equivalently considering the effect of constraining transit to $\bar{x}_B$ in panel (a) of Figure A12. Given the fixed supply of housing, the constraint of public transit from x_B in the baseline to $\bar{x}_B$ shifts the mass of poor subway commuters $\bar{x}_B - x_B$ to the poor car commuting region. This increases congestion to the left of x_B for all car commuters, steepening the slope of the bid-rent curves to the left of x_B more for the rich than for the poor due to the high VOT among the rich. In the extreme case of removing all subway from the city, the bid-rent curve for the rich would become even steeper as it gets closer to the city as the congestion worsens. Only the rich will occupy the city center. The results are consistent with Glaeser et al. (2008) which calibrates this class of models to corroborate the narrative that better public transportation leads more poor to live in the city center.

Driving Restriction. Finally, we consider the effect of raising the cost of driving via a driving restriction. In practice, the driving restriction only bans a portion of the cars from driving each day. We assume that during those days, residents need to take the subway to work. This would imply that the car commuters would need to pay the fixed cost of two modes, while the variable cost would be a weighted sum of the two modes. Panel (b) of Figure A12 shows the effect of the driving restriction. Due to the increase in the fixed cost for car commuters, the intercept of the bid-rent curves shifts down, more so for the poor than for the rich (the denominator being larger for the rich). The change in the slope for the car commuters is subjects to two countervailing forces. On the one hand, the added variable cost (due to the higher variable cost when using the subway) will increase the slope. On the other hand, congestion reduction to the left of x'_B will reduce the slope, more so for the rich than for the poor. The first force likely dominants. The impact of the policy is that the rich reduce car commuting ($x'_A - x_A$) by more than the poor ($x'_C - x_C$).

Welfare. Given the stylized nature of this model, the welfare implications from this stylized example are probably less informative to real world policy applications than the empirical exercises performed in Section 7 in the main text. That said, Figures A11 and A12 yield an important observation: key welfare effects follow the movement up and down of equilibrium bid-rent functions from capitalization of changes in the transportation system. Put simply, an approximation to total welfare is the sum of equilibrium rents paid:[39]

[39] To be precise, this provides the sum of willingness to pay as reflected by equilibrium rents. As we consider the effects of different transportation policies on rents, their differences can be interpreted as an approximation to differences in welfare. The rationale for

$\int_0^{\bar{x}} p^*(s)ds$.

This can be visualized as the area under the equilibrium bid-rent envelope in the figures: when the envelope is lower than the baseline under a given policy scenario, the aggregate rent (and thus welfare) is lower, and visa versa. Considering Figure A12 panel (b), the effect of the driving restriction seems to lower the sum of rents by more than it is increased as the envelope is lower beyond x'_B. This accounts for a larger area than the increase in the level of the equilibrium bid-rent to the left of x'_B. In contrast, the overall effect of congestion pricing in Figure A11 panel (b) is to increase the equilibrium bid-rents envelope at nearly all points in the city because congestion reduction flattens bid-rents and the redistribution of the toll revenues increases their value for subway commuters. This points to an important general equilibrium effect of transportation policies in cities, where their benefits are capitalized into the housing market and provide additional welfare gain relative to a driving restriction beyond a partial equilibrium framework on the transportation sector as shown in Figure 3 in the main text. Our simulations suggest the welfare impacts of these capitalization effects (and subsequent sorting) may be larger than the direct effects on transportation choice itself.

C Reduced-form Evidence

This section presents reduced-form evidence of the impact of the car driving restriction policy (CDR) on the housing market. We examine the price gradient with respect to subway proximity as well as household sorting behavior.

Price Gradient w.r.t. Subway Proximity A number of confounding factors could undermine identification of the causal relationship between housing price gradients and the driving restriction policy. For example, if amenities improve over time in locations near subway stations more than in locations that are farther away from subway stations, this would result in a larger price increase for homes close to subway stations, leading to an overestimation of the true impact of the policy. On the other hand, if changes in dis-amenities such as congestion or noise follow the aforementioned pattern, we would underestimate the impact of the policy. Causal identification requires an assumption that the housing price gradient with respect to the distance to the nearest subway station would be unchanged in the absence of the driving restriction policy. We test the plausibility of this assumption by examining trends in price gradients in the periods leading up to the policy based on an event study framework:

$$\text{Price}_{jt} = \sum_{k=-24}^{24} \beta_k \times \text{Dist}_{jt} \times \mathbb{1}(t=k) + \mathbf{x}'_{jt}\gamma + \varepsilon_{jt} \tag{A9}$$

this approach follows the Henry George Theorem since transportation system investments can be thought of as spatially heterogeneous public goods (Stiglitz, 1977; Arnott and Stiglitz, 1979). If transportation policies create or remove welfare reducing distortions, this will be reflected in relative changes in the bid-rent curves. Albouy and Farahani (2017) show that the value of infrastructure could be underestimated using this approach and argue for a more broad method to incorporate imperfect mobility, federal taxes, and non-traded production.

where j denotes a home and t denotes a month. The outcome variable is the unit price (¥1000 per m^2). We allow the slope of the price gradient β's to vary over time. The regression includes a flexible set of controls ($\mathbf{x}_{jt}$) that include neighborhood fixed effects, year by month fixed effects, city district by year fixed effects. We also control for complex-level attributes such as complex age, the floor area ratio and green space ratio, the complex land area, the number of units and buildings in the complex, and home management fee (HOA fee). Standard errors are clustered at the neighborhood level to allow for correlations among the homes in the same neighborhood (e.g., due to unobservables).

Panel (a) of Appendix Figure A9 shows the coefficient estimates of β_k that vary by quarter. There does not appear to be a pre-existing trend before the policy, alleviating the concern of the time-varying and location-specific unobservables. While there is not a clear relationship between subway proximity and housing price before the policy, there is a clear downward shift in the slope of price gradient. Moreover, the negative relationship between subway distance and housing price becomes stronger over time. The increasingly larger impact over time after the policy could be driven by the fact that the policy uncertainty is reduced over time and enforcement is tightened gradually.

One additional concern for identification is that subway expansion may result in network externalities so that the benefit of a new station occurs not just to those living or working nearby but to all who use that station. Since subway construction takes a long time with a significant lead time of public announcements, there should have been a steepening of the slope *before* the driving restriction if our results are driven by the subway expansion. To further address this issue, we include in the regressions a measure of subway density that is constructed as the inverse distance weighted number of subway stations from a given location following Li et al. (2019). This measure can be considered as the number of subway stations per unit area centered around a given housing unit and it increases as the subway network expands.

Appendix Table A1 provides the regression results for seven specifications. The parameter of interest is the interaction between subway distance and the policy dummy. Adding neighborhood fixed effects to control for neighborhood amenities significantly changes the main coefficient on subway distance and the interaction coefficient from Column (1) to Column (2). Further including district by year fixed effects and the rich set of complex-level variables in Columns (3) and (4) barely changes the results. Column (4) corresponds to the specification for the event study in Panel (a) of Appendix Figure A9. Column (5) is a weighted regression to make the sample more representative of the universe housing transactions in Beijing, where the weighting procedure is described in Section A.2. The last two columns include the subway density measure. The results from different specifications are qualitatively the same: the driving restriction increases the price premium for homes that are closer to subway.

Panel (b) of Appendix Figure A9 provides a falsification test by randomizing the treatment status (before or after the driving restriction policy) of each transaction while keeping the share of post-policy transactions fixed. The figure shows the histogram of the coefficient estimates of the interaction between distance and treatment dummy from 500 iterations. The estimate from the true sample (-0.075) lies far away from the histogram. This alleviates the concern that the estimated impact might be driven by unobservables.

Our analysis so far assumes a linear relationship between housing price and subway proximity. To relax this assumption, we specify a piece-wise linear relationship between housing pricing and subway distance. Appendix Figure A10 presents the price gradient estimates for three distance bands for the pre- and post-policy periods separately. The marginal impact of subway proximity is the strongest for properties within five km but tapers offer after ten km, implying a convex price function. In addition, consistent with our previous analysis, the relationship between subway proximity and housing price is strengthened after the policy for all distance bands but especially for properties within five to ten km.

D Estimation Details

Estimating Travel Mode Choices To recover preference parameters in travel model choices, we use simulated maximum likelihood. The notation follows that in Section 4.2 of the main text. The log likelihood function is defined as:

$$\begin{aligned} ln\mathscr{L}(\gamma,\eta,\theta) &= \sum_i \sum_{m=1}^{M_i} \mathbb{I}_m^i lnR_{ijm}(\gamma,\eta,\theta) \text{ where} \\ R_{ijm}(\gamma,\eta,\theta) &= \frac{1}{H}\sum_{h=1}^{H} \frac{\exp\left(\bar{u}_{ijmh}(\gamma,\eta,\theta)\right)}{\sum_k \exp\left(\bar{u}_{ijkh}(\gamma,\eta,\theta)\right)} \end{aligned}$$

where i denotes survey respondents, j denotes home locations, m is the travel mode, and M_i includes modes available to commuter i. The indicator function $\mathbb{I}_j^i$ takes value one if commuter i chooses travel mode m. The mode choice probability is denoted by R_{ijm} and calculated by averaging over $H = 100$ vectors of Halton simulation draws. Utility $\bar{u}_{ijmh}(\gamma,\eta,\theta)$ is similar to equation (3) but without the error term: $\bar{u}_{ijmh} = \theta_{imh} + \gamma_{1ih}\cdot time_{ijm}(v_{ij}) + \gamma_2 \cdot cost_{ijm}/y_i + \mathbf{w}_{ijm}\eta$, where h (and $\{\theta_{imh},\gamma_{1ih}\}$) refers to the hth Halton draw. In the estimation, we leverage the fact that we observe the round trips for work commute and use the same simulation draw for the two trips by the same commuter to capture the mode-specific preference for each individual.

The parameter estimates $\hat{\gamma},\hat{\eta},\hat{\theta}$ maximize the simulated log-likelihood defined above. Once we have these estimates for the commuting preference, we plug them in the housing transaction data and calculate ease-of-commuting EV for each i and j pair based on the observed housing and job locations via:

$$EV_{ij}^t = \frac{1}{H}\sum_{h=1}^{H} \log\left(\sum_m \exp\left(\bar{u}_{ijmh}(\hat{\gamma},\hat{\eta},\hat{\theta})\right)\right)$$

and include it as a housing attribute in the estimation of housing demand below. Importantly, we construct the EV term separately for husband and wife and include both in the housing demand.

Estimating Housing Demand The housing demand model is estimated using a two-step procedure. The notation follows Section 4.4 of the main text. Nonlinear parameters θ_1 is estimated via simulated Maximum

Likelihood with a nested contraction mapping in the first step, and linear parameters θ_2 are estimated in the second step via linear IV/GMM. The log likelihood function is defined as:

$$
\begin{aligned}
ln\mathscr{L}(\theta_1,\delta_j) &= \sum_i\sum_j \mathbb{I}^i_j w_i lnP_{ij}(\theta_1,\delta_j), \text{ where} \\
P_{ij}(\theta_1,\delta_j) &= \frac{1}{Q}\sum_{q=1}^{Q}\frac{\exp[\mu_{ijq}(\theta_1)+\delta_j]}{\sum_k \exp[\mu_{ikq}(\theta_1)+\delta_k]},
\end{aligned}
$$

The indicator function $\mathbb{I}^i_j$ takes value one if household i chooses housing j and w_i is the weight of household i (obtained from the entropy balancing in Section A.2 to make the mortgage data more representative of home buyers in Beijing). The housing demand choice probability is denoted as P_{ij} and calculated by averaging over $Q=200$ vectors of Halton simulation draws. Parameters θ_1 are non-linear parameters (since the log-likelihood function is a nonlinear function of these parameters) that characterize household preference heterogeneity.

We search for θ_1 and population average utilities δ_j to maximize the log-likelihood function, subject to the constraint that the model predicted housing demand based on δ_j can replicate observed housing demand (as in Berry et al. (1995)):

$$
\begin{aligned}
&\max_{\theta_1,\{\delta_j\}_j} ln\mathscr{L}(\theta_1,\delta_j) \qquad \text{s.t.} \\
\sum_{i\in C^{-1}(j)}\frac{1}{Q}\sum_{q=1}^{Q} w_i \frac{\exp[\mu_{ijq}(\theta_1)+\delta_j]}{\sum_k \exp[\mu_{ikq}(\theta_1)+\delta_k]} &= w_j,\ \forall j\in J
\end{aligned}
\tag{A10}
$$

The left-hand-side of constraint (A10) is model predicted housing demand for property j. The first summation $\sum_{i\in C^{-1}(j)}$ aggregates simulated choice probabilities over all households whose choice set contains property j. The second summation averages over $Q=200$ vectors of Halton simulations draws to simulate household i's probability of choosing property j. The right-hand-side is the observed housing demand for property j (which is 1 weighted by the entropy weight w_j).

We follow the literature and use the following contraction mapping to solve for $\{\delta_j\}_j$ that satisfies constraint (A10):

$$
\begin{aligned}
\delta_j^{d+1} &= \delta_j^d + \ln(w_j) - \ln D_j(\theta_1,\delta_j^d), \text{where} \\
D_j(\theta_1,\delta_j^d) &= \sum_{i\in C^{-1}(j)}\frac{1}{Q}\sum_{q=1}^{Q} w_i \frac{\exp[\mu_{ijq}(\theta_1)+\delta_j^d]}{\sum_k \exp[\mu_{ikq}(\theta_1)+\delta_k^d]}
\end{aligned}
$$

where $d+1$ is the $d+1$-th iteration, w_j is observed property j's demand, and $D_j(\theta_1,\delta_j^d)$ is model-predicted demand in iteration d. Our model assumes a closed city – all households choose a place to live in Beijing – and there is no outside option, hence we normalize the population-average utility of the home with the lowest unit price to zero (results are the same) regardless of which home we use for the normalization). As the unobserved housing attributes ξ_j that are correlated with price and EV terms are absorbed by property fixed effects δ_j, the simulated MLE produces consistent estimates on household specific parameters θ_1.

After obtaining $\hat{\theta}_1$ and $\{\hat{\delta}_j\}_j$, we recover the linear parameters θ_2 via IV:

$$\hat{\delta}_j(\theta_2) = \alpha_1 p_j + \mathbf{x}_j \bar{\beta} + \xi_j$$

where the IVs are discussed in Section 4.4 of the main text.

E The Simulation Algorithm

Simulation Algorithm The notation in this section follows that in the main text. Household-trip characteristics $\{\mathbf{w}\}$, housing attributes $\{\mathbf{X}\}$, and commuting distance are fixed at the observed level. Demand parameters are denoted by: $\{\gamma_1, \gamma_2, \eta, \beta, \alpha, \phi, \theta, \xi\}$. We fix the random Halton draws throughout the simulation.

We now describe the algorithm for counterfactual simulations. The goal of each simulation is to find the equilibrium vector of traffic density and housing prices. The simulation algorithm has both the outer loop and the inner loop. The outer loop searches for driving speeds and traffic density that clear the traffic sector, while the inner loop searches for housing prices that clear the housing market. In counterfactual analyses that shut down sorting, there is no inner loop.

The algorithm starts with an initial vector of housing price $\mathbf{p}^0$ and driving speed $\mathbf{v}^0$ (and traffic density d^0). We use the observed levels as the starting point. Repeat the following steps until convergence:

1. Based on density level d^t and price vector $\mathbf{p}^t$ for iteration t (d^0 and $\mathbf{p}^0$ for the first iteration):

(a) Update the driving speed for every household:

$$\frac{v_{ij}^t - v_{ij}^0}{v_{ij}^0} = e^v * \frac{d^t - d^0}{d^0},$$

where v_{ij}^t is updated driving speed for household i's work commute from home j and e^v is the speed-density elasticity -1.1.

(b) Use v_{ij}^t from step (a) to revise each commuter's driving time: $time_{ijk,drive}^t = \frac{dist_{ijk,\text{drive}}}{v_{ijk}^t}$ (as well as the commuting time via taxi: $time_{ijk,taxi}^t$), where k denotes member k in household i;

(c) Update the ease-of-commuting measure EV for every household member:

$$EV_{ijk}^t = \frac{1}{H}\sum_{h=1}^{H} \log\left(\sum_m \exp\left(\theta_{hm} + \gamma_{1ih} time_{ijk,m}^t + \gamma_2 \frac{\text{cost}_{ijk,m}}{\text{hourly wage}_{ik}} + \mathbf{w}_{ijk,m}\eta\right)\right)$$

where h is a Halton simulation draw for the random coefficient of travel time and the random coefficients of each travel mode, m stands for travel mode, and θ_{hm} and γ_{1ih} are simulated random coefficients for travel time and mode dummies.

In a similar manner, update member k's driving probability:

$$R^t_{ijk,driving} = \frac{1}{H}\sum_{h=1}^{H}\frac{\exp\left(\theta_{h,drive}+\gamma_{1ih}time^t_{ijk,drive}+\gamma_2\frac{\text{cost}_{ijk,drive}}{\text{hourly wage}_{ik}}+\mathbf{w}_{ijk,drive}\eta\right)}{\sum_m \exp\left(\theta_{hm}+\gamma_{1ih}time^t_{ijk,m}+\gamma_2\frac{\text{cost}_{ijk,m}}{\text{hourly wage}_{ik}}+\mathbf{w}_{ijk,m}\eta\right)}$$

If the counterfactual analysis incorporates sorting, continue with steps (d)-(e). Otherwise skip them and move to step (f);

(d) Given the updated EV term, search for a new housing price vector $\mathbf{p}^t$ that clears the housing market with housing demand equal to housing supply:

$$\sum_{i\in C^{-1}(j)}\frac{1}{Q}\sum_{q=1}^{Q}w_i\frac{\exp\left(\alpha_i p^t_j+\mathbf{X}_j\beta_i+\sum_k\phi_{kq}EV^t_{ijk}+\xi_j\right)}{\sum_{s\in C(i)}\exp\left(\alpha_i p^t_s+\mathbf{X}_s\beta_i+\sum_k\phi_{kq}EV^t_{isk}+\xi_s\right)} = S_j,\ \forall j\in J \tag{A11}$$

where the left-hand-side (LHS) is the simulated demand for property j and the right-hand-side is the housing supply S_j (under fixed housing supply, S_j is fixed at w_j, the weight for property j). The first summation of the LHS is over households whose choice set includes property j, denoted as $C^{-1}(j)$. The second summation aggregates over $Q = 200$ Halton draws of random coefficients. Each household has weight w_i to make the sample more representative of home buyers in Beijing. The coefficients ϕ_{kq} are member k's simulated random coefficients for the EV term.

As the model is a closed city with no outside options, the market clearing condition pins down the housing price vector $\mathbf{p}^t$ up to a constant. We normalize the average housing price (the mean of vector $\mathbf{p}^t$) to be the same as the average price observed in the sample.

For counterfactual analyses that allow the housing supply to respond to housing price changes, we use a constant supply-price elasticity at 0.53 following Wang et al. (2012):

$$\frac{S^t_j-w_j}{w_j} = 0.53*\left(\frac{p^t_j-p^0_j}{p^0_j}\right)$$

Solving equilibrium price vector $\mathbf{p}^t$ requires iterating the market clearing condition (A11) many times. This is the inner loop as illustrated below.

(i) Suppose that the price vector is $\mathbf{p}^l$ in iteration l ($l = 1$ for the first iteration);

(ii) Update the price vector:

$$\mathbf{p}^{l+1}_j = \mathbf{p}^l_j+[log(S_j)-log(D^l_j(\mathbf{p}^l_j))]/k$$

where $\mathbf{p}^{l+1}$ is the updated price vector, S_j is the observed housing supply, $D^l_j(\mathbf{p}^l_j)$ is the model predicted demand in iteration l (the LHS of equation (A11)), and k is a pre-set constant that controls the step size of each iteration.

(iii) If $\left\|\mathbf{p}^{l+1}-\mathbf{p}^l\right\| < \varepsilon^{\text{inner}}_{tol}$ where ε_{tol} is a pre-set tolerance level, stop. Otherwise, return to step (ii). In our simulation, this algorithm always converges. Let $\mathbf{p}^t = \mathbf{p}^{l+1}$, the fixed point from the inner loop.

(e) At the new equilibrium housing price $\mathbf{p}^t$, revise the housing demand choice probability:

$$P_{ij}^t = \frac{1}{Q}\sum_{q=1}^{Q} \frac{\exp\left(\alpha_i p_j^t + \mathbf{X}_j\beta_i + \sum_k \phi_{kq} EV_{ijk}^t + \xi_j\right)}{\sum_{s\in C(i)} \exp\left(\alpha_i p_s^t + \mathbf{X}_s\beta_i + \sum_k \phi_{kq} EV_{isk}^t + \xi_s\right)}$$

(f) Update the traffic density using the revised probability to drive and take taxis (and the revised probability that household i chooses property j in the case of sorting):

$$\tilde{d} = \sum_i w_i \sum_j P_{ij}^t \left[\sum_k (R_{ijk,drive}^t \times dist_{ijk,\text{drive}} + R_{ijk,taxi}^t \times dist_{ijk,\text{taxi}})\right]$$

where the summation inside the large brackets $[\cdot]$ aggregates over commuting member k within each household. If sorting is shut down, the housing demand choice probability P_{ij}^t is fixed at initial levels.

2. If $\left\|\tilde{d} - d^t\right\| < \varepsilon_{tol}^{\text{outer}}$ where $\varepsilon_{tol}^{\text{outer}}$ is a pre-set tolerance level, stop. Otherwise, set $d^{t+1} = \varphi d^t + (1-\varphi)\tilde{d}$ for some $\varphi \in (0,1)$ and return to step 1.

Ring-specific Density The discussion above assumes that the traffic density is city-wide. To accommodate ring-specific traffic density, we start with apportioning each trip to the relevant ring road segment according to the radius of each ring. For example, if the radius of ring road segment between the kth and k+1th ring is $r_{k,k+1}$, a trip that originates in ring 5 and ends in ring 3 will contribute $\frac{r_{3,4}}{r_{3,4}+r_{4,5}}$ to the 3rd-4th ring road traffic density and $\frac{r_{4,5}}{r_{3,4}+r_{4,5}}$ to the 4th-5th ring road traffic density. Accordingly, we break down the driving distance of each trip $dist_{ijk,\text{drive}}$ to the ring-specific driving distance $\{dist_{ijks,\text{drive}}\}_s$, where s denotes a ring-road segment.

The simulation algorithm is the same as above, except that we now have ring-road specific traffic speed v_{ijs}^t and ring-road specific traffic density d_s. Specifically, in step (a) above, we update the ring-road specific speed v_{ijs}^t via:

$$\frac{v_{ijs}^t - v_{ijs}^0}{v_{ijs}^0} = e^v * \frac{d_s^t - d_s^0}{d_s^0}.$$

In step (b) above, Use v_{ijs}^t from step (a) to revise each member k's driving time spent in ring-road segment s. The total driving time is the sum of the ring-road specific driving time: $time_{ijk,drive}^t = \sum_s \frac{dist_{ijks,\text{drive}}}{v_{ijks}^t}$ (as well as the commuting time via taxi: $time_{ijk,taxi}^t$). In step (f) above, we aggregate traffic demand to ring-road specific density measures as follows,

$$\tilde{d}_s = \sum_i \sum_j P_{ij}^t \left[\sum_k (R_{ijk,drive}^t \times dist_{ijks,\text{drive}} + R_{ijk,taxi}^t \times dist_{ijks,\text{taxi}})\right]$$

The rest of the simulation steps remain the same.

Assumptions for the Welfare Analysis We make some assumptions in the welfare analysis. First, we use a 30-year time horizon for our welfare analysis. The assumption on the time horizon affects the welfare magnitude (a shorter period is associated with lower welfare gains), but does not affect the comparison across different transportation policies. Second, to be conservative, we only include commuting trips (which account for 60% of all trips and 75% of total travel distance in 2014) and ignore non-commuting trips in calculating the benefits of subway expansion. Lastly, we abstract away from distortionary taxes and assume instead that subways' construction costs are financed via a non-distortionary uniform head tax. Similarly, when congestion toll revenues are recycled, they are distributed evenly across households in a lump sum manner.

To facilitate comparison, we calibrate the congestion charge at ¥1.13/km to achieve the same level of congestion reduction as under driving restriction with the 2008 subway network. We include (lifetime) subway and bus fares as part of the government revenue in the counterfactual analyses, though they account for a negligible fraction of the total welfare. For example, the subway fare is ¥2 per ride. Total lifetime subway fares paid by a household is roughly ¥2,400 (¥2 per ride * 8% likelihood to take subway * 500 rides/year * 30 years), much smaller than the net welfare per household as reported in the main text. Lastly, consumer surplus reported in the counterfactual analyses refers to the discounted lifetime consumer surplus over the property tenure, as discussed in Section 5.2.

We account for both the capital and operating costs of the congestion pricing system based on the system in Singapore. Singapore's electronic road pricing scheme launched in 1998 costed $110 million to set up with an annual operating cost of $18.5 million.[40] Singapore is upgrading the system to be satellite-based and the setup cost would be about $370 million but the operating cost is expected to be lower than the current system. We assume that the congestion pricing system in Beijing would be satellite-based and we scale both the setup and operating costs up by the population of Beijing relative to that of Singapore. To facilitate policy comparison, we calculate the 30-year total discounted cost per household (assuming 7.2 mil. households in Beijing). This amounts to ¥3000, about 2.5% of the total toll revenue per household during the same period.

We include the construction and operating costs of the new subway lines built between 2008 and 2014 The subway construction cost is ¥245.23 billion (Li et al., 2019), implying about ¥34,000 per household. Based on an annual operating cost per household per year of ¥1246[41], the total operating cost during a 30-year period is ¥69,000 per household. Together, the discounted construction and operating cost amounts to ¥103,000 per household during a 30-year period in the welfare analyses.

[40] https://www.zdnet.com/article/singapore-readies-satellite-road-toll-system-for-2021-rollout/ and https://nyc.streetsblog.org/wp-content/uploads/2018/01/TSTC_A_Way_Forward_CPreport_1.4.18_medium.pdf.

[41] See http://fgw.beijing.gov.cn/zb/ggjt/gg/201410/P020141013443109115783.pdf.

F Figures and Tables

Figure A1: Subway Network Expansion in Beijing

(a) 09/28/1999

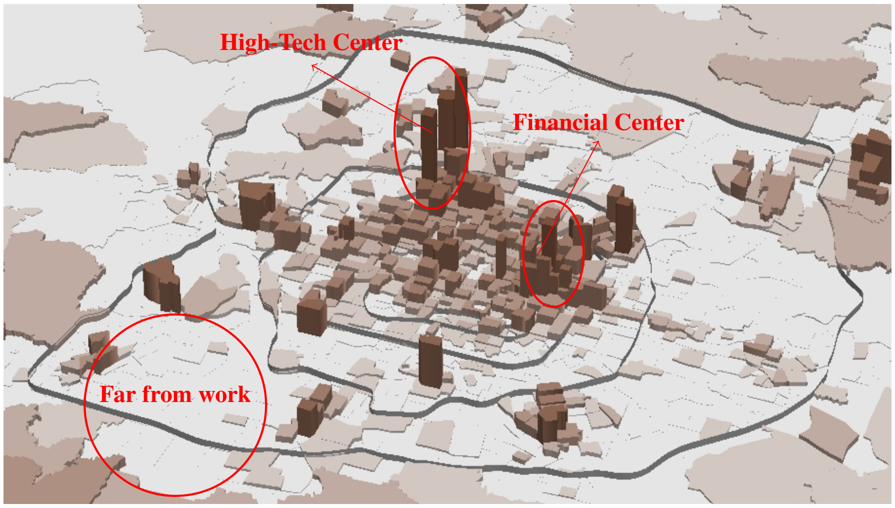
(b) 07/19/2008

(c) 12/28/2014

(d) 12/28/2019

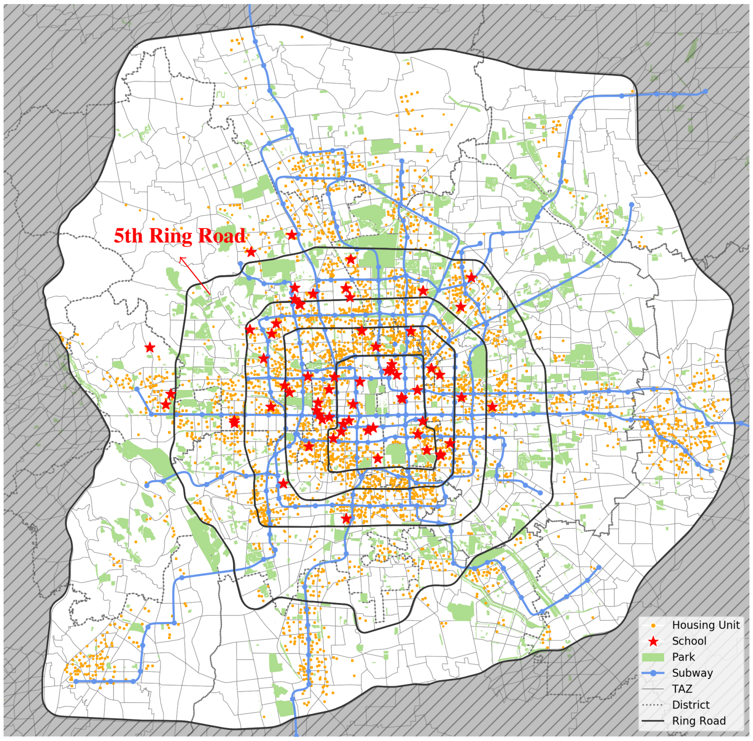
Note: Subway expansion from 1999 to 2019 in Beijing expanded from 2 lines to 22 lines. From 2007 to 2018, 16 new subway lines were built with a combined length of over 500km. By the end of 2019, the Beijing Subway is the world's longest and busiest subway system with a total length of nearly 700km and daily ridership over 10 million.

Figure A2: Job Density



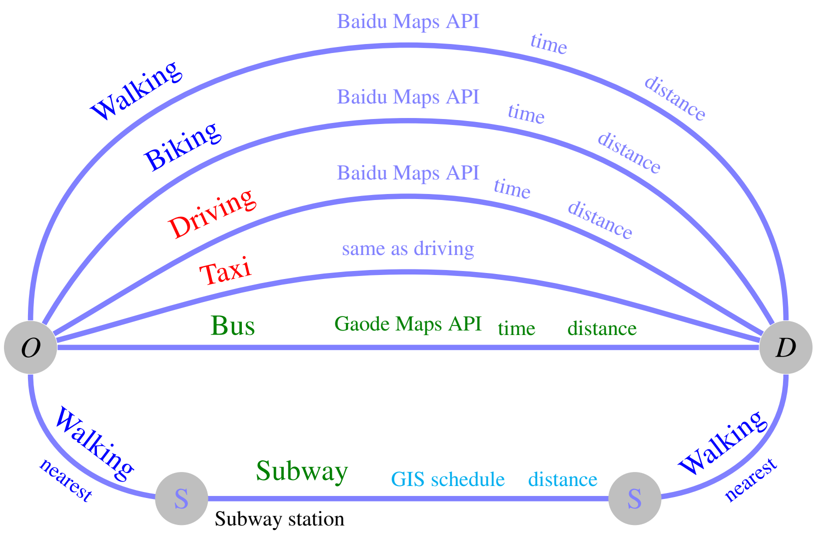
Note: This figure plots work density by TAZ based on work locations from the mortgage data. Darker colors/taller shapes indicate greater work density.

Figure A3: Housing, Amenities, and Transportation Network



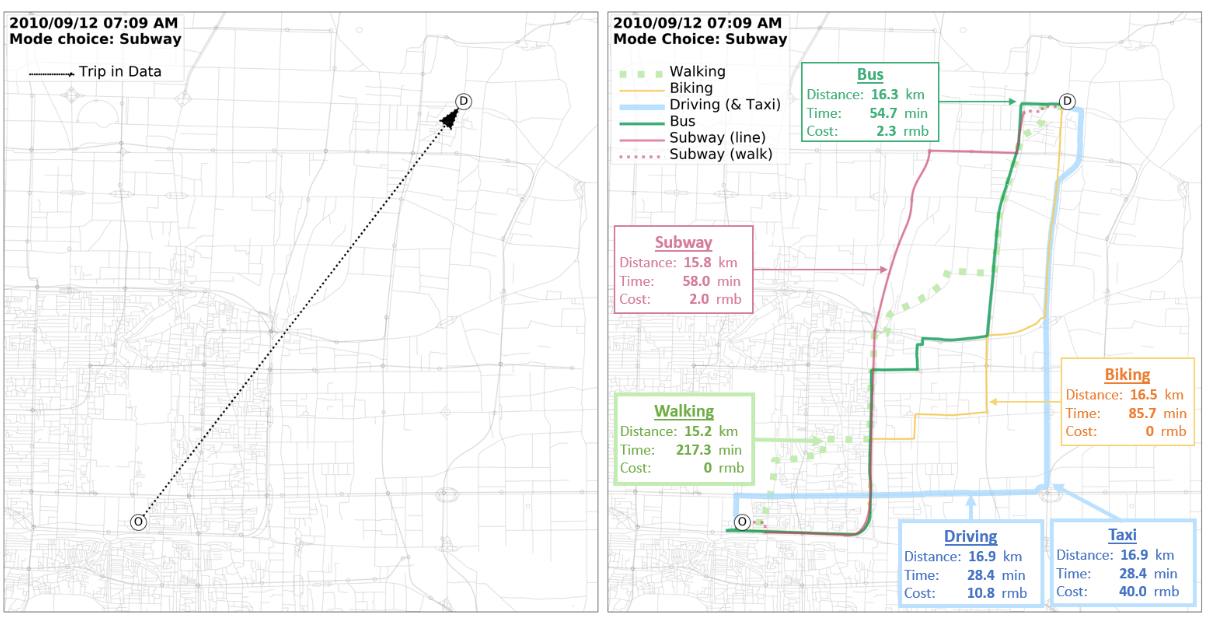
Note: The figure shows the home locations in the mortgage data overlaid with ring roads (black lines), subway lines in blue (as of 2015), government-designated key schools (red stars), and government-designated parks (green area). The outermost black line traces out the 6th ring road.

Figure A4: Construction of the Travel Choice Set



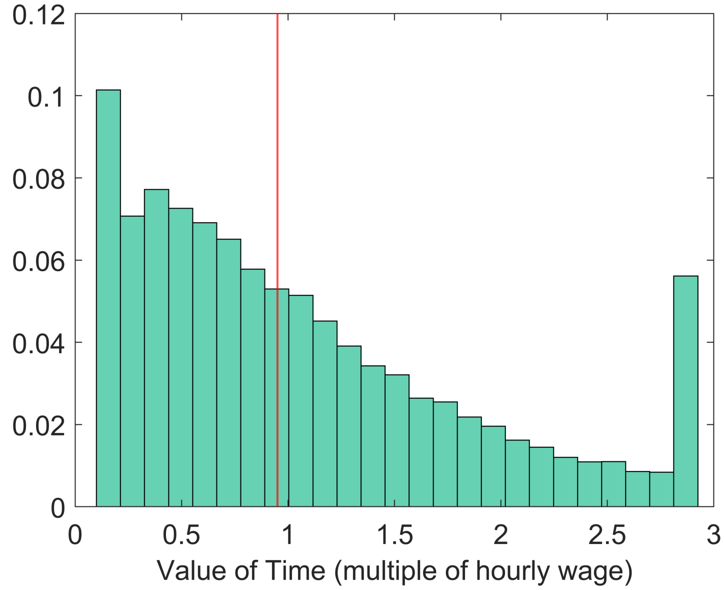
Note: Travel time and distance using walk, bike, car, and taxi are constructed using the Baidu Maps API based on the survey reported departure time and day of the week. Bus travel time and distance are constructed using Gaode Maps API because it provides the number of transfers and walking time between bus stops. The travel time for bus, car and taxi are adjusted based on the historical traffic congestion condition in the survey month-year. For trips via subway, the walking time and distance to and from subway stations are provided by Baidu Maps API. Subway transit distance and time from the origin subway station to the destination station are calculated using GIS based on the historical subway network and subway time tables in 2010 and 2014.

Figure A5: An Example of Travel Routes



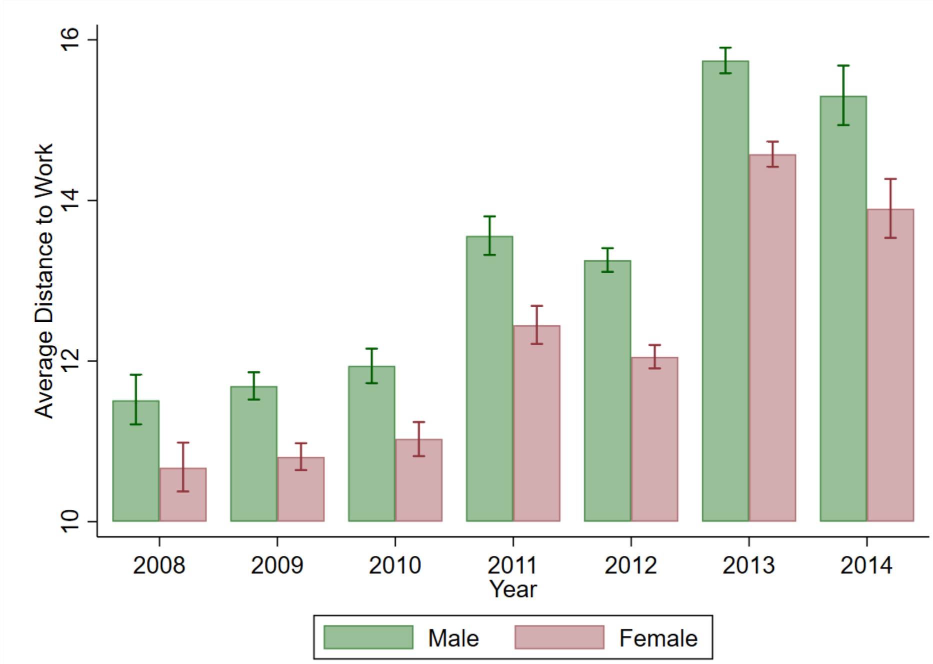
Note: the figure shows travel time and cost of six routes for a particular trip that started at 7:09am on 9/12/2010. The chosen mode was subway. The left panel shows the the straight-line direction of travel, while the right panel shows the time, monetary cost, and distance for each travel mode and the corresponding route constructed by Baidu API, Gaode API and GIS.

Figure A6: Implied Value of Time Distribution from the Mode Choice Estimation



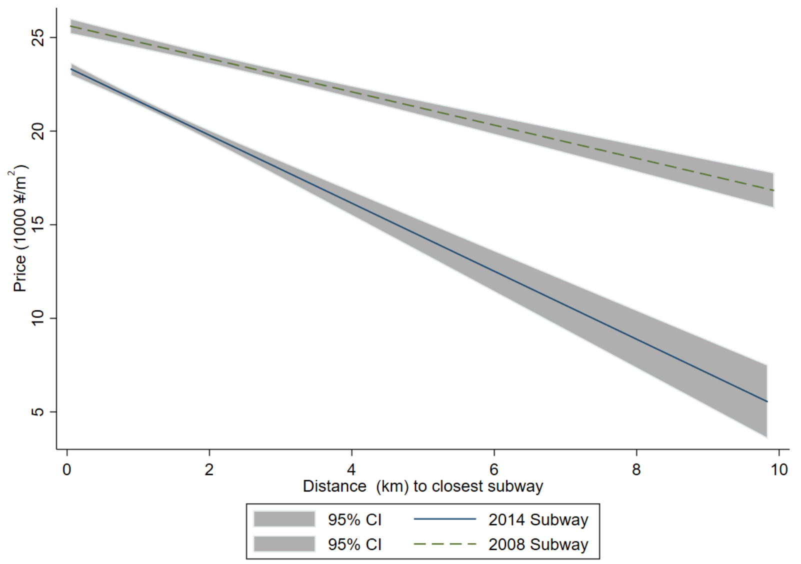
Note: The figure plots the estimated distribution of value of time (VOT) in terms of hourly wage that is based on Column (6) of Table 3. VOT is measured by the ratio of the preference on travel time over the preference on monetary travel cost. The former one has a winsorized (at the 5th and 95th percentile) chi-square distribution with three degrees of freedom, while the latter one is inversely related to income. The red line shows the average VOT (95.6% of the hourly wage). The median VOT is 84.6% of the hourly wage.

Figure A7: Distance to Work by Gender in km



Note: the figure displays the average distance to work by year for male (green bars) and female (red bars) household members, separately. The whiskers denote 95% intervals. Males have longer commutes than females. The increasing commuting distance over time reflects the expansion of the city and its transportation infrastructure.

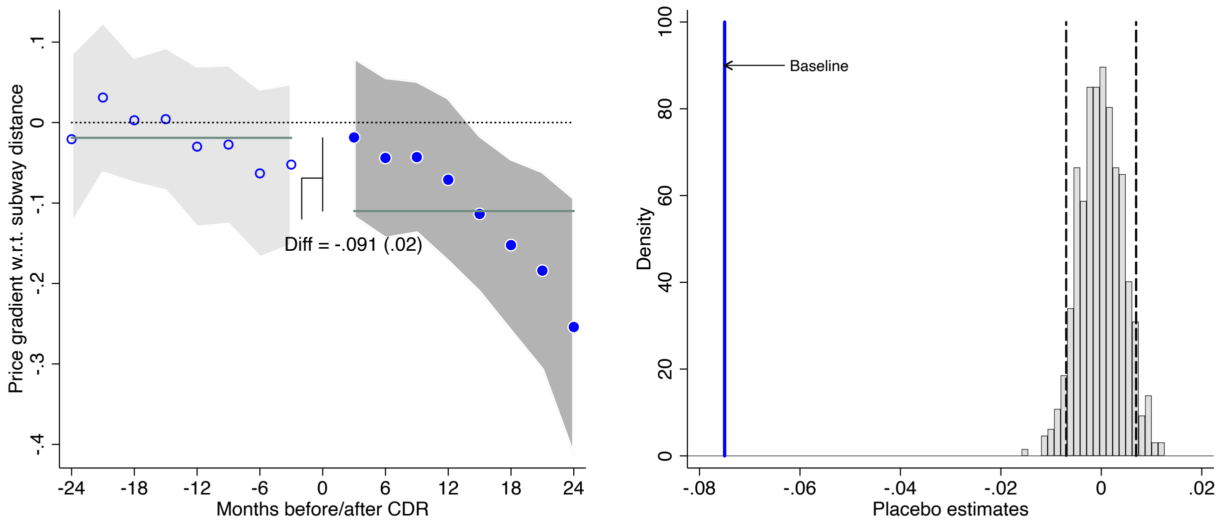
Figure A8: Price Gradient under the 2008 and 2014 Subway Network



Note: This plot shows the simulated bid- rent curve with respect to subway distance under the 2008 and 2014 network, respectively. The results corresponds to Columns (1) and (4) of Table 6. The gradient of the bid-rent curve under the 2014 subway system (-¥1900/m^2 per km) is steeper than the 2008 subway system (-¥700/m^2 per km), reflecting households' higher WTP for proximity to subway stations when the subway system is more desirable. The bid-rent shifts down under the 2014 subway system that reaches to cheaper homes farther away from the city center.

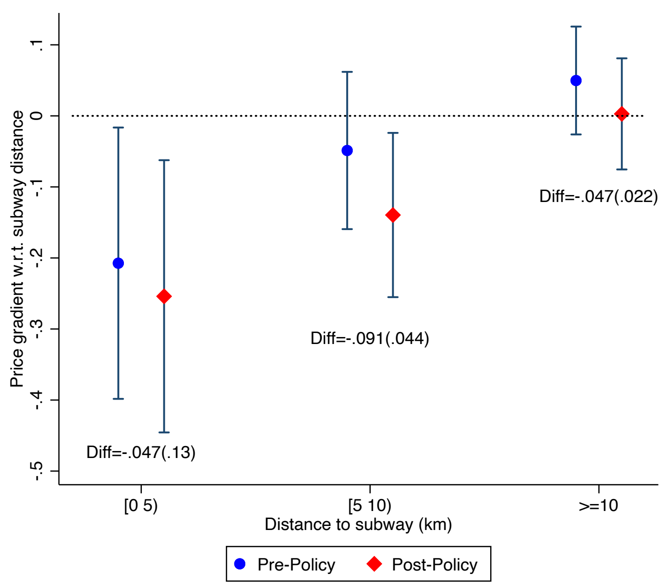
Figure A9: Event Study and Falsification Test

(a) Event Study (b) Falsification Test



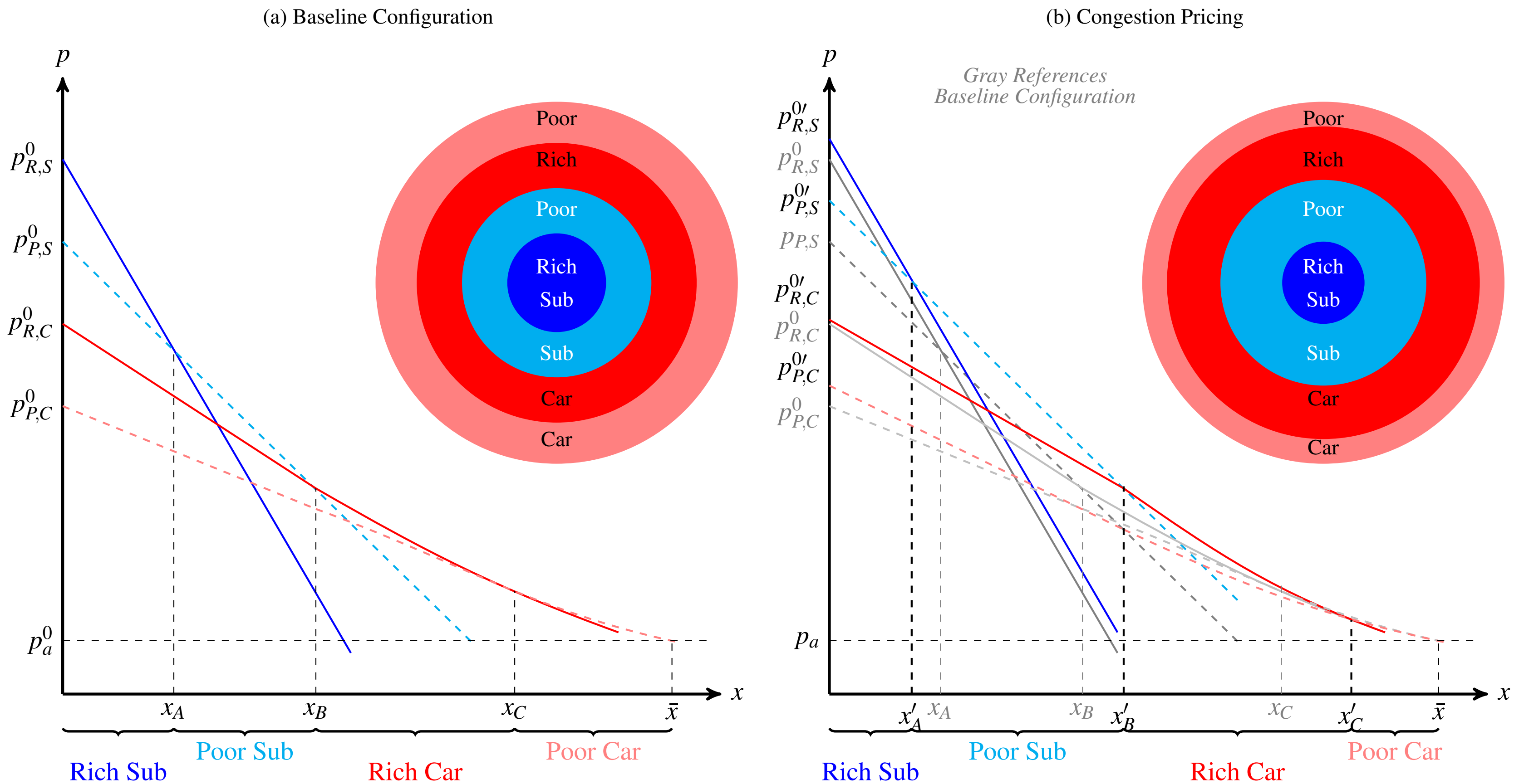
Note: Panel (a) shows estimates from the event study of Beijing's driving restriction on the price gradient w.r.t the subway distance. The blue dots report the estimates by quarter relative to the start of the driving restriction in July 2008. Panel (b) reports the distribution of 500 coefficient estimates from placebo tests with randomized event time. Dashed vertical lines indicate the 95% confidence interval, while the solid vertical line indicates the baseline estimate from Column (4) in Table A1. The regressions in both panels include neighborhood fixed effects, year by month fixed effects, city district by year fixed effects, and complex characteristics.

Figure A10: Price Gradient by Distance Bands



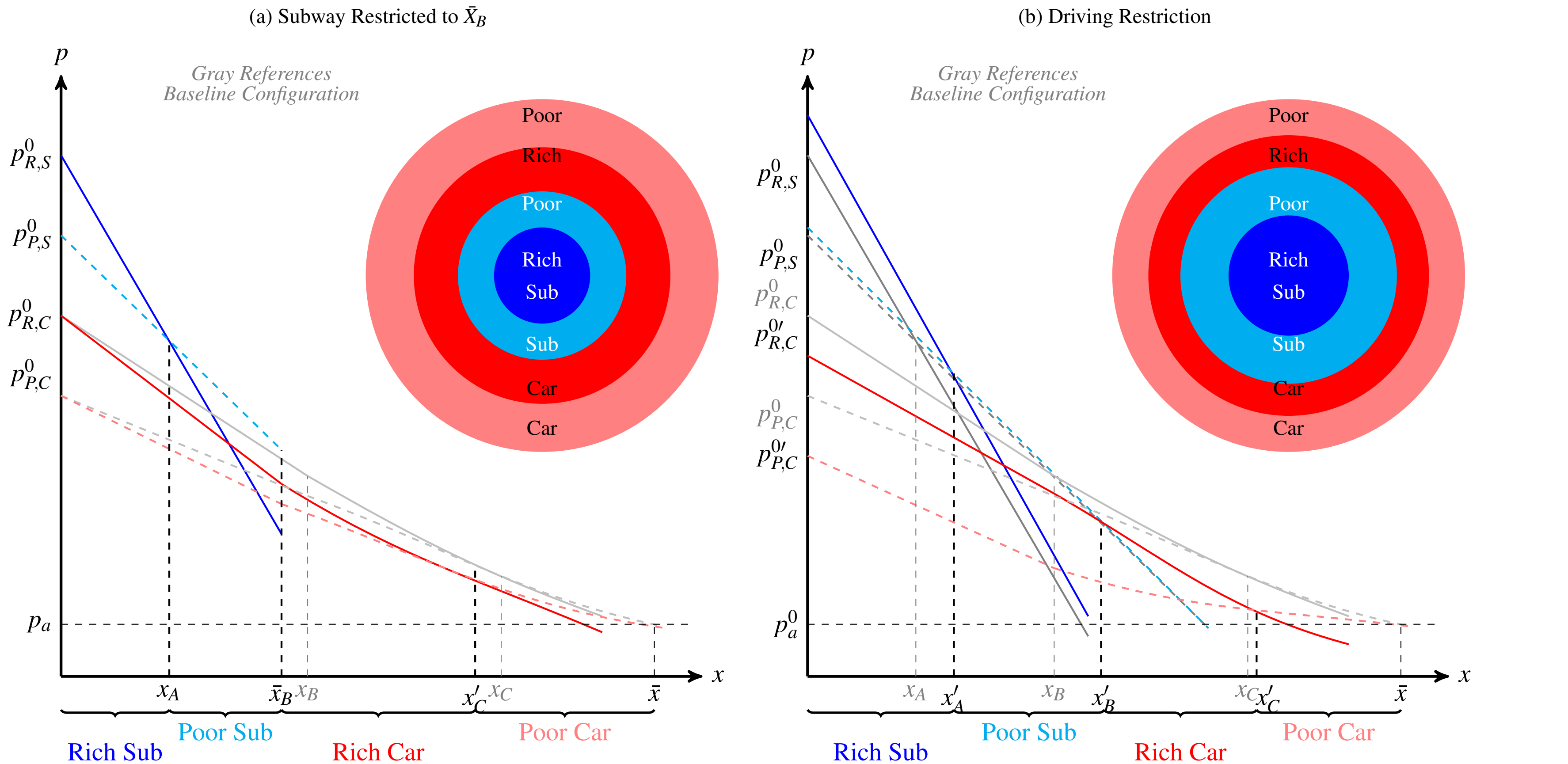
Note: This figure shows the price gradient estimates and their 95% confidence intervals by subway distance bands for the pre-policy (blue dots) and post-policy (red diamond) periods, respectively. It also displays the differences between the pre- and post-policy price gradient estimates and the standard errors of the differences. The controls include neighborhood fixed effects, year by month fixed effects, city district by year fixed effects, and complex characteristics. The standard errors are clustered at the neighborhood level.

Figure A11: Spatial Equilibrium with Two Modes & Income Heterogeneity: Baseline & Congestion Pricing



Note: The x-axis denotes the distance to city center and the y-axis is the rental price. Concentric circles show the equilibrium pattern of housing location where the linear city is drawn as radially symmetric. Commuters are in two income groups (Rich, R, denoted by solid lines, and Poor P, denoted by dashed lines) and choose from two modes: car (C, in red) and subway (S, in blue). The model is calibrated to generate an urban configuration as indicated. Bid-rent curves are steepest for the rich because of their higher value of time. Bid-rent curves are flatter for car commuting because the variable time cost is assumed to be smaller. Values indicated by $p^0_{d,m}$ are the value of bid-rent curves at the origin, while those with a $\prime$. Gray curves, dashed lines and text in panel (b) reference the baseline configuration in panel (a). Panel (b) shows the effect of congestion pricing in a per-kilometer basis. It increases the y-intercept through lump-sum remittance of toll revenue back to all households. For car commuting, there are offsetting effects: congestion is lower with the toll because there are fewer car commuters, but longer commutes face larger total tolls, making bid-rent curves steeper.

Figure A12: Spatial Equilibrium with Two Modes & Income Heterogeneity: Subway & Driving Restriction



Note: The x-axis denotes the distance to city center and the y-axis is the rental price. Concentric circles show the equilibrium pattern of housing location where the linear city is drawn as radially symmetric. Commuters are in two income groups (Rich, R, indicated with solid lines, and Poor P, indicated with dashed lines) and choose from two modes: car (C, in red) and subway (S, in blue). The model is calibrated to generate an urban configuration as indicated. Bid-rent curves are steepest for the rich because of their higher value of time. Bid-rent curves are flatter for car commuting because the variable time cost is assumed to be smaller. Values indicated by $p^0_{d,m}$ are the value of bid-rent curves at the origin, while $p^{0\prime}_{d,m}$, are the new values at the origin after the policy change. Gray curves reference the baseline configure in panel (a) of Figure A11. Panel (a) shows the effect of restricting the subway network to $\bar{x}$. Comparing this panel to panel (a) of Figure A11 shows the effect of expanding the subway network from $\bar{x}_B$ to x_B. There is no effect of expansion on the bid-rents for subway except that they do not extend beyond $\bar{x}_B$ in panel (a) of Figure A11 because the subway has not been built that far. Because expansion allows a larger share of commuters to use the subway, here only from the poor, it induces lower congestion, flattening out the bid-rent curves for driving, for the rich more than the poor because of value of time differences. Panel (b) considers a driving restriction policy, which induces increases in both the fixed and variable costs of driving not remitted to households (unlike congestion pricing). Car commuters will need to incur the fixed cost of both driving and and using subway, and when they cannot drive, they will have to use subway which has a higher variable cost due to time. The bid-rent curves for car commuters move down and become steeper because of the change in fixed cost and variable cost of commuting, respectively. The increase in commuting costs is larger for the rich due to their high VOT than for the poor, leading to a larger movement of the rich away from driving to subway.

Table A1: Effect of Driving Restriction on Price Gradient w.r.t. Subway Proximity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subway Distance (km)	-0.199	0.004	0.003	0.001	0.005	0.021	0.017
	(0.021)	(0.052)	(0.050)	(0.048)	(0.076)	(0.047)	(0.078)
Subway Distance (km) × CDR	-0.125	-0.080	-0.081	-0.075	-0.126	-0.063	-0.118
	(0.030)	(0.019)	(0.019)	(0.019)	(0.033)	(0.019)	(0.033)
Subway Density						0.019	0.011
						(0.007)	(0.012)
Year x Month FE	Y	Y	Y	Y	Y	Y	Y
Neighborhood FE	N	Y	Y	Y	Y	Y	Y
District x Year FE	N	N	Y	Y	Y	Y	Y
Complex-level Controls	N	N	N	Y	Y	Y	Y
Weighted	N	N	N	N	Y	N	Y
Adjusted R^2	0.248	0.495	0.503	0.520	0.773	0.522	0.774

Note: The dependent variable is housing price (¥1000/m^2), with a sample mean of ¥10,078/m^2. The sample spans 24 months before and after the start of car driving restriction policy (CDR) in July 2008. We remove observations from July to September 2008. The policy was more aggressive during this period whereby half of the vehicles were restricted on a given weekday (due to the 2008 Olympics in August), and changed to one weekday per week from October 2008 onward. The number of observations is 23,917. Subway distance is the distance (in km) from the housing unit to the nearest subway station. Subway density is constructed at the TAZ level as the inverse distance weighted number of subway stations from the centroid of an TAZ. There are 150 neighborhoods, 16 district, and 2804 complex in the sample. The complex-level attributes include complex age, the floor area ratio, the green space ratio, the land area of the complex, home management fee (HOA fee), and the number of building and apartment units in the complex. Weights are constructed via entropy balancing in Appendix Section A.2.

Table A2: Housing Demand Without EV Terms - Linear Parameters

Variables	OLS (1)	OLS (2)	IV1 (3)	IV2 (4)	IV2+IV3 (5)	All (6)
Price (¥mill.)	-2.073	-2.062	-7.101	-4.224	-5.031	-5.356
	(0.180)	(0.176)	(1.646)	(0.535)	(0.432)	(0.418)
Ln(home size)	-3.590	-3.657	5.102	0.104	1.512	2.079
	(0.248)	(0.251)	(2.937)	(0.932)	(0.761)	(0.765)
Building age	-0.032	-0.026	-0.144	-0.076	-0.095	-0.103
	(0.006)	(0.006)	(0.040)	(0.012)	(0.010)	(0.010)
Floor area ratio	0.017	0.001	-0.009	-0.007	-0.009	-0.009
	(0.031)	(0.023)	(0.036)	(0.021)	(0.025)	(0.027)
Ln(dist. to park)	0.167	0.052	-0.513	-0.196	-0.285	-0.321
	(0.059)	(0.054)	(0.225)	(0.073)	(0.075)	(0.081)
Ln(dist. to key school)	0.631	0.555	-0.034	0.312	0.223	0.187
	(0.060)	(0.091)	(0.213)	(0.086)	(0.089)	(0.091)
Year-Month-District FE	Y	Y	Y	Y	Y	Y
Neighborhood FE		Y	Y	Y	Y	Y
First-stage Kleinberg-Paap F			9.9	10.5	14.2	14.2
Avg. Housing Demand Price elasticity	3.09	3.10	-1.94	0.94	0.13	-0.19

Note: The number of observations is 79,894. The dependent variable is the recovered population-average utilities $\{\delta_j\}_j$ when EV is excluded from housing attributes. The first two columns and the last four present OLS and IV estimates, respectively. The floor area ratio is total floor area over the complex's parcel size and measures complex density. Distance to key school is the distance to the nearest key elementary school. Column (3) use IV1 as price instruments, i.e. the number of homes that are within 3km from a given home, outside the same complex, and sold in a two-month window. Columns (4), (5), and (6) use IV2, i.e. the average attributes of these homes (building size, age, log distance to park, and log distance to key school). Column (5) also includes IV3, the interaction between IV2 and the winning odds of the licence lottery. The winning odds decreased from 9.4% in January 2011 to 0.7% by the end of 2014. Column (6) uses all IVs. Standard errors are clustered at the neighborhood level.

Table A3: Housing Demand - Nonlinear Parameters with Alternative Sampling

	0.5% Sample (1)		1% Sample (2)	
	Para	SE	Para	SE
Demographic Interactions				
Price (¥mill.) * ln(income)	1.153	0.018	1.030	0.016
Age in 30-45 * ln(distance to key school)	-0.459	0.011	-0.420	0.010
Age $>$ 45 * ln(distance to key school)	-0.122	0.024	-0.123	0.021
Age in 30-45 * ln(home size)	1.681	0.034	1.486	0.029
Age $>$ 45*ln(home size)	3.011	0.070	2.746	0.061
EV_{Male}	0.831	0.064	0.755	0.006
EV_{Female}	0.976	0.069	0.893	0.006
Random Coefficients				
$\sigma(\mathrm{EV}_{Male})$	0.333	0.154	0.379	0.013
$\sigma(\mathrm{EV}_{Female})$	0.408	0.142	0.482	0.012
Log-likelihood	-128,976		-168,808	

Note: The table replicates Table 4 and reports MLE estimates of housing demand's non-linear parameters using mortgage data from 2006-2014 with 77,696 observations. Column (1) constructs households' choice set using a 0.5% random sample of all houses sold during a two-month window around the purchase date of the chosen home. Column (2) reproduces our preferred specification in Column (3) of Table 4 that uses a 1% random sample.

Table A4: Housing Demand - Linear Parameters with Alternative Sampling

Variables	0.5% Sample (1)	1% Sample (2)
Price (in 1 million RMB)	-7.417	-6.596
	(0.590)	(0.534)
Ln(property size)	4.355	3.879
	(1.073)	(0.969)
Building age	-0.139	-0.132
	(0.014)	(0.013)
Complex FAR	-0.019	-0.023
	(0.037)	(0.034)
Ln(dist. to park)	-0.442	-0.424
	(0.114)	(0.103)
Ln(dist. to key school)	0.321	0.288
	(0.128)	(0.118)
Year-Month-District FE	Y	Y
Neighborhood FE	Y	Y
First-stage Kleinberg-Paap F	14.22	14.22
Avg. Price elasticity	-1.64	-1.44
Avg. Price elasticity CI	[-2.83, -0.44]	[-2.52, -0.35]

Note: this table replicates Table 5 and reports IV estimates of linear parameters in housing demand. Column (1) constructs households' choice set using a 0.5% random sample of all houses sold during a two-month window around the purchase date of the chosen home. It uses all three sets of price IVs. Column (2) reproduces our preferred specification in Column (6) of Table 5 that uses a 1% random sample. Standard errors clustered at the neighborhood level.

Table A5: The Speed Traffic Density Elasticity Estimate

Region	(1) 2-3 Ring Roads	(2) 3-4 Ring Roads	(3) 4-5 Ring Roads	(4) 5-6 Ring Roads	(5) All
Log of Density (IV)	-1.250	-1.185	-1.287	NA	-1.099
	(0.148)	(0.111)	(0.417)		(0.089)
Log of Density (OLS)	-0.583	-0.645	-0.362	-0.542	-0.554
	(0.065)	(0.046)	(0.043)	(0.048)	(0.027)
Observations	45,152	49,351	29,241	32,926	156,670
Average speed (km/h)	28.00	30.39	32.86	31.20	30.3

Note: This table presents 2SLS results on the speed-density relationship by ring-road segments (e.g., between the 2nd and 3rd ring roads). The segment within the 2nd ring road is omitted due to the lack of observation. The dependent variable is ln(speed in km/h) and the key explanatory variable is log(traffic density in the number of cars/lane-km). The IVs are based on the driving restriction policy which has a preset rotation schedule using the last digit of the license plate number. They include a policy indicator for days when vehicles with a license number ending 4 or 9 are restricted from driving and interactions between this variable and hour-of-day dummies. Our sample consists of road segment by hour during peak hours within the 6th ring road in 2014. We focus on the top quintile observations with traffic density larger than 35 cars per lane-km. The average speed for these observations is 30km/h, close to the city-wide average speed during peak hours and more relevant for our analysis on commuting trips. The control variables include temperature (C^o), wind speed (km/h), visibility (km), dummies for wind directions and sky coverage at the hourly level. The time and spatial fixed effects include day-of-week, month-of-year, hour-of-day, holiday, and monitoring stations fixed effects. Parentheses contain standard errors clustered by road segments. Significance: $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

Table A6: Model Prediction on Changes in Housing Price Gradient due to Driving Restriction

	(1)	(2)	(3)
Subway Distance	-0.725	-0.302	
	(0.066)	(0.159)	
Subway Distance × CDR	-0.034	-0.034	-0.034
	(0.001)	(0.001)	(0.001)
Neighborhood FE	N	Y	N
home FE	N	N	Y
Adjusted R^2	0.329	0.400	0.999

Note: The analysis is based on the mortgage data's 2014 cohort with 7,136 observations. Using the parameter estimates from the sorting model, we simulate the equilibrium housing prices under the 2008 network for two scenarios: with and without the car driving restriction (CDR). We then regress the simulated housing prices in ¥$1{,}000/m^2$ on subway distance (in km) as in Table A1. Subway distance is the observed distance from the housing unit to the nearest subway station based on the 2008 subway network. The driving restriction steepens the price gradient with respect to subway access, consistent with results in Table A1 based on observed data. The magnitude is somewhat smaller, because the reduced-form result reflects a short-run response while the structural simulation incorporates long-run equilibrium adjustments (especially the rebound effects). The two samples are also different. The reduced-form analysis uses two years before and after CDR's initial implementation date while the structural analysis uses the 2014 cohort. Standard errors clustered at the neighborhood level.

Table A7: Simulation Results with Household Sorting and Variable Housing-Supply

	2008 Subway Network						2014 Subway Network					
	(1) No Policy		(2) Driving restriction		(3) Congestion pricing		(4) Subway Expansion		(5) + Driving restriction		(6) + Congestion pricing	
			Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)	
Income relative to the median	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Panel A: travel mode shares in percentage points and average speed												
Drive	41.65	21.44	-7.19	-3.41	-3.49	-5.37	-2.32	-1.80	-8.61	-4.69	-5.32	-6.53
Subway	9.02	10.77	1.31	0.73	0.87	1.02	5.00	6.57	6.17	6.97	5.52	7.25
Bus	22.44	30.47	1.79	0.61	0.52	1.13	-1.52	-2.48	0.31	-1.53	-0.76	-1.00
Bike	15.96	24.01	1.59	0.79	0.76	1.74	-0.86	-1.78	0.42	-1.12	-0.19	-0.25
Taxi	2.20	1.32	1.19	0.55	0.63	0.58	-0.20	-0.15	0.86	0.31	0.38	0.33
Walk	8.74	11.99	1.31	0.73	0.71	0.90	-0.10	-0.36	0.86	0.05	0.37	0.19
Speed (km/h)	21.49		3.83		3.97		1.13		4.68		5.09	
Panel B: sorting outcomes												
Distance to work (km)	18.56	15.66	0.02	0.02	-0.29	-0.19	0.76	0.61	0.85	0.62	0.43	0.44
Distance to subway (km)	5.33	4.30	-0.05	0.01	-0.11	-0.07	-4.13	-3.45	-4.14	-3.44	-4.14	-3.45
Panel C: welfare changes per household (thousand ¥)												
Consumer surplus (+)			-226.9	-31.6	-85.1	-70.7	187.4	105.3	-42.7	71.8	86.8	34.4
Toll revenue (+)					136.4	136.4					128.9	128.9
Subway cost (–)							103.0	103.0	103.0	103.0	103.0	103.0
Net welfare			-226.9	-31.6	51.4	65.7	84.4	2.3	-145.7	-31.2	112.7	60.3

Note: the table replicates Table 6 with sorting but allows housing supply to adjust with a price elasticity of 0.53. The simulations use the 2014 cohort (households who purchased homes in 2014) and are based on parameters reported in Column (6) of Table 3, Column 3 of Table 4, and Column (6) of Table 5. Appendix E explains the simulation procedure. Column (1) reports results when no policy was in place. Columns (2) to (6) present differences from Column (1). Driving restriction prohibits driving in one of five work days. Congestion pricing is fixed at ¥1.13 per km as in Table 6. High-income household are those with income above the median. Toll revenue is recycled uniformly across households. Subway cost includes the construction and operation costs that are equally distributed among 7.2 million households. Net welfare is consumer surplus plus recycled revenue and minus subway costs.

Table A8: Simulation Results without Household Sorting

	2008 Subway Network						2014 Subway Network					
	(1) No Policy		(2) Driving restriction		(3) Congestion pricing		(4) No Policy		(5) Driving restriction		(6) Congestion pricing	
	Baseline levels		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)	
Income relative to the median	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Panel A: travel mode shares in percentage points and average speed												
Drive	41.02	21.02	-6.49	-2.99	-2.87	-5.01	-1.26	-1.07	-7.68	-4.08	-4.69	-6.11
Subway	9.43	11.24	0.83	0.26	0.45	0.54	3.61	5.08	4.73	5.47	5.11	6.77
Bus	22.31	30.07	1.88	0.98	0.77	1.65	-1.35	-2.03	0.51	-1.05	-0.63	-0.60
Bike	16.08	24.08	1.48	0.73	0.64	1.72	-0.81	-1.57	0.55	-0.85	-0.32	-0.31
Taxi	2.17	1.32	1.24	0.55	0.65	0.57	-0.08	-0.07	0.99	0.41	0.40	0.34
Walk	8.98	12.26	1.06	0.47	0.37	0.53	-0.10	-0.34	0.91	0.09	0.13	-0.08
Speed (km/h)	21.49		3.82		3.61		1.76		5.39		5.09	
Panel B: sorting outcomes												
Distance to work (km)	18.56	15.66										
Distance to subway (km)	5.33	4.30					-4.13	-3.45	-4.13	-3.45	-4.13	-3.45
Panel C: welfare changes per household (thousand ¥)												
Consumer surplus (+)			-227.3	-31.0	-110.7	-83.2	207.7	117.0	-32.2	82.1	83.2	37.2
Toll revenue (+)					138.8	138.8					127.9	127.9
Subway cost (–)							103.0	103.0	103.0	103.0	103.0	103.0
Net welfare			-227.3	-31.0	28.1	55.6	104.7	14.0	-135.2	-20.9	108.1	62.1

Note: the table replicates Table 6 but shuts down household sorting. In other words, travel mode choices adjust and clear the traffic sector but households do not change residential locations. Column (1) reports results when no policy was in place. Columns (2) to (6) present differences from Column (1). Driving restriction prohibits driving in one of five work days. Congestion pricing is set at ¥1.13 per km as in Table 6. High-income household are those with income above the median. Toll revenue is recycled uniformly across households. Subway cost includes the construction and operation costs that are equally distributed among 7.2 million households. Net welfare is consumer surplus plus recycled revenue and minus subway costs.

Table A9: Simulation Results with Ring-Road Specific Density Measures

	2008 Subway Network						2014 Subway Network					
	(1) No Policy		(2) Driving restriction		(3) Congestion pricing		(4) Subway Expansion		(5) + Driving restriction		(6) + Congestion pricing	
			Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)	
Income relative to the median	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Panel A: travel mode shares in percentage points and average speed												
Drive	41.66	21.44	-7.18	-3.40	-3.52	-5.39	-2.17	-1.66	-8.53	-4.62	-5.25	-6.41
Subway	9.02	10.77	1.29	0.70	0.86	0.97	4.62	6.05	5.79	6.44	5.25	6.81
Bus	22.44	30.47	1.78	0.60	0.58	1.23	-1.54	-2.53	0.31	-1.57	-0.74	-1.04
Bike	15.95	24.01	1.60	0.80	0.78	1.78	-0.78	-1.63	0.53	-0.93	-0.10	-0.11
Taxi	2.20	1.32	1.19	0.55	0.62	0.57	-0.17	-0.11	0.89	0.36	0.38	0.35
Walk	8.74	11.99	1.31	0.74	0.68	0.85	0.03	-0.12	1.02	0.32	0.47	0.39
Speed (km/h)	21.50		3.83		3.81		1.48		5.07		5.29	
Panel B: sorting outcomes												
Distance to work (km)	18.56	15.66	0.01	0.01	-0.16	-0.07	0.37	0.17	0.42	0.16	0.17	0.10
Distance to subway (km)	5.33	4.30	-0.03	0.03	-0.02	0.02	-4.13	-3.44	-4.13	-3.44	-4.13	-3.44
Panel C: welfare changes per household (thousand ¥)												
Consumer surplus (+)			-227.6	-32.8	-104.0	-74.6	215.6	98.9	-16.5	63.4	99.2	26.4
Toll revenue (+)					135.7	135.7					125.2	125.2
Subway cost (–)							103.0	103.0	103.0	103.0	103.0	103.0
Net welfare			-227.6	-32.8	31.7	61.1	112.6	-4.1	-119.5	-39.6	121.4	48.6

Note: this table replicates Table 6 but incorporates ring-road specific traffic densities: density within the 3rd ring road, between 3rd and 4th ring roads, between 4th and 5th ring roads, and between 5th and 6th ring roads. Column (1) reports results when no policy was in place. Columns (2) to (6) present differences from Column (1). Driving restriction prohibits driving in one of five work days. Congestion pricing is set at ¥1.13 per km as in Table 6. High-income household are those with income above the median. Toll revenue is recycled uniformly across households. Subway cost includes the construction and operation costs that are equally distributed among 7.2 million households. Net welfare is consumer surplus plus recycled revenue and minus subway costs.

Table A10: Simulation Results using Preferred Specifications but without Random Coefficients

	2008 Subway Network						2014 Subway Network					
	(1)		(2)		(3)		(4)		(5)		(6)	
	No Policy		Driving restriction		Congestion pricing		Subway Expansion		+ Driving restriction		+ Congestion pricing	
	Baseline levels		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)		Δs from (1)	
Income relative to the median	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Panel A: travel mode shares in percentage points and average speed												
Drive	47.17	28.26	-9.05	-5.38	-2.38	-3.25	-0.43	-0.35	-9.40	-5.67	-2.81	-3.56
Subway	6.38	7.99	1.22	0.66	0.30	0.35	0.92	1.32	2.32	2.11	1.25	1.83
Bus	19.67	27.85	3.45	2.09	0.89	1.39	-0.24	-0.51	3.08	1.47	0.63	0.77
Bike	16.70	23.54	2.68	1.66	0.68	1.02	-0.14	-0.31	2.46	1.29	0.52	0.64
Taxi	1.10	0.96	0.26	0.12	0.08	0.06	-0.02	-0.02	0.23	0.09	0.06	0.04
Walk	8.98	11.40	1.44	0.85	0.42	0.44	-0.08	-0.13	1.31	0.71	0.34	0.29
Speed (km/h)	22.36		4.59		4.59		0.16		4.67		4.77	
Panel B: sorting outcomes												
Distance to work (km)	18.66	15.66	-0.04	0.08	-0.59	-0.13	0.15	0.05	0.17	0.11	-0.46	-0.08
Distance to subway (km)	5.43	4.36	-0.06	0.06	-0.12	0.12	-4.24	-3.51	-4.24	-3.51	-4.25	-3.50
Panel C: welfare changes per household (thousand ¥)												
Consumer surplus (+)			-1447.6	-338.0	-797.6	-394.2	118.5	60.5	-1308.5	-274.0	-669.9	-327.4
Toll revenue (+)					390.7	390.7					386.0	386.0
Subway cost (–)							103.0	103.0	103.0	103.0	103.0	103.0
Net welfare			-1447.6	-338.0	-406.9	-3.6	15.5	-42.5	-1411.5	-377.0	-386.9	-44.4

Note: the table replicates Table 6 but removes random coefficients in both travel and housing choices from the preferred specification. The model without random coefficients produces low VOT, counter-intuitive substitution patterns, and very different welfare predictions from Table 6. Column (1) reports results when no policy was in place. Columns (2) to (6) present differences from Column (1). Driving restriction prohibits driving in one of five work days. Congestion pricing is set at ¥1.13 per km as in Table 6. High-income household are those with income above the median. Toll revenue is recycled uniformly across households. Subway cost includes the construction and operation costs that are equally distributed among 7.2 million households. Net welfare is consumer surplus plus recycled revenue and minus subway costs.