Kyoto University,
Graduate School of Economics
Discussion Paper Series

The Effects of Mass Transit System on Urban Population Distribution:
Evidence from Wuhan

Se-il Mun, Lei Qin, Yue Zhou

Discussion Paper No. E-23-003

Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

September, 2023

The Effects of Mass Transit System on Urban Population Distribution: Evidence from Wuhan

Se-il Mun* Lei Qin* Yue Zhou†

September 29, 2023

Abstract

This paper aims to evaluate empirically how station spacing affects the density along the transit line and the compactness of the urban area. We derive the population density equation as a function of station spacing, based on urban economics model of residential land use. We estimate the population density equation using data for grids in Wuhan, China. Based on the estimated equation, we conduct counterfactual simulations for several cases of station spacing to evaluate the extent to which shorter station spacing contributes to land use compactness.

Keywords: Mass Transit System; Compact land use; Station spacing; Population distribution

JEL classifications: R14; R21; R31; R42

*Graduate School of Economics, Kyoto University. Address: Yoshida-honmachi, Sakyo, Kyoto 606-8501, Japan

†Beijing ByteDance Technology Co., Ltd. Address: Haidian District, Beijing 100086, China

1 Introduction

The development of a mass transit system (MTS) is an essential instrument in compact city policies, which aim to reduce energy consumption and greenhouse gas emissions (OECD 2012). We observe that population density is higher along the transit line, which enhances the viability of the transit system through increased ridership. One distinctive feature of mass transit systems is that passengers require access to stations, which are limited points along the route. Households are willing to pay higher prices for housing near the stations, so developers tend to construct residences with higher density near stations to accommodate more dwellings per unit of land, while land use density is lower in areas farther from the station. In essence, the closer the station spacing, the denser the land use along the transit line should be. Note that the spacing between stations should be determined by the planner of the transit system, taking into account the impact on land use. The primary objective of this paper is to evaluate how station spacing affects the density along the transit line and the compactness of the urban area.

We develop a monocentric city model of land use in which workers use mass transit to commute from residential locations to the central business district (CBD). Commuters incur costs to access the nearest transit station, in addition to the fare and linehaul time cost for the transit line from the station to the CBD. For empirical analysis, we construct the model for discrete space and derive the equation of population density as a function of station spacing. We estimate the population density equation using data for grids in Wuhan, China. Based on the estimated equation, we conduct counterfactual simulations for several cases of station spacing to evaluate the extent to which shorter station spacing contributes to land use compactness.

Numerous studies have examined the relationship between public transit and land use, with the most common approach being to estimate housing prices as a function of proximity to transit (e.g., Gibbons and Machin (2005), Billings (2011), and Dubé et al. (2013)). A primary issue in the literature has been controlling for the two-way interaction as follows: while proximity to public transit affects land use, the proximity is a result of station location planning that considers land use. For instance, locations with good

access to public transit attract residential demand, so population density should be higher in the vicinity, and planners tend to choose station placement in such densely populated areas. In other words, access to public transit is endogenous, making the ordinary least square (OLS) estimation method biased in most applications.

Our study contributes to the literature in the following ways. First, we are the first to discuss that MTS's compactness, which is the interval spacing of metro stations, affects intracity population distribution based on a spatial equilibrium model. Second, we provide a methodology to empirically investigate the effect of transit infrastructure improvement, in which aggregated landsat data can be easily applied. Lastly, we provide a novel instrument to address the endogeneity of station spacing.

This paper is structured as follows. Section 2 reviews the existing literature. Section 3 presents the residential land use model with MTS for discrete space, while Section 4 depicts the empirical data of Wuhan, China. Section 5 elucidates the identification and estimation results, Section 6 discusses the policy counterfactuals that illustrate the causal effect of policy changes. Finally, Section 7 concludes the paper.

2 Literature Review

Vast works of literature shed light on the impact of MTS on intracity land use (Tan et al. 2019; Asahi et al. 2022), housing price (Baum-Snow and Kahn 2000; Bowes et al. 2001; Gibbons and Machin 2005; Billings 2011; Dubé et al. 2013; Tan et al. 2019), population distributions (Baum-Snow, Kahn, and Voith 2005; Baum-Snow, Brandt, et al. 2017), as well as other economic issues (Bertaud et al. 2004; Brooks et al. 2019; Tyndall 2021; Pogonyi et al. 2021). Almost all empirical studies affirm positive promotion of dwellers' amenities, alongside housing prices or other economic variables due to the improvement of transportation infrastructure. The impact on population distribution or city structure, on the other hand, is determined by social planners. For example, the settlement of the U.S. emphasizes low-density suburban development and high automobile dependence, while Western European countries esteem more compact urban growth

(Bertaud et al. 2004). Besides, an intracity rail improvement spurs a more centralized city (Baum-Snow, Kahn, and Voith 2005), while highway configurations disperse population and GDP from a central city to suburban areas (Baum-Snow, Brandt, et al. 2017).

The location of transit station is affirmed to be endogenous by massive existing literature (Ihlanfeldt et al. 1998; Baum-Snow and Kahn 2000; Holzer et al. 2003; Billings 2011; Tyndall 2021). An influential explanation is that public transportation station locations correlate to local economic trends (Ihlanfeldt et al. 1998). For instance, stations are located in high streets with abundant amountS of entertainment or commercial service (Gibbons and Machin 2005); On the other hand, affluent neighborhoods oppose rail infrastructure due to concerns about the rise of local crime rate (Gibbons 2004; Gibbons and Machin 2005; Kahn 2007).

The two main approaches to address this endogeneity in practice are (i) fixed effect (FE) estimation method or case-specific Difference-in-Differences (DID) estimation method, and (ii) Instrument Variables (IV) or generalized method of moments (GMM) estimation method. For example, for (i), Pogonyi et al. (2021) utilized a two-way FE method (together with IV) to investigate economic activity. An influential DID approach provided by Gibbons and Machin (2005) regards newly constructed stations in London in the late 1990s as a quasi-experiment and utilizes this transport innovation to investigate the before-and-after housing price change compared with control neighborhoods that do not receive the treatment of transport innovation. This idea inspires a lot of empirical studies (Dubé et al. 2013; Zhu et al. 2016; Tan et al. 2019). Other DID approaches, e.g., Billings (2011) and Heilmann (2018) utilize proposed but eventually not constructed rail line corridor as control neighborhoods. For (ii), the station placement endogeneity is addressed by IV of income information (Petitte et al. 1999), demographic characteristics (Baum-Snow and Kahn 2000), intracity spatial characteristics (Tyndall 2021) as well as others (Baum-Snow, Brandt, et al. 2017). Ahlfeldt et al. (n.d.) provide a sophisticated GMM application for accessing structural parameters of their proposed model, which incorporates agglomeration and dispersion forces and an arbitrary number of

heterogeneous locations of city blocks.

However, two-way FE or DID may be infeasible to deal with the endogeneity of public transportation stations due to the ambiguity of the source of endogeneity, in the sense that the error component can be both time-variant and individual-variant. Specifically, suppose the goal is to investigate the impact of public transportation on housing price, and the government builds metro lines by the criterion of the premium of land price. In this case, due to simultaneity, the error component is time-variant and individual-variant. Hence, the result of a two-way FE estimation is biased and inconsistent without careful discussion in general.

The validity of DID methods is also often questioned due to the potential violation of the corresponding parallel trends. For example, some DID settings regard newly constructed stations within old metro lines as exogenous policy changes to identify parameters of interest. But as Billings (2011) pointed out, these stations are often determined endogenously to be located in more desirable and faster-appreciating neighborhoods. Another influential DID setting utilizes proposed but not constructed corridors as the control group. This provides a nice identification strategy, but other kinds of transportation supply (e.g., bus, taxi, or Uber) will enter this area as compensation for the lack of MRTs. Consequently, unconditional DID or conditional DID lacking this measure may fail to identify the average treatment effect of the critical variable.

On the other hand, In IV or GMM case, finding exogenous instruments in practice takes work. For example, it was prevalent using the endogenous variable's lagged value as instrument, but Reed (2015) explains that this instrument is legitimate under severe conditions, and he suggests researchers circumvent using this instrument in empirical studies. Bellemare et al. (2017) further discusses the identification of lagged explanatory variables under more general endogeneity forms and concludes that these variables should not be used for identification purposes without careful arguments on substantive grounds. Brooks et al. (2019) supplement empirical evidence that explains why such instruments might be invalid.

The rareness of authoritative data limits empirical research on MRT in China. Empirical research in China sometimes needs proxy or synthetic variables potentially measured with error. Wilhelm (2019) and

Lee et al. (2020) provide a nonparametric test for measurement error with help of the existence of repeated measures. Conley (1999) provides a nonparametric and semi-positive variance-covariance matrix estimator for the GMM estimation of spatially correlated variables.

Our work is most closely related to Tan et al. (2019), who study the effects of metro lines on land use and housing price in Wuhan. They also suffer from the lack of authoritative data. They collect high-resolution images of Wuhan from *Google Maps* of before-and-after new metro stations were built to create a set of panel data for a DID estimation. They use the same landsat data of Xu (2017) as ours. The differences between our work and Tan et al. (2019) is, they study the impact of MRT on the change of land use and housing price using synthetic data mentioned above and transaction data in a micro perspective, while we investigate the effect of MRT on intracity population distribution using landsat data in an aggregate perspective. Unlike previous studies, our paper aims to evaluate the effect of the compactness of metro station design, which attracts rare (if any) attention from urbaneconomic studies. Moreover, we provide a model that can easily apply the landsat data, which is much easier and cheaper than microdata to acquire.

3 Theoretical Model

We first introduce the continuous commuting cost model proposed by Mun et al. (2018). Consider a linear city where an employment center (CBD) at the end of left and labeled as x_0. Distance from the CBD is denoted by x. We assume that transport modes available for commuting in this city are mass transit and walking. The mass transit system connects residential locations and the CBD. A distinctive feature of the mass transit system is that passengers are allowed to get on and off only at a limited number of points, i.e., stations. The stations are labeled by $1, 2, \ldots, i, i+1, \ldots, I$. The commuting cost is given by

$$t(x) = \begin{cases} (a_1 + f)x_i + a_2(x - x_i) & \text{for } x_i \leq x < \hat{x}_i \\ (a_1 + f)x_{i+1} + a_2(x_{i+1} - x) & \text{for } \hat{x}_i \leq x < x_{i+1}, \end{cases} \tag{1}$$

where f and a_1 are monetary (fare) and non-monetary (time) costs for line haul per distance, and a_2 is access cost per distance. We assume $a_2 > a_1 + f$, which means that walking is more costly than riding the transit. The location where the commuting costs of the two routes (via ith station and via $(i+1)$th station) are equalized is given at

$$\hat{x}_i = x_i + (\frac{f + a_1 + a_2}{2a_2})(x_{i+1} - x_i)$$

$$= \phi x_{i+1} + (1 - \phi)x_i, \tag{2}$$

where $\phi \equiv (f + a_1 + a_2)/(2a_2) \in (0, 1)$. Thus, $\hat{x}_i$ is the critical boundary for people who live between x_i and x_{i+1} to decide which station to ride. This linear city is illustrated in Panel A of Figure 1 with the critical boundary between x_i and x_{i+1}.

3.1 Discrete space model

We transform the continuous linear city model into the discrete-space city model by distributing the city space into a finite number of aligned zones. The city is represented by I squared grids with length s. The mass transit system straightly passes through the center of each grid. The CBD is coordinated as the original point so that for a resident lived at (x, y) of this city, $x \in [0, Is]$ and $y \in [-(1/2)s, (1/2)s]$. Residents of each grid are uniformly distributed. For now we make a restrictive assumption on discrete city structure.

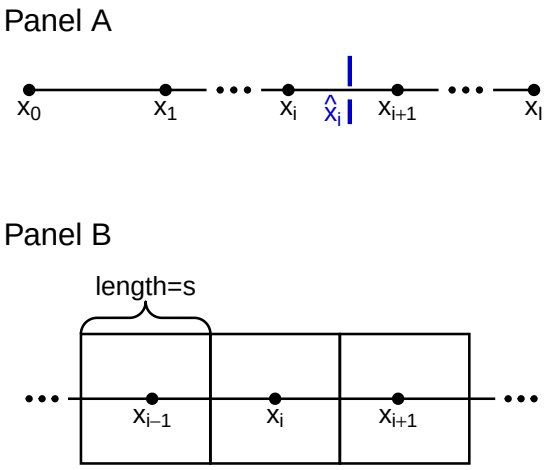

Figure 1: Structure of theoretical cities. Panel A illustrates a continuous linear city with the critical boundary between x_i and x_{i+1}. Panel B reveals a discrete-space city under Assumption 1.

Assumption 1. *There is one and only one station in each grid.*

Assumption 1 ensures a well-formed city structure that is essential to derive an analytical expression. The discrete city model is illustrated as Panel B of Figure 1.

Consider the which-station-to-ride problem for dwellings between x_i and x_{i+1}. The commuting cost of people live in (x, y) where $x \in [x_i, x_{i+1})$, $-\frac{1}{2}s, y \in [-\frac{1}{2}s, \frac{1}{2}s]$, is again given by Equation 1, where the access cost term is replaced by $a_2\rho((x - x_i), y)$ for people who ride x_i, and $a_2\rho((x - x_{i+1}), y)$ for people who ride x_{i+1}, with $\rho(\cdot)$ a specific distance function. To avoid complexity, we adopt Manhattan distance: $\rho(x, y) = |x| + |y|$. In this scenario, the critical boundary is irrelevant to the y coordinate: it is the vertical line passing through $\hat{x}_i$ given by Equation 2.

Now we deduce the commuting cost for a well-formed gird. Let i be a representative grid such that the critical boundary between x_i and x_{i+1} locates inside grid i, and the critical boundary between x_{i-1} and x_i locates exactly at (i-1)s[1]. We define station x_i's average interval spacing $d_i = \frac{1}{2}(x_{i+1} - x_{i-1})$, which is the main theoretical focus of this paper. Let t_i be the average commuting cost of dwellers who live in grid i:

[1]The location of critical boundary between x_{i-1} and x_i does not alter the result in the proceeding discussion.

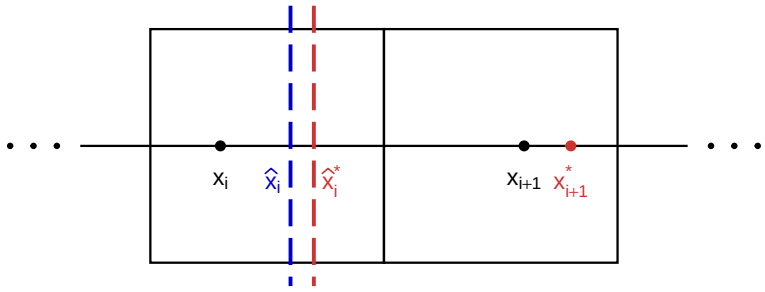

Figure 2: Representative gird i under Assumption 1 that the critical boundary lies in grid. x_j denotes jth station, $\hat{x}_j$ the critical boundary between the jth and $j+1$th station, and A^* the after-change variable of symbol A.

$$t_i = \int_{(i-1)s}^{\hat{x}_i} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \left[(a_1+f)x_i + a_2\rho(x-x_i,y)\right] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x \tag{3}$$
$$+ \int_{\hat{x}_i}^{is} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \left[(a_1+f)x_{i+1} + a_2\rho(x-x_{i+1},y)\right] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x.$$

Now let d_i increases Δd with x_{i-1} fixed. This statement is equivalent to shifting x_{i+1} to the right by $2\Delta d$. Consequently, we have $d^* = d_i + \Delta d$, $x^*_{i+1} = x_{i+1} + 2\Delta d$ and $\hat{x}^*_i = \phi x^*_{i+1} + (1-\phi)x_i$, where A^* represents the after-change status of the variable A. Such a change is illustrated by Figure 2. Consider dwellers who live in $[\hat{x}^*_i, is)$. Their commuting costs increase since both line haul and access costs increase, and the value is given by

$$\Delta t_i(\hat{x}_i^*, is) = \int_{\hat{x}_i^*}^{is} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \left[(a_1 + f)(x_{i+1}^* - x_{i+1}) + a_2(x_{i+1}^* - x_{i+1}) \right] \frac{1}{s^2} \, \mathrm{d}y \, \mathrm{d}x$$

$$= -\frac{8a_2\phi^2}{s}(\Delta d)^2 + \frac{4a_2\phi(is - \hat{x}_i)}{s}\Delta d. \tag{4}$$

Notice that the value is positive when $\hat{x}_i^*$ is in $[\hat{x}_i, is)$. In this circumstance, the commuting cost for dwellers in $[\hat{x}_i^*, is)$ increases when x_i remains in grid i.

Now turn to the people who live in $[\hat{x}_i, \hat{x}_i^*)$. They used to ride x_{i+1} rather than x_i, indicating that the latter cost more. Now they ride x_i since x_{i+1} is shifted further, meaning that their commuting cost rises as well. The increment is given by

$$\Delta t_i(\hat{x}_i, \hat{x}_i^*) = \int_{\hat{x}_i}^{\hat{x}_i^*} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \left[(a_1 + f)(x_i - x_{i+1}) + a_2\big(|x - x_i| - |x - x_{i+1}|\big) \right] \frac{1}{s^2} \, \mathrm{d}y \, \mathrm{d}x$$

$$= \frac{4a_2\phi^2}{s}(\Delta d)^2 > 0. \tag{5}$$

In the meanwhile, the commuting cost remains unchanged for residents in $[(i-1)s, \hat{x}_i)$.

Summing up the above, the total increment of average commuting cost due to the expansion of interval spacing is

$$\Delta t_i = \phi_1(\Delta d)^2 + \phi_2 \Delta d, \tag{6}$$

where $\phi_1 \equiv -4a_2\phi^2/s < 0$ and $\phi_2 \equiv 4a_2\phi(is - \hat{x}_i)/s > 0$. Dividing both sides of Equation 6 by Δd and letting $\Delta d \to 0$ yields

$$\frac{\mathrm{d}t_i}{\mathrm{d}d_i} = \phi_2 > 0, \tag{7}$$

which leads to one of the most important implications of this model: increasing the spacing of stations leads to a higher commuting cost. The reverse is also true: decreasing the interval spacing leads to a lower commuting cost. One can deduce this assertion through a reverting process: the station in grid $i+1$ was at $\hat{x}_i^*$ originally and it was shifted to $\hat{x}_i$. In this circumstance, Equation 6 represents the decrease in commuting cost. To sum up, we have the following proposition.

Proposition 1. *Suppose that Assumption 1 holds. Then the commuting cost within a grid area decreases (increases) as the metro station design becomes more compact (less compact).*

The proof of Proposition 1 along with calculation details of this section are in Appendix A. It is worth pointing out that there are two ways to increase the interval spacing of station x_i in grid i: shifting x_{i+1} to the right (x_{i+1} farther from CBD) as we showed, or shifting x_{i-1} to the left (x_{i-1} closer to CBD). The latter case leads to a different version of the total change of commuting cost resembling Equation 6, but the first derivative of the total commuting cost with respect to interval spacing is positive as in the former case. The relative sizes of the former and latter cases cannot be determined because their values depend on different settings of parameters. To sum up, Proposition 1 holds in both cases, although their marginal effects can diverge, and no clear dominance emerges when comparing the relative magnitudes of the two cases' marginal effects in general. Exploring this anisotropy's theoretical properties or manifesting it with empirical data has gone beyond the scope of this paper. The remaining part focuses on the x_{i+1} farther from CBD case, as discussed earlier.

3.2 The demand side

Now we proceed to formulate the residential location model from Proposition 1. A resident seeks to maximize the utility, $U(z, q)$, where z, q respectively represent the consumption of consumer goods and housing size. The budget constraint is $z + qr_i = y - t_i$, where r_i, y are housing rent in grid i and income respectively. We follow the standard bid rent approach of Fujita (1989). The bid rent for a housing unit in grid i is defined as

$$\Psi_i(u) = \max_{z,q}\{\frac{y - t_i - z}{q} | U(z, q) = u\}. \tag{8}$$

By solving the utility constraint for z, we have $z = Z(q, u)$. Then the bid rent function becomes

$$\Psi_i(u) = \max_{q}\{\frac{y - t_i - Z(q, u)}{q}\}. \tag{9}$$

The bid max lot size is obtained as $q_i = q(y - t_i, u)$. We can show the following properties of the bid rent and bid-max lot size: $\partial\Psi_i/\partial t_i < 0, \partial q_i/\partial t_i > 0$.

Housing units are produced by developers, with land and non-land (building materials) inputs. We assume a constant returns to scale technology for housing production. In this case, output is represented by floor-area ratio ρ, i.e., the ratio of floor to the whole area. The cost function for housing production is defined by $K(\rho)$, where $K' > 0$ and $K'' > 0$ are assumed. The profit of a developer per unit land area of grid i is

$$\pi_i^H = r_i\rho - K(\rho) - p_i, \tag{10}$$

where p_i is land rent. Assuming that the housing market is perfectly competitive, each developer is a price-taker. The condition for profit maximizing floor-area ratio is $r_i - K'(\rho) = 0$, and the solution is

denoted by $\rho_i = \rho(r_i)$. the floor area ratio is increasing with housing rent: $\partial\rho_i/\partial r_i = 1/K''(\rho_i) > 0$. Perfect competition implies that developers yield zero profit. So the land rent is determined by $p_i = r_i\rho_i - K(\rho_i)$.

Land market clearing condition is $\rho_i s^2 = n_i q_i$, where n_i is number of residents in grid i. Thus we obtain the population density as follows:

$$\frac{n_i}{s^2} = \frac{\rho_i}{q_i}. \tag{11}$$

Now we assess the effect of the interval spacing of stations on the population density. Compelling the equilibrium condition $r_i = \Psi_i(u)$:

$$\frac{\partial n_i/s^2}{\partial d_i} = \left(\frac{\partial \rho_i}{\partial r_i}\frac{\partial r_i}{\partial t_i}\frac{1}{q_i} - \frac{\partial q_i}{\partial t_i}\frac{\rho_i}{{q_i}^2}\right)\frac{\partial t_i}{\partial d_i} < 0. \tag{12}$$

Equation 12 depicts an essential implication of our model: reducing (expanding) the interval spacing of stations leads to a corresponding decrease (increase) in population density. The following assertion summarizes the preceding discussion.

Proposition 2. *Suppose that the prerequisites of Proposition 1 are satisfied. Then the population density of a grid area rises (falls) with a more compact (less compact) metro station design.*

Proposition 2 proposes a novel perspective on how interval spacing of metro stations (i.e., degree of compactness of metro station design) affects intracity demographic distribution. Mun et al. (2018) provide a sibling version of Proposition 2 in a continuous linear city model.

To discuss the impact of interval spacing on population density further, we specify Cobb-Douglas form utility function for the residents: $U(z, q) = \alpha \log z + \beta \log q$, where $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$. We further assume $\alpha > \beta$ by empirical evidence. The cost of housing production is specified as $K(\rho) = \lambda\rho^\gamma$, where $\lambda > 0$, $\gamma > 1$. With these specifications, one can express the population in terms of interval spacing:

$$n_i = \kappa(\zeta_0 + \zeta_1 d_i + \zeta_2 {d_i}^2)^{\delta}, \tag{13}$$

where $\kappa > 0$, $\delta > 1$, $\zeta_0 > 0$, $\zeta_1 < 0$, $\zeta_2 > 0$. We call Equation 13 the *demand side* equation, since it depicts the effects of metro stations' interval on population density. We put the calculation procedure of obtaining Equation 13 in Appendix B.

3.3 The supply side

Suppose that the transit system is designed by the social planner whose objective is to minimize the sum of transportation costs and construction costs of the transit system:

$$T = \sum_{i=1}^{J} (n_i t_i + c\frac{1}{d_i}) \tag{14}$$

where c is a constant and $1/d_i$ represents the density of stations. The government regards the population within grid i as exogenous. Minimizing T w.r.t. d_i (ignoring the interaction between d_i's) yields

$$d_i = \left(c^{-1}\phi_2 n_i\right)^{-1/2} \tag{15}$$

where Equation 7 is invoked. We call Equation 15 the *supply side* of this economy, since it depicts the impact of population density on interval spacing: as population density increases, the social planner tends to reduce the interval spacing of stations (i.e., increasing station density).

3.4 Relaxing restrictions on discrete city structure

Assumption 1 restricts the city structure that can be violated in reality. It is reasonable to replace this restriction with a more realistic one. We define a vacant grid as a grid that has no metro station, and a stacked grid as a grid equipped with two metro stations. We assume that the grid length s is small enough

to ensure no grid with three or more stations.

Assumption 1′. *There are neither two vacant grids in a row nor two stacked grids in a row.*

To evaluate the effect of metro station density on these two kinds of grids, we define $x_{k(i)}$ the nearest station of grid i's representative point from the left, and $d_i^+ = (x_{k(i)+1} - x_{k(i)})$ the generalized interval spacing between the nearest right and left stations of grid i.

For a vacant grid, $x_{k(i)}$ and $x_{k(i)+1}$ sit in the left and right neighbor grid respectively under Assumption 1′. It is possible that the critical boundary $\hat{x}_{k(i)}$ lies outside the grid i given extreme parameters, for example, if $x_{k(i)}$ is close to $(i-1)s$ and $x_{k(i)+1}$ locates near $(i+1)s$, then $\hat{x}_{k(i)} \equiv \phi x_{k(i)+1} + (1-\phi)x_{k(i)} \approx (i + 2\phi - 1)s$ is in $(i+1)$th grid for $\phi > 1/2$. We assume no such cases by compelling the following restriction:

Assumption 2 (No-over-vacant)**.** *The distance between station $x_{k(i)+1}$ and $x_{k(i)}$ satisfies $x_{k(i)+1} - x_{k(i)} \in (\underline{B}, \bar{B})$, where $\underline{B} \equiv \big((i-1)s - x_{k(i)}\big)/\phi$ and $\bar{B} \equiv \big(is - x_{k(i)}\big)/\phi$.*

Assumption 2 is just the algebraic transform of $(i-1)s < \hat{x}_{k(i)} < is$. It implies that the interval spacing is bounded above and below. Under Assumption 1′, for a vacant grid there is a natural upper bound $3s$ and a lower bound s, Assumption 2 further restricts the metro-relative parameters so that the critical boundary locates in grid i for analytic convenience. This restriction also has a clear meaning in practice for commuting cost minimization with respect to construction costs by social planner, in the sense that if it is violated, then dwellers in grid i will suffer from high access costs. A well-behaved vacant grid is illustrated in Panel A of Figure 3. Now we increase d_i^+ by shifting $x_{k(i)+1}$ to the right by sufficiently small Δd^+, resulting in a shift of critical boundary $\hat{x}_{k(i)}$ to $\hat{x}^*_{k(i)} \equiv \hat{x}_{k(i)} + \Delta d^+$. This is illustrated in Panel A of Figure 3. Consequently, the welfare of dwellers of both $[\hat{x}_{k(i)}, \hat{x}^*_{k(i)})$ and $[\hat{x}^*_{k(i)}, is)$ decrease based on discussions analogue to Equation 5 and Equation 4 respectively. Hence for the vacant grid case, the commuting cost is increasing with the interval spacing as the scenario of Assumption 1, and consequently the conclusion of

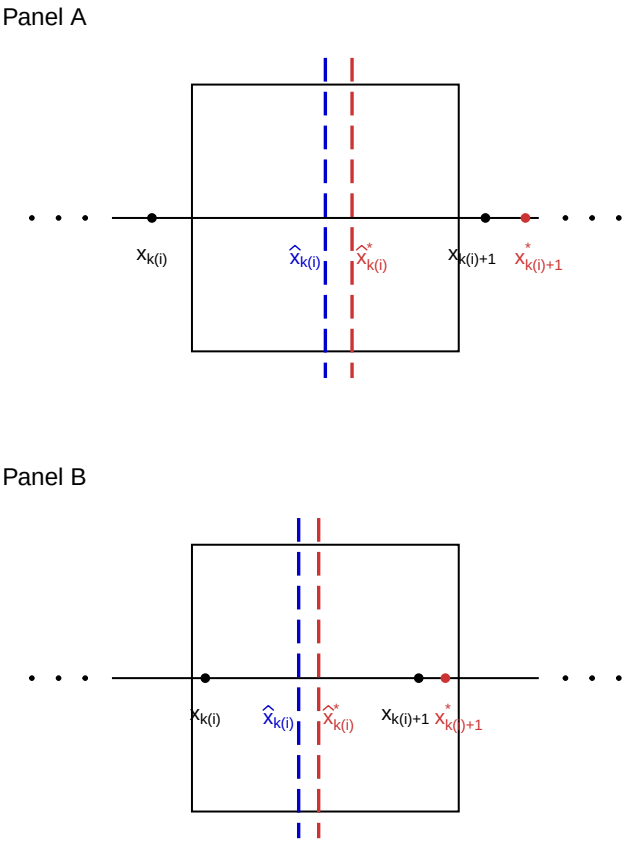

Figure 3: Well-behaved grids under Assumption 1′. Panel A illustrates a vacant grid that satisfies Assumption 2. Panel B illustrates a stacked grid that satisfies Assumption 3. x_j denotes jth station, $\hat{x}_j$ the critical boundary between the jth and $(j+1)$th station, and A^* the after-change variable of symbol A.

Proposition 1 and Proposition 2 can apply to vacant grid case.

In the case of stacked gird, both $x_{k(i)+1}$ and $x_{k(i)}$ are in gird i. Consider the situation that $x_{k(i)+1}$ shifts to the right by a small Δd^+. Consequently, the welfare of dwellers in $[\hat{x}_{k(i)}, is)$ changes. For dwellers in $[\hat{x}_{k(i)}, \hat{x}^*_{k(i)})$, their commuting cost increases because they change their riding decision due to the distantness of $(k(i)+1)$th station. Recall that in Equation 1, the commuting cost is the sum of line haul cost and access cost. This shift increases the line haul cost for dwellers in $[\hat{x}^*_{k(i)}, is]$. The access cost, for dwellers in $[\hat{x}^*_{k(i)}, x_{k(i)+1})$ is increasing due to distantness, for dwellers in $[x_{k(i)+1}, x^*_{k(i)+1})$ is unchanged due to symmetry, and for dwellers in $[x^*_{k(i)+1}, is]$ is decreasing due to nearness. The improvement of commuting condition for $[\hat{x}^*_{k(i)+1}, is)$ leads to analytical complexity, and hence we compel another structural restriction

to simplify the discussion.

Assumption 3 (No-over-stacked)**.** *The distance between station* $x_{k(i)+1}$ *and* $x_{k(i)}$ *satisfies* $(1-\phi)\left(x_{k(i)+1}-x_{k(i)}\right) > is - x_{k(i)+1}$ *as* $\Delta d^{+} \to 0$.

Assumption 3 ensures that the length of $[\hat{x}^{*}_{k(i)}, x_{k(i)+1})$ overwhelms the length of $[x^{*}_{k(i)+1}, is)$. This restriction implies that if two stations are stacked into one grid, they are intentionally placed at a considerable distance from each other by social planner. This arrangement is made to extend their transportation influence over a broader area. It means that in a well-behaved stacked grid i, the former station should be located near $(i-1)s$, while the latter should be close to is. Resembling the former, Assumption 3 restricts the metro-relative parameters to normalize a well-behaved stacked grid for analytical purposes. If Assumption 3 holds, then the total commuting cost of the stacked grid is increasing with greater Δd^{+}. A well-behaved stacked grid is illustrated in Panel B of Figure 3.

As the conclusion of the theoretical part, the following propositions summarize the discussions of well-behaved vacant and stacked grids. The proof resembles earlier discussions and hence omitted.

Proposition 1′. *Suppose that Assumption 1′ holds. Furthermore, (i) Assumption 2 is satisfied for vacant grids, and (ii) Assumption 3 is satisfied for stacked grids. Then the commuting cost within a grid area decreases (increases) with a more compact (less compact) metro station design.*

Proposition 2′. *Suppose that the prerequisites of Proposition 1′ are satisfied. Then the population density of a grid area rises (falls) with a more compact (less compact) metro station design.*

4 Data Description

We choose Wuhan, one of the greatest city of the central region of China, as our empirical research object. This section illustrates some basic information of Wuhan and details of our data.

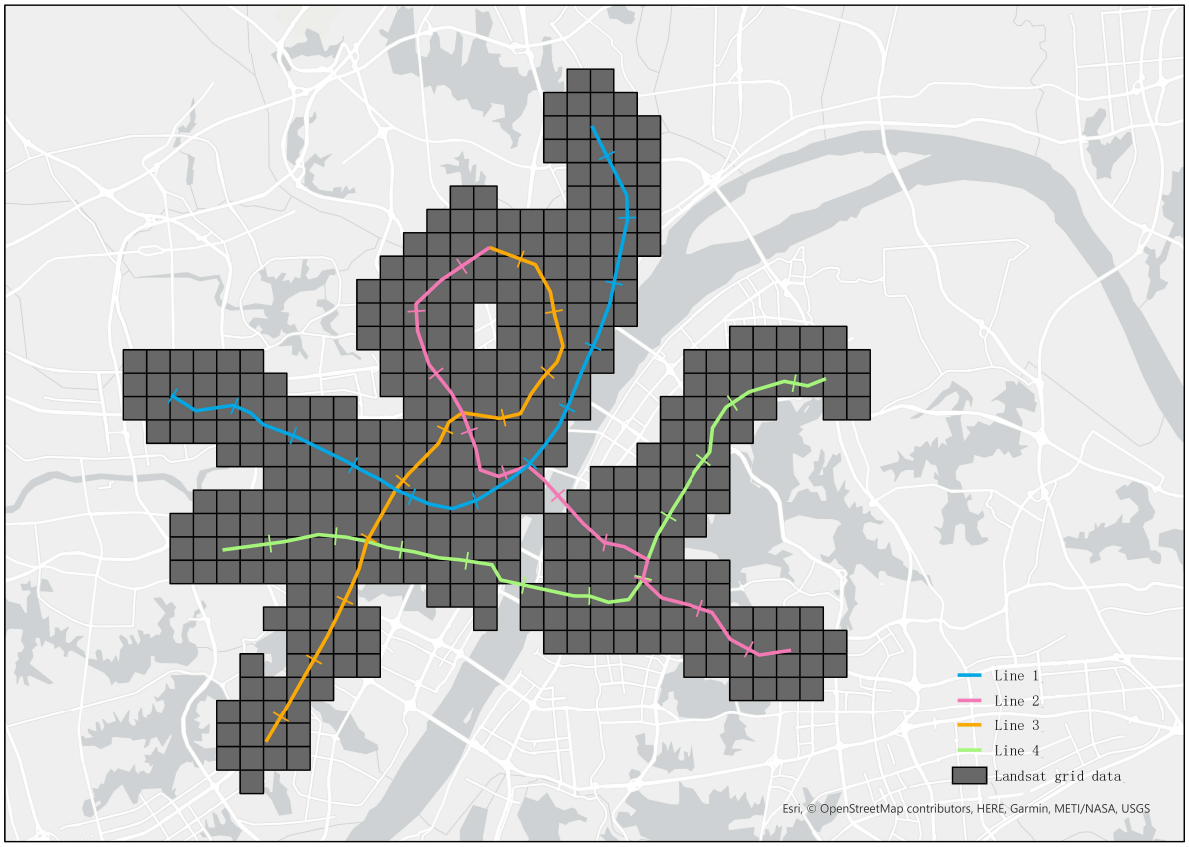

Figure 4: Observations and metro lines in this research.

4.1 Overview

Wuhan is situated in the central region of China, positioned at latitude 29°58′—31°22′ N and longitude 113°41′—115°5′ E. As the capital of Hubei Province, the Wuhan metropolitan area covers an expanse of $8569.15\,\mathrm{km}^2$, while the urban core of Wuhan city itself spans $905.62\,\mathrm{km}^2$. As of 2019, there were 3.30 million households and a registered population of 9.06 million, with individuals aged 18-59 constituting 62.13% of the population. The total employment in the region amounted to 6.23 million, distributed across primary, secondary, and tertiary sectors at rates of 7.97%, 36.96%, and 55.07%, respectively. Over the decade from 2009—2019, Wuhan's gross domestic product (GDP) experienced an average annual growth rate of 13.38%, culminating at 1,622.321 billion CNY (234.843 billion current USD) by the close of 2019. This economic performance positioned Wuhan as the 8th largest city in mainland China in terms of GDP.

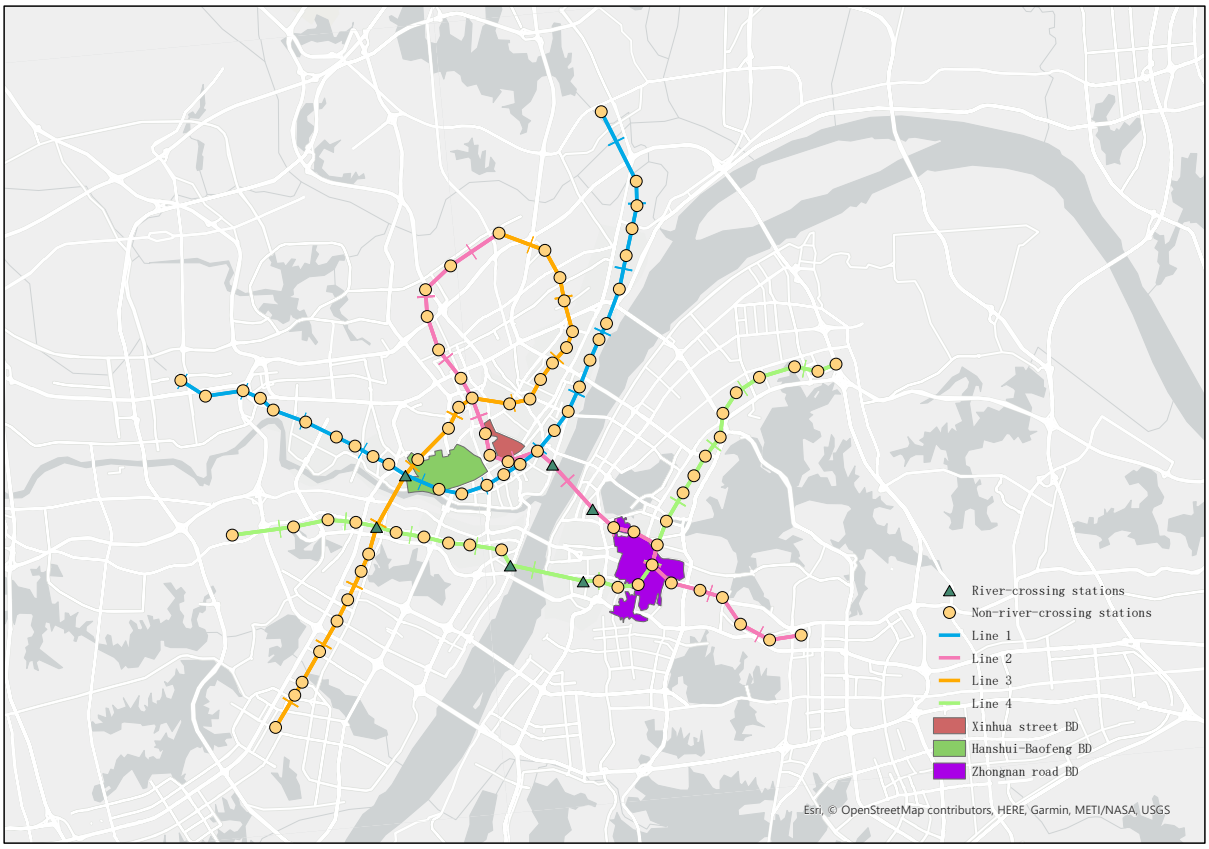

Figure 5: Metro stations of river-crossing and non-river-crossing, metro lines and identified BDs. River-crossing stations: Jianghan Road-Jiyuqiao (Yangtze River) of Line 2; Lanjiang Road-Fuxing Road (Yangtze River) of Line 4; Zongguan-Wangjiawan (Han River) of Line 3.

Sitting at the confluence of the Yangtze River and the Han River in the Yangtze Plain, Wuhan enjoys a significant advantage in terms of nationwide transportation. The ferry routes passing through Wuhan connect it to the west, linking dozens of cities in the interior such as Chongqing and Yichang. To the east, it serves as a gateway to one of the world's largest megalopolis areas: the Yangtze River Delta. Additionally, as a key stop along the Beijing–Guangzhou railway, one of China's most important north-south rail corridors, Wuhan serves as a vital link between the other two highly developed megalopolises in China: the Beijing-Tianjin-Hebei Metropolitan Area and the Pearl River Delta Economic Zone. This central position has earned Wuhan the renowned title of the 'Nine Provinces' Thoroughfare'.

business district	scale (people)	density $\text{people}/\text{km}^2$	ratio of total	ratio of major urban area
Xinhua Street	56,793	48,571.73	1.087%	4.527%
Hanshui-Baofeng	144,301	33,016.23	2.763%	11.502%
Zhongnan Road	124,748	16,421.65	2.389%	9.944%
Total	325,842	24,804.46	6.239%	25.973%

Table 1: Descriptions of employment information of identified business centers of Wuhan in 2013. The total employment scale of Wuhan in 2013 is 5,222,400 people (data from the Wuhan Statistical Yearbook-2014). The employment scale of major urban area of Wuhan is 1,254,535 people (data from Zou (2018)). Representative stations for Xinhua Street BD: Youyi Road (Line 1), Zhongshan Park (Line 2). Representative stations for Hanshui-Baofeng BD: Taipingyang (Line 1), Qiaokou Road (Line 1). Representative station for Zhongnan Road BD: Zhongnan Road (Line 2 & 4).

4.2 Wuhan's employment center

We consider the employment centers in Wuhan as 'business centers' (referred to as BDs) as used in the theoretical part. Building on the work of Zou (2018), we identify three BDs from among four sub-districts. Zou (ibid.) identified four sub-districts as employment centers in the third economic census of Wuhan in 2013 are: Xinhua Street (located in Jianghan District), Hanshui Bridge (in Qiaokou District), Baofeng Street (in Qiaokou District), and Zhongnan Road (in Wuchang District). We group Hanshui Bridge and Baofeng Street as a single BD due to their geographic proximity. Table 1 provides a summary of the employment data for these identified BDs.

4.3 Wuhan's metro

Wuhan metro is the mass transit system operated by Wuhan Metro Group Co., Ltd. By the end of 2019, the metro was constituted of 228 stations of 9 lines, and $334.42\,\text{km}$ of route length. Due to the limitation of the data, we focus on the area covered by the Line 1, 2, 3, and 4. Elaborated metro lines are detailed in Table 2 and illustrated in Figure 4 and Figure 5.

Segment description	Date opened	No. of stations	Route length (km)
Line 1: phase i	2004.7.28	10	9.769
Line 1: phase ii	2010.7.29	16	18.494
Line 1: Hankou north extension	2013.5.28	3	5.555
Line 2: phase i	2012.12.28	21	27.152
Line 3	2015.12.28	15	29.660
Line 4: phase i	2013.12.28	15	15.429
Line 4: phase ii	2014.12.28	13	17.974

Table 2: Descriptions of metro lines we used in empirical research.

4.4 The data

The Resource and Environment Science and Data Center of China has published Landsat data for a $1\,\mathrm{km}\times 1\,\mathrm{km}$ grid of population distribution covering the entire mainland China for the years 2005 and 2015 (Xu 2017). Areas predominantly covered by hydrology have been excluded from the dataset. To address the slight misalignment of representative points between 2015 and 2005, we have adjusted the 2005 population figures. This adjustment assumes a uniform distribution of population within each grid and calculates a geographical average based on the 2015 representative points. Our selection of observation grids is based on the grids that intersect with a $2\,\mathrm{km}$ buffer zone around metro lines, as illustrated in Figure 5.

To perform regression analysis, covariates for each observation must be computed or synthesized. We opt to designate the geographic center of each observation as its representative point. The metric "metro proximity" (l) is defined as the distance from the nearest station on the selected metro lines to the representative point of each observation. The primary explanatory variable, "interval spacing" (d), for each observation represents the average (geographical, not architectural) distance between its nearest station and both the former and latter stations. In the case of a terminus station, d simply corresponds to the distance from the former station. For interchange stations, d is calculated as the average distance between both lines. The "commuting time" (t) is determined as the shortest riding duration from the nearest station to the representative stations listed in Table 2. We sourced coordinates and interval length data between stations from the official website of Wuhan Metro Group Co., Ltd.[2]. The computation of l is carried out using the

[2] http://www.wuhanrt.com/ (in Chinese)

geometric processing software *ArcGIS*, while t is obtained from the navigation application *Gaode Maps*.

Our objective is to alleviate individual heterogeneity that may introduce bias into our estimation coefficients. To achieve this, we control for variables that capture entity-specific characteristics. One such variable, denoted as op, is crafted to represent the time difference between December 2015 and the month when the nearest metro station to each observation became operational. For instance, if $op = 10$, it signifies that the nearest station opened in February 2015.

Additionally, we seek to account for the developmental level of each observation, as this information is not readily available from authoritative sources. We address this by synthesizing a developmental index through the following procedure. Initially, we collect data on the total length of the main roads within the observation areas, designated as $roadlength$. Subsequently, we calculate the number of government facilities within a circular region with radius α_1 centered at each observation's representative point, denoted as $\#government$. The government data is sourced from Wuhan's point-of-interest (POI) dataset, and both stages of this process are conducted using *ArcGIS*.

The resulting synthetic developmental index (sdi) is formulated as follows:

$$sdi = \alpha_2[1 - q(roadlength)] + (1 - \alpha_2)[1 - q(\#government)], \tag{16}$$

Here, $q(\cdot)$ returns the quantile rank of $\cdot$ among all observations (e.g., $q(x_i) = 0.06$ indicates that the value of x for the ith observation ranks in the top 6% of values), and $\alpha_2 \in [0, 1]$. It is clear that sdi lies in the interval [0, 1], with higher values indicating a higher degree of development.

We elucidate the rationale behind our selection of variables for synthesizing the developmental index (sdi). Road length is widely recognized as a robust proxy for assessing development. Our government-related point-of-interest (POI) data comprises government agencies at various administrative levels (provincial, municipal, and county), government-affiliated social groups (e.g., fire control and service centers for governmental activities), civil society organizations (such as welfare institutes, disabled persons' federation

offices, Red Cross branches, and chambers of commerce), and legal institutions (including police departments, courts, and security checkpoints).

The benefit of utilizing government-related POI data lies in two key aspects: (i) The quantity of such data serves as a reflection of an area's significance, and (ii) it is less susceptible to alteration due to demographic shifts. Conversely, introducing POI variables like restaurants, entertainment venues, medical institutions, or shopping services, which are highly susceptible to demographic adjustments, can introduce endogeneity issues into our estimation due to simultaneity.

To ensure robustness, we generate multiple versions of the developmental index (sdi) by experimenting with different values of $\alpha_1 \in \{1.6, 2.0, 2.5, 3.0\}$ and $\alpha_2 \in \{0.33, 0.5, 0.67\}$. All of these versions exhibit similar characteristics and yield results with only minor variations. Consequently, we opt for the parameters $(\alpha_1, \alpha_2) = (2.0, 0.5)$ as our final choice.

To assess the reliability of our sdi composition, we conduct a nonparametric test for detecting measurement errors following the methodology of Lee et al. (2020). The results confirm that the sdi generated with specific parameters is not affected by measurement errors, reinforcing its suitability as a proxy in our estimation. A summary of the variables employed in the regression is provided in Table 3.

5 Empirical Results

In this section, we elucidate the identification and estimation process of our regression model and present the empirical findings.

5.1 Methodology and Identification

Our primary objective is to empirically substantiate that the increase in population density is attributed to the increase of station density (i.e., reduce of interval spacing), aligning with the theoretical discussions outlined in Equation 13. We estimate the following regression equation:

$$pop_{2015} - pop_{2005} = \beta_0 + \beta_1 d + \mathbf{x}'\boldsymbol{\beta} + \varepsilon \tag{17}$$

Here, pop_t represents the logarithm of population density in year t, $\mathbf{x}$ encompasses the necessary covariates, and ε denotes the error term. For simplicity, we omit the subscript i denoting individual observations. Equation 17 can be viewed as a constrained IV regression with a linear restriction, where the coefficient of the lagged population (pop_{2005}) is constrained to be 1.

Let us elucidate the structure of Equation 17: (i) It carries an explicit economic interpretation: the R.H.S. variables account for the differences in population distribution. (ii) Acknowledging potential measurement error in Landsat data (Kawaguchi et al. 2017), restricting the coefficient of pop_{2005} to 1 ensures the unbiasedness and consistency of the regression. (iii) The variable d is predetermined in 2015 but not in 2005. Constraining the coefficient of pop_{2005} to unity safeguards the identification of β_1. (iv) This constraint also partially alleviates time-invariant unobservable individual heterogeneity among entities.

While prior research suggests that metro station locations are endogenous, this consideration is reasonable when studying geographic data of administrative-based units (e.g., tracts, counties, cities, or provinces) (Baum-Snow and Kahn 2000; Tyndall 2021). However, in this study, demographic data is observed at a Landsat-based level ($1\,\mathrm{km} \times 1\,\mathrm{km}$ grid) rather than traditional administrative units. The key distinction lies in the way our data is structured: the coordinates of observation representative points can be regarded as randomly assigned. Consequently, station proximity (l) in our research is exogenous, as it is solely determined by the distance from a representative point to its nearest station. This exogeneity arises from the fact that it is implausible to assume that the social planner would select station locations based on representative points dictated by Landsat data, rather than considering geographical and economic characteristics. In essence, the existence of Landsat-based data provides observations akin to those derived from a randomized experiment.

However, the key variable of this study, interval spacing d, is afflicted by endogeneity since it is inherently determined by station locations. To address this, we propose an instrument to met the endogeneity

issue: whether the nearest station crosses rail tracks over a river. A river-crossing station is associated with a higher value of d because its location is constrained by the width of the river, rendering it impractical to construct a station directly over the river. Stations crossing other types of hydrology, such as lakes or reservoirs, are subject to selection bias. The government can avoid constructing metro lines that cross a lake by adjusting station locations to minimize costs. However, constructing river-crossing stations is inevitable for connecting the built-up areas of Wuhan due to its specific hydrological characteristics. River-crossing stations are highlighted in Figure Figure 5. Descriptive statistics for this instrument are in Table 3.

We aim to provide a clear explanation of the parameters to be estimated and what they represent. We establish certain conditions for the causal interpretation of Equation 17:

Assumption 4. *The distribution of pop_{2005} conditional on non-metro covariates is the same across observations.*

This assumption requires that the observations follow the same distribution before any treatment. This condition is standard in empirical research, and our data appears to meet this requirement, given that they originate from adjacent areas within the city of Wuhan.

Assumption 5. *The distribution of pop_{2015} conditional on $(d, \mathbf{x}_1)$ is the same across observations.*

This assumption relies on two factual premises: (i) the process of urbanization in the sampled area was concluded by 2005. This assertion safeguards us against overestimating the treatment effect of metro-related variables due to unidentifiable exogenous positive demographic shocks. The proportion of grids designated as urban or rural construction land is 47.5% in 2005 across all observations, signifying a relatively high level of urbanization. The fact that this proportion only slightly increases to 50.3% in 2015 further supports this assertion. (ii) There were no regime changes in the regional development plan from 2005 to 2015. If this premise were untrue, no time-invariant parameter could effectively assess regional dummies. Under Assumption 4 and 5, the coefficients of metro-relative variables in Equation 17 can be identified as marginal

effects and consistently estimated using the IV method.

The identification of marginal effects hinges on the fact that "almost all" metro stations were constructed between 2005 and 2015. However, there are stations, specifically those belonging to Phase i of Line 1, that were built before 2005. We retain these data in our dataset for two reasons: firstly, this line covers grids containing the BD area and features crucial interchange stations that cannot be disregarded. Secondly, it only impacts approximately 4.5% of the observations. Consequently, this issue does not significantly undermine the causal interpretation of metro-relative variables. The primary risk is the potential underestimation of the marginal effects of metro-relative variables.

A substantial subset of our observations did not experience treatment by the MTS before 2015. Consequently, conventional panel data estimation methods cannot be directly applied to assess the data's performance in this research context. To address this limitation, we expand our dataset by incorporating population data for the year 2010. For the untreated observations at year t (i.e., those observations where the nearest metro station had not been constructed by year t), we set the metro-relative variables equal to the maximum values observed in the year 2015. In our research, higher values of metro-relative variables signify poorer transit accessibility.

With this expanded dataset, we employ the following FE estimation:

$$pop_{it} = \theta_1 d_{it} + \mathbf{x}_{it}'\boldsymbol{\theta} + \boldsymbol{\Theta}_t + \boldsymbol{\varepsilon}_{it}, \tag{18}$$

In this equation, pop_{it} represents the logarithm of population, d_{it} stands for the interval spacing of metro stations, $\mathbf{x}_{it}$ encompasses the necessary covariates, $\boldsymbol{\Theta}_t$ accounts for the time fixed effect, and $\boldsymbol{\varepsilon}_{it}$ represents the error terms. Here, i varies from 1 to 400, while t takes on the values of 2015, 2010, and 2005, corresponding to individual and time indexes.

One limitation of this approach is that Equation 18 cannot eliminate the individual fixed effects due to the issue of multicollinearity. This multicollinearity arises because we assume that unobservable variables

related to incomplete nearest station construction take on their maximum values in the year 2015. Consequently, while the estimates obtained from Equation 18 contribute supplementary evidence to our analysis, they do not fully resolve the challenge stemming from the absence of pretreatment data.

5.2 Parameter Estimates

5.2.1 Baseline estimates

Table 4 reports the OLS and IV estimates of the demand side equation. We use Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors of Conley (1999) to calibrate the spatial dependence.

First, we compare the OLS and IV estimates under the same covariate settings (columns 1 and 5, 2 and 6, 3 and 7, and 4 and 8). These comparisons reveal that endogeneity trends lead to an underestimation of the marginal effect of d. This suggests that the government tends to construct stations with smaller values of d in areas with higher current population density, rather than considering future population levels.

Second, we find that the coefficients for l and t are insignificant in columns 1 and 5 but have the expected signs. One plausible explanation for this phenomenon is that regions near Business Districts (BDs) experienced earlier development and occupation, which discouraged new citizens from settling. Consequently, new migrants had to accept residing areas with higher values of l and/or t. This hypothesis gains support when we stratify the observations by the population level in 2005, as columns 2 and 6 reveal the statistical significance of l and t. This empirical finding provides support for Assumption 5 by shedding light on the state of urbanization in the observed area during the pre-treatment period.

Third, to test the robustness of our estimates under various covariate circumstances, we introduce additional dummy variables into columns 7 and 8. Remarkably, the estimates of d in columns 7 and 8 remain robust, leading us to choose column 6 as the baseline estimate.

The first stage estimation of columns 5-8 is significantly positive at 0.1% level, reinforcing our assertion regarding suitability of the instrument. The Cragg-Donald Wald F Statistic of baseline estimation is 70.711,

and the Kleibergen-Paap Wald F Statistic is 136.459, both strongly rejecting that the instrument is weak[3].

Based on the baseline estimation, we find that reducing the interval spacing between metro stations by $100\,\mathrm{m}$ results in a 4.48% increase in the population density of a region. This outcome aligns with the predictions of our theoretical framework. Conversely, being situated $100\,\mathrm{m}$ farther away from a station leads to a 1.08% decrease in population density.

The relatively lower level of statistical significance (in comparison to column 2) may be attributed to the potential loss of efficiency introduced by the IV method. Nevertheless, the empirical evidence underscores that the interval spacing of metro stations exerts a significantly stronger causal effect on aggregate population density than station proximity, and this effect is characterized by higher statistical significance. This observation emphasizes that the influence and importance of interval spacing (metro station density) on metropolitan structures have been underestimated and overlooked.

Furthermore, an additional minute of commuting time to the nearest Business District (BD), as depicted in Table Table 1, corresponds to a 1.91% decrease in population density. The coefficient for the open time of the nearest station is -0.0077, indicating that older stations hinder population growth. When we consider that a 1% increase in the synthetic development index results in approximately a 3.6% decrease in population density, it becomes evident that regions with earlier and more extensive development exhibit slower population growth. This observation suggests that by 2005, the level of urbanization in Wuhan had reached a high point.

5.2.2 FE estimates

Table 5 presents the FE estimates. In columns 1-5, we examine the performance of metro-relative variables without controlling for district dummies. Across all covariates settings in these columns, we consistently observe negative estimates for the metro-relative variables, aligning with our predictions.

[3]The corresponding Stock-Yogo critical value of 10% maximal IV size is 16.38.

Columns 6-8 estimate the marginal effects while incorporating district dummies. Comparing columns 6 and 7 with columns 4 and 5, respectively, we notice an improvement in goodness of fit and a reduction in estimated marginal effects. This suggests the crucial role of district dummies in this empirical analysis, as was the case in the baseline regressions[4].

In column 7, the estimated marginal effect of interval spacing is -0.0233, while estimates of other metro-relative variables remain negative, in accordance with our predictions. Column 8 introduces the synthetic development index (sdi) into the regression. However, this inclusion necessitates the restrictive assumption that the development index, representing the relative importance of each grid, remains constant over time. Therefore, we believe that column 8 yields inferior results compared to column 7.

It is worth pointing out that the estimates in column 7 of Table 5 closely resemble the baseline IV results (column 6 of Table 4) under similar covariate settings. Notice that Table 5 reports estimates over a 5-year span.

Furthermore, FE regressions demonstrate statistical significance with fewer covariate settings compared to OLS regressions (columns 1-4 in Table 4), and their results align more closely with the IV estimates. Therefore, in the empirical analysis of MRT, we generally conclude that the FE estimation method (including a specific DiD approach) outperforms OLS. However, it's worth emphasizing that in this specific research, FE may not accurately estimate the marginal effects of metro-relative variables due to the absence of pre-treatment measures. Nevertheless, the empirical evidence presented in Table 5 provides corroboration that supports the baseline implications.

6 Policy Counterfactuals

In this section, we construct a discrete linear city model, as outlined in Section 3.1, and carry out policy experiments based on the data and previous empirical findings. Our chosen representation of the linear city

[4]Estimates without district dummies are omitted due to a substantial decrease in goodness of fit and unreliable point estimates.

is the north-west segment of Line 4, extending from Zhongnan Road BD to the terminal station, Wuhan Railway Station. This segment, a relatively linear portion of our dataset, encompasses 13 metro stations and 53 Landsat observations.

To create the discrete linear city for use in policy experiments, we must flatten the selected line in reality. This process is akin to straightening a chain of interconnected rigid sticks, where each stick maintains a fixed length that cannot be extended or shortened while its position is adjusted. The varying lengths of the sticks correspond to the spacing between metro stations, the junctions between sticks represent intermediate stations, and the endpoints of this chain represent the BD and the terminal station. Once this flat representation is established, we generate adjacent grid cells of $1\,\mathrm{km} \times 1\,\mathrm{km}$ that intersect the flat line along its x-axis. Figure 6 visually illustrates this procedure using a four-station example.

The subsequent part of this section involves deriving policy counterfactual scenarios for these grid cells. We achieve this by altering the positions of metro stations and subsequently assessing the resulting changes in aggregate population, following the parameters presented in column 6 of Table 4.

6.1 Overall fitness

Before delving into the policy analysis, our initial step involves assessing how well the theoretical population predictions align with the real data. One challenge we face is the absence of real-world counterparts to the fabricated grids, as illustrated in Panel B of Figure 6, rendering their demographic characteristics unobservable. To address this issue, we employ a k-nearest-neighborhood (kNN) method to estimate the population of the fabricated grids using observed demographic data, thus treating the kNN population as a representation of reality.

To elaborate, consider the representative point of the second grid in Panel B of Figure 6, denoted as I_2, which lies between x_1 and x_2. We calculate its geographic coordinates as a convex combination of the coordinates of x_1 and x_2, where the coefficient of the convex combination is determined by the relative

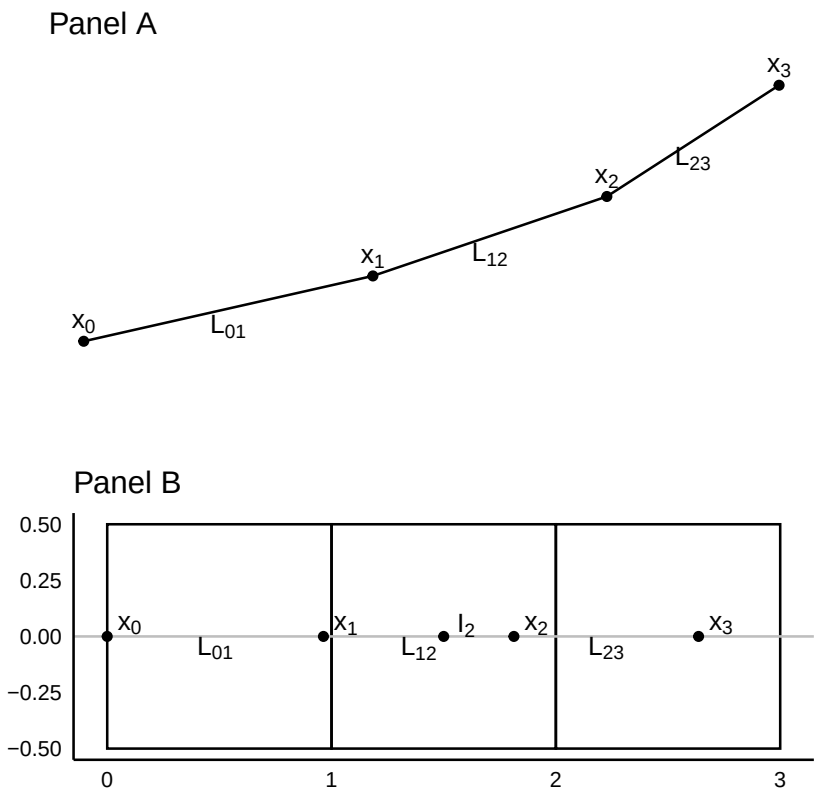

Figure 6: Four-station example of how we recover a discrete city for purpose of policy counterfactual analysis from an observed data of city. Panel A represents a fictitious city with four metro stations, where x_i is the ith station, I_j the representative point (geometric center) of grid j and L_{ij} the geographic distance between station i and j. Panel B illustrates the corresponding discrete city of Panel A, where the metro line is straightened, matching the x-axis with L_{ij} unchanged.

distances between I_2 and both x_1 and x_2. Once the coordinates of the representative point are determined, we identify k representative points in the real data that are geometrically closest to x_2. Subsequently, we assign the arithmetic average of the data from these points to I_2 –namely the method of kNN.

The theoretical prediction for the fabricated grids is recovered through a functional transformation of Equation 17:

$$\widehat{pop_{2015}} = exp\left[\hat{\beta}_0 + \hat{\beta}_1 d + \mathbf{x}'\hat{\boldsymbol{\beta}}\right] \cdot pop_{2005}. \tag{19}$$

Here, the L.H.S. of Equation 19 represents the prediction, while the right-hand side variables, including

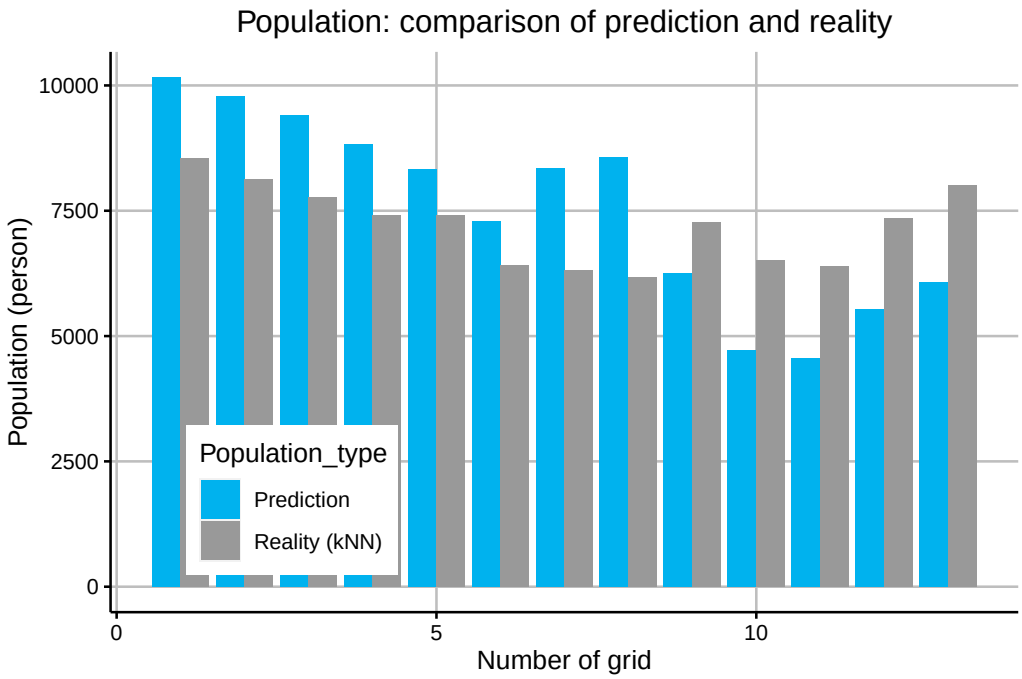

Figure 7: Theoretical and real population of selective line. kNN tuning parameter: $k = 4$. Root of mean squared prediction error: 1669.49 people. Average prediction accuracy: 77.25%.

metro-relative variables d, t, and op from real data, and the recalculated l from the linear grid-based city. Grid-relative covariates sdi, reg, and $popquan$ are also derived using the kNN approach. The respective parameters $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\boldsymbol{\beta}}$ are obtained from column 6 in Table 4, and the error term is omitted. We compare the predicted population distribution along the fabricated grids with the actual data distribution, as illustrated in Figure 7.

We now appraise the concordance between kNN-derived population estimates and our theoretical predictions. The results of our theoretical predictions align with our commuting time theory and the empirically estimated parameters in two significant ways: First, we observe a consistent downward trend in population density as the distance from the CBD increases, attributed to longer commuting times. Second, our theoretical predictions capture local deviations where the population of grids exceeds the expected values due to their proximity to metro stations. Additionally, the agreement between our predictions and the kNN-derived reality is quite satisfactory, with a root mean squared prediction error of 1669.49 people and an average prediction accuracy of 77.25%. One potential limitation is the tendency to overestimate the population in

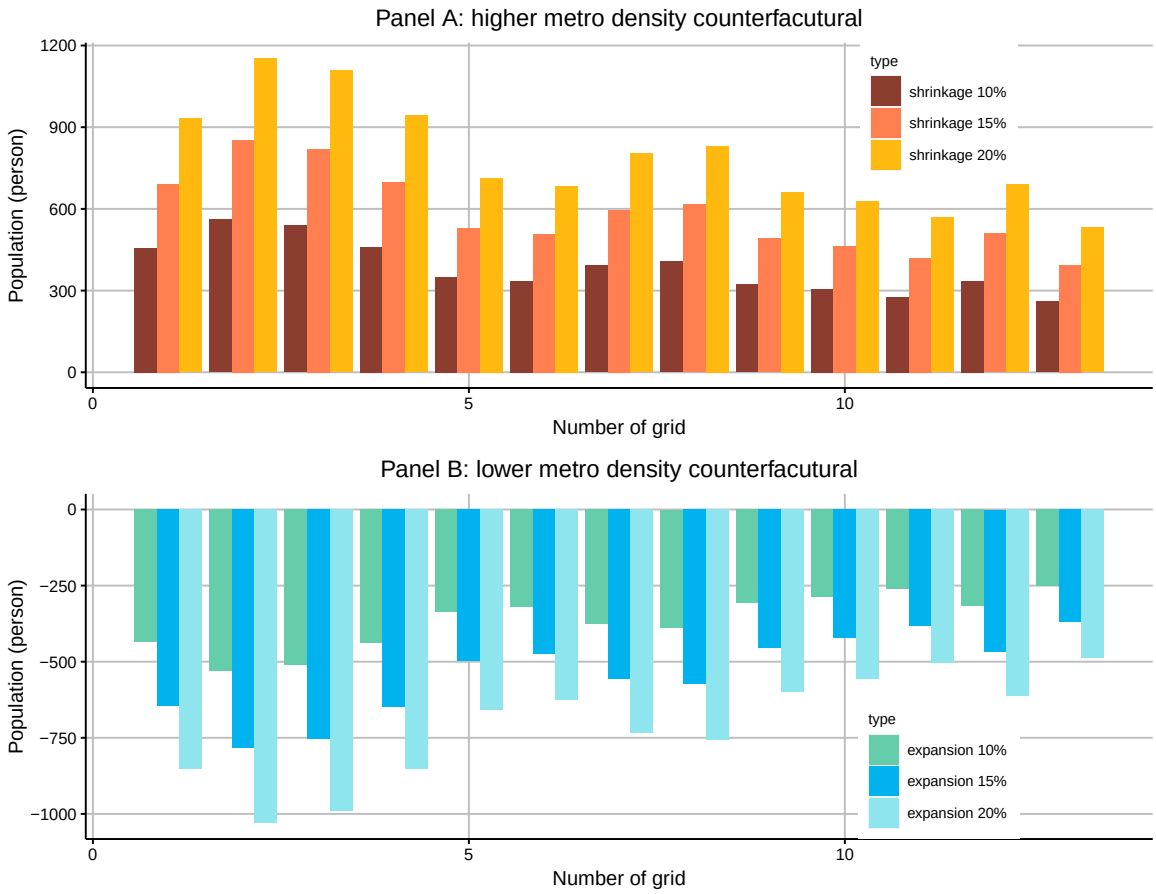

Figure 8: Result of policy counterfactuals. Panel A illustrates the counterfactuals of higher metro density. Panel B illustrates the counterfactuals of lower metro density.

the initial grids and underestimate it in the later grids of the sequence.

6.2 Counterfactual: interval spacing

In this section, we explore the counterfactual scenario of altering interval spacing (d) while keeping all other factors constant (*ceteris paribus*). To achieve this, we utilize the 13 stations along the selected line mentioned earlier to construct a theoretical city that closely resembles Panel B of Figure 6.

The counterfactual scenario assumes a metro station design that uniformly adjusts d of each grid within this theoretical city by a consistent percentage, denoted as $a\%$, while keeping all other metro-related variables unchanged. We measure the impact of this policy change by examining the resulting change in $\widehat{pop}_{2015}$.

In Panel A of Figure 8, we compare the counterfactuals to the baseline prediction for each grid, considering interval spacing reductions ranging from 10% to 20%. As expected, a denser metro station design leads to a more significant increase in population density. It is not surprising that all counterfactuals demonstrate that grids located closer to the CBD experience a greater increase in population density compared to remote grids. This phenomenon is due to the fact that grids near the CBD already have a higher population density and therefore benefit more from improvements in metro station design.

We introduce the concept of the 'effective city length,' defined as the length of the policy counterfactual city that equalizes the population with the baseline city. Specifically, let N represent the total population of the baseline, and N_i^* denote the population of the ith grid in the policy counterfactual. We identify the grid number, denoted as i^*, that satisfies $N_{i^*-1}^* < N$ and $N_{i^*}^* \geq N$. The effective city length is then calculated as $L^* = i^* - 1 + l^*$, where l^* is chosen to satisfy $\sum_i^{i^*-1} N_i^* + N_{i^*}^* l^* = N$. It's important to note that this concept assumes a uniform population distribution within the grids of the policy counterfactuals. For the case of contraction, the effective city lengths for 10%, 15%, and 20% contractions are respectively $12.210\,\mathrm{km}$, $11.813\,\mathrm{km}$, and $11.413\,\mathrm{km}$.

In Panel B of Figure 8, we present the counterfactuals for the expansion of station intervals. Similar to the contraction scenarios, we observe a hierarchical decrease in population density with increasing degrees of expansion. Grids near the CBD also experience a more significant decrease in population density due to the deterioration of the metro station design. Using the concept of effective city length, it takes $13.816\,\mathrm{km}$, $14.235\,\mathrm{km}$, and $14.661\,\mathrm{km}$ for the 10%, 15%, and 20% expansion counterfactuals, respectively, to equalize the total population with the baseline. Since the number of grids exceeds 13 in the baseline, we assume that the additional grids in the policy counterfactuals have the same population density as the last grid in the baseline.

7 Concluding Remark

This paper delves into the causal impact of metro design compactness on urban population density. To illuminate this pioneering investigation, we establish a comparative statics partial equilibrium model. In our model, dwellers' behavior is formalized as utility maximization with respect to housing size and other goods, while government behavior is assumed to be the minimization of social commuting cost. This model illustrates how the compactness of metro design plays a pivotal role in determining local commuting costs and, consequently, population density.

To obtain a closed-form solution, we streamline certain theoretical assumptions and impose constraints on the theoretical structure of an ideal city. The central implication of our model, as outlined in Equation 13, unveils the causal relationship between metro design compactness and demographic distribution, aligning with intuitive expectations.

For empirical validation, we gather demographic and transportation data from Wuhan, China, spanning the years 2005 to 2015. We develop a regression design that mirrors the central implication of our theory while mitigating inconsistencies arising from the measurement error of Landsat data. We employ a unique instrument to address endogeneity concerns regarding metro station placements, which are simultaneously determined within the social equilibrium framework and linked to the hydrological structure of Wuhan city.

The results of our Instrument Variable (IV) estimation indicate that a 100-meter increase in metro station average spacing leads to a 4.48% reduction in local population density. This finding aligns with both our theoretical framework and intuitive expectations, demonstrating robustness across various covariate settings.

We employ parameters from reduced-form estimation to estimate the aggregate local population, closely resembling the city model within our discrete city theory. Subsequently, we leverage this population estimate for policy counterfactual analysis. To reconcile the collected data with the theoretical counterfactual city, we employ a k-nearest-neighbor (kNN) method.

The counterfactual scenarios elucidate that a denser metro station design results in higher population density, and conversely, a sparser design leads to lower population density. Introducing the concept of 'effective city length,' as previously defined, we observe that a 10%, 15%, and 20% reduction in metro station intervals shrinks the current city length to $12.210\,\mathrm{km}$, $11.813\,\mathrm{km}$, and $11.413\,\mathrm{km}$, respectively, to accommodate existing residents. Conversely, for expansion counterfactuals of 10%, 15%, and 20%, the effective city lengths increase to $13.816\,\mathrm{km}$, $14.235\,\mathrm{km}$, and $14.661\,\mathrm{km}$, respectively.

References

Ahlfeldt, Gabriel M. et al. (n.d.). "The Economics of Density: Evidence From the Berlin Wall". In: *Econometrica* 83 ().

Asahi, Kenzo, Andrea Herrera, and Hugo E. Silva (2022). "The Effects of Land-Use Regulation and Transport Infrastructure on the Distribution of Economic Activity". In: *SSRN*. URL: https://ssrn.com/abstract=4187215.

Baum-Snow, Nathaniel, Loren Brandt, et al. (2017). "Roads, Railroads, and Decentralization of Chinese Cities". In: *The Review of Economics and Statistics* 99.3, pp. 435–448. URL: https://doi.org/10.1162/REST_a_00660.

Baum-Snow, Nathaniel and Matthew E. Kahn (2000). "The Effects of New Public Projects to Expand Urban Rail Transit". In: *Journal of Public Economics* 77.2, pp. 241–263. ISSN: 0047-2727. URL: https://www.sciencedirect.com/science/article/pii/S0047272799000857.

Baum-Snow, Nathaniel, Matthew E. Kahn, and Richard Voith (2005). "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000". In: *Brookings-Wharton Papers on Urban Affairs*, pp. 147–206. ISSN: 15287084. URL: http://www.jstor.org/stable/25067419.

Bellemare, Marc F., Takaaki Masaki, and Thomas B. Pepinsky (2017). "Lagged Explanatory Variables and the Estimation of Causal Effect". In: *The Journal of Politics* 79.3, pp. 949–963. URL: https://doi.org/10.1086/690946.

Bertaud, Alain and Harry W. Richardson (2004). "Transit and Density: Atlanta, the United States and Western Europe". In: *Urban Sprawl in Western Europe and the United States*. Ed. by Chang-Hee Christine Bae and Harry W. Richardson. Routledge.

Billings, Stephen B. (2011). "Estimating the Value of a New Transit option". In: *Regional Science and Urban Economics* 41.6, pp. 525–536. URL: https://www.sciencedirect.com/science/article/pii/S0166046211000500.

Bowes, David R. and Keith R. Ihlanfeldt (2001). "Identifying the Impacts of Rail Transit Stations on Residential Property Values". In: *Journal of Urban Economics* 50, pp. 1–25. URL: https://www.sciencedirect.com/science/article/pii/S0094119001922144.

Brooks, Leah and Byron Lutz (2019). "Vestiges of Transit: Urban Persistence at a Microscale". In: *The Review of Economics and Statistics* 101.3, pp. 385–399. URL: https://doi.org/10.1162/rest_a_00817.

Conley, T.G. (1999). "GMM Estimation with Cross Sectional Dependence". In: *Journal of Econometrics* 92.1, pp. 1–45. URL: https://www.sciencedirect.com/science/article/pii/S0304407698000840.

Dubé, Jean, Marius Thériault, and François Des Rosiers (2013). "Commuter Rail Accessibility and House Values: The Case of the Montreal South Shore, Canada, 1992–2009". In: *Transportation Research Part A: Policy and Practice* 54, pp. 49–66. URL: https://www.sciencedirect.com/science/article/pii/S0965856413001377.

Fujita, Masahisa (1989). *Urban Economic Theory*. Cambridge University Press.

Gibbons, Stephen (2004). "The Costs of Urban Property Crime". In: *The Economic Journal* 114.499, F441–F463. URL: http://www.jstor.org/stable/3590167.

Gibbons, Stephen and Stephen Machin (2005). "Valuing Rail Access Using Transport Innovations". In: *Journal of Urban Economics* 57.1, pp. 148–169. URL: https://www.sciencedirect.com/science/article/pii/S0094119004001020.

Heilmann, Kilian (2018). "Transit access and neighborhood segregation. Evidence from the Dallas light rail system". In: *Regional Science and Urban Economics* 73, pp. 237–250. URL: https://www.sciencedirect.com/science/article/pii/S0166046218301066.

Holzer, H. J, J. M Quigley, and S. Raphael (2003). "Public Transit and the Spatial Distribution of Minority Employment: Evidence from a Natural Experiment". In: *Journal of Policy Analysis and Management* 22.3, pp. 415–441. URL: https://escholarship.org/uc/item/6n2766kx.

Ihlanfeldt, Keith R. and David L. Sjoquist (1998). "The Spatial Mismatch Hypothesis: A Review of Recent Studies and Their Implications for Welfare Reform". In: *Housing Policy Debate* 9.4, pp. 849–892. URL: https://doi.org/10.1080/10511482.1998.9521321.

Kahn, Matthew E. (2007). "Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems". In: *Real Estate Economics* 35.2, pp. 155–182. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6229.2007.00186.x.

Kawaguchi, Daiji and Norifumi Yukutake (2017). "Estimating the Residential Land Damage of the Fukushima Nuclear Accident". In: *Journal of Urban Economics* 99, pp. 148–160. URL: https://www.sciencedirect.com/science/article/pii/S0094119017300232.

Lee, Young Jun and Daniel Wilhelm (2020). "Testing for the Presence of Measurement Error in Stata". In: *The Stata Journal* 20.2, pp. 382–404. URL: https://doi.org/10.1177/1536867X20931002.

Mun, Se-il and Asato Taguchi (2018). "Design of Mass Transit System and Urban Spatial Structure". In.

OECD (2012). *Compact City Policies: A Comparative Assessment*. OECD Green Growth Studies. OECD Publishing.

Petitte, Ryan A. and Stephen L. Ross (1999). "Commutes, Neighborhood Effects, and Compensating Differentials: Revisited". In: *Journal of Urban Economics* 46.1, pp. 1–24. URL: https://www.sciencedirect.com/science/article/pii/S0094119098921179.

Pogonyi, Csaba G., Daniel J. Graham, and Jose M. Carbo (2021). "Metros, Agglomeration and Displacement. Evidence from London". In: *Regional Science and Urban Economics* 90, p. 103681. URL: https://www.sciencedirect.com/science/article/pii/S0166046221000417.

Reed, William Robert (2015). "On the Practice of Lagging Variables to Avoid Simultaneity". In: *Oxford Bulletin of Economics and Statistics* 77.6, pp. 897–905. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12088.

Tan, Ronghui et al. (2019). "The Effect of New Metro Stations on Local Land Use and Housing Prices: The Case of Wuhan, China". In: *Journal of Transport Geography* 79, p. 102488. URL: https://www.sciencedirect.com/science/article/pii/S0966692317300509.

Tyndall, Justin (2021). "The Local Labour Market Effects of Light Rail Transit". In: *Journal of Urban Economics* 124, p. 103350. ISSN: 0094-1190. URL: https://www.sciencedirect.com/science/article/pii/S0094119021000322.

Wilhelm, Daniel (2019). *Testing for the presence of measurement error*. eng. cemmap working paper CWP48/19. London. URL: http://hdl.handle.net/10419/211141.

Xu, Xinliang (2017). *1 km Grid Population Distribution Databases for China*. Demographic Resource. Resource and Environment Data Cloud Platform. DOI: https://doi.org/10.12078/2017121101. URL: https://www.resdc.cn/.

Zhu, Yi and Mi Diao (2016). "The Impacts of Urban Mass Rapid Transit Lines on the Density and Mobility of High-Income Households: A Case Study of Singapore". In: *Transport Policy* 51, pp. 70–80. URL: https://www.sciencedirect.com/science/article/pii/S0967070X16301470.

Zou, Yanjiao (2018). "Characteristics and Formation Mechanisms of Employment Spatial in Wuhan". MA thesis. East China Normal University.

Appendices

A Proof of Proposition 1

It is suffice to show Equation 4 and Equation 5, the remaining is fundamental calculus and algebraic.

Details of Equation 4.

$$\Delta t_i(\hat{x}_i^*, is) = \int_{\hat{x}_i^*}^{is} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \Big[(a_1+f)(x_{i+1}^* - x_{i+1}) + a_2(x_{i+1}^* - x_{i+1})\Big] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x$$

$$= \int_{\hat{x}_i^*}^{is} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \Big[(a_1+f)\cdot 2\Delta d + a_2 \cdot 2\Delta d)\Big] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x$$

$$= \int_{\hat{x}_i^*}^{is} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \Big[2\underbrace{(a_1+a_2+f)}_{=2a_2\phi}\Delta d)\Big] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x$$

$$= (is - \hat{x}_i^*)\frac{4a_2\phi}{s}\Delta d$$
$$= -\frac{8a_2\phi^2}{s}(\Delta d)^2 + \frac{4a_2\phi}{s}(is - \hat{x}_i)\Delta d,$$

the last equality follows from $\hat{x}_i^* = \hat{x}_i + 2\phi\Delta d$.

Details of Equation 5.

$$\Delta t_i(\hat{x}_i, \hat{x}_i^*) = \int_{\hat{x}_i}^{\hat{x}_i^*} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \Big[(a_1+f)(x_i - x_{i+1}) + a_2(2x - x_{i+1} - x_i)\Big] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x$$

$$= \int_{\hat{x}_i}^{\hat{x}_i^*} \int_{-\frac{1}{2}s}^{\frac{1}{2}s} \Big[\underbrace{(a_2 - a_1 - f)(x_{i+1} - x_i)}_{A} + \underbrace{2a_2(x - x_{i+1})}_{B}\Big] \frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x,$$

where

$$\int_{\hat{x}_i}^{\hat{x}_i^*}\int_{-\frac{1}{2}s}^{\frac{1}{2}s} A\frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x = \underbrace{\frac{2\phi(a_2-a_1-f)}{s}(x_{i+1}-x_i)\Delta d}_{C},$$

$$\int_{\hat{x}_i}^{\hat{x}_i^*}\int_{-\frac{1}{2}s}^{\frac{1}{2}s} B\frac{1}{s^2}\,\mathrm{d}y\,\mathrm{d}x = \frac{2a_2}{s}\int_{\hat{x}_i}^{\hat{x}_i^*} x\,\mathrm{d}x - \frac{4a_2\phi}{s}x_{i+1}\Delta d$$

$$\begin{aligned}
&= \frac{a_2}{s}(\hat{x}_i^{*2}-\hat{x}_i^2) - \frac{4a_2\phi}{s}x_{i+1}\Delta d \\
&= \frac{4a_2\phi^2}{s}(\Delta d)^2 + \underbrace{\frac{4a_2\phi}{s}\hat{x}_i\Delta d}_{D} - \frac{4a_2\phi}{s}x_{i+1}\Delta d,
\end{aligned}$$

the last equality is because $\hat{x}_i^* = \hat{x}_i + 2\phi\Delta d$. Summing C and D:

$$\begin{aligned}
C+D &= \frac{2\phi(a_2-a_1-f)}{s}(x_{i+1}-x_i)\Delta d + \frac{4a_2\phi}{s}\big[\phi x_{i+1} + (1-\phi)x_i\big]\Delta d \\
&= \frac{2\phi}{s}\big[(a_2-a_1-f)x_{i+1} - (a_2-a_1-f)x_i + (a_1+a_2+f)x_{i+1} + (a_2-a_1-f)x_i\big]\Delta d \\
&= \frac{4a_2\phi}{s}x_{i+1}\Delta d,
\end{aligned}$$

the second equality is because $2a_2\phi \equiv a_1+a_2+f$ and $2a_2(1-\phi) \equiv a_2-a_1-f$. Summing the remaining part yields that

$$\Delta t_i(\hat{x}_i, \hat{x}_i^*) = \frac{4a_2\phi^2}{s}(\Delta d)^2.$$

B Derivation of Equation 13

From the Cubb-Douglas utility function we have $z = e^{u/\alpha} q^{-\beta/\alpha}$. Substituting this equation into Equation 9:

$$\Psi_i(u) = \max_q \{ \frac{(y - t_i)}{q} - e^{\frac{u}{\alpha}} q^{-\frac{1}{\alpha}} \}. \tag{B.1}$$

The FOC yields

$$\begin{aligned} &- (y - t_i) q_i^{-2} + \frac{1}{\alpha} e^{\frac{u}{\alpha}} q_i^{-\frac{1+\alpha}{\alpha}} = 0 \\ &\Rightarrow q_i = \alpha^{-\frac{\alpha}{\beta}} e^{\frac{u}{\beta}} (y - t_i)^{-\frac{\alpha}{\beta}}. \end{aligned} \tag{B.2}$$

Substituting Equation B.2 back into Equation B.1, we have the bid rent

$$\begin{aligned} \Psi_i &= \alpha^{\frac{\alpha}{\beta}} e^{-\frac{u}{\beta}} (y - t_i)^{\frac{1}{\beta}} - \alpha^{\frac{1}{\beta}} e^{-\frac{u}{\beta}} (y - t_i)^{\frac{1}{\beta}} \\ &= \alpha^{\frac{\alpha}{\beta}} \beta e^{-\frac{u}{\beta}} (y - t_i)^{\frac{1}{\beta}}. \end{aligned} \tag{B.3}$$

By maximising condition of Equation 10, we have $r_i = K'(\rho_i) = \lambda\gamma\rho_i^{\gamma-1}$. Combining this equation and Equation B.3 and compelling the equilibrium condition $\psi_i = r_i$:

$$\rho_i = (\frac{r_i}{\lambda\gamma})^{\frac{1}{\gamma-1}} = (\lambda\gamma)^{-\frac{1}{\gamma-1}} \alpha^{\frac{\alpha}{\beta(\gamma-1)}} \beta^{\frac{1}{\gamma-1}} e^{-\frac{u}{\beta(\gamma-1)}} (y - t_i)^{\frac{1}{\beta(\gamma-1)}}.$$

Combining the preceding equation and Equation 11 yields

$$n_i = \frac{\rho_i s^2}{q_i} = \underbrace{(\lambda\gamma)^{-\frac{1}{\gamma-1}} \alpha^{\frac{\alpha\gamma}{\beta(\gamma-1)}} \beta^{\frac{1}{\gamma-1}} e^{-\frac{u\gamma}{\beta(\gamma-1)}} s^2}_{\kappa>0} (y - t_i)^{\frac{1+\alpha(\gamma-1)}{\beta(\gamma-1)}}.$$

Denote $\delta \equiv [1 + \alpha(\gamma - 1)]/\beta(\gamma - 1)$ and notice that $\delta > 1$ since $\alpha(\gamma - 1)/\beta(\gamma - 1) > 1$. Now we summarize that

$$n_i = \kappa(y - t_i)^{\delta}. \tag{B.4}$$

Finally, rewriting Equation 6 as follows:

$$t_i - \overline{t_i} = \phi_1(d_i - \overline{d_i})^2 + \phi_2(d_i - \overline{d_i}),$$

where $\overline{t_i} \equiv t_i - \Delta t$ and $\overline{d_i} \equiv d_i - \Delta d$ respectively represent commuting time and interval spacing of metro stations before change. This equation yields that

$$t_i = \phi_1 {d_i}^2 + (\phi_2 - 2\phi_1\overline{d_i})d_i + (\overline{t_i} + \phi_1\overline{d_i}^2 - \phi_2\overline{d_i}).$$

We obtain Equation 13 by substituting the preceding equation into Equation B.4, where

$$\zeta_0 \equiv y - (\overline{t_i} + \phi_1\overline{d_i}^2 - \phi_2\overline{d_i}) > 0,$$

$$\zeta_1 \equiv -(\phi_2 - 2\phi_1\overline{d_i}) < 0,$$

$$\zeta_2 \equiv -\phi_1 > 0.$$

Variables	Description	Min	Max	Median	Mean	S.d.	Units
Dependent							
pop_{2015}	Logarithm of population in 2015	5.91	10.34	8.98	8.78	1.01	log person
Key Explanatory							
d	Average interval spacing of stations	7.70	33.11	12.59	14.65	5.94	hundred meters
Metro-relative							
Covariates							
pop_{2005}	Logarithm of population in 2000	2.68	4.04	3.63	3.55	0.32	log person
l	Distance from grid center to nearest station	0.34	24.71	10.34	10.93	5.92	hundred meters
t	commuting time from nearest station to nearest BD	0	38	21.00	20.49	9.80	minute
op	open time of the nearest station from Dec. 2015	0	137	24	30.26	32.00	month
sdi	synthetic development index	0.05	0.99	0.56	0.55	0.23	-
District Dummy							
$reg1$	=1 if in Wuchang district, otherwise 0	0	1	0	0.10	0.30	-
$reg2$	=1 if in Qingshan district, otherwise 0	0	1	0	0.03	0.17	-
$reg3$	=1 if in Qiaokou district, otherwise 0	0	1	0	0.09	0.28	-
$reg4$	=1 if in Jianghan district, otherwise 0	0	1	0	0.07	0.25	-
$reg5$	=1 if in Jiangan district, otherwise 0	0	1	0	0.16	0.36	-
$reg6$	=1 if in Huangpi district, otherwise 0	0	1	0	0.05	0.21	-
$reg7$	=1 if in Hongshan district, otherwise 0	0	1	0	0.16	0.36	-
$reg8$	=1 if in Hangyang district, otherwise 0	0	1	0	0.16	0.36	-
$reg9$	=1 if in Dongxihu district, otherwise 0	0	1	0	0.11	0.32	-
Population							
Level Dummy							
$popquan1$	=1 if pop_{2005} lies between $1st$ and $2nd$ quantile	0	1	0	0.25	0.43	-
$popquan2$	=1 if pop_{2005} lies between $2nd$ and $3rd$ quantile	0	1	0	0.25	0.43	-
$popquan3$	=1 if pop_{2005} lies above $3rd$ quantile	0	1	0	0.25	0.43	-
Land Use Dummy							
$cst1a$	=1 if is labeled as urban construction land in 2015	0	1	1	0.52	0.50	-
$cst2a$	=1 if is labeled as rural construction land in 2015	0	1	0	0.01	0.07	-
CBD Dummy							
$xinhua$	=1 if nearest CBD is Xinhua Street	0	1	0	0.33	0.47	-
$hanshui$	=1 if nearest CBD is Hanshui-Baofeng	0	1	0	0.39	0.49	-
Instrument							
$cross.river$	=1 if the nearest station connects rail tracks that are across a river	0	1	0	0.07	0.25	-

Table 3: Descriptive statistics of variables we used in empirical research.

Variables	population increment ($pa - pb$) by OLS				population increment ($pa - pb$) by IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
interval spacing (d)	-0.0397	-0.0324	-0.0330	-0.0314	-0.0818	-0.0448	-0.0473	-0.0462
	(0.0089)	(0.0093)	(0.0090)	(0.0092)	(0.0146)	(0.0143)	(0.0132)	(0.0147)
dis. to nearest station (l)	-0.0071	-0.0122	-0.0107	-0.0110	-0.0033	-0.0108	-0.0091	-0.0090
	(0.0061)	(0.0060)	(0.0060)	(0.0064)	(0.0065)	(0.0064)	(0.0064)	(0.0069)
commuting time (t)	-0.0053	-0.0208	-0.0202	-0.0216	-0.0005	-0.0191	-0.0182	-0.0182
	(0.0072)	(0.0071)	(0.0071)	(0.0067)	(0.0079)	(0.0076)	(0.0076)	(0.0080)
open time (op)	-0.0062	-0.0078	-0.0075	-0.0076	-0.0062	-0.0077	-0.0074	-0.0074
	(0.0015)	(0.0013)	(0.0012)	(0.0012)	(0.0015)	(0.0013)	(0.0012)	(0.0012)
development index (sdi)	-0.5728	-0.2887	-0.4864	-0.4911	-0.8088	-0.3575	-0.5664	-0.5547
	(0.2009)	(0.1846)	(0.1855)	(0.1814)	(0.2165)	(0.1951)	(0.1932)	(0.1877)
(intercept)	1.2605	1.7514	1.7935	1.9816	1.8871	1.9272	1.9964	2.1341
	(0.3644)	(0.3893)	(0.3900)	(0.4414)	(0.4102)	(0.3988)	(0.3912)	(0.4430)
District Dummy	YES	YES	YES	YES	YES	YES	YES	YES
Population Level Dummy	NO	YES	YES	YES	NO	YES	YES	YES
Land Use Dummy	NO	NO	YES	YES	NO	NO	YES	YES
CBD Dummy	NO	NO	NO	YES	NO	NO	NO	YES
adjusted R-squared	0.5724	0.6222	0.6378	0.6365	0.5640	0.6095	0.6251	0.6259

Table 4: OLS and IV estimates of the demand side estimation. HAC standard errors of Conley (1999) in parentheses. Number of observations: 400.

Variables	population increment ($deltapop$) by FE							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
interval spacing d	-0.0669	-0.0615	-0.0453	-0.0422	-0.0423	-0.0231	-0.0233	-0.0223
	(0.0063)	(0.0065)	(0.0066)	(0.0065)	(0.0064)	(0.0063)	(0.0062)	(0.0063)
dis. to nearest station l		-0.0212		-0.0160	-0.0164	-0.0076	-0.0079	-0.0072
		(0.0053)		(0.0049)	(0.0049)	(0.0041)	(0.0041)	(0.0040)
commuting time (t)			-0.0291	-0.0278	-0.0299	-0.0145	-0.0187	-0.0172
			(0.0032)	(0.0032)	(0.0033)	(0.0031)	(0.0033)	(0.0040)
open time (op)					-0.0028		-0.0034	-0.0035
					(0.0011)		(0.0011)	(0.0011)
development index sdi								0.1369
								(0.1394)
Time FE	YES	YES	YES	YES	YES	YES	YES	YES
District Dummy	NO	NO	NO	NO	NO	YES	YES	YES
adjusted within R-squared	0.2062	0.2249	0.2911	0.3015	0.3047	0.4336	0.4379	0.4381

Table 5: FE estimates by of the demand side estimation. HAC standard errors of Conley (1999) in parentheses. Number of observations: 1200.