Heterogeneous Effects of Urban Transport Infrastructure on Population Distribution: The Role of Educational Access*

Jiawei Mo†

April 26, 2023

Abstract

This article investigates the heterogeneous effects of urban transport infrastructure on population distribution within a city. I focus on the case of Xiamen—a coastal city in China—where two bridges and a tunnel have been built to promote population growth on the city's periphery. I first show that although population share increased substantially on the bridges-connected periphery, no significant growth in the population share was observed on the tunnel-connected periphery. This pattern is surprising, given that the reduction in the commuting distance enabled by the tunnel is more than five times as large as that enabled by the bridges. I then calibrate a quantitative urban model to demonstrate the importance of access to high-quality schools in explaining the distinct effects of the infrastructures. Counterfactual exercises suggest that increasing educational resources on the tunnel-connected periphery or relaxing the restriction on cross-district school enrollment may facilitate the intended population growth effect of the tunnel.

Keywords: Urban Transport Infrastructure; Population Distribution; Educational Access.

JEL classification: R12, R42.

*I thank the editor (Gabriel Ahlfeldt), two anonymous referees, Jianyong Fan, J. Vernon Henderson, Songlin Li, Ming Lu, Mengyu Yang, Qinghua Zhang, seminar/conference participants at the 2019 Asian Meeting of the Econometric Society, the 8th RUSE International Workshop, Xiamen University, Shanghai Jiao Tong University, and The Chinese University of Hong Kong (Shenzhen) for helpful discussions, suggestions, and comments. Financial support from the National Natural Science Foundation of China (No. 72103001) and the seed fund of the School of Economics, Peking University, are gratefully acknowledged. Any remaining errors are my own.

†School of Economics, Peking University, Beijing, China; email: jwmo@pku.edu.cn.

1 Introduction

Modern transport infrastructure plays an important role in shaping cities—yet the impact of transport infrastructure on population distribution remains uncertain. Transport infrastructure can either centralize or decentralize population distribution, and proximity to transport networks could benefit or hurt the development of different locations. Literature has expanded extensively on the heterogeneous effects of transport infrastructure in both inter-city (e.g., Baum-Snow, 2007; Faber, 2014; Donaldson and Hornbeck, 2016; Baum-Snow et al., 2017; Banerjee et al., 2020; Jedwab and Storeygard, 2022) and intra-city (e.g., Small and Verhoef, 2007; Heblich et al., 2020; Balboni et al., 2020; Tsivanidis, 2022; Zárate, 2022) studies.

This paper examines the heterogeneous effects of urban transport infrastructure on the population distribution within a city and how specific amenities may influence the effects of the transport infrastructure. In particular, I study a case of Xiamen—a coastal city in China—where two bridges in 2008 and an undersea tunnel in 2010 were constructed to connect the city center with different parts of the periphery. Xiamen provides a unique and attractive setting for the current study. First, the geography of Xiamen naturally separates the city center from the periphery, which helps identify the impact of the bridges and tunnel by their locations. Second, unlike subways or railroads, no residential or work locations are connected along the bridges or tunnel except at the two ends, and thus simplifies the discussion throughout the transport line. Third, since Xiamen had few alternatives for commuting vehicles during the sample period, the road network played a critical role in commuting and the bridges and tunnel significantly impacted the road network. Lastly, the local government in Xiamen had clear objectives for the new infrastructures: to reduce congestion cost between the center and periphery and to promote population growth on the periphery.

I first present stylized facts related to the heterogeneous effects of the bridges and tunnel. Exploiting high-resolution point-of-interest data, I demonstrate that the population share increased substantially on the bridges-connected periphery but no significant change of population share was observed on the tunnel-connected periphery from 2004 to 2014. This pattern is surprising. On the one hand, conditional on the affected travel paths, the tunnel reduces the average commuting distance by 41%, which is more than five times as large as the bridges do. On the other, since the undersea tunnel was the largest transport infrastructure project ever undertaken, the local government had high expectations for its positive effect on the development of the periphery.

I then employ a difference-in-differences (DID) approach to investigate the impact of transport infrastructure on population distribution in Xiamen. I find a positive and statistically significant impact of the bridges on population share in locations that are close to the bridges. A 1% decrease in distance to the bridges is associated with a 0.38% increase in population share after 2008. By contrast, the impact of the tunnel is small and statistically insignificant. I verify the parallel pre-trends assumption of DID and examine the dynamic effects of the bridges and tunnel. The impact of the bridges on population share becomes pronounced after 2011 and is statistically significant during 2012-2016. By contrast, the impact of the tunnel is statistically insignificant during the whole sample period from 2000 to 2021. Results are robust after controlling for the impact of other transport infrastructure improvements, such as the first connection to high-speed railway in 2010, the capacity expansion of the international airport in 2014, and the increase in new lines for the traditional railway station in 2010, among others. I also find consistent evidence when using a non-parametric estimator proposed by de Chaisemartin et al. (2022).

To explain the distinct effects of the bridges and tunnel, I develop and calibrate a quantitative urban model following Ahlfeldt et al. (2015). I incorporate China's institutional background in the model. Specifically, I assume fixed land use in the city rather than flexible land use transfer between industrial and residential use, given the strict land use regulations in China. More importantly, I use educational access as a proxy for location amenity. In China, access to high-quality public schools plays a critical role in residents' location choices. Following the *gravity potential* approach in the geography literature (Talen and Anselin, 1998), I measure educational access by aggregating the number of available top schools—i.e., demonstration schools ranked by the provincial government—within a given area adjusted by the friction of distance between the schools and the place of residence. To highlight the unique role of educational access in understanding the heterogeneous effects of the infrastructures, I also use access to top-tier public hospitals and top-rated scenic spots as proxies for amenity and compare them with the baseline model using educational access.[1] The calibrated baseline model matches fairly well with the observed population distribution

1. See Section 2.3 for precise definitions of top schools, top-tier public hospitals, and top-rated scenic spots.

in Xiamen, both before and after operation of the bridges and tunnel, and reflects the distinct effects of the new transport infrastructure on the periphery.

Two features of educational access are essential in the quantitative analysis. First, the spatial distribution of top schools in Xiamen is highly unequal, with most concentrated on the island and bridges-connected periphery, and only a few on the tunnel-connected periphery. Second, the availability of public schools is residence-based and spatially bounded by the school's enrollment zone. These two features jointly contribute to the insignificant growth of population share on the tunnel-connected periphery. To illustrate this point, I show that alternative models that use either hospitals or scenic spots as amenity could not generate the insignificant growth of population share on the tunnel-connected periphery. Although hospitals and scenic spots are also unequally distributed, they are not spatially bounded by the district. The tunnel would greatly expand access to hospitals and scenic spots in the city center for residents who choose to live on the tunnel-connected periphery. However, the tunnel fails to improve educational access on the tunnel-connected periphery due to the cross-district enrollment restriction. I also compare the channels of productivity and educational access and show that productivity alone could not explain the observed pattern on the tunnel-connected periphery. To strengthen my argument on the important role of educational access, I provide supporting evidence that the growth of the share of married residents and the average number of children per household were substantially smaller in the tunnel-connected district than in the bridges-connected district from 2010 to 2020.

I conduct two counterfactual exercises to evaluate whether and how educational policies might enhance the population growth effect of the tunnel. The first exercise is to increase top schools on the tunnel-connected periphery. I find a substantial increase in population share on the tunnel-connected periphery when I upgrade five existing schools that are the closest to the peripheral end of the tunnel. The population growth effect of the tunnel is more prominent and statistically significant when I upgrade 10 existing schools as top schools. In the second exercise, I retain the distribution of top schools and allow for cross-district school enrollment. With this policy, the tunnel would greatly expand access to top schools in the city center for residents who choose to live on the tunnel-connected periphery. The simulation demonstrates a substantial increase in the population share on the periphery that is linked to the tunnel. Lastly, I discuss the potential costs associated with the counterfactual educational policies.

Related Literature. This paper contributes to a sizable literature on the heterogeneous effects of transport infrastructure on population distribution. Empirical studies find both population concentration and dispersion due to transport infrastructure improvements. For example, a national highway system may either centralize or decentralize population distribution within a country. Using US data between 1950 and 1990, Baum-Snow (2007) finds a decentralizing effect of the new access to highways, which contributed to declining population in the central city. Such decentralizing effects are also found in Spain (Garcia-López et al., 2015); Africa (Jedwab and Storeygard, 2022); and China (Baum-Snow et al., 2017). However, Faber (2014) and Baum-Snow et al. (2020), again using Chinese data, provide evidence of a centralizing effect whereby connection to the national highway system increased the population of the central city but had negative impacts on the periphery. Allen and Arkolakis (2022) find large heterogeneous welfare effects of different parts of US Interstate Highway System based on a general equilibrium geographic framework. While these contradictory findings may be a consequence of sampling (Baum-Snow et al., 2020), no consensus has been reached so far on the effects of transport infrastructure on population distribution, depending on the data, time period, and methodology used.[2] This paper complements the literature by exploring a unique setting in Xiamen city in which a highly developed island is connected to a less developed mainland area via new bridges and a tunnel. I directly compare the distinct effects of the bridges and tunnel on population growth on the periphery, and provide new evidence that specific amenities such as educational access may impact the effects of the transport infrastructure.

My work also links to a burgeoning literature on the distributional impact of transport infrastructure within cities (Small and Verhoef, 2007; Ahlfeldt et al., 2015; Heblich et al., 2020; Balboni et al., 2020; Tsivanidis, 2022; Zárate, 2022). For instance, Balboni et al. (2020) examine the differential impacts of the Dar es Salaam BRT system on the welfare of low- and high-income residents. Zárate (2022) studies how transit improvements reallocate workers from informal to formal sectors using the opening of new subway lines in Mexico City. This paper adds to this literature by investigating the distributional effects of transport infrastructure across space within a city. Departing from standard quantitative urban models, in which amenities are recovered from

2. There is also no consensus on the effects of transport infrastructure on economic growth. For example, Donaldson and Hornbeck (2016) find that railroad networks contributed to economic growth using US data between 1870 and 1890. Nevertheless, Banerjee et al. (2020) find that proximity to transport networks has no effect on per capita GDP growth in China during 1992-2007.

the model, I construct various proxies for location amenities using observed data and explore the relationship between educational access and the distinct effect of transport infrastructure.[3]

This study contributes to the education literature by examining the role of educational access in the heterogeneous effects of transport infrastructure on residential location choices. Prior research has established that improving access to high-quality education can enhance student performance and increase the economic returns to schooling (Card, 1993; Talen, 2001). Moreover, it has been demonstrated that improving school quality, increasing school choices, and investing in school facilities can increase local housing prices (Cellini et al., 2010; Fack and Grenet, 2010; Chan et al., 2020). This study extends previous literature by highlighting the potential influence of educational access on the effect of transport infrastructure on residential location choices. Policymakers need to consider this interactive effect when determining the optimal locations and capacity of schools, in addition to minimizing distances and accounting for the endogenous location decisions of residents (Epple et al., 2018; Loumeau, 2023). By doing so, policymakers can maximize the total welfare derived from public investment in education and transportation.

The remainder of the paper is organized as follows. Section 2 describes the geography of Xiamen and the background of the urban transport infrastructure. Section 3 presents stylized facts and reduced-form evidence related to the heterogeneous effects of the infrastructure. Section 4 develops a quantitative urban model for the case of Xiamen. Section 5 describes details of model calibration and illustrates the importance of educational access in explaining the heterogeneous effects of the infrastructure. Section 6 conducts counterfactual exercises to evaluate different educational policies and their impacts on the effects of transport infrastructure. Section 7 concludes.

2 Background and Data

This section describes the geographical and institutional background of Xiamen city in China, which is critical for the empirical analysis. The geography of Xiamen provides a unique and attractive setting to study the heterogeneous effects of urban transport

3. Gaigné et al. (2022) explicitly measure location amenities using the number of outside geocoded pictures taken by residents at a certain location.

infrastructure on population distribution. It is also related to the government's intention for the construction of the new transport infrastructure. Data sources and key variables are then provided.

2.1 Geography of Xiamen

Xiamen is a prefecture-level city in Fujian province and one of the first five special economic zones in China. The city is located in the middle of the west coast of the Taiwan Straits and the center of the Golden Triangle area of southern Fujian. With a total area of 1,700.61 km^2, Xiamen had a population of 5.18 million in 2020, which is comparable to Hong Kong (1,106.66 km^2 and 7.42 million residents). Urban residents in Xiamen accounted for 89.4% of the population, and the gross regional domestic production of Xiamen was 638.4 billion yuan (92.5 billion US dollar) in 2020.[4]

The geography of Xiamen is highly diverse, with moutains and hills in the northwest; high plains, low plains, terraces, sea alluvial plains, and tidal flats in the middle; and two islands in the south—Xiamen Island and Gulang Island (Figure 1(A)). The city center is located on Xiamen Island, and peripheral regions are on the mainland. Importantly, the sea separates the city center from the periphery, which naturally requires transport infrastructure that connects the island with the mainland.

Xiamen administers six districts. The Siming and Huli districts are situated on Xiamen Island (which includes Gulang Island), and the Jimei, Haicang, Tong'an, and Xiang'an districts are on the mainland (Figure 1(B)).

2.2 Transport Infrastructure and Government Intention

This paper focuses on two bridges—the Jimei and Xinglin bridges—and one tunnel—the Xiang'an undersea tunnel—in Xiamen, as depicted in Figure 1(A). The two new bridges have been in operation since 2008 and the tunnel opened in 2010. Other bridges were constructed before 2000. The new bridges connect Huli district with Jimei district, both of which are in the northern part of the island. The tunnel links the east part of the island to Xiang'an district.

The local government in Xiamen had clear objectives for the bridges and tunnel.

4. Data are from the Yearbook of Xiamen Special Economic Zone 2021.

The bridges were planned to serve two purposes. First, the government intended to reduce traffic congestion between the island and the mainland. Only one bridge connected the two regions before operation of the new ones, and increasing traffic volume required new transport infrastructure in the northern part of the island. Second, the government aimed to promote population growth and economic development in the northern mainland.[5]

Regarding the tunnel, the government intended to promote economic integration between the island and Xiang'an district, which is the least developed region in the city. Local officials had high expectations for the positive effects of the tunnel on the development of Xiang'an. For example, the government described the tunnel as the largest transportation infrastructure project ever undertaken in the city, and highlighted the tremendous reduction in travel time between the island and Xiang'an district, which decreased from 1.5 hours to 9 minutes between the two ends of the tunnel.[6] Local officials also addressed the great success of the tunnel in overcoming the global challenges of underwater construction.[7]

2.3 Data

This paper exploits point-of-interest (POI) data in the geographic information system (GIS). The basic unit of analysis is an area of around 1 km^2 based on POI data on population from the LandScan database at a spatial resolution of 30 arc-seconds (0.008333333 decimal degrees). To determine the land use of each location, I map POI data on population onto the land use map.[8] There are 266 residential locations (land for residential use) and 214 work locations (land for industrial and commercial use) in Xiamen. Figure 2 shows the distribution of the two types of locations in Xiamen. All other variables are averaged or mapped to location level. I obtain the longitude and

5. See news reports (in Chinese) on local officials' comments on the opening of the Jimei and Xinglin bridges, respectively.

6. See official news report (in Chinese) on the opening of Xiang'an Tunnel.

7. See official news report (in Chinese) on the construction of Xiang'an Tunnel.

8. Land use map is available in the Urban Comprehensive Planning of Xiamen (2011-2020) provided by the Xiamen Municipal Natural Resources and Planning Bureau. By mapping POI data onto the map of land use, locations that match the land for residential use are defined as residential/housing locations, and locations that fall on land for industrial or commercial use are defined as work locations. If both residential and industrial (commercial) land appear in the same location, the definition of the location is determined by the land use that accounts for more than half of the area. For more details of the land use map, please refer to the official website (in Chinese).

latitude of the variables from Baidu Geocoding API using the corresponding detailed address. Data sources and descriptions are provided as follows.

Population. The high-resolution population distribution data are from the LandScan database, which has provided global population data at a spatial resolution of 30 arc-seconds since 2000. The sample period is from 2000 to 2021.[9] The data represent an average population distribution over 24 hours and accounting for all human activities during the entire day. Thus, the value of each location is not just the number of residents but also number of people who do business, travel, or participate in other activities.[10] Figure 3 presents the population distribution in Xiamen in 2007, before operation of the bridges or tunnel. Residents are concentrated on the island, and especially in the southwest part of the island, where the traditional city center and popular tourist attractions are located.

Road distance. Location-to-location distance is calculated based on a detailed road map in Xiamen, as shown in Figure 4. The road map is obtained from OpenStreetMap data for 2014. I assume that the only change in the road network in Xiamen during the sample period was the introduction of the two new bridges in 2008 and one tunnel in 2010. Based on the changed road map, I calculate the change in road distance between any two locations that are affected directly or indirectly by the new infrastructure.[11] Appendix B provides details on calculating the lowest-cost-path distance with the network analysis in ArcGIS.[12]

9. The population data are generated using spatial modeling techniques developed by Oak Ridge National Laboratory, with spatial data as the primary input. For more details, please refer to LandScan Population Data.

10. Comparing the values of cells between years should be done with attention because the data generation process is updated annually to incorporate new spatial maps and new imagery analysis techniques, which may result in measurement errors between years and the data versions. This measurement error is assumed to be uncorrelated with the change in transportation network. I smooth the population data by averaging the raw data in the previous, current, and subsequent years. An alternative data source is the WorldPop database, which is more comparable between years. However, a substantial share of land area reports missing population values in Xiamen in all years in the WorldPop data.

11. One of the main purposes of building the new bridges in Xiamen was to relieve congestion when commuting between the city center and the periphery. To reflect the congestion alleviation of the two new bridges, I manually increase the commuting distance by 1 km for each commuting path between the city center and the bridge-connected periphery before operation of the new bridges (see details in Appendix B).

12. When calculating distance to residential or work locations, the longitude and latitude of the locations are set as the center point of the 1-km^2 cell. Note that the distance measure can be easily

Schools. Educational resource is an important type of location amenity. In China, proximity to high-quality schools plays a critical role in residents' decisions on residential locations. Prior studies use historical key schools (*zhongdian xuexiao*) ranked by the municipal government (Zhang and Chen, 2018) or tournament performance (Chan et al., 2020) as a measure of school quality in China. This paper employs a measure similar to key schools. I choose primary and middle schools in Xiamen that are on the first and second lists of demonstration schools for compulsory education reform in Fujian Province, which were published by the local government in 2018; there are 82 schools on the lists. I define these schools as *top schools* because they are selected and ranked by the Education Department of Fujian Province. All were built before 2004. I assume that the distribution of top schools did not change during the sample period. See Appendix Table A7 for the full list of top schools in Xiamen.

Hospitals. Medical service is another type of location amenity that may be taken into consideration by residents in China. I use 12 top-tier public hospitals that were built before 2008 in Xiamen as a proxy for medical service.[13] I obtain the list of public hospitals from the National Health Commission of the People's Republic of China. Appendix Table A8 provides the list of top-tier public hospitals in Xiamen.

Scenic spots. The natural environment may affect residents' location choices. I use top-rated scenic spots in Xiamen as a proxy for natural attractiveness. In China, the official rating system for national tourist attractions divides scenic spots into five levels from 5A to 1A, and the highest level is 5A. I obtain the list of A-class scenic spots in Xiamen in 2021 from the Bureau of Culture and Tourism in Xiamen. Since Xiamen has only one 5A-rated scenic spot, I choose scenic spots that were ranked 4A or above. Appendix Table A9 provides the list of 11 top-rated scenic spots in Xiamen.

---

converted to a time measure by assuming the average speed of automobiles, which are the most common vehicles in the city. I do not calculate the distance with metro rail, since there was no subway in Xiamen until 2018.

13. In China, most residents prefer public to private hospitals. Public hospitals are divided into 3 tiers, and each tier consists of 3 grades: A, B, and C. Of all ranks of hospitals in China, Tier 3 Grade A hospitals have the best medical resources. During the sample period, there were 8 Tier 3 Grade A, 2 Tier 3 Grade B, and 2 Tier 3 Grade C hospitals in Xiamen that were built before 2008. Notably, I do not include 5 Tier 3 hospitals that were newly built after 2013, but the results hold when I include the new hospitals because the spatial distribution of the new hospitals is similar to those built before 2008.

Productivity. Total factor productivity (TFP) for each work location is calculated based on the Annual Survey of Industrial Enterprises (ASIE) maintained by the National Bureau of Statistics of China (NBS) from 2000 to 2007, which covers all state-owned enterprises (SOEs) and non-SOEs whose annual sales are no less than RMB 5 million. Following Ackerberg et al. (2015), I use intermediate inputs to control for the unobserved productivity in a Cobb-Douglas value-added production function. The location's TFP is obtained by averaging the productivity of firms within 1 km. If no manufacturing firm is found within 1 km, I match with the closest firm's TFP. I assume that the distribution of firm productivity did not change during the sample period. For robustness, I also use data from the Second National Economic Census in 2008, which cover all firms in secondary and tertiary industries.[14] Given that the Economic Census data contain limited variables, with which I can't recover each firm's TFP, I use the annual wage per worker as a proxy for productivity.

Land supply. To obtain spatial data on floor ratio, I manually collect the information from daily reported land transactions by the Ministry of Land and Resources. Land transaction data are from the China Land Market website and cover land transactions as early as 1992 in Shenzhen city. However, land transaction information was not complete and not well organized until announcement of the land registration regulation in early 2008, and transaction-level data are close to national data only after 2007 (Mo, 2018).[15] I exploit the information from 2007 to 2015 and use the floor ratio of the nearest land as the proxy for the sample location.

Summary statistics. Table 1 presents summary statistics of the main variables. The average population size in the residential locations increased from 2,970 to 5,333 during 2000-2021.[16] The decline of the maximum population share (from 11.87% to 6.86%) reflects population dispersion from the traditional center (the southwest part of the island) to other areas. With the new bridges and tunnel, the population share has a

14. Appendix Table A1 describes the industry composition of firms in Xiamen in 2008. Although the manufacturing sector accounts for only 6.74% of all firms, it contributes to 37.34% of total employment.

15. To verify the completeness of transaction-level land data, I compare the aggregate data with the reported national data (Appendix Table A10).

16. Note that the total size of the population in residential locations is smaller than the population in Xiamen reported by the government. The population in LandScan is generated from satellite data and allocated to all locations, including areas for working, public services, mountains, rivers, and so on, many of which are not the sample units in the paper.

smaller standard deviation (from 0.97 to 0.66). The distribution of Xiamen's residents became more dispersed in general after operation of the new transport infrastructure. Also, the commuting distance between residential and work locations decreased by 1.47 km on average with the new infrastructure. Mean distance from residential locations to top-tier hospitals and top-rated scenic spots decreased by 2.57 and 1.0 km with the new infrastructure, respectively. Notably, distance to top schools is not affected by the new bridges and tunnel due to cross-district enrollment restriction in China.

3 Stylized Facts and Reduced-Form Evidence

This section documents the stylized facts that motivate the quantitative analysis and provide reduced-form evidence on the heterogeneous effects of different transport infrastructures on population distribution in Xiamen.

3.1 Stylized Facts

Fact 1. *The average reduction in commuting distance by the tunnel is more than five times as large as that by the bridges.*

Table 2 compares the effects of commuting-distance reduction by different infrastructures. While construction of the bridges affects more commuting paths than that of the tunnel (10,772 observations vs. 7,962), the average reduction in commuting distance by the tunnel is more than five times as large as that by the bridges (9.64 vs. 1.79). The maximum distance reduction by the tunnel is 39.35 km, which is sizeable given that the mean distance between any two locations without new infrastructure is 23.28 km. On average, the tunnel reduces commuting distance by around 41% (9.64/23.28). Combining the impacts of both the bridges and tunnel, the mean distance reduction is 6.32 km—more than one-fourth of the mean distance.

Fact 2. *The population share increased substantially on the bridges-connected periphery, whereas no significant change in population share was observed on the tunnel-connected periphery from 2003 to 2015.*

To demonstrate the changes in population distribution in Xiamen after operation of the bridges and tunnel, I adopt the 5-year average population share to minimize the measurement error of population data, as stated in Section 2.3. The population share (π_i) in residential location i is defined as the population in i divided by the total population in all residential locations. The change in population share after operation of the bridges and tunnel is calculated as follows:

$$\Delta\pi_i = \frac{1}{5}\left(\sum_{t=2011}^{2015} \pi_{it} - \sum_{t'=2003}^{2007} \pi_{it'}\right), \tag{1}$$

where the period between 2008 and 2010 is not included because the infrastructures were constructed during this period. The 5-year window is chosen to account for the time during which the new transport infrastructure had effects on residents' migration. Patterns are similar if the window period is changed to 3 or 4 years.

Figure 5 depicts the changes in population share for residential locations. I normalize the change in population share by its standard deviation (sd) and classify the changes in five categories: (1) large decrease of more than 1 sd; (2) small decrease of more than 0.1 sd; (3) unchanged between -0.1 and 0.1 sd; (4) small increase of more than 0.1 sd; (5) large increase of more than 1 sd. The threshold for "unchanged" is set to narrow the area of interest. About 61% of locations are in the unchanged group.

Two patterns are worth noting. First, in the peripheral inland area, population share increased in the southern region of Jimei district, which is close to the two bridges. By contrast, almost no change in population share was observed in Xiang'an district that connects with the tunnel. This is surprising, given that Xiang'an district enjoys the greatest reduction in commuting distance from the periphery to the city center with operation of the tunnel. Second, regarding the population distribution on the island, population share decreased in Siming district, where the most densely populated region is located. Population share increased in Huli district, which is directly connected to the bridges and tunnel.

Fact 3. *The distribution of top schools, top-tier hospitals, and top-rated scenic spots is highly unequal in Xiamen, with most concentrated on the island and bridges-connected periphery and only a few on the tunnel-connected periphery.*

Figure 4 presents the distribution of top schools, top-tier hospitals, and top-rated scenic spots (defined in Section 2.3) in Xiamen during the sample period. Top schools and top-tier hospitals are highly concentrated on the island, and some of the top schools and top-rated scenic spots are located on the bridges-connected periphery in Jimei district. By contrast, much fewer top schools and top-tier hospitals—and no top-rated scenic spots—are found on the tunnel-connected periphery in Xiang'an district.

For comparison, the distribution of productivity is relatively equal in Xiamen, as shown in Figure 6; work locations with relatively high productivity are not necessarily concentrated on the island. There are high-productivity firms on both the bridges- and tunnel-connected periphery.[17]

One aspect is worth noting. While schools, hospitals, and scenic spots are all unequally distributed, residents' access to schools differs from access to hospitals or scenic spots, in that access to schools is spatially bounded by administrative districts. The new bridges and tunnel would not affect access to schools, due to the separation of districts by Xiamen's geography, but may increase access to hospitals and scenic spots by reducing the travel cost from the periphery to the city center.

3.2 Reduced-Form Evidence

In this subsection, I provide reduced-form evidence that the new bridges increased population share on the connected periphery whereas the new tunnel did not have a significant impact on the population distribution on the corresponding periphery. Specifically, I employ a DID approach to examine the impact of transport infrastructure on the population distribution in Xiamen. The empirical specification is given as follows:

$$\ln \pi_{it} = \alpha + \beta^{\zeta} \ln Dist_i^{\zeta} \times Post_t^{\zeta} + \lambda_i + \lambda_t + \mu_{it}, \tag{2}$$

where π_{it} is the population share in residential location i in year t. $Dist_i^{\zeta}$ is the road distance to the nearest end of infrastructure $\zeta \in \{Bridges, Tunnel\}$ and $Dist_i^{Bridges} = \min\{Dist_i^{JimeiBridge}, Dist_i^{XinglinBridge}\}$.[18] The parameter β^{ζ} captures the impact of in-

17. While Figure 6 only contains manufacturing firms, I show that the average wage per worker in all industries also has a relative equal distribution on the periphery using data from the 2008 Economic Census. For more discussion, see Section 5.3.

18. Road distance is calculated based on a road network with the new bridges and tunnel. Notably, distance to the closest end of infrastructure ζ is not affected by operation of the new infrastructure when I restrict the sample to connected peripheral areas.

frastructure ζ on population share. $Post_t^{\zeta}$ is a dummy variable that equals 1 after year 2008 when $\zeta = Bridges$ and year 2010 when $\zeta = Tunnel$, and 0 otherwise. I control for location and year fixed effects, λ_i and λ_t, respectively. μ_{it} is the error term. I cluster standard errors at the district-year level.

To estimate the causal response in a DID setting with a continuous treatment proxied by the distance to the transport infrastructure, I make three assumptions. First, I consider the impact of the two bridges as one treatment and that of the tunnel as another treatment, assuming that these treatments are separate and independent. With this assumption, the estimation in each treatment does not involve heterogeneity in treatment timing.[19] Second, I assume that for all values of the treatment (i.e., the distance to the respective infrastructure), the average change in population share across all locations, if they had been assigned that value of the treatment, is the same as the average change in population share for locations that *experienced* that value of the treatment. Notably, this parallel-trends assumption is stronger than the one commonly applied in the DID setting with a binary treatment (Callaway et al., 2021). Third, I assume that residents who live closer to the nearest end of the new bridges or tunnel were more affected by the respective infrastructure.

Table 3 reports the baseline results. The first two columns are the results with respect to the bridges, and I find a positive and statistically significant impact of the bridges on population share in locations that are close to them. Specifically, a 1% decrease in distance to the bridges is associated with a 0.377% increase in population share after 2008 for the full sample. For instance, relative to a mean location that is 17.73 km from the bridges, a location that is 10 km from the bridges experienced a 24.1% $(= \exp(-0.377 * (\ln 10 - \ln 17.73)) - 1)$ increase in population share after operation of the bridges. The estimated elasticity doubles when I restrict the sample to Jimei district, which is directly connected to the bridges, as shown in Table 3 Column 2.

By contrast, the impact of the tunnel is negative and statistically insignificant,

19. It is important to note that the estimated coefficient on one treatment may be contaminated by the effect of the other treatment in a two-way fixed-effect regression (de Chaisemartin and D'Haultfoeuille, 2022a). Unfortunately, a reliable estimator for the case with several continuous treatments without stayers is not yet available (de Chaisemartin and D'Haultfoeuille, 2022b). Given that controlling for the other treatment may result in a more biased estimator than not controlling for it, as shown in de Chaisemartin and D'Haultfoeuille (2022a), I choose the estimation with only one treatment as the preferred specification.

as shown in Columns 3 and 4 in Table 3. Neither the full sample nor the locations in Xiang'an district experienced significant change in population share when getting closer to the tunnel after 2010. Overall, the growth of population share in Xiamen is decreasing in distance to the bridges and increasing in the distance to the tunnel, as shown in the last column. As a comparison, growth of the population share in the city center is more related to distance to the tunnel than to the bridges, as indicated by the first column in Appendix Table A2, in which I restrict the sample to Siming and Huli districts.

It is worth noting that the results for population size (log) are almost the same as those for population share (log), since the changes in total population are absorbed by year fixed effects. Appendix Table A3 reports the results for population size, and the estimates are identical to the baseline results.

Parallel pre-trends and dynamic effects. I estimate the following equation to verify the parallel pre-trends assumption of DID and examine the dynamic effects of the bridges and tunnel:

$$\ln \pi_{it} = \alpha + \sum_{s=2000}^{2021} \beta_s^{\zeta} \ln Dist_i^{\zeta} \times T_{st}^{\zeta} + \lambda_i + \lambda_t + \mu_{it}, \qquad (3)$$

where T_{st}^{ζ} is a dummy variable that takes value 1 if $s = t$ and zero otherwise. The coefficient β_s^{ζ} measures differences in the response of population share to the new infrastructure by comparing locations with different distances to infrastructure ζ in year s. Figure 7 plots the estimates of β_s^{ζ} for the bridges and tunnel, respectively. The impact of the bridges on population share becomes pronounced after 2011 and is statistically significant during 2012-2016.[20] The dynamic effect of the bridges rules out the possibility that earlier operation of the bridges in 2008 led to the large relocation of residents before operation of the tunnel in 2010, and consequently no significant effect of the tunnel is observed. By contrast, the impact of the tunnel is statistically insignificant during the whole period. Overall, the parallel pre-trends assumption holds for the bridges and tunnel.

20. For the bridges, the positive and significant estimates during 2000-2002 may be the confounding impacts of the existing Xiamen Bridge or simply due to measurement error, since a substantially larger confidence interval is observed before 2002.

Other transport infrastructure improvements. In Table 4, I control for three other transport infrastructure improvements during the sample period that may correlate with operation of the bridges and tunnel. I focus on the population share in two peripheral areas, Jimei and Xiang'an districts. The road distance to other transport infrastructures is calculated based on a road network with the new bridges and tunnel. Figure 4 displays the spatial distribution of the additional transport infrastructures.

First, Xiamen has been connected by high-speed rail since operation of the New Xiamen North Railway Station on April 26, 2010. The first high-speed railway line was called the Fuzhou-Xiamen line and officially opened on that day. The New Xiamen North Railway Station is located in the north of Jimei district, which is around 9.2 and 10.5 km to the Jimei and Xinglin bridges, respectively. Columns 1 and 4 in Table 4 report the results after controlling for the road distance to the high-speed railway station interacted with the corresponding year dummy ($Post2010$).

Second, Gaoqi International Airport, which was built in 1983, substantially increased its capacity in December 2014 after opening its fourth terminal. The airport is located in Huli district and is around 6 km from either the Jimei or Xinglin bridge and 8.5 km to the Xiang'an undersea tunnel. Columns 2 and 5 report the results with interaction of the road distance to the airport and the corresponding year dummy ($Post2014$).

Lastly, Xiamen Station in the city center, which was built in 1957, added a new line—Fuzhou-Xiamen section of Hangzhou-Shenzhen line—on April 26, 2010, and increased its capacity in 2012. Xiamen station is located in Siming district and is around 20 km to either the bridges or the tunnel. Columns 3 and 6 control for road distance to the center station interacted with the corresponding year of improvement ($Post2010$).

As shown in Table 4, results are robust after controlling for the three additional transport infrastructure improvements. The new bridges have positive and significant impacts on the population share of the neighboring locations in Jimei district, whereas the tunnel has no significant impact on the locations in Xiang'an district. In addition, the impact of the high-speed railway station is statistically significant in both districts with opposite signs; the airport and center station do not have a significant effect.

Transport infrastructures that changed the road network were mainly constructed and operated after 2015. For example, Hai Xiang, the main road spanning from the west to the east of Xiang'an district, was opened in 2016 and further connected the district with other regions in Xiamen. The first subway line in Xiamen was opened on December 31, 2017. In 2018, construction of the Xiang'an bridge—the second transport infrastructure connecting the island with Xiang'an district—was officially started. These projects may change the shortest path between locations. In Appendix Table A4, I exclude observations after 2015 and the results are robust.

Non-parametric estimation. One concern is that the two-way fixed-effect (TWFE) estimator in a DID setting with a continuous treatment may generate undesirable weights for heterogeneous treatment effect parameters, which may lead to misleading results (Callaway et al., 2021). To address this concern, I provide robustness checks using a non-parametric estimator proposed by de Chaisemartin et al. (2022), as shown in Table A5 in the Appendix.[21] Consistent with the baseline results, I find a statistically significant impact of the bridges on the population share in Jimei district, and the effect of the tunnel on the population share in Xiang'an district is statistically insignificant. It is important to note that another concern is the potential "selection bias" that may arise from heterogeneous treatment effect functions across different treated units. Unfortunately, as discussed in Callaway et al. (2021), a practical solution to this issue is not yet available.

Further discussion on endogeneity. The baseline results in Table 3 do not necessarily reflect a causal effect of the transport infrastructure. The location choices of the bridges and tunnel are not random, as stated in Section 2.2. However, typical instrumental variables used in studies on intercity transport infrastructure, such as historical routes (Baum-Snow et al., 2017; Banerjee et al., 2020) or simulated optimal transport networks (Faber, 2014), could not be used in Xiamen's case. Exogenous shocks within the city, such as Berlin's division and reunification (Ahlfeldt et al., 2015), are also unavailable.

For the causality concern, my argument is as follows. First, the impact of the

21. I apply the first estimand proposed by de Chaisemartin et al. (2022), which is equivalent to the W_{DID}^{x} estimand in de Chaisemartin and DHaultfoeuille (2018). de Chaisemartin et al. (2022) demonstrate that their estimator can be extended to the cases where there are no "stayers" (i.e., untreated locations in this paper), provided there are "quasi-stayers". I define quasi-stayers as the locations that are more than 20 km away from the respective transport infrastructure. This distance is close to the mean distance and is far enough to minimize the effects of the infrastructure while including a sufficient number of locations.

bridges on population growth in the bridge-connected periphery would be overestimated (underestimated) if there were other favorable (unfavorable) policy shocks in Jimei district related to distance to the bridges. However, I am not aware of any important policies related to distance to the bridges during 2008 to 2015.

Second, the insignificant population growth in the tunnel-connected periphery would be the result of unfavorable policy shocks that prevent residents from moving to Xiang'an district. I argue that this is unlikely to be the case, because local government has focused more on the urban development of Xiang'an district than that of other districts in the past decade. For instance, the growth rate of investment in fixed assets in Xiang'an district from 2010 to 2020 was 539% and the value increased from 13.5 to 86.4 billion yuan. The other five districts had a growth rate below 228%.[22] Another alternative explanation is that, since Xiang'an was the least developed region in Xiamen, it would take years for residents to change their beliefs in that region. This hypothesis is related to the effects of amenity on agents' location choices, which is the main focus of the paper. I argue that access to top schools, which is spatially bounded and difficult to change in the short run, contributes to the heterogeneous effects of the transport infrastructures.

4 A Quantitative Urban Model

To unveil the mechanism behind the heterogeneous effects of the bridges and tunnel as presented in the reduced-form evidence, I develop a quantitative urban model following Ahlfeldt et al. (2015). I simplify two assumptions to focus on the heterogeneous impacts of the changes in transportation network on residents' location choices. First, I fix the land use map. The locations of residential and industrial land are exogenously given. Second, technology is a linear production function with one input factor. The wage is hence equal to the productivity in different locations. The model simplification aims to fit the institutional background in the case of Xiamen in China. I show that the quantitative model can generate the heterogeneous effects of the bridges and tunnel as typically observed in the data.

22. Data are from the Yearbook of Xiamen Special Economic Zone 2021.

4.1 Preference

Consider a city with I residential and J work locations. These locations are exogenously determined by land use planning by local officials and will not change in the short run. The utility of resident o living in location i and working in location j is given by

$$u_{ijo} = \frac{z_{ijo}B_i}{d_{ij}}(\frac{c_{ijo}}{1-\beta})^{1-\beta}(\frac{h_{ijo}}{\beta})^{\beta}, \tag{4}$$

where c_{ijo} and h_{ijo} are the consumption of final goods and housing area, respectively. The parameter $\beta \in (0,1)$ is the share of housing expenditure. Utility increases in residential amenity (B_i) and idiosyncratic preference (z_{ijo}) drawn from a Fréchet (cumulative) distribution, $F(z) = e^{-z^{-\epsilon}}$, where $\epsilon > 1$ is the shape parameter. A greater ϵ implies a smaller variation in the distribution and consumer preferences among locations are less diverse.

The transportation network affects residents' utility through commuting cost. The disutility of commuting from residential location i to work location j is defined as $d_{ij} = e^{\kappa\tau_{ij}} \in (1,\infty)$, where $\tau_{ij} > 0$ is the travel distance measured in kilometers and $\kappa > 0$ governs the scale of commuting cost. A larger value of κ means a stronger effect of the transportation network on population distribution. Travel distance is calculated based on a lowest-cost path derived from a given transportation network ψ. The change in commuting cost due to the new transport infrastructure ζ is denoted as

$$\Delta_\zeta d_{ij} = d_{ij}(\psi_\zeta) - d_{ij}(\psi_0) = e^{\kappa\tau_{ij}(\psi_\zeta)} - e^{\kappa\tau_{ij}(\psi_0)} \leq 0, \tag{5}$$

where ψ_0 and ψ_ζ are the transportation networks before and after the introduction of new infrastructure ζ, respectively. Commuting cost decreases if the lowest-cost travel route is affected by infrastructure ζ. Otherwise, it remains unchanged.

Given the budget constraint $c_{ijo} + q_i h_{ijo} = w_j$, the indirect utility is as follows:

$$u_{ijo} = z_{ijo}B_i q_i^{-\beta} w_j d_{ij}^{-1}, \tag{6}$$

where w_j is the wage earned in work location j and q_i is the rental price of housing in residential location i. The price of consumption goods is normalized as one.

4.2 Educational Access

In China, access to high-quality public schools is an important type of location amenity that affects residents' location choices. Like in the US and many other countries, the availability of public primary schools in urban China is residence-based and spatially bounded by the school's enrollment zone (Chan et al., 2020; Zhang and Chen, 2018). In practice, the choices of primary and middle schools in Chinese cities often depend on the address of residents' *hukou* (household registration status).

To measure the educational access of residential locations, I follow the well-known *gravity potential* approach in the geography literature (Talen and Anselin, 1998), which aggregates the number of available public facilities within a given area adjusted by the friction of distance:

$$B_i = \sum_{k}^{N_e} \frac{1}{\tau_{ik}} I\{k \in \Omega_i\}, \tag{7}$$

where τ_{ik} is the lowest-cost-path distance (km) from residential location i to school k based on the road map and N_e is the total number of top schools in Xiamen.[23] The indicator function $I\{k \in \Omega_i\}$ takes value 1 if school k belongs to the set of available public schools for the i^{th} location (Ω_i) and zero otherwise. To model school enrollment restriction, I assume that Ω_i is bounded by the district to which location i belongs. This measure also corresponds to the closeness centrality in the network literature (see, e.g., Bloch et al., 2016). Appendix Figure A1 shows the spatial distribution of educational access in Xiamen.

Three remarks are in order. First, I assume equal enrollment capacity for each school in the sample, since data on school capacity is unavailable. Second, I assume that utility from educational access is decreasing in the distance between residential and school locations. This is supported by empirical findings that distance to schools matters for student achievement (Talen, 2001). Third, operation of the new bridges and tunnel does not affect the distance to schools because of the cross-district enrollment restriction. Nonetheless, educational access would be changed by the new infrastructures when relaxing the enrollment restriction, as shown later in the counterfactual exercises in Section 6.2.

23. In the robustness checks of model fit in Section 5.2, I show that the model prediction is the same when using either kilometer or meter as the distance unit in the measure of educational access, and simulated results are robust when using a higher order of spatial frictions, i.e., $B_i = \sum_k \frac{1}{\tau_{ik}^2} I\{k \in \Omega_i\}$.

Alternative proxies for location amenity. To demonstrate the critical role of educational access in understanding the distinct effects of the bridges and tunnel, I compare the model using educational access (which I call the baseline model) with the model using access to top-tier public hospitals or top-rated scenic spots as proxies for amenity (which I call the alternative model). Specifically, access to public facility $g \in \{hospital, scene\}$ is defined as follows:

$$B_i^g = \sum_s^{N_g} \frac{1}{\tau_{is}^g}, \tag{8}$$

where τ_{is}^g is the lowest-cost-path distance from residential location i to the s^{th} public facility g and N_g is the total number of public facilities g in Xiamen.

As mentioned in Fact 3 in Section 3.1, the spatial distributions of schools, hospitals, and scenic spots are all highly unequal. However, educational access differs from the other two types of access, in that educational access is spatially bounded by the district of residence and thus the bridges and tunnel do not affect it. By contrast, the choices of public hospitals and scenic spots (and arguably other public facilities such as supermarkets and coffee shops) are not spatially restricted. As a result, the operation of the tunnel would substantially enhance access to these unequally distributed and spatially unrestricted amenities for locations on the tunnel-connected periphery.

4.3 Population Distribution

The derivation of population distribution within the city relies on two favorable properties of Fréchet distribution. First, a monotonic transformation of Fréchet distributed random variable (FDRV) is Fréchet distributed, by which we obtain Fréchet distributed indirect utility (u_{ij}). Second, the maximum of a sequence of FDRVs is Fréchet distributed. Thus, the utility obtained from the realization of potential location choices, $u = max\{u_{ij}\}_{\forall i,j}$, is Fréchet distributed. Given these two properties, the probability that a resident chooses to live in location i and work in location j is[24]

$$\pi_{ij} = Pr[u_{ij} \geq \max\{u_{rs}\}; \forall rs \neq ij] = \frac{(B_i q_i^{-\beta} w_j d_{ij}^{-1})^\epsilon}{\sum_i^I \sum_j^J (B_i q_i^{-\beta} w_j d_{ij}^{-1})^\epsilon} \equiv \frac{\phi_{ij}}{\Phi}. \tag{9}$$

24. See Appendix A for the proof.

It can be proved that the share of population living in location i equals the overall probability that a resident lives in i, which is obtained by summing the probabilities across work locations:

$$\pi_i = \sum_j^J \pi_{ij} = \Phi^{-1}(B_i q_i^{-\beta})^\epsilon \sum_j^J (w_j d_{ij}^{-1})^\epsilon. \tag{10}$$

The population distribution in different residential locations depends on two factors: the housing-price-discounted amenity and distance-discounted wages.[25] Residents are more likely to live in residential location i with better amenity, lower housing price, and closer to high-wage work locations. Similarly, the overall probability that a resident works in location j is obtained by summing the probabilities across residential locations.

4.4 Production

To focus mainly on residents' choices in housing and work locations, I simplify the production side by assuming perfect competition, homogeneous output, and a linear production function with only labor inputs: $y_j = A_j l_j$.[26] Therefore, wage is equal to the average labor productivity in the corresponding location:

$$w_j = A_j. \tag{11}$$

Location-specific productivity is assumed to be unchanged in the short run. Thus, a change of transportation network does not change the wage distribution in the city. It is easy to extend the model so that the transportation network affects wages by

25. In equilibrium, Φ is a constant determined by the reservation utility in other cities, which will be shown later in the spatial equilibrium condition.

26. To simplify the analysis, I have not included land use by firms in the production function. One potential concern is that the tunnel-connected periphery may allocate more land for production and less for housing, resulting in increasing housing prices and hindering residents from moving to Xiangan district. However, I argue that this hypothesis is unlikely for several reasons. First, land use by firms is determined based on the land use plan made by the local government every 10 years, and I did not find a substantial difference in the share of land use by firms between the tunnel-connected periphery and the bridges-connected periphery. Second, the average price of residential land from 2010 to 2015 was 9,239 and 12,146 RMB/m^2 in Xiangan and Jimei districts, respectively, indicating that residential land prices were actually lower in Xiangan compared to Jimei district, which does not support the hypothesis. Lastly, industrial land in Xiangan district was primarily located in the northeastern part of the mainland, which is far away from the tunnel, further suggesting that the tunnel is unlikely to have a significant impact on land use by firms in Xiang'an district.

introducing agglomeration forces in productivity, as discussed in Appendix D.1.

4.5 Housing Market

Housing price is an important aspect of changes in labor mobility in response to changes in the transportation network in the short run. Following standard assumptions in urban models (Roback, 1982; Glaeser et al., 2006), the rental price of housing in different locations is determined by the wages of who chooses to live in that location, given a fixed housing supply. Housing demand in residential location i is given by

$$H_i^d = \sum_j^J h_{ijo} \pi_{ij} L, \tag{12}$$

where L is the total population in the city. Housing supply (H_i) is exogenously given. In equilibrium, housing demand equals housing supply: $H_i^d = H_i$. Combining the housing market clearing condition with location choices equation (9), individual housing demand derived from the utility (4), and the wage equation (11), the equilibrium rental price of housing in location i is derived as follows:

$$q_i = \left[\beta L (\Phi H_i)^{-1} B_i^{\epsilon} \sum_j^J A_j^{\epsilon+1} d_{ij}^{-\epsilon} \right]^{1/(1+\beta\epsilon)}. \tag{13}$$

It is worth mentioning that Ahlfeldt et al. (2015) use another approach to model the equilibrium housing price. They introduce a no-arbitrage condition between commercial and residential land use after accounting for land use regulations. Since prices for commercial and residential land are positively related, it is possible that the housing (residential land) price in location i is negatively related to the wage in i, given firms' trade-off between paying wages and paying rent for commercial land.[27] This could be true in the long run when land use could be fully adjusted. But it may not apply in the short run, especially in countries where land use is highly regulated, as in China.

27. This one-to-one mapping between land price and wage greatly simplifies the model solution of spatial equilibrium compared with the one-to-many mapping, as assumed in this paper.

4.6 Spatial Equilibrium

The spatial equilibrium of labor mobility requires the expected utility in the city be equal to the reservation utility ($\bar{u}$), or the outside option, in other cities:[28]

$$\mathbb{E}(u) = \gamma \left[\sum_{i}^{I} \sum_{j}^{J} (B_i q_i^{-\beta} w_j d_{ij}^{-1})^{\epsilon} \right]^{1/\epsilon} = \bar{u}, \tag{14}$$

where $\gamma = \Gamma((\epsilon - 1)/\epsilon)$ is a constant and $\Gamma()$ is the Gamma function. Therefore, $\Phi \equiv \sum_i^I \sum_j^J (B_i q_i^{-\beta} w_j d_{ij}^{-1})^{\epsilon}$ is a constant in equilibrium determined by the reservation utility and the shape parameter.

Substituting housing price equation (13) and wage equation (11) into the equilibrium condition of labor mobility, I solve for the total population in the city:

$$L = \frac{1}{\beta}(\frac{\gamma}{\bar{u}})^{\frac{1}{\beta}} \left\{ \sum_{r}^{I} \sum_{s}^{J} (B_r A_s d_{rs}^{-1})^{\epsilon} \left[H_r^{-1} B_r^{\epsilon} \sum_{k}^{J} A_k^{1+\epsilon} d_{rk}^{-\epsilon} \right]^{\frac{-\beta\epsilon}{1+\beta\epsilon}} \right\}^{\frac{1+\beta\epsilon}{\beta\epsilon}} . \tag{15}$$

Instead of assuming a fixed total population, I allow residents to move in or out of the city in the model because the total population in Xiamen will change in response to significant improvements in the transportation network. I compare the welfare implications of these two assumptions in Appendix D.2.

The equilibrium of the model is characterized by the model's parameters $\{\beta, \epsilon, \kappa\}$, reservation utility $\bar{u}$, vectors of exogenous productivity, amenity, housing supply $\{\mathbf{A}, \mathbf{B}, \mathbf{H}\}$, and the commuting matrix $\{\mathbf{d}_{\mathbf{I*J}}^{\psi}\}$, given transportation network ψ. Given these inputs, I solve for the endogenous vectors of (residential) population shares, wages, housing prices $\{\pi, \mathbf{w}, \mathbf{q}\}$, and total population L.

5 Calibration and The Role of Educational Access

This section calibrates the quantitative urban model based on the case of Xiamen and discusses how educational access leads to the distinct effects of the new bridges and

28. See Appendix A for the proof.

tunnel on the population distribution on the periphery. I first show that the calibrated model with educational access generates patterns similar to those observed in the data. I then compare with the simulated results of three alternative models: model 1 with the access to hospitals; model 2 with the access to scenic spots; and model 3 with equal amenity and heterogeneous productivity. None of these alternative models generate the patterns as observed in the data. I also provide supporting evidence for the role of educational access in residents' location choices.

5.1 Calibration

Three parameters need to be calibrated: the share of housing expenditure (β), shape parameter of Fréchet distribution (ϵ), and scale of commuting cost (κ). For β, I calculate the expenditure share of housing consumption of total consumption reported by the Xiamen Municipal Bureau of Statistics. The share of housing expenditure is around 0.259 to 0.264 during 2014-2017.[29] I set $\beta = 0.26$.

Regarding the structural parameters ϵ and κ, I identify the optimal values by solving the following objective function which minimizes the weighted sum of square differences between observed and simulated changes in the population distribution in Xiamen.

$$\min_{\{\epsilon,\kappa\}} \sum_i w_i \left(\frac{\triangle\pi_i}{\sigma} - \frac{\triangle\hat{\pi}_i}{\hat{\sigma}} \right)^2, \tag{16}$$

where $\triangle\pi_i$ and $\triangle\hat{\pi}_i$ are the observed and simulated changes in the average population share in residential location i from periods 2003-2007 to 2011-2015, respectively, as defined in Equation (1) in Section 3.1. For $\triangle\hat{\pi}_i$, I simulate the population share both before and after operation of the new infrastructure by manually changing the road network. σ and $\hat{\sigma}$ are the standard deviations of $\triangle\pi_i$ and $\triangle\hat{\pi}_i$, respectively. I normalize the changes in population share by the standard deviation of the changes, because the variance of simulated changes is systematically smaller than that of the observed data. The weight $w_i = \ln(100/\tau_{i\zeta})$, where $\tau_{i\zeta}$ is the shortest-path distance from residential location i to either end of the new bridges or tunnel. I put larger weights on locations

29. There is a change in statistical accounts for household expenditure since 2014, which contains detailed and precise data on consumption expenditure. I assume the share of housing expenditure is stable during the sample period. The simulated population distribution is not sensitive to the value of β.

closer to the new infrastructure. I take the natural logarithm to smooth distribution of the inverse of the distance and multiply it by 100 to avoid negative weights.

I solve Equation (16) by iteration and set initial values as $\epsilon_0 = 6.83$ and $\kappa_0 = 0.01$, which are the estimates in Ahlfeldt et al. (2015). To compute standard errors for the structural parameters, I employ a bootstrap approach following Gandhi et al. (2020).[30] Note that models with different proxies for amenity, as discussed in Section 4.2, correspond to different estimates of the structural parameters.

Table 5 reports the calibrated values of parameters for the baseline model with access to top schools and two alternative models with access to top-tier public hospitals and top-rated scenic spots, respectively. Estimates for the baseline model are $\epsilon = 9.582$ and $\kappa = 0.025$, while those in the two alternative models are $\epsilon = 14.857$ and $\kappa = 0.019$ for alternative model 1 with hospitals, and $\epsilon = 13.483$ and $\kappa = 0.022$ for alternative model 2 with scenic spots. Notably, standard errors in the alternative models are substantially greater than those in the baseline model. The reason is that the alternative models are unable to generate the pattern on the tunnel-connected periphery observed in the data, which I will discuss later in Section 5.3.

5.2 Model Fit

This subsection presents the performance of the baseline model with educational access. Figure 8 plots the correlation between the simulated results and the observed data in terms of three outcome variables: the (log) population distribution before and after operation of the new bridges and tunnel and changes in the distribution (normalized by the standard deviation of the changes) in Xiamen. All three variables exhibit positive and statistically significant correlation between the simulated and observed data.[31] The pairwise correlation coefficients for the three subfigures are 0.5063, 0.5769, and 0.3741,

30. I solve the optimization problem using the modified Newton-Raphson method with STATA 17.0. For bootstrap standard errors, I replace the data with a random sample with replacement drawn from the full sample and re-estimate ϵ and κ 500 times.

31. Appendix Table A6 reports the estimates of a simple linear regression of observed data on simulated data. I also find high correlation between the observed and simulated land price in Xiamen both before and after operation of the bridges and tunnel, as shown in Appendix Figure A4. The location-level land price is the average price of land that was within 1 km road distance from the corresponding residential location. If no land transaction is found within 1 km, then the land price in that location is missing. I cannot compare the change in the land price because only a few locations have land prices both before and after operation of the new infrastructures.

respectively.

Figure 9 presents the spatial distribution of simulated changes in population share. Two patterns match fairly well with the observed data in Figure 5. First, population share increases on the bridges-connected periphery in Jimei district but is unchanged on the tunnel-connected periphery in Xiang'an district. Second, regarding the island area, population share increases in the eastern part, which is linked with the tunnel, and decreases in the western part in the traditional city center.

Figure 10 further compares the mean value of the simulated versus actual change in population share normalized by its standard deviation (SD) in the peripheral areas. To focus mainly on locations that are most affected by the new infrastructure, I choose residential locations within 5 km road distance from the nearest end of either the bridges or tunnel as *target locations*.[32] Appendix Figure A2 shows the spatial distribution of target locations. The mean change in population share on the targeted bridges-connected periphery is 0.65 SD in the baseline model and 0.67 SD in the data, while the respective mean on the targeted tunnel-connected periphery is 0.005 SD in the baseline model and -0.02 SD in the data. Overall, the baseline model with educational access reproduces the heterogeneous effects of the new transport infrastructure on the periphery, which fit the data reasonably well.[33]

Robustness. The baseline model calibrates the structural parameters using information from before and after the changes in the transport infrastructure. To check the robustness of the calibrated model, one may be interested in the model performance when only using information prior to the construction of the bridges and tunnel. To this end, I propose an alternative objective function to calibrate the key parameters, which only applies pre-construction information. The detailed process of calibration is provided in Appendix C. Model performance is shown in Appendix Figures A5 and A6. First, the model reasonably matches the level of and changes in the population share in the observed data. Second, although the prediction of the mean change in population

32. The model can still produce distinct effects of the bridges and tunnel when I increase the distance to 10 km, but it would be less fitted with the data since the impact of the infrastructure becomes trivial when locations are far from the infrastructure, and other uncontrolled factors may dominate the effects in the observed data.

33. Appendix Figure A3 reports the average changes in population share in different regions in Xiamen, including the city center and other locations, based on the observed data and different models. The baseline model predicts relatively high population growth in the city center compared with the observed data.

share on the targeted bridges-connected periphery is not as close to the data as the prediction in the baseline model, it successfully reproduces the distinct effects of the bridges and tunnel on the peripheral population distribution.

I also conduct two robustness checks to examine the sensitivity of model performance to the measure of educational access. First, I use meters instead of kilometers as the distance unit in the measure of educational access. Results are the same as those in the baseline model, as shown in Appendix Figure A7, because any non-zero constant multiplied by the distance will be cancelled out when deriving the equilibrium population share. Second, I use a higher order of spatial frictions to measure educational access, i.e., $B_i = \sum_k \frac{1}{\tau_{ik}^2} I\{k \in \Omega_i\}$. Results are generally robust, as shown in Appendix Figure A8.

5.3 The Role of Educational Access

In this subsection, I highlight the important role of educational access in the distinct effects of the new transport infrastructure on the population distribution on the periphery. I first compare baseline results with those using access to top-tier public hospitals or top-rated scenic spots as proxies for amenity. I show that models that use hospitals or scenic spots as the amenity could not generate the insignificant growth of population share on the tunnel-connected periphery. I then compare the channels of productivity and educational access and show that productivity alone also could not generate the pattern on the tunnel-connected periphery. I argue that the cross-district enrollment restriction is the critical feature of educational access that erodes the population growth effect of the tunnel on the periphery. I also provide supporting evidence that in the past decade, changes in the share of married residents and the average number of children per household were substantially smaller in Xiang'an district, where the tunnel connects, than in Jimei district, where the bridges connect.

Educational access vs. alternative proxies for amenity. One concern is that educational access may only be a proxy for other amenities such as coffee shops, supermarkets, hospitals, and so on. While it is not possible to consider all types of amenities, I choose two representative examples—hospitals and scenic spots—for comparison. As discussed in Fact 3 in Section 3.1, the spatial distributions of top-tier public hospitals and top-rated scenic spots are highly unequal, and much fewer are located on the tunnel-connected periphery, which is similar to the pattern for top schools.

With the calibrated parameters in Table 5, I simulate changes in population distribution using top-tier public hospitals and top-rated scenic spots as proxies for amenities. The measure for access to the hospitals and scenic spots follows the discussion in Section 4.2. The last two groups in Figure 10 show the mean values of the simulated changes in population share on bridges- and tunnel-connected peripheral areas using the two alternative proxies for the amenity. In sharp contrast to the baseline results, the mean changes in population share on the targeted tunnel-connected periphery are above 0.6 SD in two alternative models—much higher than the baseline model and the data. Meanwhile, the prediction for the targeted bridges-connected periphery in both alternative models is below 0.2 SD—much lower than the benchmarks.[34] Overall, the two alternative models do not fit the patterns for target locations and cannot explain the insignificant growth of population share on the tunnel-connected periphery. My argument is that whereas many other amenities are also spatially unequal, residents who choose to live on the tunnel-connected periphery would still benefit from operation of the tunnel since it would greatly expand their access to these amenities in the city center. Nonetheless, educational resources such as top schools are spatially bounded, which causes the tunnel to fail to improve educational access on the tunnel-connected periphery.

Educational access vs. productivity. Another concern is that the baseline results in Figure 10 may be driven by the distribution of productivity rather than educational access. To check this, I compare the baseline results with the predictions in two counterfactual models: one with the actual distribution of firm productivity and equal educational access and the other with the actual educational access and equal productivity. Specifically, I set $B_i = 1$ for all residential locations in the first counterfactual and $A_j = 1$ for all work locations in the second counterfactual. All other settings are the same as in the baseline model, including the parameters.

The last two groups in Figure 11 report the counterfactual results. While both counterfactual scenarios predict high growth of population share on the bridges-connected periphery, they differ in terms of the results for the tunnel-connected periphery. The first scenario, with only the productivity channel, predicts a more than 1.4 SD increase

34. Predictions for the city center and other regions by the two alternative models are qualitatively similar to those in the baseline model (see Appendix Figure A3).

in population share on the tunnel side, which is much different from the prediction of near-zero growth in the second scenario, with only the channel of educational access.[35] Comparing these two counterfactual results with the baseline results, it is unlikely that the productivity channel drives the main results on the tunnel-connected periphery.

A related concern is that the current measure of productivity only includes the manufacturing sector, which ignores the service sector and may not reflect the true distribution of location productivity. I use the annual wage per worker from the 2008 National Economic Census as a proxy for productivity and check the robustness of the baseline model.[36] As stated in Section 2.3, the Economic Census covers all firms in secondary and tertiary industries. Appendix Figure A11 shows the simulated results using data from the Economic Census. Estimates of the parameters are $\epsilon = 6.901$ and $\kappa = 0.024$.

Two main findings are worth noting. First, results from the baseline model with both wage and educational access fit well with the observed data, which suggests that the baseline results are robust to the alternative measure of productivity, which includes firms in the service sector. Second, the wage channel alone could not explain the insignificant growth of population share on the tunnel-connected periphery, which is similar to the results in Figure 11. Taken together, the results are robust when I incorporate firms in both secondary and tertiary industries, and the productivity channel is not the key driver of the distinct effects of the new infrastructure.

Supporting evidence. If educational access is an important factor that contributes to the distinct effects of the new infrastructure, we might observe that residents with children were less likely to move to Xiang'an district even after operation of the tunnel, compared with residents with no children. To check this hypothesis, I compare changes in the share of married residents and changes in the average number of children per household in different districts in Xiamen from 2010 to 2020, using data from the Sixth (2010) and Seventh (2020) National Population Census. I assume that married

35. Appendix Figure A9 shows the results in all regions, including the city center and other locations.

36. The location-level wage is the weighted average of the wages of firms within 1 km from the work location. I use employment share as the weight. If no firm is found within 1 km, I then match with the closest firm's wage. The location-level average wage computed from the Economic Census is highly correlated with that from the ASIE data with only the manufacturing sector, as shown in Appendix Figure A10.

residents are more likely to have children than unmarried.

Figure 12 reports the statistics. First, Jimei district experienced a 6.5-percentage-point increase in the share of married residents, from 60.6% to 67.1%, which is much higher than the mean change (4.33) in the whole city. By contrast, Xiang'an district increased by only 0.9 percentage points in the share of married residents, from 72.2% to 73.1%. Second, Jimei district experienced a 0.2-percentage-point increase in the average number of children per household, from 0.92 to 1.12, which was again higher than the mean change (0.15) in the whole city. In sharp contrast to other districts, Xiang'an district experienced a negative change in the number of children, decreasing from 1.34 to 1.27. While many factors could contribute to the demographic changes in Xiamen, Figure 12 supports the hypothesis that educational access may be an important factor in the location choices of residents with children. As long as residents care much about educational access for their children—as is true in China—they would be less willing to move to Xiang'an district, which is short on top schools, regardless of the huge reduction in travel distance to the city center.

6 Counterfactual Exercises

The quantitative analysis suggests that the unequal distribution of educational resources and the cross-district enrollment restriction could jointly reduce residents' incentives to migrate to the tunnel-connected periphery despite the substantial commuting benefits of using the tunnel. In this section, I conduct several educational policy exercises to see whether and how these policies may promote population growth around the tunnel, as intended.

I examine two types of policies to alter educational access on the tunnel-connected periphery. One policy is to increase the number of top schools on the tunnel-connected periphery, and the other is to relax school enrollment regulations and allow cross-district school entrance. Both policies are implemented before the construction of the bridges and tunnel. I investigate the effectiveness of these two options based on the calibrated quantitative model and discuss the potential costs associated with each.

6.1 Increasing Top Schools on the Periphery

In the first exercise, I choose a certain number of primary and middle schools that were already at the peripheral end of the tunnel during the sample period but not on the list of top schools and upgrade them to top schools. Instead of building new schools in arbitrary locations, investing more educational resources in existing schools may be a more reasonable choice for counterfactual analysis.

Figure 13 demonstrates the counterfactual changes in population share for target locations when there is an increase in top schools on the tunnel-connected periphery before operation of the bridges and tunnel. The first category in Figure 13 shows results from the observed data. The second and third categories show results when I upgrade 5 and 10 existing schools that are closest to the peripheral end of the tunnel, respectively.[37] The counterfactual increase in population share on the bridges-connected periphery is slightly smaller than in the observed data, but the counterfactual increase in population share on the tunnel-connected periphery is enhanced substantially when I increase by 5 top schools. The population growth effect of the tunnel is even more prominent and statistically significant when I increase by 10 top schools.

A possible way to improve the educational quality of existing schools on the periphery is to reallocate educational resources across districts and schools. For instance, Beijing launched a teacher rotation program in recent years to improve educational equity and quality. Teachers in primary and middle schools who have worked at the same school for more than 6 years and are more than 5 years from retirement may be required to rotate to a school in another district and share their experience with their new colleagues and students.[38] Local governments may provide incentives for qualified teachers to move to schools on the periphery, which might complement the effect of the urban transport infrastructure that aims to promote population growth on the periphery.

37. Appendix Figure A12 shows the locations of counterfactual top schools in the analysis. Appendix Figure A13 reports counterfactual results in all regions.

38. See, e.g., "Beijing's teacher rotation policy aims to improve equity," China Daily, 26 August 2021.

6.2 Relaxing the Cross-district School Enrollment Restriction

The second educational policy is to relax the cross-district school enrollment restriction and allow qualified students to choose from all 82 top schools in the city regardless of where they live. Given this assumption, the choice set of schools is no longer bounded by the district, and hence expands greatly for all residential locations. The counterfactual educational access in location i is now given by $\tilde{B}_i = \sum_k^{N=82} \frac{1}{\tau_{ik}}$. Notably, the counterfactual educational access would be affected by the new transport infrastructure, because there are cross-district paths to schools via the bridges or tunnel.

The last category in Figure 13 presents the counterfactual change in the population share for target locations when I relax the cross-district school enrollment restriction both before and after operation of the bridges and tunnel. I observe the largest increase in population share on the tunnel-connected periphery of all the counterfactual exercises, which is about a 2-SD increase in the population share. Meanwhile, the mean change in population share on the bridges-connected periphery is about 0.5 SD, which is the smallest of all the settings. Overall, residents may be more willing to migrate to the tunnel-connected periphery after operation of the tunnel, provided that their children have the opportunity to enroll in high-quality schools in the city center.

6.3 Comparing the Two Counterfactual Policies

The two counterfactual policies discussed above correspond to "moving teachers" and "moving students" across districts, respectively. Moving teachers to the periphery does not require a substantial expansion of school capacity in certain top schools. However, the costs of moving teachers are twofold. First, local governments or target schools on the periphery need to provide incentives for high-quality teachers to move in, and it may take years to improve the educational quality on the periphery. Second, local governments may also need to compensate schools for the loss of high-quality teachers. This problem would be less serious if target schools on the periphery could attract talented teachers from other cities.

By contrast, moving students by allowing cross-district enrollment does not require direct compensation to teachers or students. However, other costs of moving students are also present. For instance, certain top schools might have excess demand far beyond the current enrollment capacity. Local governments or certain top schools might have to invest more resources to expand school capacity or propose new admission mechanisms for screening candidates. The expanded class size might, in turn, lower the educational quality. Moreover, many students might have to travel a long distance to such schools, which might increase risks on the road and the burden on parents.

In sum, the counterfactual analysis based on the quantitative model offers the possibility of employing educational policies to facilitate the intended population growth effect of the new transport infrastructure. Nevertheless, the costs and benefits of such policies must be fully taken into account before putting them into practice.

7 Conclusion

The heterogeneous effects of transport infrastructure on population distribution are well documented in the literature. However, it is less clear how a specific amenity might affect the impact of the transport infrastructure on population distribution within a city. This paper intends to fill this gap by exploiting a unique setting in Xiamen, a coastal city in China, in which a highly developed island is connected to a less developed mainland area via new bridges and a tunnel.

The local government in Xiamen had clear objectives for the new transport infrastructure: to ease increasing congestion between the city center and the periphery and to promote population growth and economic development on the periphery. While the bridges and tunnel serve similar functions, I demonstrate that only the bridges-connected periphery experienced a significant increase in the population share; the tunnel-connected periphery did not attract significant inflow of residents during the sample period. This pattern was unexpected, as the tunnel reduces commuting distance by 41%, which is more than five times as large as the bridges do. The local government also had high expectations for the positive effect of the tunnel on peripheral development.

To shed light on the mechanism behind the distinct effects of the bridges and tunnel, I develop and calibrate a quantitative urban model that fits China's institutional background. The calibrated model implies that the uneven distribution of top schools and the cross-district enrollment restriction in Xiamen might be the main factors that hinder population growth on the tunnel-connected periphery. Counterfactual exercises suggest that increasing the number of top schools on the tunnel-connected periphery or relaxing the restriction on cross-district school enrollment may strengthen the impact of the tunnel and boost population growth on the periphery, as intended by the local government.

This paper focuses on the location choices of residents and assumes that firms are fixed in the short run. Nonetheless, firms may also respond to the change in the transportation network in the long run. Modeling the location choices for both firms and residents may involve the challenge of "spatial impossibility," as first proposed by Starrett (1978) and highlighted by Proost and Thisse (2019). Further research may require more assumptions regarding firms' decisions, such as firms' idiosyncratic preferences for locations, to fully analyze the impacts of urban transport infrastructure on population distribution.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica 83*(6), 2127–2189.

Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies 89*(6), 2911–2957.

Balboni, C., G. Bryan, M. Morten, and B. Siddiqi (2020). Transportation, gentrification, and urban mobility: The inequality effects of place-based policies. *Working paper*.

Banerjee, A., E. Duflo, and N. Qian (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics 145*, 102442.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 775–805.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics 99*(3), 435–448.

Baum-Snow, N., J. V. Henderson, M. A. Turner, Q. Zhang, and L. Brandt (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics 115*, 103124.

Bloch, F., M. O. Jackson, and P. Tebaldi (2016). Centrality measures in networks. *Working paper*.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. *NBER Working Paper 4483*.

Cellini, S. R., F. Ferreira, and J. Rothstein (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics 125*(1), 215–261.

Chan, J., X. Fang, Z. Wang, X. Zai, and Q. Zhang (2020). Valuing primary schools in urban China. *Journal of Urban Economics 115*, 103183.

de Chaisemartin, C. and X. D'Haultfoeuille (2022a). Two-way fixed effects and differences-in-differences estimators with several treatments. *NBER Working Paper 30564*.

de Chaisemartin, C. and X. D'Haultfoeuille (2022b). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *NBER Working Paper 29691*.

de Chaisemartin, C., X. D'Haultfoeuille, F. Pasquier, and G. Vazquez-Bare (2022). Difference-in-differences estimators for treatments continuously distributed at every period. *arXiv preprint arXiv:2201.06898*.

de Chaisemartin, C. and X. DHaultfoeuille (2018). Fuzzy differences-in-differences. *The Review of Economic Studies 85*(2), 999–1028.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A 'market access' approach. *The Quarterly Journal of Economics*, 1–60.

Epple, D., A. Jha, and H. Sieg (2018). The superintendent's dilemma: Managing school district capacity as parents vote with their feet. *Quantitative Economics 9*(1), 483–520.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's national trunk highway system. *Review of Economic Studies 81*(3), 1046–1070.

Fack, G. and J. Grenet (2010). When do better schools raise housing prices? evidence from Paris public and private schools. *Journal of Public Economics 94*(1-2), 59–77.

Gaigné, C., H. R. Koster, F. Moizeau, and J.-F. Thisse (2022). Who lives where in the city? Amenities, commuting and income sorting. *Journal of Urban Economics 128*, 103394.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Garcia-López, M., A. Holl, and E. Viladecans-Marsal (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics 85*, 52–67.

Glaeser, E. L., J. Gyourko, and R. E. Saks (2006). Urban growth and housing supply. *Journal of Economic Geography 6*(1), 71–89.

Heblich, S., S. J. Redding, and D. M. Sturm (2020). The making of the modern metropolis: Evidence from London. *The Quarterly Journal of Economics 135*(4), 2059–2133.

Jedwab, R. and A. Storeygard (2022). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 19602010. *Journal of the European Economic Association 20*(1), 1–38.

Loumeau, G. (2023). Locating public facilities: Theory and micro evidence from Paris. *Journal of Urban Economics 135*, 103544.

Mo, J. (2018). Land financing and economic growth: Evidence from Chinese counties. *China Economic Review 50*, 218–239.

Proost, S. and J. F. Thisse (2019). What can be learned from spatial economics? *Journal of Economic Literature 57*(3), 575–643.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 1257–1278.

Small, K. A. and E. T. Verhoef (2007). *The Economics of Urban Transportation*. Routledge.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory 17*, 21–37.

Talen, E. (2001). School, community, and spatial equity: An empirical investigation of access to elementary schools in West Virginia. *Annals of the Association of American Geographers 91*(3), 465–486.

Talen, E. and L. Anselin (1998). Assessing spatial equity: An evaluation of measures of accessibility to public playgrounds. *Environment and Planning A 30*(4), 595–613.

Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: Evidence from Bogotás Transmilenio. *University of California, Berkeley (mimeo)*.

Zárate, R. D. (2022). Spatial misallocation, informality, and transit improvements: Evidence from Mexico City. *World Bank Policy Research Working Paper*.

Zhang, M. and J. Chen (2018). Unequal school enrollment rights, rent yields gap, and increased inequality: The case of Shanghai. *China Economic Review 49*, 229–240.

8 Tables

TABLE 1
Summary Statistics

Variables	Obs.	Mean	SD	Min	Max
Average population size (2000-2010)	2,926	2,970	7,888	1	106,161
Average population size (2011-2021)	2,926	5,333	9,335	1	80,330
Average population share (2000-2010,%)	2,926	0.376	0.966	0.0001	11.87
Average population share (2011-2021,%)	2,926	0.376	0.662	0.0001	6.86
Commuting distance without new infrastructure (km)	56,924	23.28	11.52	0.90	55.78
Commuting distance with new infrastructure (km)	56,924	21.81	10.22	0.90	54.15
Distance to bridges (km)	266	17.73	6.604	1.502	33.35
Distance to tunnel (km)	266	19.75	8.238	0.603	36.10
Distance to schools (km)	21,812	20.53	9.89	0.16	52.87
Distance to hospitals without new infrastructure (km)	3,192	23.29	12.21	0.59	51.07
Distance to hospitals with new infrastructure (km)	3,192	20.72	10.05	0.59	49.91
Distance to scenic spots without new infrastructure (km)	2,926	23.74	11.91	0.42	54.45
Distance to scenic spots with new infrastructure (km)	2,926	22.75	10.82	0.42	54.01
Total factor productivity (log)	214	4.303	0.491	3.314	6.958
Floor ratio	266	2.475	1.477	0.01	10.68

Notes: This table provides summary statistics of the main variables. Average population size is rounded to the nearest integer. Commuting distance refers to the lowest-cost path distance between residential and work locations. Distance to bridges (tunnel) refers to the lowest-cost path distance between residential locations and the nearest end of the bridges (tunnel). Distance to schools, hospitals, and scenic spots refers to the lowest-cost path distance between residential and public locations, respectively. New infrastructure refers to the new bridges and tunnel. Note that distance to schools is not affected by the new infrastructure due to the cross-district enrollment restriction. The average number of available schools is 12.4. The observation for commuting distance is $266 * 214$.

TABLE 2
Summary of the Changes in Commuting Distance with the New Urban Infrastructures (km)

Variables	Obs.	Mean	S.d.	Min	Max
Change in commuting distance with only bridges	10,772	-1.79	1.80	-7.33	-0.002
Change in commuting distance with only tunnel	7,962	-9.64	7.28	-39.35	-0.001
Change in commuting distance with both	13,313	-6.32	7.01	-39.35	-0.001

Notes: This table presents summary statistics for changes in commuting distance between residential and work locations with the introduction of bridges or/and tunnel on the road map. Least-cost paths between location pairs that are not affected by the bridges or tunnel are not included in the table. See the calculation of road distance in Section 2.3.

TABLE 3
Distance to Infrastructures and Population Share

Dependent Variable: Pop. Share (log)	(1) Full Sample	(2) Periphery (Jimei)	(3) Full Sample	(4) Periphery (Xiang'an)	(5) Full Sample
DistanceToBridges(log)*Post2008	-0.377***	-0.793***			-0.510***
	(0.074)	(0.185)			(0.082)
DistanceToTunnel(log)*Post2010			0.064	0.050	0.257***
			(0.068)	(0.039)	(0.061)
Location FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	5,852	1,188	5,852	1,298	5,852
R-squared	0.898	0.912	0.895	0.898	0.899

Notes: This table reports the baseline results of the impacts of the new infrastructures on population distribution in Xiamen. The full sample includes residential locations in all districts, while periphery (Jimei) and periphery (Xiang'an) refer to the sample in Jimei and Xiang'an districts, respectively. Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE 4

Robustness: Control for Other Transport Infrastructure Improvements

Dependent Variable: Pop. Share (log)	(1)	(2)	(3)	(4)	(5)	(6)
		Periphery (Jimei)			Periphery (Xiang'an)	
DistanceToBridges(log)*Post2008	-0.885***	-0.745***	-0.466***			
	(0.237)	(0.182)	(0.118)			
DistanceToTunnel(log)*Post2010				0.024	0.046	-0.043
				(0.044)	(0.043)	(0.117)
DistanceToHighSpeedStation(log)*Post2010	0.156*			-0.276**		
	(0.091)			(0.116)		
DistanceToAirport(log)*Post2014		-0.090			0.017	
		(0.192)			(0.036)	
DistanceToCenterStation(log)*Post2010			-0.647			0.335
			(0.418)			(0.286)
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,188	1,188	1,188	1,298	1,298	1,298
R-squared	0.895	0.894	0.895	0.892	0.892	0.892

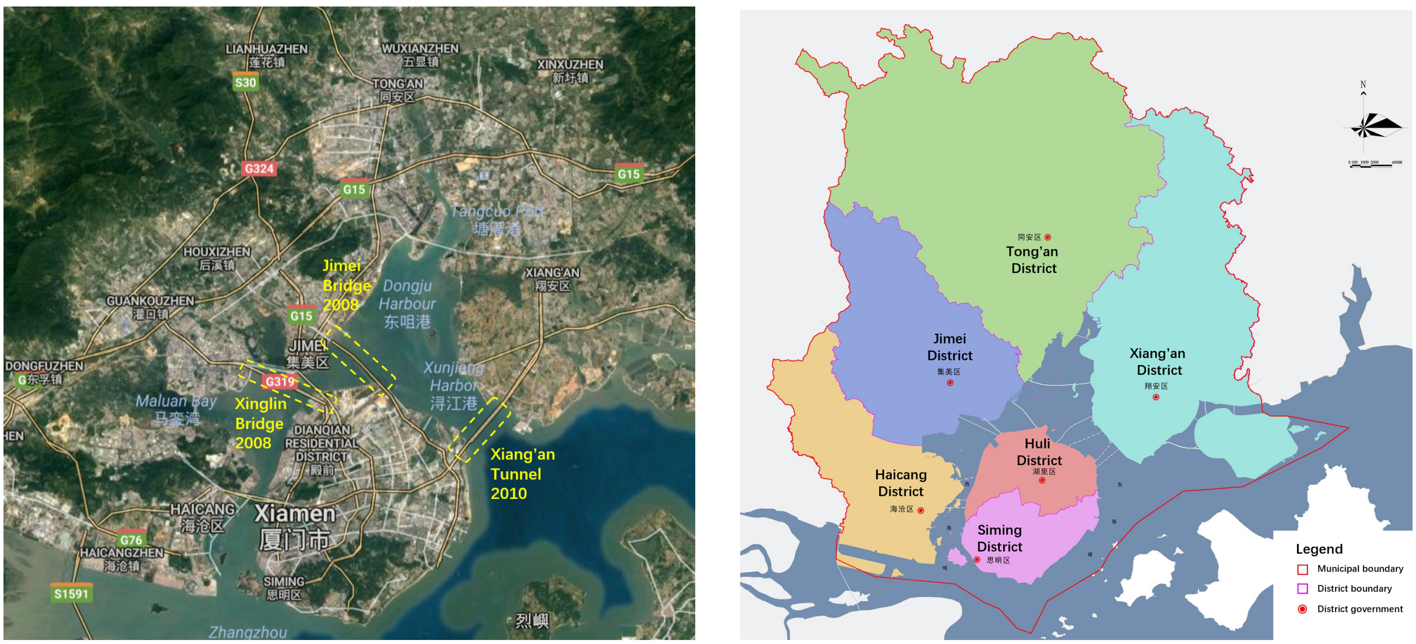
Notes: This table presents the robustness of the impacts of the new infrastructures on population distribution in Xiamen by controlling for the impacts of other transport infrastructure improvements during the sample period. Periphery (Jimei) and periphery (Xiang'an) refer to the sample in Jimei and Xiang'an districts, respectively. Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE 5
Calibrated Model Parameters

Param.	Description	Baseline Model (School)	Alternative 1 (Hospital)	Alternative 2 (Scene)
ϵ	Shape parameter of Fréchet distribution	9.582	14.857	13.483
		(2.122)	(10.333)	(9.310)
κ	Scale parameter of commuting cost	0.025	0.019	0.022
		(0.004)	(0.032)	(0.007)
β	Share of housing expenditure		0.26	

Notes: This table reports estimates of the three parameters in the baseline model with educational access and two alternative models with access to top-tier hospitals and top-rated scenic spots, respectively. See Section 4.2 for definitions of the three models. Bootstrap standard errors are shown in parentheses.

9 Figures



(A) Geography (B) Administrative Boundaries

FIGURE 1
Geography and Administrative Boundaries of Xiamen

Notes: The dashed boxes highlight the new bridges and tunnel with the corresponding operating years. The landscape of Xiamen is from Google Earth.

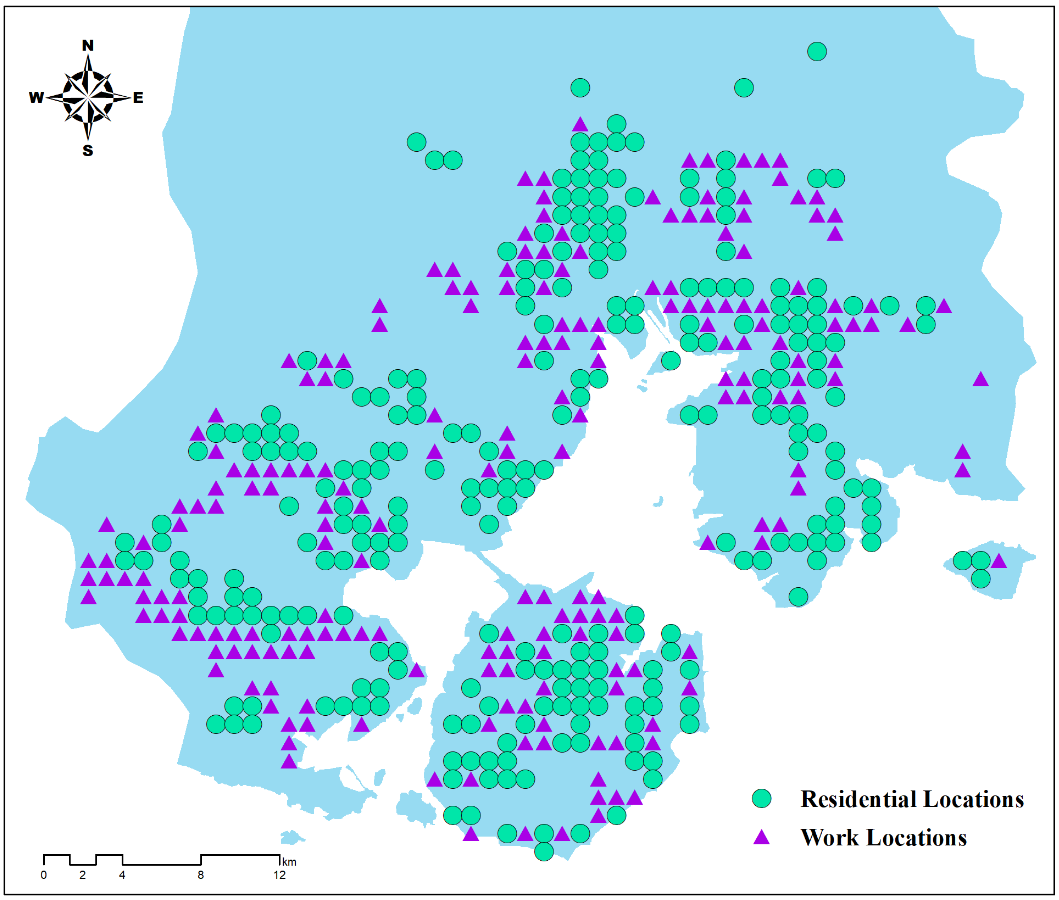


FIGURE 2
Residential and Work Locations in Xiamen

Notes: This figure presents the distribution of residential and work locations in Xiamen. Each location is an area of around 1 km^2 based on the POI data from LandScan database. The definitions of residential and work locations are provided in Section 2.3.

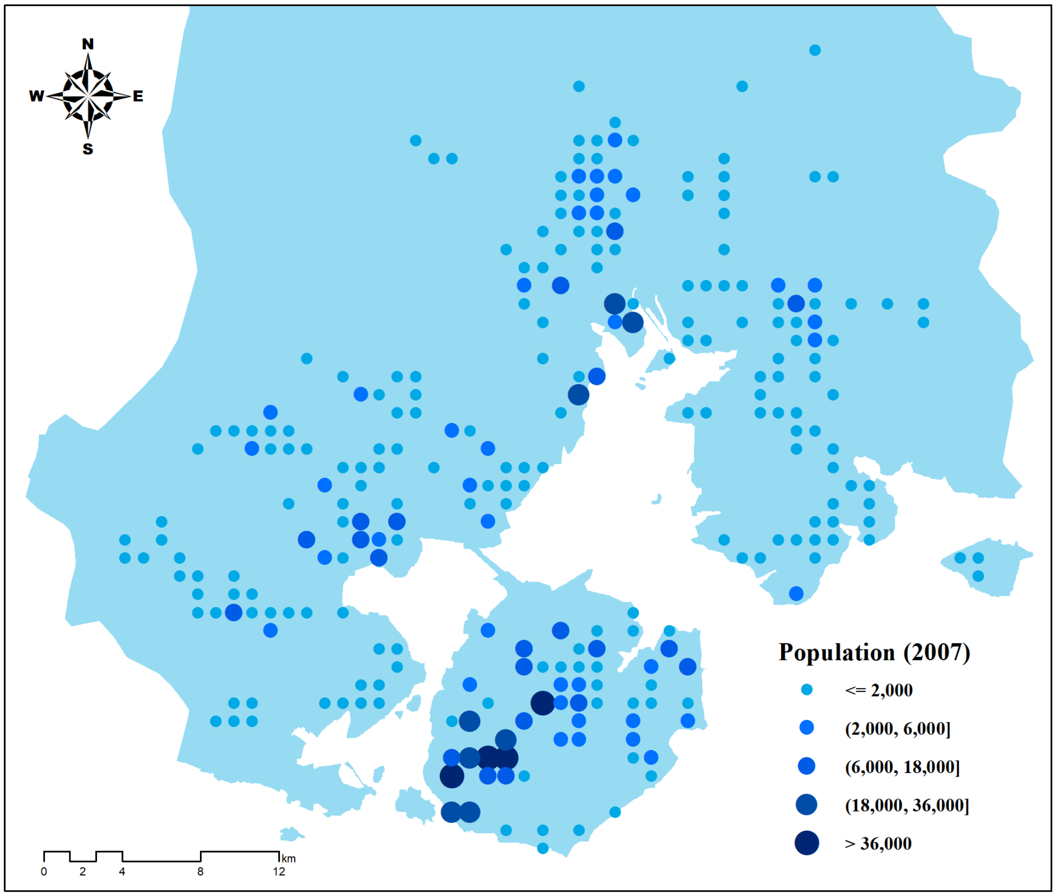


FIGURE 3
Population Distribution in Xiamen, 2007

Notes: This figure plots the population of residential locations in Xiamen in 2007. Each residential location is an area of around 1 km^2 based on the POI data from LandScan database.

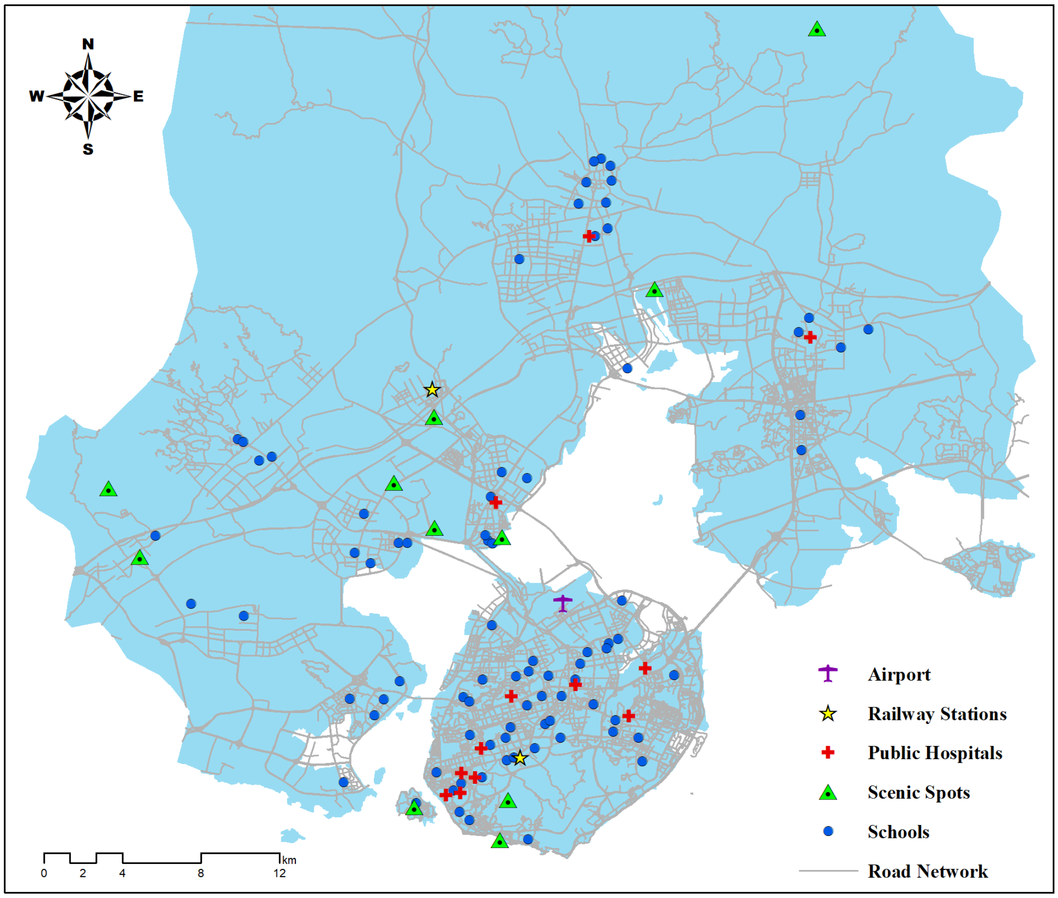


FIGURE 4
Distribution of Public Locations in Xiamen

Notes: This figure plots the distribution of various types of public locations in Xiamen, including Gaoqi International Airport, the New Xiamen North Railway Station (high-speed railway on the north mainland), the Xiamen Station (on the island), 12 top-tier public hospitals, 11 top-rated scenic spots, and 82 top primary and middle schools. Definitions of the corresponding hospitals, scenic spots, and schools are provided in Section 2.3. The road network is from OpenStreetMap in 2014.

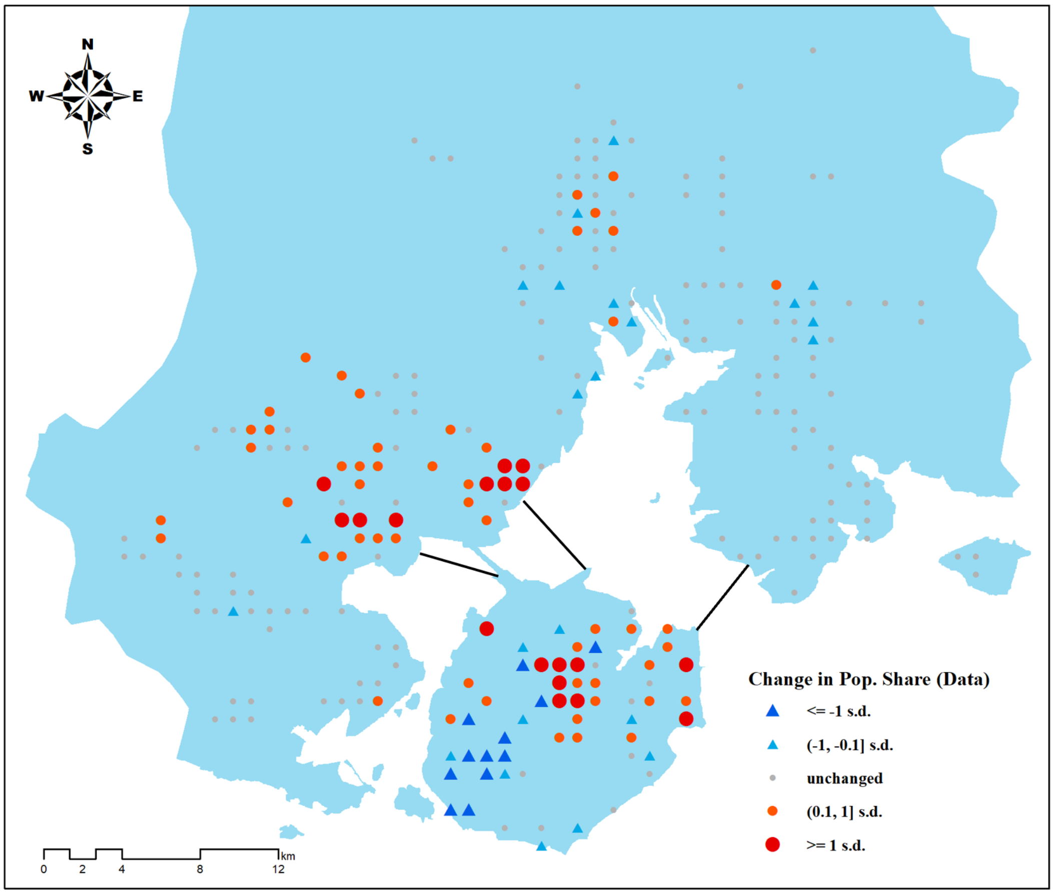


FIGURE 5
Changes in the Spatial Distribution of Residents in Xiamen

Notes: This figure shows changes in the population share in Xiamen from 2003-2007 to 2011-2015. Change in the population share is normalized by its standard deviation and the calculation is provided in Section 3.1. The two solid lines in the northern part of the island are the Jimei Bridge (right) and Xinglin Bridge (left), respectively. The solid line in the eastern part of the island is the Xiang'an Undersea Tunnel.

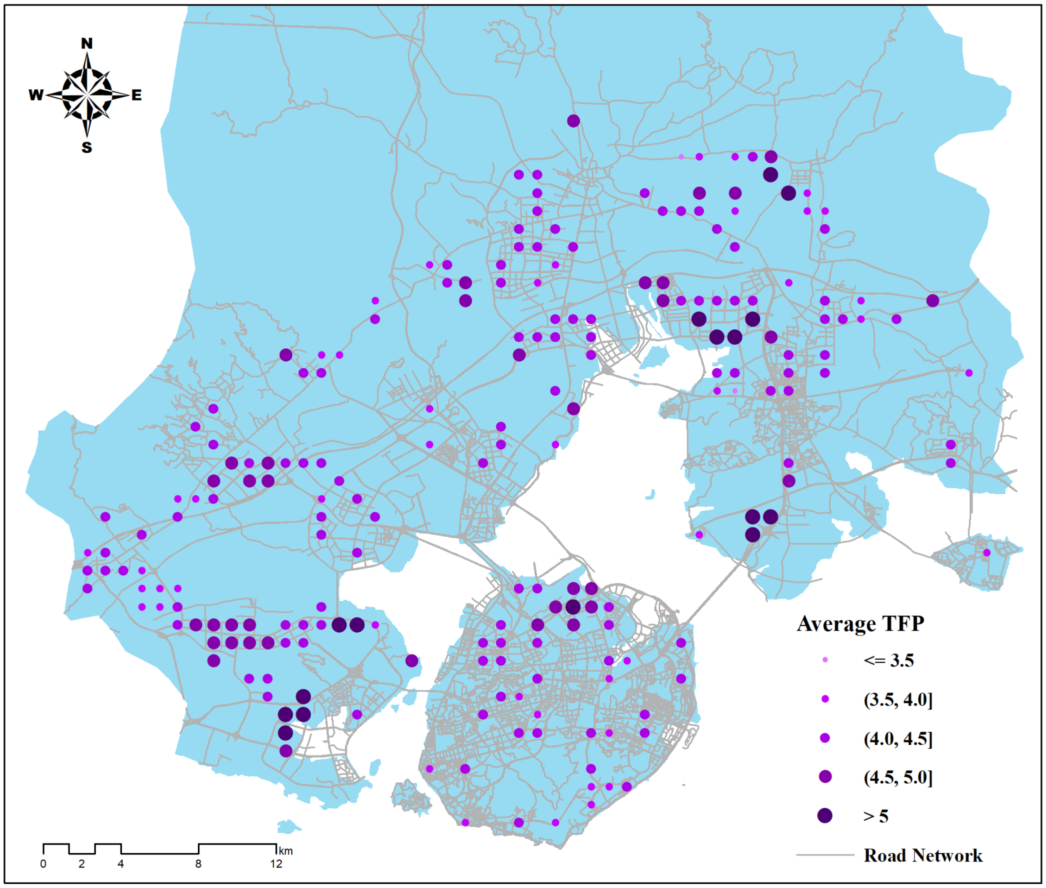
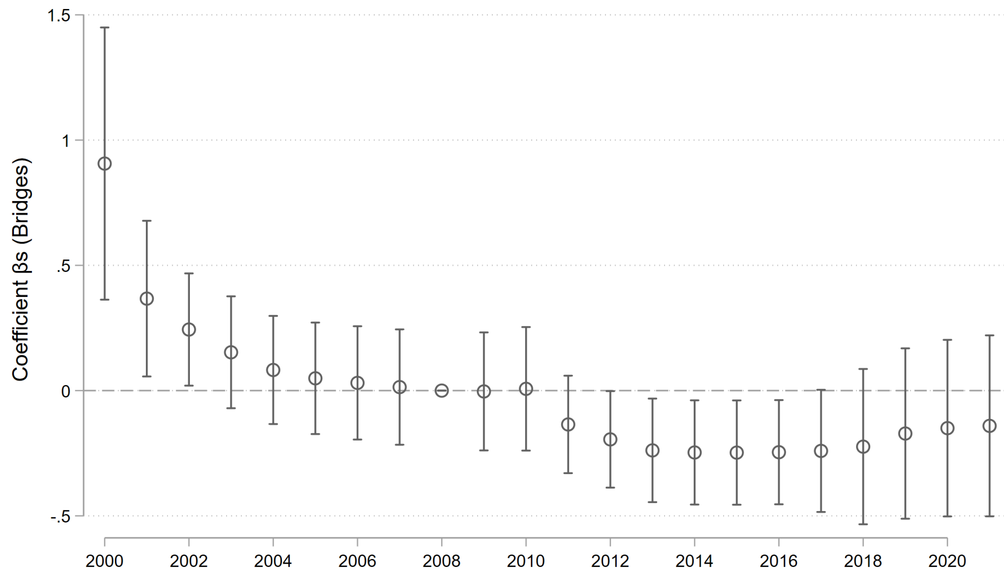
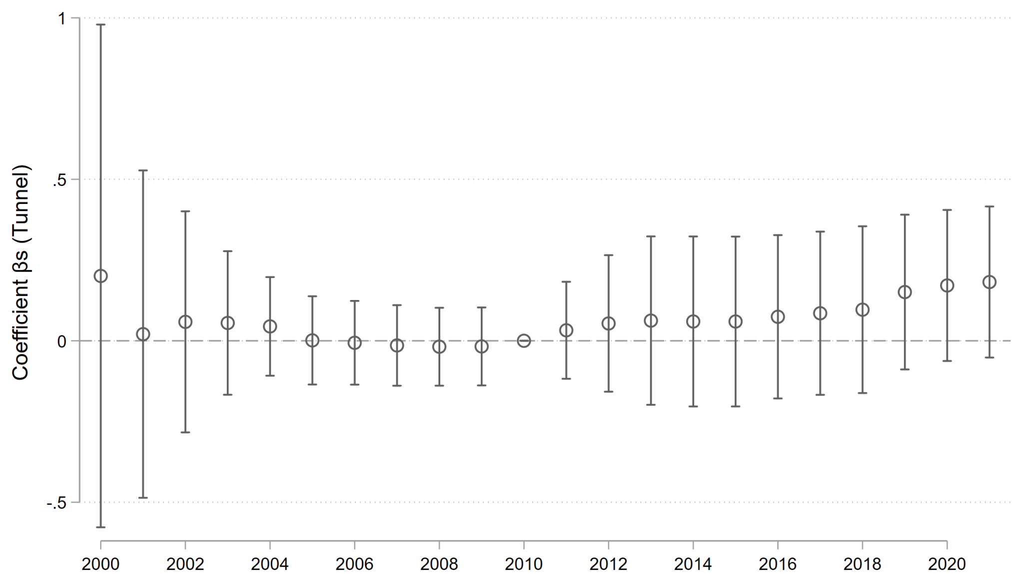


FIGURE 6
Distribution of Total Factor Productivity in Xiamen

Notes: This figure shows the distribution of total factor productivity (TFP) of manufacturing firms in Xiamen from 2000 to 2007. The calculation of average TFP is described in Section 2.3.



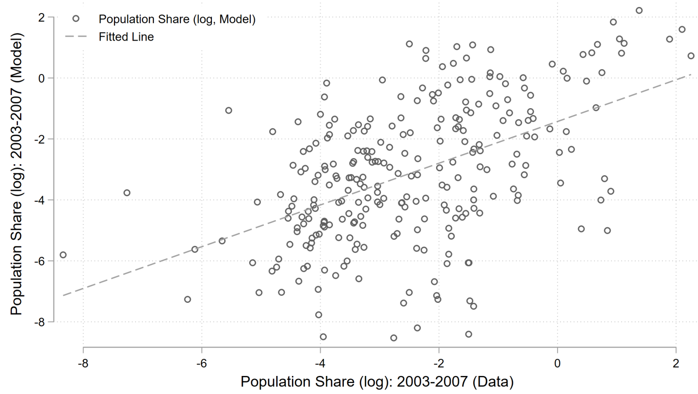
(A) Bridges



(B) Tunnel

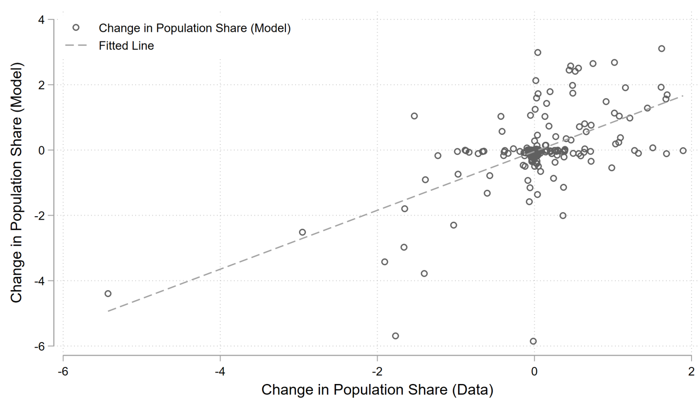
FIGURE 7
Dynamic Effects of Urban Transport Infrastructure

Notes: This figure plots the dynamic effects of the bridges and tunnel on population share. I set 2008 and 2010 as the base years for results for the bridges and tunnel, respectively. Coefficients and 90% confidence intervals are estimated following Equation (3) with the full sample.



(A) Average Population Share during 2003-2007

(B) Average Population Share during 2011-2015



(C) Change in Population Share

FIGURE 8
Model Fit: Observed vs. Simulated Population Share

Notes: This figure plots the correlation between the observed and simulated population share in Xiamen. The simulation is based on the baseline model using educational access as the proxy for amenity. I set $\beta = 0.26$, $\epsilon = 9.582$, $\kappa = 0.025$. The change in population share is normalized by its standard deviation.

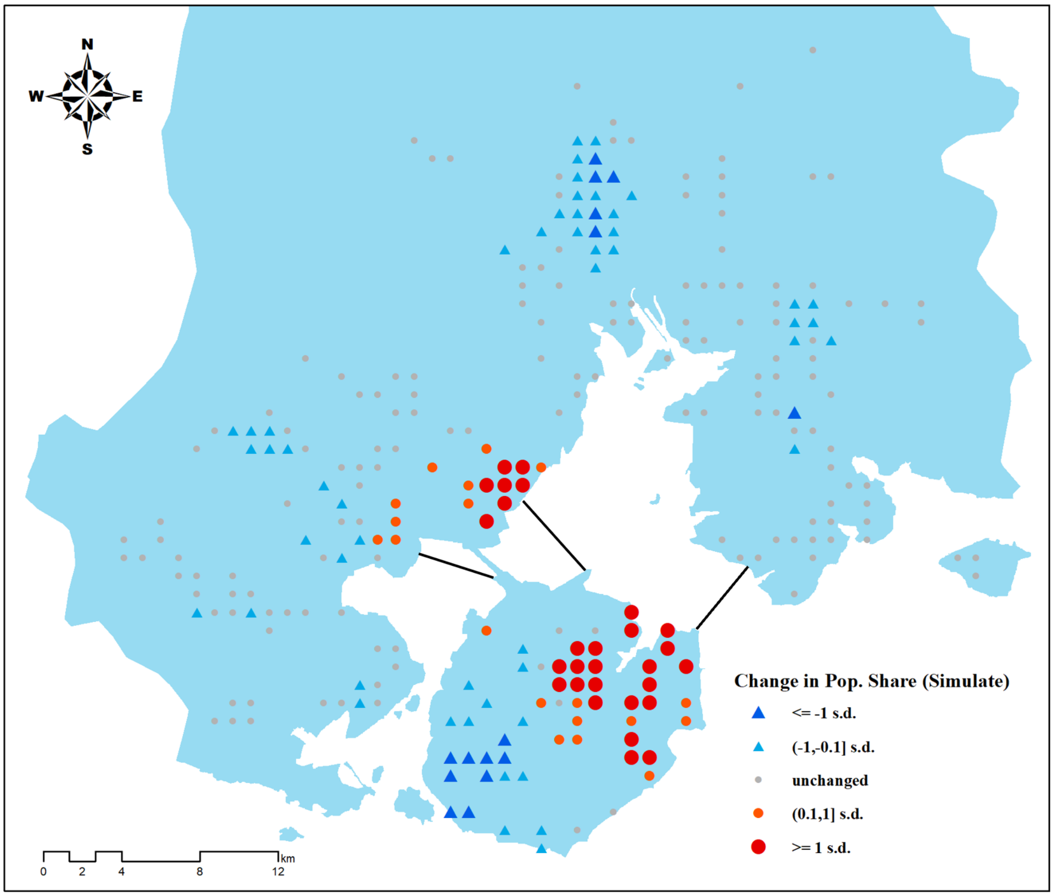


FIGURE 9
Simulated Changes in Spatial Distribution of Residents in Xiamen

Notes: This figure depicts simulated changes in the population share in Xiamen after operation of the bridges and tunnel. The simulation is performed with the baseline model using educational access as the proxy for location amenity. The parameters are set as follows: $\beta = 0.26$, $\epsilon = 9.582$, $\kappa = 0.025$. The value is normalized by the standard deviation of the simulated changes in the population share.

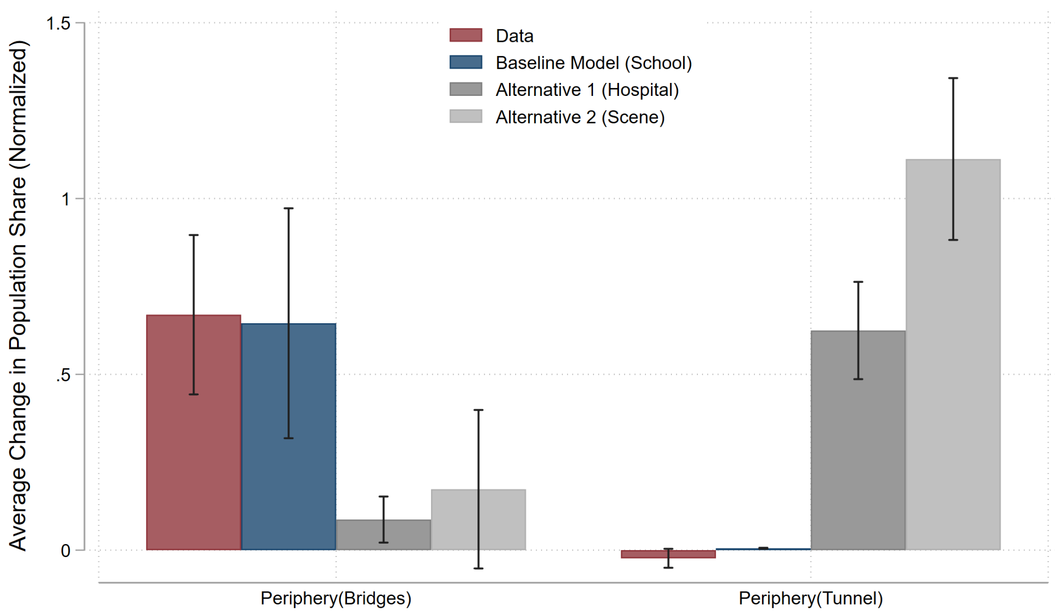


FIGURE 10
Model Fit: Average Change in Population Share on the Periphery

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) on the bridges- and tunnel-connected peripheral areas based on the observed data, the baseline model with educational access, and two alternative models with access to top-tier public hospitals and top-rated scenic spots, respectively. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The confidence interval is at the 95% level.

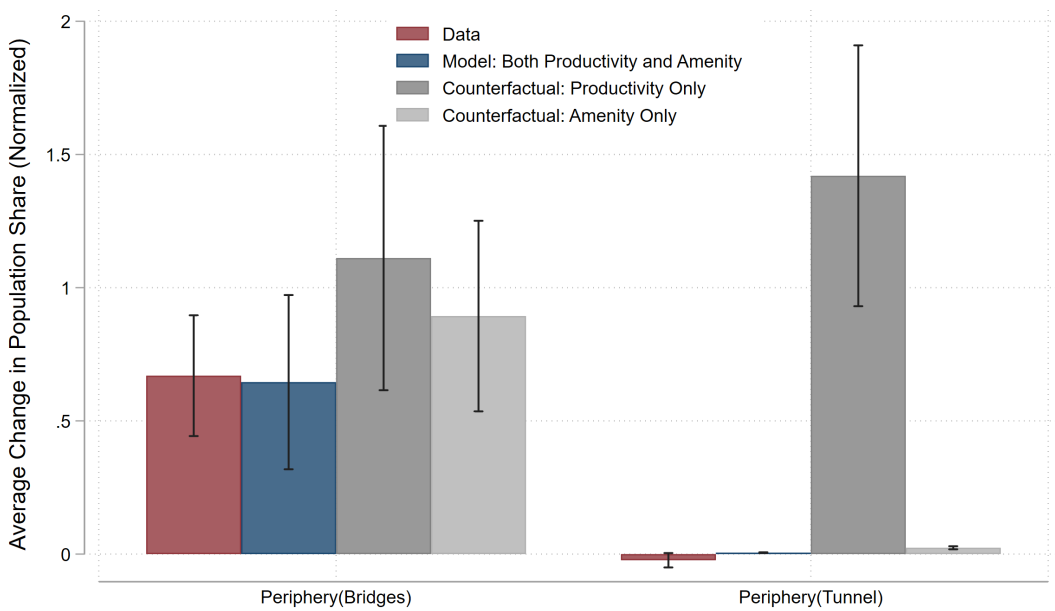
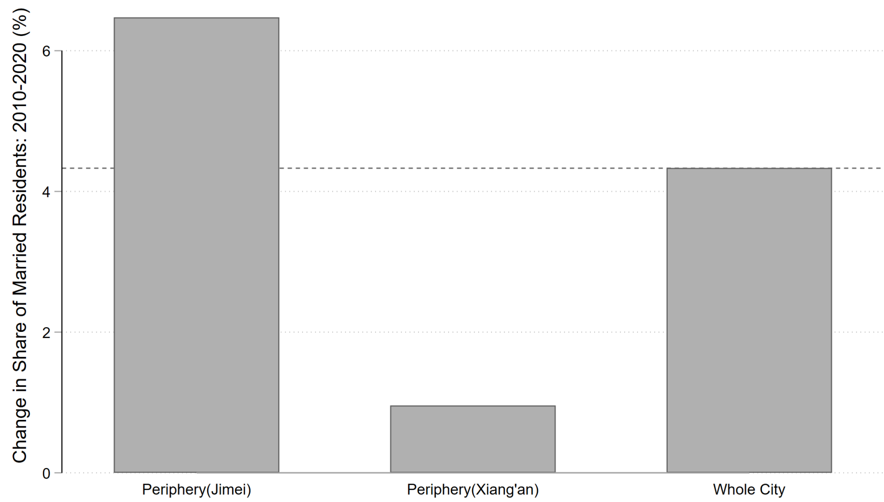
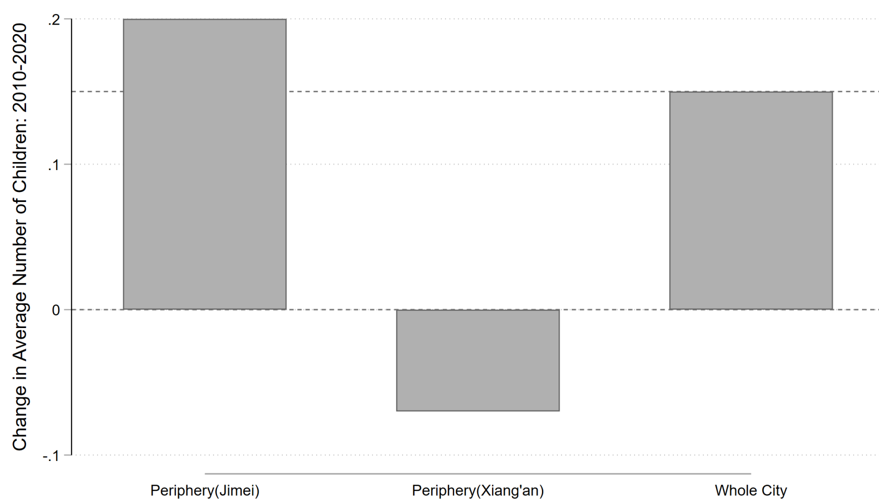


FIGURE 11
Model Decomposition: Productivity vs. Educational Access

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) on bridges- and tunnel-connected peripheral areas based on the observed data, the baseline model with educational access, and two counterfactual models with actual productivity and equal amenity for the first counterfactual (productivity only) and actual educational access and equal productivity for the second (amenity only). Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The confidence interval is at the 95% level.



(A) Change in Share of Married Residents



(B) Change in Average Number of Children

FIGURE 12
Changes in Share of Married Residents and Average Number of Children: 2010-2020

Notes: This figure shows the changes in the share of married residents and average number of children per household in different regions in Xiamen from 2010 to 2020. Periphery (Jimei) and Periphery (Xiang'an) refer to data for Jimei and Xiang'an districts, respectively. Data are from the Sixth (2010) and Seventh (2020) National Population Census.

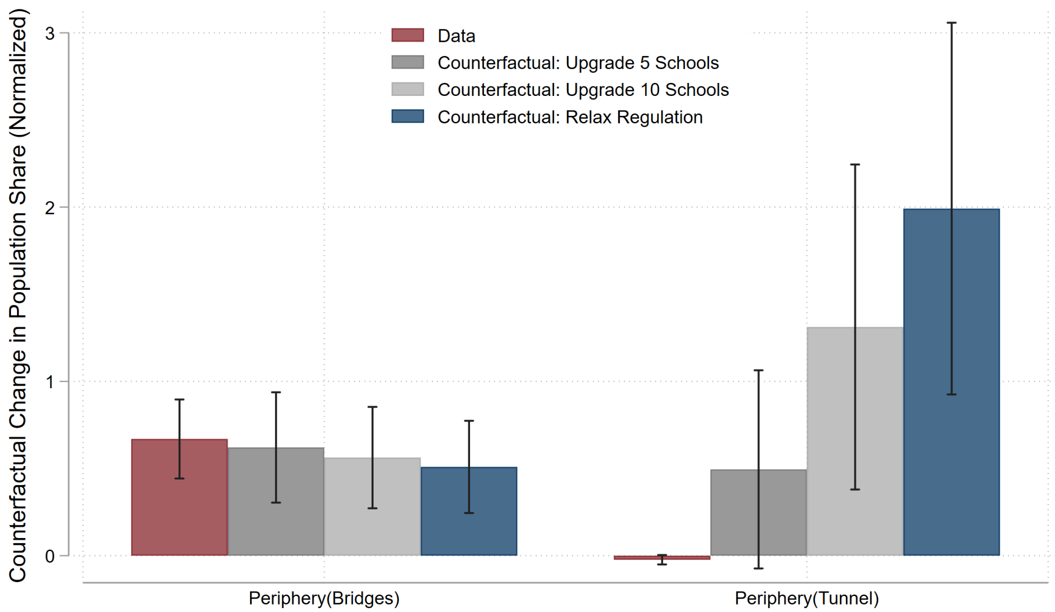


FIGURE 13
Counterfactual Exercises

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) on bridges- and tunnel-connected peripheral areas based on the observed data and three counterfactual models. The first and second counterfactual exercises choose 5 and 10 existing schools, respectively, that are the closest to the peripheral end of the tunnel and upgrade them to top schools. The third counterfactual exercise allows students to choose from all 82 top schools in the city regardless of where they live. All counterfactual assumptions are made both before and after operation of the bridges and tunnel. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The confidence interval is at the 95% level.

Appendix for "Heterogeneous Effects of Urban Transport Infrastructure on Population Distribution: The Role of Educational Access"

(For Online Publication)

Jiawei Mo*

In this appendix, I collect the analyses, discussions, figures, and tables omitted from the main text.[1]

Appendix A Proofs

A.1 Proof of Population Distribution (π_{ij})

Proof. Given the indirect utility equation (6) and the Fréchet distribution of idiosyncratic preference z_{ijo}, we obtain a Fréchet (cumulative) distribution of utility working in j and living in i as follows:

$$G_{ij}(u) = Pr[u_{ijo} \leq u] = Pr[z_{ijo} \leq (B_i q_i^{-\beta} w_j d_{ij}^{-1})^{-1} u] = e^{-\phi_{ij} u^{-\epsilon}},$$

where $\phi_{ij} \equiv (B_i q_i^{-\beta} w_j d_{ij}^{-1})^{\epsilon}$. Let U be the maximum of $\{u - ijo\}_{ij}$ for every $i - j$ pair with respect to the draws by individual o. The distribution of U is also Fréchet distributed, satisfying

$$Pr[U \geq u] = 1 - G(u) = 1 - \prod_i^I \prod_j^J G_{ij}(u).$$

*School of Economics, Peking University; email: jwmo@pku.edu.cn.

1. This note is not self-contained; it is the online appendix of the paper "Heterogeneous Effects of Urban Transport Infrastructure on Population Distribution: The Role of Educational Access."

Then we have the cumulative distribution of U:

$$G(u) = e^{-\Phi u^{-\epsilon}}, \qquad \Phi \equiv \sum_i^I \sum_j^J (B_i q_i^{-\beta} w_j d_{ij}^{-1})^\epsilon.$$

The probability of choosing the $i-j$ pair is identical to the probability that the utility drawn from the $i-j$ pair is no less than all other pairs:

$$\begin{aligned}
\pi_{ij} =& Pr[u_{ij} \geq max\{u_{rs}\}; \forall rs \neq ij] \\
=& \int_0^\infty \prod_r \prod_s^{rs \neq ij} G_{rs}(u) dG_{ij}(u) \\
=& \int_0^\infty e^{-\sum_r \sum_s^{rs \neq ij} \phi_{rs} u^{-\epsilon}} \epsilon \phi_{ij} u^{-\epsilon-1} e^{-\phi_{ij} u^{-\epsilon}} du \\
=& \int_0^\infty e^{-\sum_r \sum_s \phi_{rs} u^{-\epsilon}} \epsilon \phi_{ij} u^{-\epsilon-1} du \\
=& \int_0^\infty \frac{\phi_{ij}}{\Phi} dG(u) \\
=& \frac{\phi_{ij}}{\Phi}.
\end{aligned}$$

Note that

$$dG_{ij}(u) = \epsilon \phi_{ij} u^{-\epsilon-1} e^{-\phi_{ij} u^{-\epsilon}} du.$$

The notation $rs \neq ij$ means any combinations of r and s except the one in which $r = i$ *and* $s = j$. ■

A.2 Proof of Expected Utility ($\mathbb{E}(u)$)

Proof. The expected utility in the city is given by

$$\mathbb{E}(u) = \int_0^\infty u dG(u) = \int_0^\infty \epsilon \Phi u^{-\epsilon} e^{-\Phi u^{-\epsilon}} du.$$

Let $y = \Phi u^{-\epsilon}$, then $dy = -\epsilon \Phi u^{-\epsilon-1} du$. We have

$$du = \frac{-dy}{\epsilon \Phi u^{-\epsilon-1}}, \qquad u = (\frac{y}{\Phi})^{-\frac{1}{\epsilon}}.$$

Replace u with y in the expected function and we obtain

$$
\begin{aligned}
\mathbb{E}(u) &= \int_{\infty}^{0} \frac{\epsilon\Phi u^{-\epsilon}}{\epsilon\Phi u^{-\epsilon-1}} e^{-y}(-dy) \\
&= \int_{0}^{\infty} ue^{-y}dy \\
&= \Phi^{\frac{1}{\epsilon}} \int_{0}^{\infty} y^{-\frac{1}{\epsilon}} e^{-y}dy
\end{aligned}
$$

Note that there is a change in the integrating range from infinity to zero. Given the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$, it is straightforward to derive the expected utility as follows:

$$
\mathbb{E}(u) = \Phi^{\frac{1}{\epsilon}}\Gamma(1 - \frac{1}{\epsilon}).
$$

■

Appendix B Calculation of Shortest-Path Distance

This appendix describes, step by step, how to obtain the shortest-path distance for an arbitrary number of points in a given road map. The software I use is ArcMap 10.3 and ArcCatalog 10.3.

First, construct straight lines between the points and their closest path on the road map in ArcMap. This step is necessary because ArcMap calculates the shortest-path distance between any two points only when the points are covered by the road. In most cases, however, our points of interest are not exactly on the given road map. This step is performed in the ArcMap with functions “Analysis Tools → Proximity → Near” and “Data Management Tools → Features → XY to Line”. Then we append the new lines to the road map using the function “Data Management Tools → General → Append.”

Second, break the road feature into lines and split the lines at the given points in ArcMap. This step is to prepare for the topology analysis in a network of points and lines. Related tools are “Data Management Tools → Features → Feature to Line” and “Data Management Tools → Features → Split Line at Point.”

Third, build the topology using the shape files of points and split roads in ArcCatalog. Import the shape files into ArcCatalog and create the topology with function “New → Topology.” Note that the rule of topology analysis is “point must be covered by line.”

Fourth, create a network dataset using the shape files of points and split roads in ArcCatalog. Given the database generated in the topology, construct the network by the function “New → Network Dataset.”

Lastly, calculate the shortest-path distance using the network analyst tools in ArcMap. Create the OD cost matrix by the function “Network Analyst Tools → Analysis → Make OD Cost Matrix Layer,” which requires input of the network dataset. Then we can derive the shortest-path distance by loading the points as origins and destinations. For example, we can obtain $N * N$ distances for N points.

To model the reduction in congestion cost by the new bridges, I manually increase the commuting distance by 1 km for each path going through the old bridge before operation of the two new ones. The 1 km increase is relatively small, given that the average travel distance is about 22 km.

Appendix C Calibration with Only Pre-Construction Information

This appendix describes the calibration of two structural parameters, ϵ and κ, using only the information prior to the construction of the bridges and tunnel. I identify the optimal values by solving the following objective function, which minimizes the sum of square differences between the observed and simulated population shares during the period 2003-2007 in Xiamen.

$$\min_{\{\epsilon,\kappa\}} \sum_i \left(\pi_i^{pre} - \hat{\pi}_i^{pre}\right)^2,$$

where π_i^{pre} and $\hat{\pi}_i^{pre}$ are the observed and simulated average population share in residential location i during the period 2003-2007. For $\hat{\pi}_i^{pre}$, I simulate the population share before the operation of the new infrastructure by removing the bridges and tunnel from the transportation network. I do not include weights in the objective function in order to focus on the overall simulation performance for the entire city before the construction of the infrastructure.

The estimated results are as follows: (1) $\hat{\epsilon}$=4.216 (baseline model with educational access), 3.011 (alternative model 1 with hospital), 2.305 (alternative model 2 with scenic spots); (2) $\hat{\kappa}$=0.435 (baseline model with educational access), 0.405 (alternative model 1 with hospital), -0.059 (alternative model 2 with scenic spots). Using the calibrated parameters, I compare the simulated results with the observed data, as shown in Appendix Figures A5 and A6.

Appendix D Discussion

This appendix discusses several issues related to the interpretations of the quantitative urban model developed in the paper. First, as agglomeration is an important force in spatial models, I address the benefits and concerns when introducing agglomeration forces. Second, I discuss the difference between population distribution and population size in the quantitative analysis.

D.1 Agglomeration

Following Ahlfeldt et al. (2015), one can introduce agglomeration forces to the model by assuming production and amenity externalities as follows:

$$A_j = a_j(\sum_{s=1}^{J} e^{\delta\tau_{js}}\pi_s)^{\lambda}, \qquad B_i = b_i(\sum_{r=1}^{I} e^{\theta\tau_{ir}}\pi_r)^{\eta}.$$

Productivity in work location j consists of two parts: the production fundamentals (a_j) and externalities induced by all work locations ($\sum_{s=1}^{J} e^{\delta\tau_{js}}\pi_s$), where π_s is the share of residents in s and $e^{\delta\tau_{js}}$ is the distance (τ_{js}) discounted factor. Parameter $\delta > 0$ governs the rate of spatial decay. Parameter λ controls for the relative importance of production externalities. Similarly, the amenity in residential location i is determined by the amenity fundamentals (b_i) and the externalities, which is the aggregation of distance-discounted population shares in all residential locations. The parameters θ and η govern the spatial decay and externality importance, respectively.

Introducing agglomeration forces brings benefits for model performance and generates better approximations to the observed data. However, there are some concerns. First, I have to give up the inputs of productivity and amenities and recover the unobserved fundamentals based on the mappings from calibrated parameters and observed vectors to the unobserved fundamentals. However, this extension does not help in understanding the factors that determine the distributions of productivity and amenity. On the contrary, the educational access in this paper delivers clear information about amenity distribution as well as policy implications at the expense of model performance.

Second, mapping from the endogenous variables and parameters to the fundamentals may not exist because of the different assumptions of the land market. Ahlfeldt et al. (2015) assume a negative relation between land prices and wages in the same location based on the Cobb-Douglas production and no-arbitrage condition between commercial and residential land, which simplifies calculation of the spatial equilibrium. This assumption applies to the case in which the study period is relatively long and land use regulations are less stringent to guarantee a flexible conversion between commercial and residential land use. Since the study period in this paper is relatively short and land use regulations are strict in Xiamen, I prefer not to use their land market assumption. Instead, I assume that the land price in location i is determined by the residents who live in i and work in all work locations. This is essentially a positive relation between land prices and wages, which is a common setting in the literature (Roback, 1982; Glaeser et al., 2006). This land price setting, together with the externalities, would introduce huge burden for model solution.

Considering the main focus of the paper—i.e., distinct effect of the new infrastructure and educational access—and the model complexity of the land price determination, I do not model agglomeration explicitly. However, the model predictions should still hold when introducing the agglomeration forces, because the amenity externalities should strengthen the gap of educational access between the city center and the tunnel-connected periphery.

D.2 Population Distribution versus Population Size

This paper focuses on the change in the population share rather than the population size in each location. A decrease in the population share does not necessarily imply a net outflow of residents, since the total population is increasing due to inflow of residents from other cities when reducing the commuting cost. However, the change in the population share can be regarded as population redistribution if we fix total city population and allow the expected utility to vary in response to the change in the transportation network. These two assumptions are identical in terms of the equilibrium population distribution—i.e., fixed expected utility and changing total population vs. changing expected utility and fixed total population. All results remain the same except the equilibrium equation (15), which will be changed as follows:

$$\mathbb{E}(u)^{\frac{1}{\beta}} = \frac{1}{\beta\bar{L}}\gamma^{\frac{1}{\beta}} \left\{ \sum_{r}^{I} \sum_{s}^{J} (B_r A_s d_{rs}^{-1})^{\epsilon} \left[H_r^{-1} B_r^{\epsilon} \sum_{k}^{J} A_k^{1+\epsilon} d_{rk}^{-\epsilon} \right]^{\frac{-\beta\epsilon}{1+\beta\epsilon}} \right\}^{\frac{1+\beta\epsilon}{\beta\epsilon}}.$$

In the latter assumption, a decrease in the population share is equivalent to a net outflow of residents. However, the increased expected utility due to the transportation improvement will attract residents to move in from other cities unless we restrict intercity migration. The setting of fixed total population and varied expected utility applies to a large spatial unit, such as nations or city clusters, and it is less suitable for analysis within a city.

Regarding the welfare implications of the two assumptions, when we fix expected utility and vary total population, we are unable to address welfare through the measure of utility. Instead, we use the change in total population as a proxy for welfare.

Appendix E Tables & Figures

TABLE A1
Industry Composition of Firms in Xiamen in 2008

Industry	Num. of Firms	Firm Share (%)	Total Employment	Employment Share (%)
Manufacturing	1,961	6.74	549,346	37.34
Retail	2,860	9.83	42,522	2.89
Finance	4,761	16.36	97,172	6.60
Wholesale	9,413	32.34	109,803	7.46
Others	10,111	34.74	672,471	45.71
Total	29,106	100	1,471,314	100

Notes: This table presents the industry composition of all firms in secondary and tertiary industries in Xiamen in 2008 based on data from the 2008 National Economic Census. Manufacturing sector includes 2-digit Chinese industry codes from 13 to 43; Retail refers to the industry with a 2-digit industry code of 65; Finance sector covers 2-digit industry codes from 68 to 74, including banking, securities, insurance, leasing, real estate, and other financial services; Wholesale refers to the industry with a 2-digit industry code of 63; Others refers to the rest of 2-digit industries.

TABLE A2
Distance to Infrastructures and Population Share in the City Center

Dependent Variable: Pop. Share (log)	(1)	(2)	(3)	(4)
	City Center			
log DistanceToBridges*Post2008	0.128	0.095	0.090	0.134
	(0.127)	(0.109)	(0.112)	(0.125)
log DistanceToTunnel*Post2010	-1.505***	-1.186***	-1.249***	-1.137***
	(0.155)	(0.150)	(0.125)	(0.144)
log DistanceToHighSpeedStation*Post2010		-1.127***		
		(0.272)		
log DistanceToAirport*Post2014			-0.465***	
			(0.065)	
log DistanceToCenterStation*Post2010				0.372***
				(0.069)
Location FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,276	1,276	1,276	1,276
R-squared	0.881	0.883	0.884	0.884

Notes: This table reports results with respect to the residential locations in the city center. Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE A3
Distance to Infrastructures and Population Size

Dependent Variable: Pop. Size (log)	(1) Full Sample	(2) Periphery (Jimei)	(3) Full Sample	(4) Periphery (Xiang'an)	(5) Full Sample
log DistanceToBridges*Post2008	-0.377***	-0.793***			-0.510***
	(0.074)	(0.185)			(0.082)
log DistanceToTunnel*Post2010			0.064	0.050	0.257***
			(0.068)	(0.039)	(0.061)
Location FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	5,852	1,188	5,852	1,298	5,852
R-squared	0.889	0.894	0.886	0.892	0.890

Notes: This table reports the robustness of the baseline results by replacing population share (log) with population size (log). Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE A4
Robustness: Excluding Sample After 2015

Dependent Variable: Pop. Share (log)	(1) Full Sample	(2) Periphery (Jimei)	(3) Full Sample	(4) Periphery (Xiang'an)	(5) Full Sample
log DistanceToBridges*Post2008	-0.356***	-0.790***			-0.440***
	(0.074)	(0.208)			(0.083)
log DistanceToTunnel*Post2010			0.024	0.026	0.191**
			(0.090)	(0.040)	(0.076)
Location FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	4,256	864	4,256	944	4,256
R-squared	0.897	0.890	0.894	0.907	0.897

Notes: This table reports the robustness of the baseline results by excluding sample after 2015, given that transport infrastructures that changed the road network were mainly constructed and operated after 2015. Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE A5
Robustness: Non-Parametric Regression

Dependent Variable: Pop. Share (log)	(1) Periphery (Jimei)	(2) Periphery (Xiang'an)
Treatment of (log) DistanceToBridges in 2008	-0.174***	
	(0.063)	
Treatment of (log) DistanceToTunnel in 2010		13.87
		(15.45)
Observations	864	944

Notes: This table reports the robustness of the baseline results using the non-parametric estimator proposed by de Chaisemartin et al. (2022). I define quasi-stayers as the locations that are more than 20 km from the respective transport infrastructure. Robust standard errors in parentheses are clustered at district-year level. Significance levels: *** 1%, ** 5% and * 10%.

TABLE A6
Model Fit: Comparing Simulated and Observed Data

	(1) log Pop Share (Data) 2003-2007	(2) log Pop Share (Data) 2011-2015	(3) Change in Pop Share (Data)
log Pop Share 2003-2007 (Model)	0.375***		
	(0.039)		
log Pop Share 2011-2015 (Model)		0.435***	
		(0.038)	
Change in Pop Share (Model)			0.374***
			(0.057)
Observations	266	266	266
R-squared	0.256	0.333	0.140

Notes: This table reports the correlation between observed and simulated data by estimating a simple linear regression model. The change in population share is normalized by its standard deviation. Standard errors in parentheses. Significance levels: *** 1%, ** 5% and * 10%.

TABLE A7
List of Top Schools in Xiamen

ID	School	Type	District	Longitude	Latitude
1	Haicang Affiliated School of Beijing Normal University	First list	Haicang	118.0296	24.45869
2	Science and Technology Middle School of Xiamen University	First list	Siming	118.1141	24.43257
3	Xiamen Second Experimental Primary School	First list	Siming	118.122	24.48532
4	Xiamen No.2 Middle School	First list	Siming	118.0627	24.44921
5	Xiamen No.6 Middle School	First list	Huli	118.0843	24.49772
6	Xiamen No.5 Middle School	First list	Siming	118.129	24.47914
7	Xiamen No.1 Middle School	First list	Siming	118.0928	24.46093
8	Xiamen No.1 Middle School Haicang Branch	First list	Haicang	117.9595	24.54046
9	Xiamen Haicang Experimental Middle School	Second list	Haicang	118.0553	24.50491
10	Xiamen Haicang Yankui Primary School	First list	Haicang	118.0323	24.49678
11	Xiamen Jimei Middle School	Second list	Jimei	118.0957	24.56939
12	Xiamen Experimental Primary School	First list	Siming	118.0835	24.45815
13	Xiamen Experimental Middle School	First list	Tong'an	118.1597	24.64838
14	Xiamen Bindong Primary School	First list	Siming	118.1037	24.47912
15	Xiamen Binlang Primary School	First list	Siming	118.1061	24.48371
16	Xiamen Caitang School	First list	Huli	118.154	24.48716
17	Xiamen Zengying Primary School	First list	Jimei	118.0419	24.55912
18	Xiamen Datong Primary School	First list	Siming	118.0721	24.46309
19	Xiamen No.9 Middle School	Second list	Siming	118.1042	24.46875
20	Xiamen No.3 Middle School	Second list	Huli	118.1164	24.51429
21	Xiamen Dongdu No. 2 Primary School	Second list	Huli	118.0871	24.49571
22	Dongshan Middle School, Xiamen city	Second list	Tong'an	118.1475	24.74432
23	Xiamen Haicang District Second Experimental Primary School	Second list	Haicang	117.9839	24.53475
24	Hongtang Primary School, Haicang District, Xiamen city	Second list	Haicang	117.9434	24.57169
25	Tianxin Island Primary School, Haicang District, Xiamen city	Second list	Haicang	118.0436	24.48943
26	Xiamen Huli Second Experimental Primary School	Second list	Huli	118.1357	24.50577
27	Xiamen Huli District Teacher advanced Education School Affiliated Primary School	Second list	Huli	118.1511	24.5224
28	Xiamen Huli Experimental Primary School	First list	Huli	118.1379	24.51311
29	Xiamen Huli Experimental Middle School	First list	Huli	118.1413	24.51836
30	Huli Middle School, Xiamen city	Second list	Huli	118.0932	24.50566
31	Xiamen Jimei No.2 Primary School	Second list	Jimei	118.0978	24.56813
32	Guankou Central Primary School, Jimei District, Xiamen city	Second list	Jimei	117.9811	24.61593
33	Guankou Middle School, Jimei District, Xiamen city	Second list	Jimei	117.9967	24.6079
34	Xiamen Jimei District Le 'an Middle School	Second list	Jimei	118.0969	24.58935
35	Lehai Primary School, Jimei District, Xiamen city	First list	Jimei	118.1135	24.59803
36	Lin Primary School in Jimei District, Xiamen city	Second list	Jimei	118.039	24.58169
37	Ningbao Primary School, Jimei District, Xiamen city	First list	Jimei	118.0548	24.5683
38	Qiaoying Primary School, Jimei District, Xiamen city	First list	Jimei	118.102	24.60071

Continued on next page

Table continued

ID	School	Type	District	Longitude	Latitude
39	Shangtang Middle School, Jimei District, Xiamen city	Second list	Jimei	117.9835	24.61458
40	Xingdong Primary School, Jimei District, Xiamen city	Second list	Jimei	118.0588	24.56842
41	Xiamen Jimei Primary School	First list	Jimei	118.0946	24.57193
42	Xiamen City Jiangtou Central Primary School	Second list	Huli	118.1293	24.49812
43	Xiamen Jin'an Primary School	Second list	Huli	118.181	24.50767
44	Xiamen Jinshang Primary School	Second list	Huli	118.1439	24.49445
45	Xiamen Kangle Second Primary School	Second list	Huli	118.1143	24.5094
46	Xiamen Kangle Primary School	First list	Huli	118.1084	24.50728
47	Xiamen Lotus Middle School	First list	Siming	118.1242	24.48682
48	Xiamen Neicuo Middle School	Second list	Xiang'an	118.27	24.66608
49	Xiamen Enlightenment Middle School	Second list	Tong'an	118.1506	24.71235
50	Qianpu North District Primary School, Xiamen city	Second list	Siming	118.1645	24.4791
51	Xiamen Qianpu South District primary school	First list	Siming	118.1665	24.46824
52	Xiamen Qunhui Primary School	Second list	Siming	118.0798	24.4548
53	Xiamen People's Primary School	Second list	Siming	118.1171	24.47428
54	Xiamen Experimental Primary School Jimei Branch	First list	Jimei	117.9907	24.60607
55	Lianqian Primary School, Siming District, Xiamen city	First list	Siming	118.153	24.48185
56	Xiamen Siming Primary School	First list	Siming	118.0827	24.445
57	Xiamen Songbai Primary School	First list	Siming	118.1135	24.49379
58	Xiamen Songbai Middle School	Second list	Siming	118.1203	24.49825
59	Datong Central Primary School, Tong'an District, Xiamen city	Second list	Tong'an	118.1406	24.73362
60	Xiamen Tongan District second Experimental Primary School	Second list	Tong'an	118.1499	24.72436
61	Xiamen Tongan District First Experimental Primary School	First list	Tong'an	118.1523	24.73442
62	Xiamen Tong'an District Teacher advanced Education School affiliated primary school	Second list	Tong'an	118.1442	24.74305
63	Xiangping Central Primary School, Tong'an District, Xiamen city	Second list	Tong'an	118.1371	24.7237
64	Yang Zhai Primary School, Tong'an District, Xiamen city	First list	Tong'an	118.1447	24.70897
65	Xiamen Wushipu Primary School	Second list	Huli	118.1234	24.50753
66	Xiamen Wucun Primary School	First list	Siming	118.1076	24.46986
67	Xiamen Woulu School	Second list	Tong'an	118.11	24.69837
68	Xiamen Xiang'an No. 1 Middle School	First list	Xiang'an	118.2428	24.67143
69	Xiamen Xiang'an District second Experimental Primary school	Second list	Xiang'an	118.238	24.6648
70	Xiangan District, Xiamen City, the first experimental Primary school	First list	Xiang'an	118.2393	24.61089
71	Ma Xiang Central Primary School, Xiang'an District, Xiamen city	Second list	Xiang'an	118.2575	24.65795
72	Xiangan District Experimental School, Xiamen City	First list	Xiang'an	118.2389	24.62697
73	Xiamen Xingnan Middle School	Second list	Jimei	118.0346	24.56389
74	Xiamen Yanwu Primary School	Second list	Siming	118.0871	24.44131
75	Xiamen Double Tenth Middle School	First list	Huli	118.1501	24.51996
76	Haicang Affiliated School of Xiamen Double 10 Middle School	Second list	Haicang	118.0481	24.49672

Continued on next page

Table continued

ID	School	Type	District	Longitude	Latitude
77	Xiamen Tongan No.1 Middle School	Second list	Tong'an	118.1517	24.74119
78	Xiamen Foreign Language School	First list	Siming	118.0874	24.48018
79	Affiliated primary school of Xiamen Foreign Language School	First list	Siming	118.0968	24.47587
80	Xiamen Foreign Language School Huli Branch	First list	Huli	118.0974	24.53064
81	Xiamen Wuyuan Second Experimental School	Second list	Huli	118.1554	24.52427
82	Xiamen Wuyuan Experimental School	First list	Huli	118.1571	24.54199

Notes: See Section 2.3 for the definition and source of the list of top schools in Xiamen. The first and second lists of school types refer to the first and second lists of demonstration schools for compulsory education reform in Fujian Province published by the local governments in 2018.

TABLE A8
List of Top-Tier Public Hospitals in Xiamen

ID	Hospital	Rank	District	Longitude	Latitude
1	Xiamen Second Hospital	Tier 3 Grade A	Jimei	118.0994	24.58689
2	Xiamen Third Hospital	Tier 3 Grade B	Tong'an	118.1422	24.70886
3	Xiamen University Affiliated Zhongshan Hospital	Tier 3 Grade A	Siming	118.0927	24.47419
4	Xiamen University Affiliated First Hospital	Tier 3 Grade A	Siming	118.0832	24.45387
5	Xiamen Fifth Hospital	Tier 3 Grade B	Xiang'an	118.2436	24.66262
6	Xiamen University Affiliated Success Hospital	Tier 3 Grade A	Siming	118.09	24.46084
7	Xiamen Hospital of Traditional Chinese Medicine	Tier 3 Grade A	Huli	118.136	24.50334
8	Xiamen Xian Yue Hospital	Tier 3 Grade A	Siming	118.1065	24.49802
9	Xiamen Maternal and Child Health Hospital	Tier 3 Grade A	Siming	118.0767	24.45293
10	Stomatological Hospital of Xiamen Medical College (CAI Tang General Hospital)	Tier 3 Grade C	Huli	118.1604	24.48917
11	Stomatological Hospital of Xiamen Medical College (Douxi Branch)	Tier 3 Grade C	Siming	118.0838	24.46286
12	Xiamen University Affiliated Cardiovascular Hospital	Tier 3 Grade A	Huli	118.1679	24.51108

Notes: See Section 2.3 for the definition and source of the list of top-tier public hospitals in Xiamen.

TABLE A9
List of Top-Rated Scenic Spots in Xiamen

ID	Scene	Rank	District	Longitude	Latitude
1	Gulangyu Scenic Area	5A	Siming	118.0621	24.44736
2	Garden Botanical Garden	4A	Siming	118.105	24.45067
3	Garden Expo Garden	4A	Jimei	118.0712	24.57544
4	Hulishan Fort	4A	Siming	118.1012	24.43217
5	Jimei Aoyuan	4A	Jimei	118.1024	24.57118
6	Sun Moon Valley Hot Spring Theme Park	4A	Haicang	117.9362	24.56193
7	Tianzhu Mountain Forest Park	4A	Haicang	117.9219	24.59335
8	Beichen Mountain Scenic Area	4A	Tong'an	118.2467	24.80428
9	Xiamen Fangte Tourist Area	4A	Tong'an	118.1721	24.68482
10	Chengyi Science and Technology Exploration Center	4A	Jimei	118.0527	24.59604
11	Xiamen Old Courtyard Scenic Area	4A	Jimei	118.071	24.62629

Notes: See Section 2.3 for the definition and source of the list of top-rated scenic spots in Xiamen.

TABLE A10
Comparison of Land Supply Data

year	obs.	land leasing revenue (bn yuan)			land leasing area ('000 ha)			land allocation area ('000 ha)		
		transaction	official	(%)	transaction	official	(%)	transaction	official	(%)
2000	1,146	0.48	59.56	0.80	1.37	48.63	2.82	1.46	80.57	1.81
2001	2,353	1.57	129.59	1.21	4.25	90.39	4.70	8.58	73.98	11.59
2002	13,389	23.86	241.68	9.87	19.37	124.23	15.59	8.41	88.05	9.56
2003	18,173	32.05	542.13	5.91	29.12	193.60	15.04	10.49	65.26	16.08
2004	37,578	113.28	641.22	17.67	50.98	181.51	28.09	19.37	62.05	31.22
2005	28,857	129.34	588.38	21.98	41.38	165.59	24.99	18.66	64.62	28.88
2006	38,373	109.93	807.76	13.61	73.40	233.02	31.50	10.45	63.79	16.38
2007	130,276	921.78	1,221.67	75.45	219.24	234.96	93.31	88.03	76.09	115.69
2008	108,519	853.94	1,025.98	83.23	190.84	165.86	115.06	97.34	62.38	156.04
2009	137,885	1,493.57	1,717.95	86.94	223.60	220.81	101.26	158.69	122.29	129.77
2010	175,494	2,758.77	2,746.45	100.45	314.89	293.72	107.21	188.05	138.27	136.01
2011	199,543	3,059.54	3,212.61	95.24	353.97	335.09	105.64	303.98	257.21	118.19
2012	188,853	2,674.65	2,690.00	99.43	330.24	332.43	99.34	352.16	377.13	93.38
2013	220,724	4,109.70	4,164.90	98.67	373.38	374.80	99.62	363.38	373.28	97.35
2014	174,321	3,086.19	3,437.74	89.77	267.66	277.35	96.51	311.46	369.83	84.22

Source: China Land Market website, Ministry of Land and Resources of China.
Notes: Observations include land leasing and land allocation. I drop 405 observations that have transaction values of more than 5 billion yuan per piece of land, or an area of more than 1,000 hectares, or per hectare land price of more than 500 thousand yuan per square meter (less than the highest record in 2015). I also drop 119,016 duplicated observations (most are before 2008).

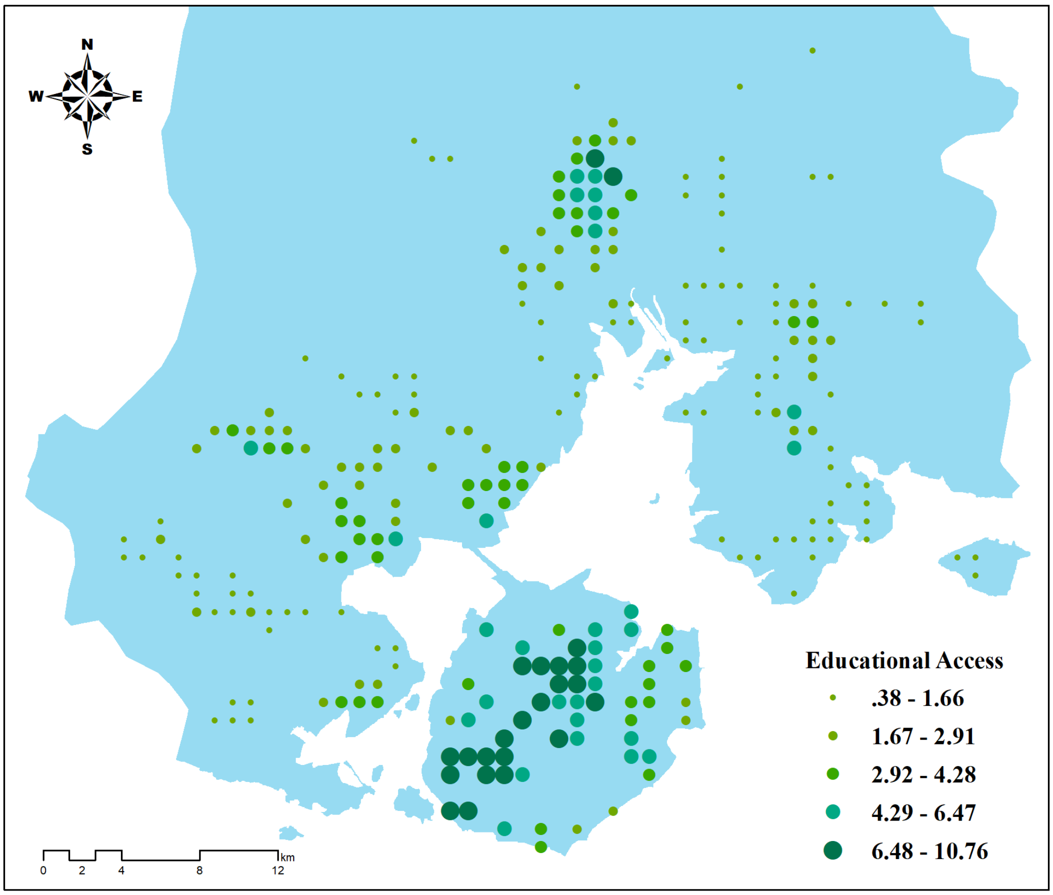


FIGURE A1
Spatial Distribution of Educational Access

Notes: This figure shows the spatial distribution of educational access of residential locations in Xiamen. The calculation of educational access is given in Section 4.2.

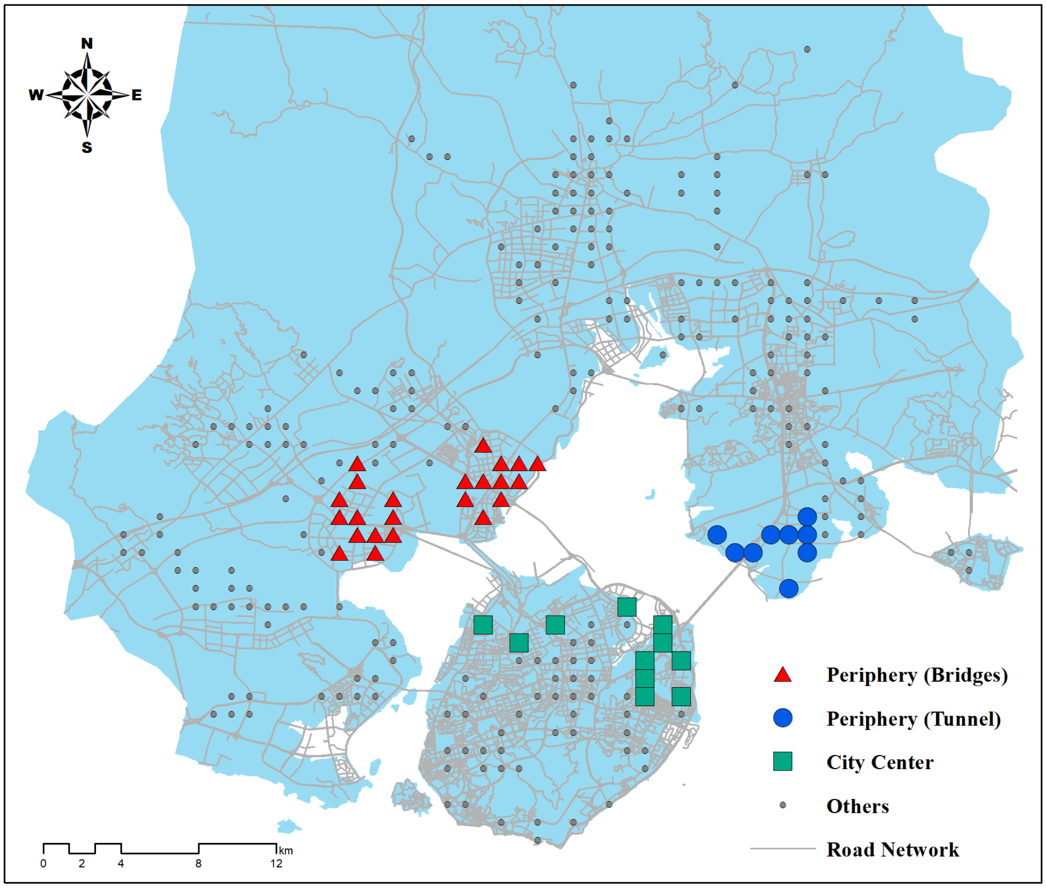


FIGURE A2
Target Locations

Notes: This figure shows the spatial distribution of target locations which are defined as the residential locations within 5 km road distance from the nearest end of either the bridges or tunnel. Triangles indicate target locations on the bridges-connected periphery; Circles refer to target locations on the tunnel-connected periphery; Squares suggest target locations in the city center.

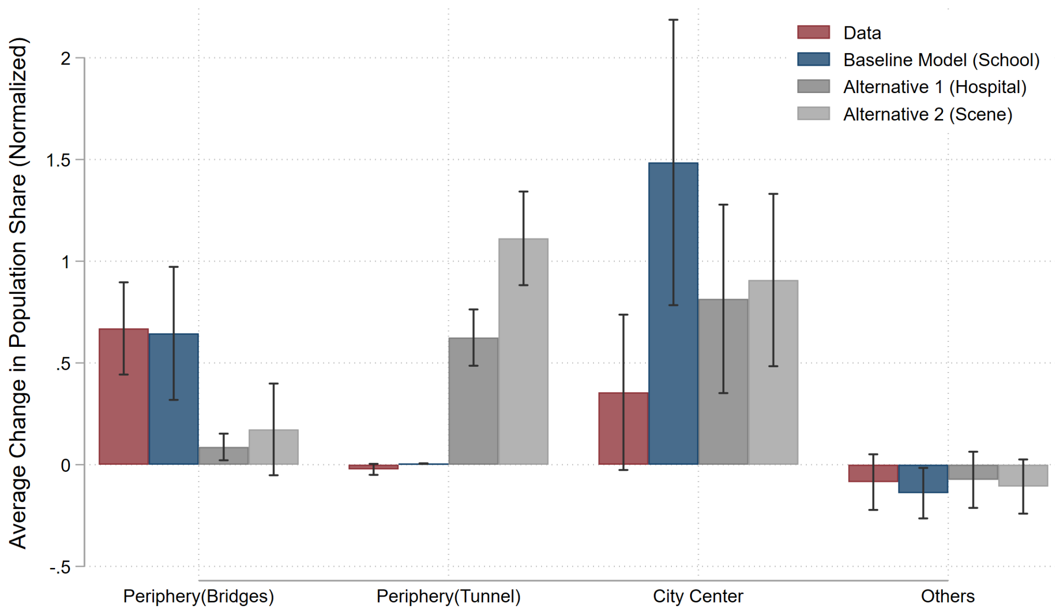


FIGURE A3
Model Fit: Average Change in Population Share in All Regions

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) in different regions in Xiamen based on the observed data, the baseline model with educational access, and two alternative models with access to the top-tier public hospitals and top-rated scenic spots, respectively. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. City Center refers to the residential locations on the island that are within 5 km road distance from either the bridges or tunnel. Others refers to the rest of the residential locations. The confidence interval is at the 95% level.

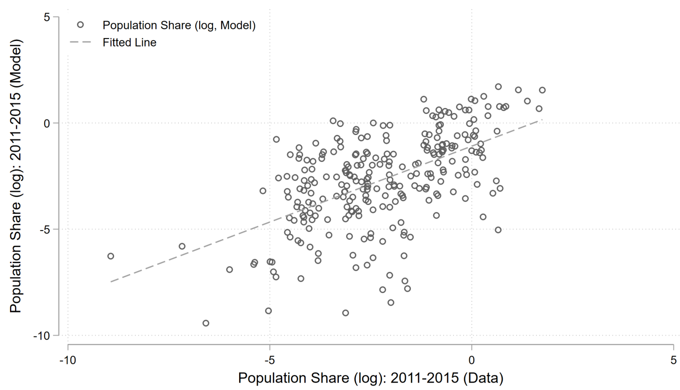
(A) Average Land Price (log) before 2010

(B) Average Land Price (log) after 2010

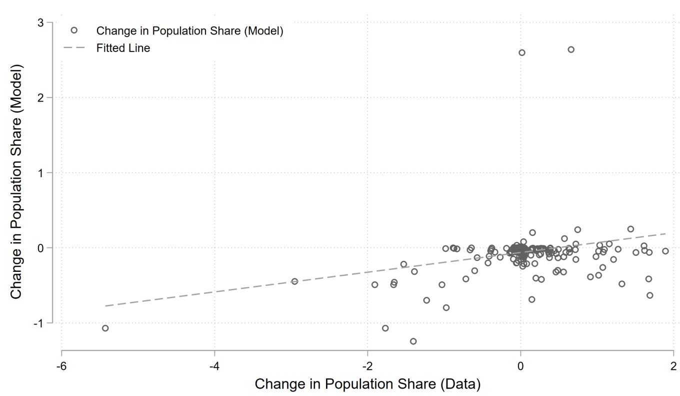
FIGURE A4
Model Fit: Observed vs. Simulated Land Price

Notes: This figure plots the correlation between the observed and simulated land price in Xiamen both before and after operation of the bridges and tunnel. The location-level land price is the average price of land that was within 1 km road distance from the corresponding residential location. If no land transaction is found within 1 km, then land price in that location is missing. Since land price data are reliable only after 2007, as explained in Section 2.3, I use the average land price in 2007-2010 as the proxy for land price before operation of the new infrastructures. The simulation is based on the baseline model using educational access as the proxy for location amenity. I set $\beta = 0.26$, $\epsilon = 9.582$, $\kappa = 0.025$.

(A) Average Population Share during 2003-2007



(B) Average Population Share during 2011-2015



(C) Change in Population Share

FIGURE A5
Robustness: Calibration with Only Pre-Construction Information

Notes: This figure shows the robustness of model fit when using only pre-construction information in calibration. It plots the correlation between the observed and simulated population share in Xiamen. The simulation is based on the baseline model using educational access as the proxy for amenity. The calibrated parameters are $\epsilon = 4.216$ and $\kappa = 0.435$. The change in population share is normalized by its standard deviation.

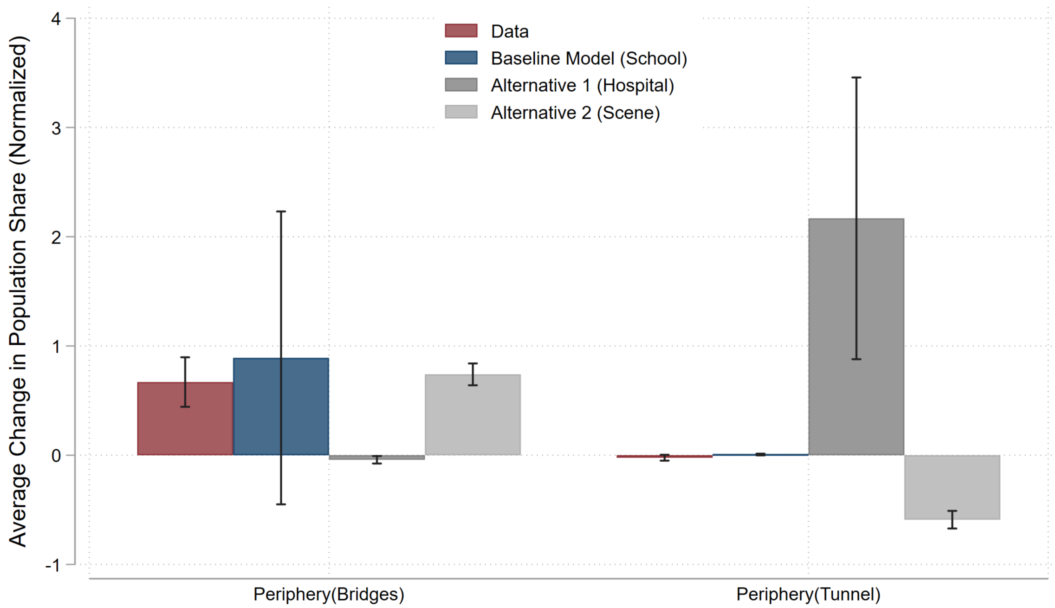


FIGURE A6

Robustness: Using Only Pre-Construction Information in Calibration

Notes: This figure shows the robustness of model fit when using only pre-construction information in calibration. It compares the average change in population share (normalized by the standard deviation of the changes) on the bridges- and tunnel-connected peripheral areas based on the observed data, the baseline model with educational access, and two alternative models with access to top-tier public hospitals and top-rated scenic spots, respectively. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The confidence interval is at the 95% level.

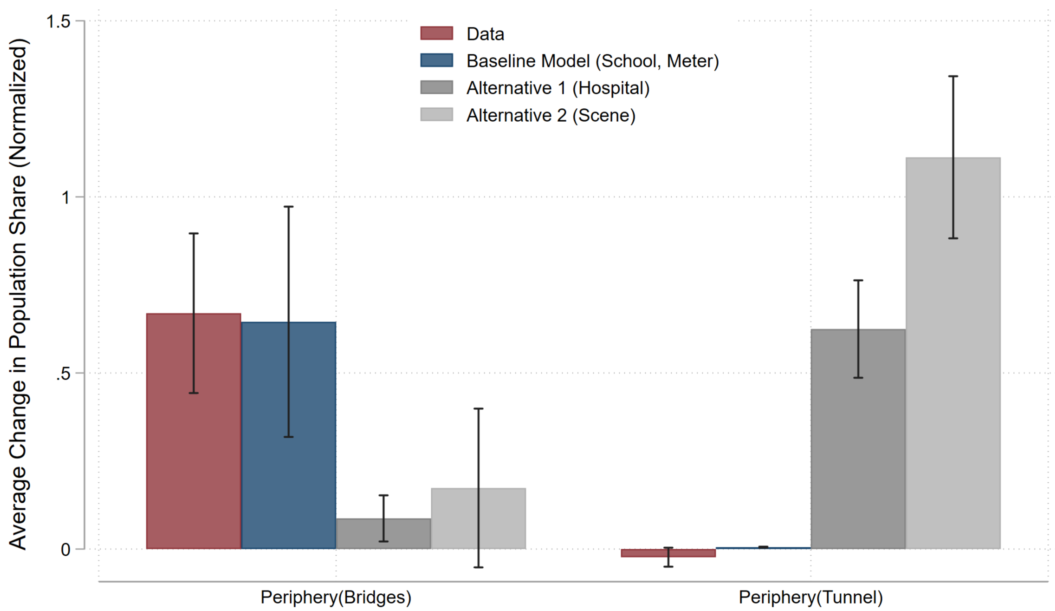


FIGURE A7
Robustness: Using Meters Instead of Kilometers in the Measure of Educational Access

Notes: This figure shows the robustness of model fit when using meters instead of kilometers in the measure of educational access. It compares the average change in population share (normalized by the standard deviation of the changes) on the bridges- and tunnel-connected peripheral areas based on the observed data, the baseline model with educational access, and two alternative models with access to top-tier public hospitals and top-rated scenic spots, respectively. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The confidence interval is at the 95% level.

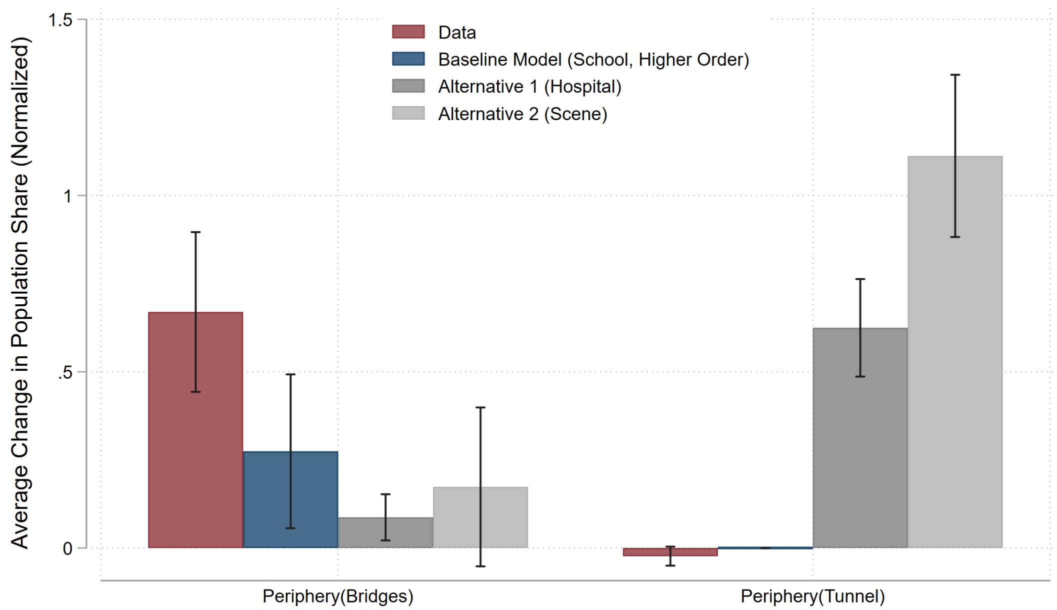


FIGURE A8
Robustness: Using Higher Order of Spatial Friction in the Measure of Educational Access

Notes: This figure shows the robustness of model fit when using second order of spatial friction in the measure of educational access. It compares the average change in population share (normalized by the standard deviation of the changes) on the bridges- and tunnel-connected peripheral areas based on the observed data, the baseline model with educational access, and two alternative models with access to top-tier public hospitals and top-rated scenic spots, respectively. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The calibrated parameters for the model with educational access are $\epsilon = 11.833$ and $\kappa = 0.042$. The confidence interval is at the 95% level.

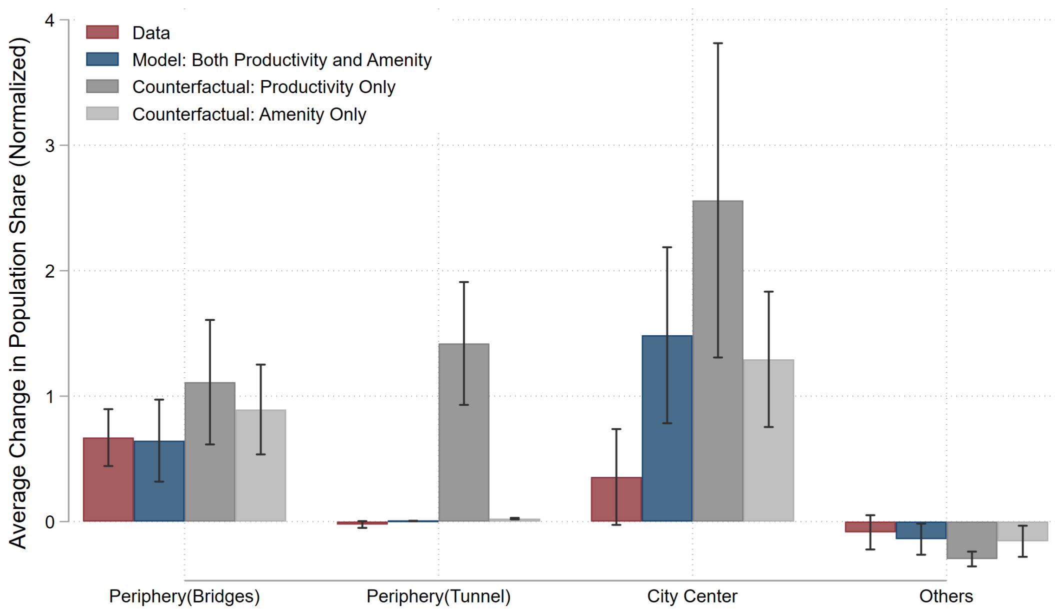


FIGURE A9
Model Decomposition: Productivity vs. Educational Access in All Regions

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) in different regions in Xiamen based on the observed data, the baseline model with educational access, and two counterfactual models with actual productivity and equal amenity for the first counterfactual (productivity only) and actual educational access and equal productivity for the second (amenity only). Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. City Center refers to the residential locations on the island that are within 5 km road distance from either the bridges or tunnel. Others refers to the rest of the residential locations. The confidence interval is at the 95% level.

FIGURE A10
Location-Level Wage Computed from Economic Census vs. ASIE Data

Notes: This figure plots the correlation between the location-level average wage computed from the Economic Census data and that from the Annual Survey of Industrial Enterprises (ASIE) data. The location-level wage is the employment-weighted average of the wages of firms within 1 km from the work location. If no firm is found within 1 km, I match with the closest firm's wage.

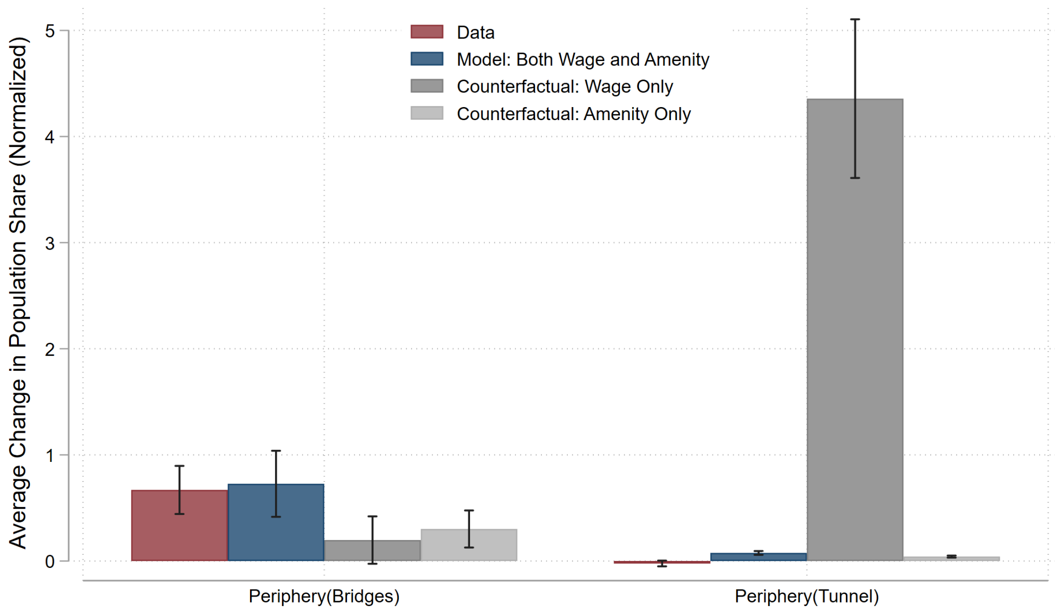


FIGURE A11
Model Decomposition: Wage vs. Educational Access using Economic Census data

Notes: This figure shows the simulated results using wage per worker from the Economic Census data as an alternative measure of firm productivity. I compare the baseline results using both wage and educational access with results based on two counterfactual models: one with actual wage and equal amenity (wage only) and the other with actual educational access and equal wage (amenity only). Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance from the bridges and tunnel, respectively. The average change in population share is normalized by the standard deviation of the changes. The confidence interval is at the 95% level. I set $\beta = 0.26$, $\epsilon = 6.901$, $\kappa = 0.024$.

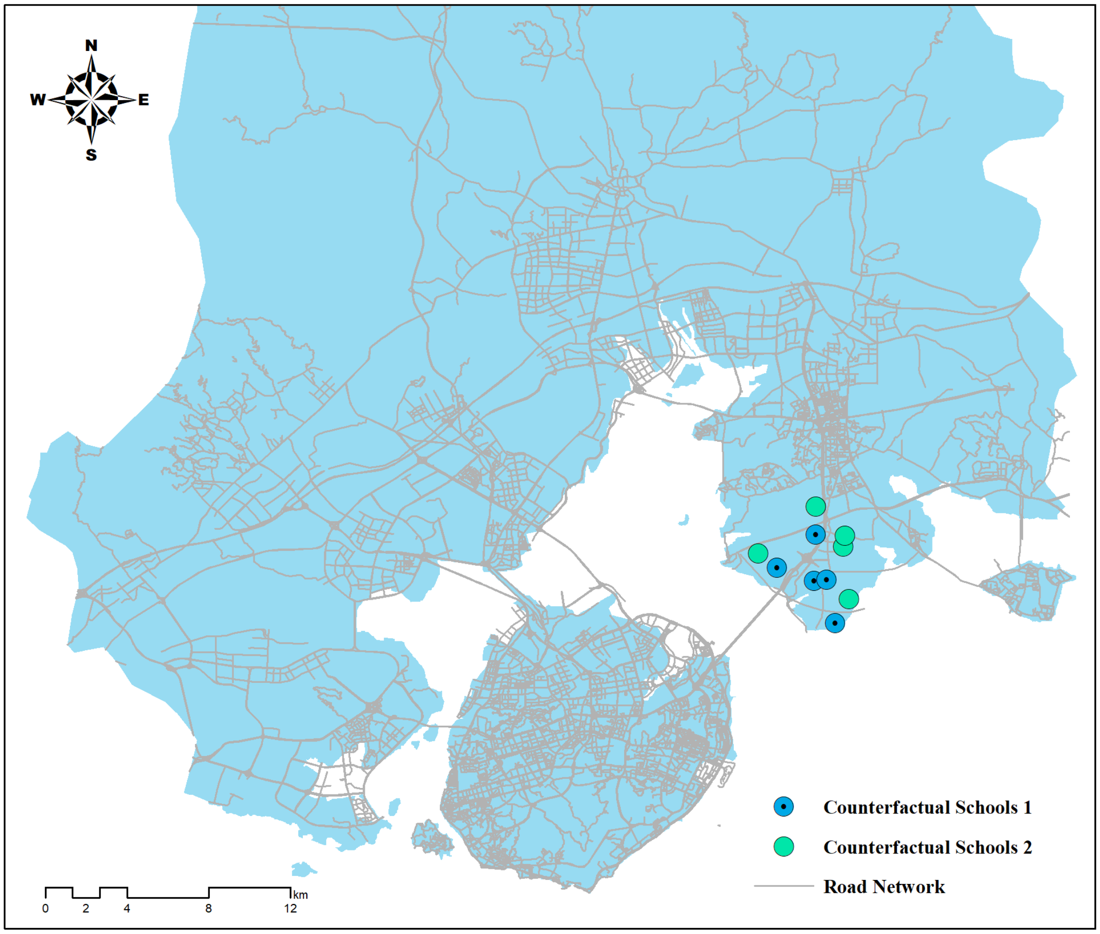


FIGURE A12
Locations of Counterfactual Top Schools

Notes: This figure shows the spatial distribution of counterfactual top schools on the tunnel-connected periphery. Counterfactual school 1 refers to the 5 existing schools that are the closest to the peripheral end of the tunnel and not on the list of top schools. These 5 schools are the top schools added in the first counterfactual exercise in which I upgrade 5 schools. Counterfactual school 2 refers to the other 5 existing schools that are close to the peripheral end of the tunnel and not on the list of top schools. The 10 schools in both categories are the top schools added in the second counterfactual exercise, in which I upgrade 10 schools.

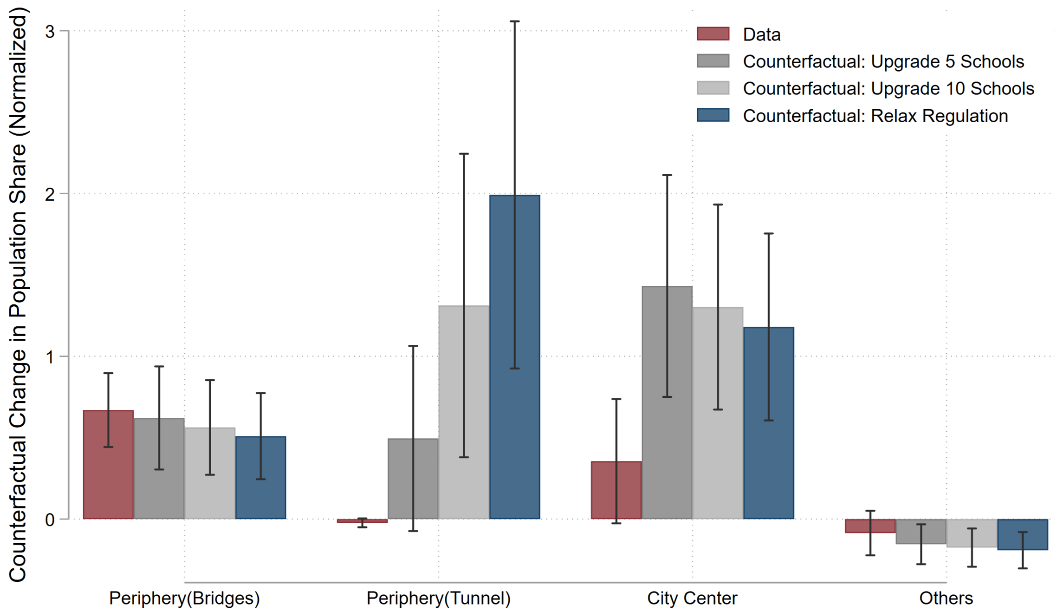


FIGURE A13
Counterfactual Exercises: All Regions

Notes: This figure compares the average change in population share (normalized by the standard deviation of the changes) in different regions in Xiamen based on the observed data and three counterfactual models. The first and second counterfactual exercises choose 5 and 10 existing schools, respectively, that are the closest to the peripheral end of the tunnel and upgrade them as top schools. The third counterfactual exercise allows students to choose from all 82 top schools in the city regardless of where they live. All counterfactual assumptions are made both before and after the operation of the bridges and tunnel. Periphery (Bridges) and Periphery (Tunnel) refer to residential locations on the periphery that are within 5 km road distance to the bridges and tunnel, respectively. City Center refers to the residential locations on the island that are within 5 km road distance to either the bridges or tunnel. Others refers to the rest of the residential locations. The confidence interval is at the 95% level.