The Economics of Traveling: A Study of High-Speed-Railway Expansion

Zijian He

Yale

March 2023

Click here for latest version

Abstract

Human mobility across cities is a crucial factor for regional development, yet its effects and mechanism remain poorly understood due to limited data on travel flows and their mixed purposes. This paper addresses this gap by using high-frequency GPS data from mobile users in China to measure travel flows across cities and to disentangle business and leisure activities, which have different implications for regional structural change. Furthermore, I develop a multi-sector economic geography model with both trade in goods and travel in humans to examine the welfare effects of China's recent expansion of the High-Speed-Railway system (HSR) since HSR only reduces people's travel costs, not goods' trade costs. The model suggests travel flows and their sensitivity to spatial functions are sufficient to evaluate welfare. I estimate the human mobility aspect of infrastructural improvement contributes to 0.3 percent of GDP. I also find significant inequality in the gains of HSR across cities and industries. These results have important implications for the ongoing policy debate on infrastructure improvement.

Keywords: transportation, infrastructure, high speed railway, travel

JEL codes: R2, R3, R41

1. Introduction

The movement of physical goods in the space has been well-studied in the trade literature (e.g., Eaton and Kortum (2002); Allen and Arkolakis (2014)). Meanwhile, people travel frequently: in 2019, the U.S. had 80 million international visitors and 2.3 billion domestic person trips.[1] Recently, infrastructure facilitating human mobility over longer distances has recently raised policymakers' attention. However, the economic implications of traveling are less clear. This paper seeks to quantify the significance of human mobility across cities on the spatial organization of economic activities. Specifically, it aims to answer questions such as how travel time influences individuals' travel decisions, how travelers alter the industrial structure of cities, and what the welfare implications are of infrastructure improvement via the promotion of travel rather than the facilitation of trade in goods.

To answer those questions, I first bring high-frequency GPS data to provide new measurements of travel patterns and their activities. For this study, travel is defined as a temporal visit to a location distant from one's primary residence. This definition distinguishes travel from other forms of mobility, such as commuting and migration. Commuting, for instance, typically occurs within city limits, with individuals traveling to and from their workplaces regularly. Migration, on the other hand, involves a permanent change of residence and is a far less frequent occurrence than travel. I first show people travel often, and travel intensity significantly decays with spatial frictions. The GPS data also infer individuals' business or leisure activities based on the location of GPS-enabled buildings, thereby can provide a high-frequency measurement of city-level business or leisure activity intensity. The variations observed in these measurements between workdays and holidays, as well as the cross-sectional variations across cities, can be used to infer comparative advantages at the city level.

The second innovation of this paper is introducing a novel method for decom-

[1] www.ustravel.org

posing bilateral travel flows into business and leisure purposes using high-frequency data. This decomposition is meaningful since the purposes of travel have different implications for the economic activities in their destinations: Leisure travelers are consumers, and their activities are primarily related to the consumer service sectors, while business travelers are workers, and their actions are primarily connected to the production and sales of the goods sectors. However, as with many trade data, we often only observe the total bilateral flows without knowing their specific purposes[2]. In this regard, the high-frequency GPS data on bilateral flows, coupled with city-level measurements of leisure/business intensity, become extremely valuable for the decomposition process. The key insight gained from this decomposition is that during holidays, as leisure intensity increases within the cities, the method attributes more bilateral travel flows to leisure purposes and vice versa.

To quantify the impact of human mobility on the spatial organization of economic activities, this study focuses on China's recent expansion of the High-Speed Railway (HSR) network. The rationale for selecting HSR is that it substantially lowers the cost of travel for individuals while keeping the cost of trade in physical goods unchanged. To make progress, I develop a multi-region multi-sector economic geography model that incorporates both trades in goods and human travel. This model enables the inference of trade in goods from business travel, thereby overcoming the challenge of data unavailability on trade in goods in China. Second, the model also shows welfare can be sufficiently summarized by travel flows and two key elasticities on the response of leisure/ business travel flows to travel cost changes. I estimate those elasticities using decomposed travel flows inferred from previous steps. In the counterfactual analysis, I replace HSR trains with traditional trains. The model suggests that the replacement of HSR trains, on average,

[2]For example, whether the goods being traded are for final consumption or intermediate usage has to rely on some assumptions. See Koopman et al. (2014); Wang et al. (2023) for recent development on this issue.

decreases city-level welfare by 0.3%, comparable to the construction cost of HSR, 0.25%. I also find significant inequality in the gains of HSR across cities and industries.

This paper is related to the literature on infrastructural evaluations. There are a bunch of papers evaluating the welfare effect of inter-city infrastructural improvement on reducing goods shipment costs, such as Allen and Arkolakis (2014, 2022), Donaldson and Hornbeck (2016), Faber (2014), Duranton and Turner (2012), e.t.c. This paper highlights another "missing" benefit of these infrastructures: increasing travel through reducing travel time.

There is also literature related to the papers evaluating the welfare effect of intra-city infrastructural on reducing the commuting cost, such as Ahlfeldt et al. (2015), Miyauchi et al. (2021) and Tsivanidis (2019). I focus on traveling rather than commuting: travel is associated with less frequency, longer distances, and potentially different purposes. Morten and Oliveira (2018) finds roads also have effects on migration. Migration, to another extreme, are long-run effect, and they are just a tiny proportion of the travel flows, so I do not study migration in this paper.

This research is also related to some reduced-form evidence on the effect of high-speed railway.Bernard et al. (2019) and Xu (2018) find that HSR boosts firm performance by reducing the search cost. Lin (2017) find empirical evidence that access to HSR, on average, increases railway ridership and employment. My research features modeling traveling and offers a structural view of those findings. In particular, I will show how different sectors benefit from reduced travel time.

Lastly, this paper relates to the literature on spatial distribution service sectors. Faber and Gaubert (2019), Eckert (2019), Fan et al. (2021) build model featuring service sectors in economic geography. However, they do not look directly into travel activities or focus on identifying the source of spatial friction for service sectors.

The remainder of this paper is organized in the following way. Section 2 introduces data and some motivating reduced-form results for the structural model. Section 3 presents the multi-region multi-sector economic geography model and

discusses sufficient welfare statistics. Section 4 shows details of the structural estimation strategy, including how to decompose bilateral travel flows into different purposes and how to infer trade in goods using business travel flows. Section 5 conducts the counterfactual analysis to quantify the welfare distribution of HSR expansion. Section 6 concludes.

2. Data and Motivating Facts

The data I used for this paper consists of four parts: trade costs, travel flows, HSR networks, and city-level economic statistics. I look at the city at the prefecture level. China has 293 cities, so the notion city here is comparable to the U.S. Metropolitan Statistical Areas (392) and is larger than the U.S. commuting zones (709).

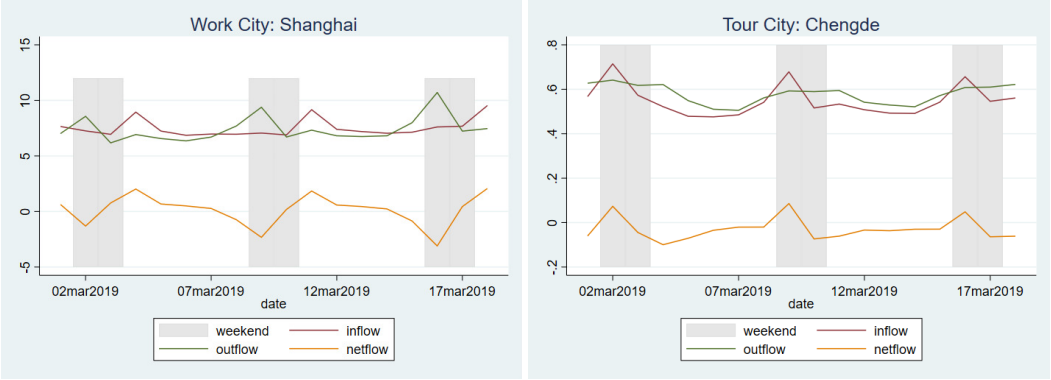
2.1 Travel Flows and City Activity Intensity

The high-frequency travel flows data comes from the Baidu location-based services (LBS). The data I used for this paper is from January 17, 2019, to March 31, 2021. They define travel as changing users' GPS location within 8 hours. Baidu aggregates daily bilateral travel flows among all pairs of cities. Since Baidu has large numbers of users based in China, the travel flows are reasonably representative. See Appendix A for a detailed discussion.

Differentiating between business and leisure travelers is crucial to understand the link between travel patterns and economic activities. While aggregate GPS data cannot distinguish between travel purposes, exploiting time variation can help to detect a city's comparative advantages regarding business or leisure services. I make sense of this approach in Figure 1 by showing that the cyclicality of travel patterns and the direction of travel flows on weekdays and weekends are critical indicators. For instance, a city like Shanghai, which is a typical business destination, experiences inflows during weekdays and outflows during weekends, while a city

like Chengde, which is a typical leisure destination, experiences outflows during weekends and inflows during weekdays. By utilizing City Activity Intensity data, this approach becomes quantitatively feasible, as I will demonstrate in the next step.

Figure 1: Travel patterns for two typical cities



Note: The comparison of travel flows between two typical cities in China

The high-frequency City Activity Intensity data utilized in this study is also obtained from Baidu LBS. Baidu leverages GPS location to infer individuals' business or leisure activities based on the types of locations where the user's GPS is located. A heatmap of GPS pinpoints within all locations provides a high-frequency measurement of city-level business or leisure activity intensity[3]. The variations observed in these measurements between workdays and holidays, combined with bilateral travel flows, can be utilized to decompose the travel purposes and infer the comparative advantages of the cities. Further discussion of this procedure can be found in Section 4.2.

[3]See Figure A.5 for a concrete example.

2.2 HSR Networks

I extracted the expansion of HSR Networks from the historical schedule table available on the official HSR website[4]. It is important to note some unique features of this context relevant to my study. First, the HSR networks have aggressively expanded since 2008 (Figure **??**). As of 2021, over three-quarters of cities are connected to these networks. Given that HSR significantly reduces travel time between cities, this expansion represents a massive change in spatial frictions.

Secondly, it is worth noting that HSR is exclusively used for passenger transportation, not for shipping goods, as shown in Figure A.6. Although some HSR trains have been used for shipping goods, HSR delivery has never played a significant role in the shipment market.

Thirdly, there might be concerns that HSR facilitates commuting, which is not the primary focus of this paper. According to Ollivier et al. (2014), while HSR commuting does exist, it is still a minor part. The majority of HSR travelers are engaged in non-commute business and leisure activities.

To provide direct evidence that HSR expansion facilitates human mobility, I look at the 31 cities that newly connect to HSR networks between 2020 and 2021. Using the travel flow data. I run the following regression:

$$\ln Gross_{i,t} = \beta Connect_{i,t} + D_t + D_i + \epsilon_{i,t},$$

where $Gross_{i,t}$ is the averaged gross travel flows of the city i in year t, $Connect$ is an indicator of whether city i is connected to HSR in year t. $D.$ are the city, year fixed effects. Table **??** presents the estimation results. I find HSR connection is associated with a 13% increase in gross travel flows.

[4]See https://www.12306.cn/index/.

2.3 Travel Costs and Transportation Modes Choices

The travel cost data utilized in this study is obtained from the Baidu Map Application programming interface (Baidu API). The route-planning technology incorporated in Baidu API generates estimates for the time and cost of a trip between any two cities in China based on the transportation mode users select, i.e., car, train, or air[5]. Compared to Geographic Information System (GIS), this route-planning method has two advantages. Firstly, it accounts for all transfer times, leading to a more accurate travel time measurement. Secondly, it provides a comprehensive overview of all available transportation alternatives, which better aligns with the data. In most GIS approaches, researchers compute the shortest travel time path, assuming perfect substitution between transportation modes, which contradicts the data. In this study, I estimate the substitution between HSR and existing transportation modes and highlight how the current infrastructure affects the benefits of HSR.

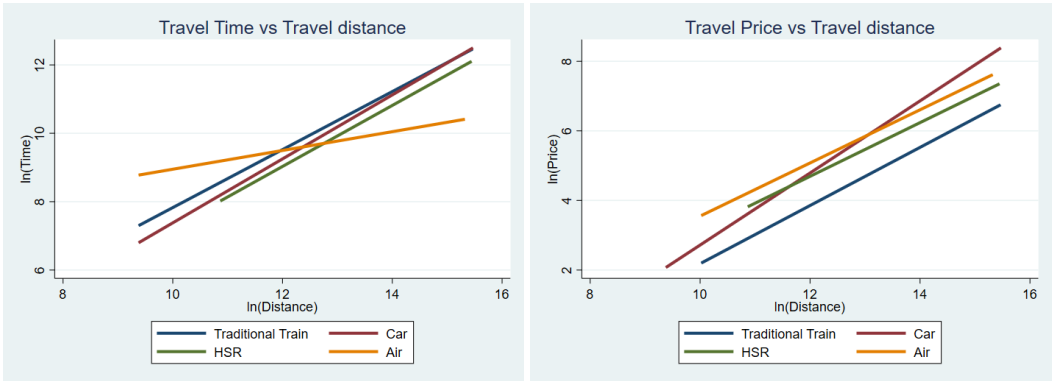
Figure 2 plots the log-linear fit of travel time and cost against the distance. For long-distance, HSR is faster and cheaper than cars. Therefore, we should anticipate HSR substitutes more for cars for longer distances.

To estimate the demand system of transportation mode choice, it is necessary to measure the market share of each transportation mode. To achieve this, the Tencent Location Base Service Data is utilized. This data allows for the tracking of users' locations in a limited number of cities, utilizing a speed tracker to infer their travel modes (car, plane, or train). As a result, it provides the ability to observe the market share of each transportation mode in different pairs of cities, where transportation time and prices can vary significantly[6]. Section 4.1 describes the demand estimation for transportation modes.

[5] see Figure A.1 for a detailed example of how I extract this information from Baidu Map API

[6] In the Appendix A, I provide additional details on the Tencent Location Base Service (LBS). However, I did not use this data in the main analysis as it only covers a limited number of cities and does not have the same level of high-frequency records as the Baidu LBS data.

Figure 2: Travel time and costs by transportation mode



Note: Unit: Distance: meters; Time: seconds; Price: RMB. The log-linear fit of travel time and travel cost on distance

3. A general equilibrium model of traveling

In this section, I present a multi-region multi-sector economic geography model that incorporates trade in goods and human travel. The model serves three primary purposes. Firstly, it provides a framework for using high-frequency data to decompose bilateral travel flows into different purposes. Secondly, it enables the use of business travel flows to infer trade flows, which are often difficult to measure in China. Finally, it generates sufficient statistics for conducting welfare analysis.

Consider an economy consisting of N cities. In each city, there are two sectors: consumer service sector (L) and goods sector (G). Services are sold locally but are available to consumers in other cities through their leisure travel. Goods are tradable. I model business travel as input for sales of remote markets. Labor is perfectly mobile across sectors but immobile across cities.

3.1 Transportation demand system

Conditional people decide to travel from i to j, the probability of choosing mode $m \in \{\text{Train, Car, Air}\}$ is specified as:

$$Pr_{ij}^{m} = \frac{\exp(\xi_m - \alpha \ln(t_{ij}^{m}))}{\sum_{l \in M} \exp(\xi_l - \alpha \ln(t_{ij}^{l}))}, \tag{1}$$

where t_{ij}^{m} are the travel time from i to j of mode m, ξ_m is the taste shifter. The expected utility takes the following form:

$$\chi_{ij} = (\sum_{l \in M} \exp(\xi_l - \alpha \ln(t_{ij}^{l})))^{-\frac{1}{\alpha}}. \tag{2}$$

(2) sheds light on how HSR improves up the existing infrastructure.

3.2 Households

The utility of household living in city i is specified as below:

$$U_i = u_i C_{i,G}^{\mu} C_{i,L}^{1-\mu},$$

where u_i is the amenity of living in city i. $C_{i,G}$ is the composite goods consumption. $C_{i,L}$ is the leisure service.

3.3 Consumer service and leisure travel

I assume perfect competition in the leisure service sectors and constant-return-to-scales production function $Y_{i,L} = L_{i,L}$. Households from city i first set up their budget on leisure consumption, then decide their destination. The leisure consumption of traveling to j is:

$$C_{i,L}(\omega) = B_j \frac{(1-\mu) w_i}{p_{j,L} \chi_{ij}^{\eta}} \epsilon_{ij}(\omega), \tag{3}$$

where B_i is the quality of leisure consumption. χ_{ij} : is the expected utility of travel . η is the elasticity of how the travel cost is transferred into the total price for the trip. $\epsilon_{ij}(\omega)$ is the idiosyncratic preference shocks following Fréchet distribution with shape parameters θ and mean 1. The price index of leisure consumption is:

$$P_{i,L}^{-\theta} = \sum_{k \in N} (\frac{\chi_{ik}^{\eta} w_k}{B_k})^{-\theta}.$$

The leisure travel share is:

$$\lambda_{ij}^{L} = \frac{(\frac{\chi_{ij}^{\eta} w_i}{B_i})^{-\theta}}{\sum_{k \in N} (\frac{\chi_{ik}^{\eta} w_k}{B_k})^{-\theta}}. \tag{4}$$

3.4 goods sectors and business travel

The composite good is a CES aggregation of varieties produced in different places:

$$C_{j,G} = (\sum_{k \in N} M_k q_{kj}^{\frac{1}{\sigma}} c_{kj}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}},$$

where σ is the substitution elasticity across different varieties. M_j is the number of firms located in city j and q_{ij} is the quality shifter. The market structure of goods is monopolistic competition. A firm that produces x_i units of output needs to incur total labor costs as:

$$l_i = F + \frac{x_i}{A_i},$$

where F is the fixed cost of operating. A_i is the firm-level productivity. Shipment of goods from i to j are subject to iceberg trade cost τ_{ij}. The firm set its optimal price as follows:

$$p_i^* = \frac{\sigma}{\sigma - 1} \frac{w_i}{A_i}.$$

. Conditional on quality q_{ij}, the gross profit generated from market j is:

$$\pi_{ij} = \frac{q_{ij}(\tau_{ij}p_i^*)^{1-\sigma}}{\sigma P_{j,G}^{1-\sigma}} E_{j,G}.$$

I model business travel b_{ij} as the input to improve quality shifter in the destination market. Potential cases corresponding to these assumptions are firms sending salesmen, engineers, and managers to increase the perception and quality of the product in the destination market. The market-specific quality production function is specified as

$$q_{ij} = \frac{b_{ij}^{\gamma}}{a}, \gamma < 1.$$

Therefore, the firm's marketing optimization problem, after setting the optimal price, becomes:

$$\max_{b_{ij}} b_{ij}^{\gamma} \frac{(\tau_{ij}p_i^*)^{1-\sigma}}{a\sigma P_{j,G}^{1-\sigma}} E_{j,G} - b_{ij} w_i \chi_{ij}^{\eta}.$$

The optimal number of business trips is:

$$b_{ij}^* = \kappa \tau_{ij}^{\frac{1-\sigma}{1-\gamma}} \chi_{ij}^{\frac{\eta}{\gamma-1}} w_i^{\frac{\sigma}{\gamma-1}} A_i^{\frac{\sigma-1}{1-\gamma}} P_{j,M}^{\frac{\sigma-1}{1-\gamma}} E_j^{\frac{1}{1-\gamma}},$$

where both trade costs and travel costs enter the optimal decision.

3.5 Gravity of travelers and the general equilibrium

In the model, the travelers for leisure and business purposes generate two gravity equations: the gravity of business traveling as

$$\lambda_{ij}^B = \frac{\tau_{ij}^{\frac{1-\sigma}{1-\gamma}} \chi_{ij}^{\frac{-\eta}{1-\gamma}} P_{j,M}^{\frac{\sigma-1}{1-\gamma}} E_{j,M}^{\frac{1}{1-\gamma}}}{\sum_k \tau_{ik}^{\frac{1-\sigma}{1-\gamma}} \chi_{ik}^{\frac{-\eta}{1-\gamma}} P_{k,M}^{\frac{\sigma-1}{1-\gamma}} E_{k,M}^{\frac{1}{1-\gamma}}}, \tag{5}$$

and the gravity of leisure traveling as

$$\lambda_{ij}^{L} = \frac{\chi_{ij}^{-\eta\theta} B_i^{\theta} w_i^{-\theta}}{\sum_k \chi_{ik}^{-\eta\theta} B_k^{\theta} w_k^{-\theta}}. \tag{6}$$

The spatial equilibrium of the economy is, given labor forces distribution $\{L_{\cdot}\}$, sectoral labor allocation and wages $\{w_{\cdot}, L_{\cdot,G}, L_{\cdot,L}\}$ that satisfies:

1. tradable goods market clear:

$$w_i L_{i,G} = \sum_j \pi_{ij} \mu w_j (L_{j,G} + L_{j,L}).$$

2. service market clear:

$$w_i L_{i,L} = \sum_j \lambda_{ji}^{L} (1-\mu) w_j (L_{j,G} + L_{j,L}).$$

3.6 Sufficient statistics of welfare

In this section, I briefly discuss the model properties and sufficient welfare statistics. This will guide us to the data requirement and key parameters to be estimated in the counterfactual analysis. First, although I do not directly observe bilateral trade data among cities, I can infer trade share using business travel share because they have a tight connection in the model:

$$\frac{\pi_{ij}}{\pi_{ik}} = \frac{E_j^{-1}}{E_k^{-1}} \frac{\chi_{ij}^{\eta}}{\chi_{ik}^{\eta}} \frac{\lambda_{ij}^{B}}{\lambda_{ik}^{B}}. \tag{7}$$

Since firms always spend a constant share in business travel, the cost of business travel $\chi_{ij}^{\eta} \lambda_{ij}^{B}$ is proportional to the trade volume $E_j \pi_{ij}$. Therefore, I can infer trade share as long as expenditure, travel cost, and business travel shares are available. Second, the short-run welfare changes for the city i, assuming labor cannot reallo-

cate among cities, are:

$$\hat{W}_i = \underbrace{(\hat{\lambda}^L_{ii})^{-\frac{1-\mu}{\theta}}}_{\text{leisure travel share}} \underbrace{(\hat{\pi}_{ii})^{\mu\frac{1-\gamma}{1-\sigma}}}_{\text{trade share}} \underbrace{(\hat{L}_{i,M})^{-\mu\frac{1-\gamma}{1-\sigma}}}_{\text{agglomeration forces}}$$

Several deviations from the ACR (Arkolakis et al. (2012)) formula are (1) leisure travel share, which is related to the consumer market access of leisure service, matters for welfare. (2) the trade share could also change due to travel cost due to tight connection in (7). (3) The third part is the agglomeration force because of the love of variety in the goods sectors. Therefore, the welfare change can be sufficiently summarized using leisure travel share, trade share, and labor reallocation, along with key structural parameters, which I will estimate in the following section.

4. Estimation

4.1 Transportation demand system estimation

I estimate transportation demand system (1) using multinomial logit regression:

$$\ln(Pr^m_{ij}) - \ln(Pr^0_{ij}) = \xi_k - \xi_0 - \alpha(\ln(t^k_{ij}) - \ln(t^0_{ij})) + \epsilon_{ijm} \tag{8}$$

Table 1 shows the estimation results. Column (1) is the baseline results. I add the monetary price in Column (2). The insignificance of price suggests passengers are not sensitive to the price of the transportation methods. From the estimated result, I back out the model-consistent measure of expected travel cost up to some constant:

$$\hat{\chi_{ij}} = (\sum_{l\in M} \exp(\hat{\xi}_l - \hat{\alpha}\ln(t^l_{ij})))^{-\frac{1}{\hat{\alpha}}} \tag{9}$$

Here, a potential concern is the transportation demand may be different for travel purposes (e.g., business or leisure). Table A.2 addresses this concern. There is no systematic difference in transportation methods between weekday and week-

Table 1: Transportation demand system multinomial logit estimation

	(1)	(2)	(3)
	lnRshare	lnRshare	lnRshare
	b/se	b/se	b/se
lnRTime	-2.220***	-2.106**	-1.599***
	(0.515)	(0.754)	(0.387)
F_mode_2	0.958	0.834	
	(0.621)	(1.037)	
lnRPrice		-0.959	-1.028
		(0.777)	(0.770)
Const.	-0.895***	-0.412	-0.054
	(0.292)	(0.509)	(0.410)
N	19	18	18
r2	0.591	0.583	0.557

Note: Train is set as the baseline. $\ln Rshare$ is the market share relative to train from i to j. $\ln RTime$ is the travel time relative to the train from i to j. $\ln RPrice$ is the travel monetary cost relative to train from i to j.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

end travel flows. So business/ leisure travelers are less likely to have completely different preferences.

4.2 Model-consistent decomposition of travel flows into different purposes

As discussed in the previous section, business and leisure travelers have different economic implications for cities' industrial structure, so it's essential to measure the city's attractiveness for business and leisure travelers separately. How-

ever, in GPS datasets, we only observe the sum of business and leisure travelers between two cities, $F_{ij,t}$. To make progress, I am going to exploit additional data, the time-series variations of the business and leisure travel activities within the city, $F^W_{ii,t}, F^L_{ii,t}$, to infer the out-of-city travelers into business and leisure purposes. Specifically, I assume the total number of people who want to travel for business and leisure varies across time. Still, conditional on business/leisure travel, the propensity to go out-of-city versus within the city only depends on city fundamentals that do not change over time. I summarize the assumptions as follows:

Assumption 1 *I assume the data-generating process takes the following form:*

$$F_{ij,t} = C_i \frac{\lambda^B_{ij}}{\lambda^B_{ii}} F^W_{ii,t} + \frac{\lambda^L_{ij}}{\lambda^L_{ii}} F^L_{ii,t} + \epsilon_{ij,t}. \tag{10}$$

In (10), C_i is the city-level ratio of business travelers versus commuters, which only depends on the industrial structure. $\lambda^B_{\cdot}, \lambda^B_{\cdot}$ are business/leisure travel flows correspond to (5) and (6). All these variables are consistent with the theory in that they only depend on city-level fundamentals that do not change over a short period of time. Therefore, based on this data-generating process, I regress $F_{ij,t}$ on $F^W_{ii,t}, F^L_{ii,t}$,

$$\hat{F}_{ij,t} = \hat{\alpha}_{ij} F^W_{ii,t} + \hat{\beta}_{ij} F^L_{ii,t},$$

then normalized the *i*-*j* coefficients by city *i*, I get the model-consistent measure of travel shares for different purposes

$$\widehat{\lambda^B_{ij|\cdot}} = \frac{\lambda^B_{ij}}{\sum_{k\neq i} \lambda^B_{ik}} = \frac{\hat{\alpha}_{ij}}{\sum_{k\neq i} \hat{\alpha}_{ik}}$$
$$\widehat{\lambda^L_{ij|\cdot}} = \frac{\lambda^L_{ij}}{\sum_{k\neq i} \lambda^L_{ik}} = \frac{\hat{\beta}_{ij}}{\sum_{k\neq k} \hat{\beta}_{ik}}.$$

4.3 Gravity estimation

Now, with model-consist measure of $\lambda^{\hat{B}}_{ij|\cdot}, \lambda^{\hat{L}}_{ij|\cdot}$, I estimate the key travel elasticities in (5) and (6) using PPML, following Silva and Tenreyro (2006).

$$\ln \widehat{\lambda^{B}_{ij|\cdot}} = \frac{\widehat{-\eta}}{1-\gamma} \ln \hat{\chi}_{ij} + \frac{\widehat{1-\sigma}}{1-\gamma} \ln \tau_{ij} + O^B_i + D^B_j + \epsilon^B_{ij},$$
$$\ln \widehat{\lambda^{L}_{ij|\cdot}} = \widehat{-\eta\theta} \ln \hat{\chi}_{ij} + O^L_i + D^L_j + \epsilon^L_{ij}.$$

Here I calculated $\hat{\chi}_{ij}$ using (9) and I proxy τ_{ij} using road distance. Table 2 shows the estimation results.

Table 2: PPML

	(1) workshare b/se	(2) leisureshare b/se
lntravelcost	-0.891***	-3.029***
	(0.052)	(0.020)
lnd	-1.274***	
	(0.037)	
Const.	19.546***	23.097***
	(0.664)	(0.497)
Origin	Yes	Yes
Destination	Yes	Yes
N	87155	87282
r2	0.691	0.664

Note: PPML estimation of gravity-equations of business and leisure travel.
* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

According to the model, those destination/ origin fixed effects of business and

leisure should reveal the comparative advantages of cities. To make sense of those fixed effects, I report the correlation between fixed effects and city-level statistics in Table 3. The destination fixed effects for business purposes positively relate to secondary and tertiary sector share. The destination fixed effects are positively related to only tertiary sector share. I also report the original fixed effects for business and leisure purposes, which capture the idea of market competition in my setting. Those two original fixed effects show the opposite correlation with the city's employment share in the consumer service sector versus producer service sectors. The analysis here is to validate the decomposition in (10) capture meaningful industrial comparative advantages of cities.

Table 3: Make sense of the fixed effects

	(1)	(2)	(3)	(4)
	lndfwork	lndfleisure	lnofwork	lnofleisure
	b/se	b/se	b/se	b/se
secondary_gdp_sh	0.021**	-0.001		
	(0.010)	(0.007)		
tertiary_gdp_sh	0.076***	0.025***		
	(0.012)	(0.008)		
lnCS_emp			-0.489***	0.117**
			(0.101)	(0.050)
lnPS_emp			0.227	-0.393***
			(0.157)	(0.078)
Const.	-1.608*	24.134***	0.165	25.353***
	(0.908)	(0.619)	(1.214)	(0.602)
N	294	294	273	285
r2	0.150	0.063	0.141	0.113

Note: Regress estimated fixed effects on city-level statistics
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

4.3.1 Other parameters

The other parameters of the model are calibrated or taken from the literature. Trade elasticity $1-\sigma$ is set to be -8 as Allen and Arkolakis (2022). γ is set as 0.2 to match the average marketing cost of china listed firms. $1-\mu$ is set as 0.15 to match the leisure expenditure share from China's yearly statistics. Table 4 presents the complete estimates of all parameters.

Table 4: Parameters estimation/ calibration

Parameters	Descriptions	Value	Estimation/ Calibration
α	elasticity of travel cost with travel time	2.2	Multinomial Logit
$-\eta\theta$	leisure travel elasticity with travel cost	-3	PPML
$-\frac{\eta}{1-\gamma}$	business travel elasticity with travel cost	-0.89	PPML
γ	marketing cost share	0.2	Calibration
$1-\sigma$	trade elasticity	-8	AA (2020)
$1-\mu$	leisure expenditure share	0.15	Consumer Service GDP share

5. Counterfactual

In this section, I conduct a counterfactual analysis to quantify the welfare gains of HSR, which will help us understand how travel costs affect regional welfare. The counterfactual replaces all the HSR in 2021 with the traditional trains, holding other transportation infrastructures and cities' fundamentals unchanged. Here, I focus on the short-term consequences so workers can only reallocate across sectors but can not migrate. In this setting, the data required to compute the counterfactual equilibrium is minimum, summarized in the following proposition.

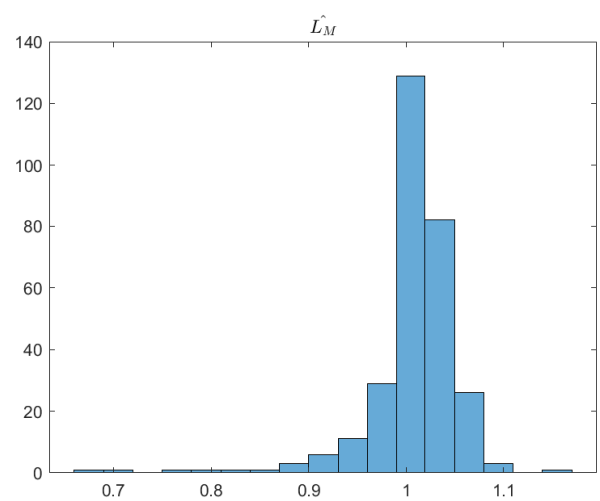
Proposition 1 *Given the allocation of current equilibrium* $\{\lambda_{\cdot}^{0}, w_{\cdot}^{0}, L_{\cdot}^{0}\}$*, and the coun-*

terfactual changes in travel cost $\{\hat{\chi_{ij}}\}$ *, the counterfactual equilibrium in changes* $\{\hat{\lambda}., \hat{w}., \hat{L}.\}$ *solves the following systems of equations:*

$$
\begin{aligned}
w_i^0 L_{i,M}^0 \hat{w}_i \hat{L}_{i,M} &= \sum_j \pi_{ij}^0 \hat{\pi}_{ij} \mu w_j^0 \hat{w}_j L_j, \\
\hat{\pi}_{ij} &= \frac{\hat{\chi}_{ij}^{\frac{\eta\gamma}{\gamma-1}} \hat{w}_i^{\frac{1-\sigma-\gamma}{1-\gamma}} \hat{L}_{i,M}}{\sum_k \pi_{kj}^0 \hat{\chi}_{kj}^{\frac{\eta\gamma}{\gamma-1}} \hat{w}_k^{\frac{1-\sigma-\gamma}{1-\gamma}} \hat{L}_{k,M}}, \\
w_i^0 L_i \hat{w}_i &= \sum_j (\pi_{ij}^0 \hat{\pi}_{ij} \mu + \lambda_{ji}^{L0} \hat{\lambda}_{ji}^L (1-\mu)) w_j^0 \hat{w}_j L_j, \\
\hat{\lambda}_{ji}^L &= \frac{\hat{\chi}_{ji}^{-\eta\theta} \hat{w}_i^{-\theta}}{\sum_k \lambda_{jk}^{L0} \hat{\chi}_{jk}^{-\eta\theta} \hat{w}_k^{-\theta}}
\end{aligned}
$$

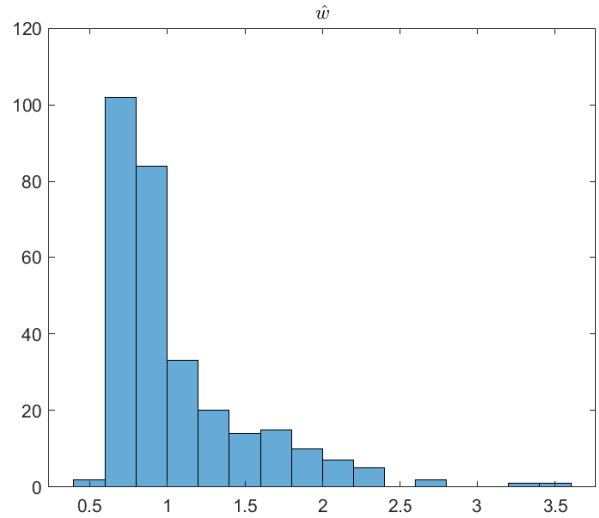
Figure 3 shows the counterfactual labor reallocation of labor across sectors. As one could expect, the change in spatial frictions can reshape the comparative advantages of cities. Compared with the reduced form approach, which shows the average effect, the structural model shows the distributional impact of HSR on different cities.

Figure 3: Counterfactual sectoral reallocation



The counterfactual results on welfare are shown in Figure 4. Most of the cities in China suffer from this counterfactual change, especially large cities, meaning that HSR is, on average, welfare-improving. But, perhaps surprisingly, I find some cities suffer from the introduction of HSR. This could occur due to worsening terms of trade. This finding is in accordance with Faber (2014), who also found welfare loss in some small cities due to China's highway expansion. According to the Mid-to-Long Railway Plan (2008), the development of HSR aimed to facilitate the flow of factors and reduce regional inequality. In this sense, the policy implication of HSR expansion could be ambiguous since HSR increases big cities' welfare at the expense of hurting some small cities.

Figure 4: Counterfactual welfare change



In terms of aggregate effects, I found the HSR, on average, improves welfare by around 0.3%. Two comments emerge from this result. First, this welfare gain is comparable with the 1.1% - 1.4 % welfare gain of U.S roads in facilitating trade (Allen and Arkolakis (2014)). However, if one believes roads reduce both trade costs and travel costs, then there should be missing considerable welfare gains due to facilitating travel. Second, although the welfare gains of HSR are not small, it is

still not far beyond the construction cost. The nominal construction cost of HSR is estimated at around 5% of the 2018 China GDP. I assume those costs are financed by raising capital with a 5% interest rate, which ends up with an effective welfare cost of HSR around 0.25%. However, the welfare gains calculated here are short-run. If HSR might also encourage migration, as in Morten and Oliveira (2018), the welfare gains of HSR could potentially be higher.

6. Conclusion

In this paper, I used new GPS data to document the traveling patterns, which are less studied by the literature. I show that traveling is frequent, and they are subject to spatial frictions. To investigate the economic implication of traveling, I further develop a quantitative model of business/ leisure traveling with their interaction with city-level economic activities. To quantify the relative importance of the model, I study China's HSR expansion. The counterfactual results suggest that spatial frictions in travel time are essential in shaping our economics. I find HSR, on average, improves welfare by 0.3% but causes unequal gains among cities and industries. However, this paper mainly focuses on the short-run analysis where migration is absent. The long-run distributional effect of infrastructural improvement is an important question for future study.

References

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

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy*," *The Quarterly Journal of Economics*, August 2014, *129* (3), 1085–1140.

_ and _ , "The Welfare Effects of Transportation Infrastructure Improvements," *The Review of Economic Studies*, November 2022, *89* (6), 2911–2957.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, February 2012, *102* (1), 94–130.

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito, "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, April 2019, *127* (2), 639–688.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach*," *The Quarterly Journal of Economics*, May 2016, *131* (2), 799–858.

Duranton, Gilles and Matthew A. Turner, "Urban Growth and Transportation," *The Review of Economic Studies*, October 2012, *79* (4), 1407–1440.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

Eckert, Fabian, "Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy," February 2019.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, July 2014, *81* (3), 1046–1070.

_ and Cecile Gaubert, "Tourism and Economic Development: Evidence from Mexico's Coastline," *American Economic Review*, June 2019, *109* (6), 2245–2293.

Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti, "Growing Like India: The Unequal Effects of Service-Led Growth," Technical Report w28551, National Bureau of Economic Research, Cambridge, MA March 2021.

Koopman, Robert, Zhi Wang, and Shang-Jin Wei, "Tracing Value-Added and Double Counting in Gross Exports," *American Economic Review*, February 2014, *104* (2), 459–494.

Lin, Yatang, "Travel costs and urban specialization patterns: Evidence from China's high speed railway system," *Journal of Urban Economics*, March 2017, *98*, 98–123.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen Redding, "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data," Technical Report w28497, National Bureau of Economic Research, Cambridge, MA February 2021.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," August 2018, p. 65.

Ollivier, Gerald, Richard Bullock, Jin Ying, and Nanyan Zhou, "High-Speed Railways in China," *China Transport Topics*, December 2014. Publisher: World Bank, Beijing.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, November 2006, *88* (4), 641–658.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," June 2019, p. 114.

Wang, Zhi, Shang-Jin Wei, and Kunfu Zhu, "Measuring Domestic Factor Content in Bilateral and Sectoral-Level Trade Flows," Technical Report w30953, National Bureau of Economic Research, Cambridge, MA February 2023.

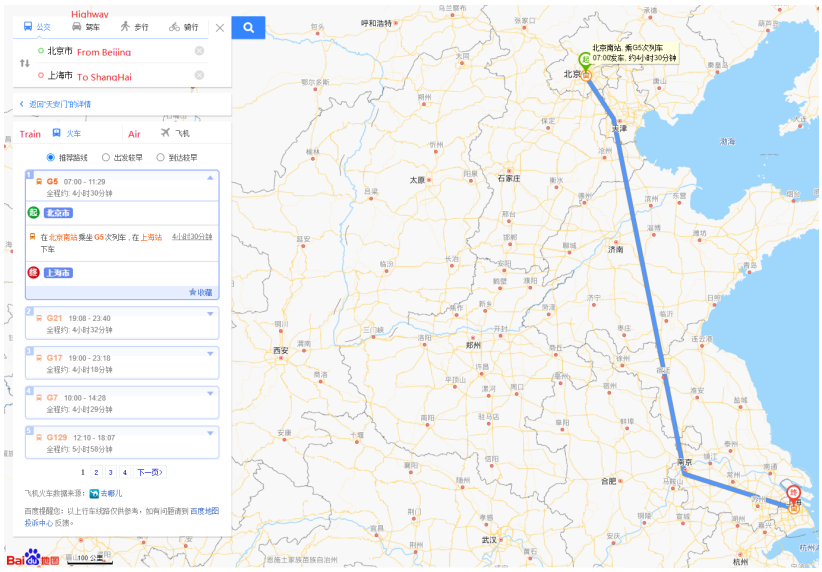
Xu, Mingzhi, "Riding on the New Silk Road: Quantifying the Welfare Gains from High-Speed Railways," October 2018.

Appendix

A. Data Appendix

In this section, I provide more details about data collection and validation processes. Figure A.1 shows the data extraction details from Baidu Map API.

Figure A.1: Baidu Map API



Note: Source: Baidu Map API. This figure shows an example of how to extract travel cost information of different modes (car, train, air) from Baidu Map API.

Figure A.2 gives more details on the information we can extract from Baidu and Tencent Location Based Services.

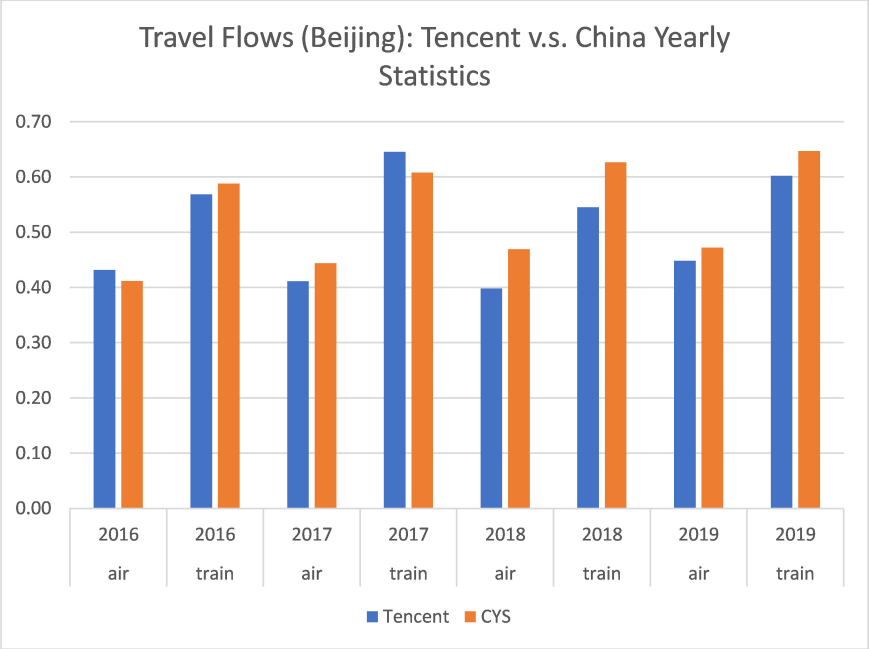
Since Baidu has a large customer base in China, the travel flows data used here are fairly representative. Figure A.3 and A.4 plot the travel flows data against China's Yearly Statistics.

Figure A.5 shows the business activity intensity in Beijing over eight weeks. It shows robust cyclicality patterns and has lower business activity intensity in town during holidays.

Figure A.2: Details of Baidu and Tencent LBS

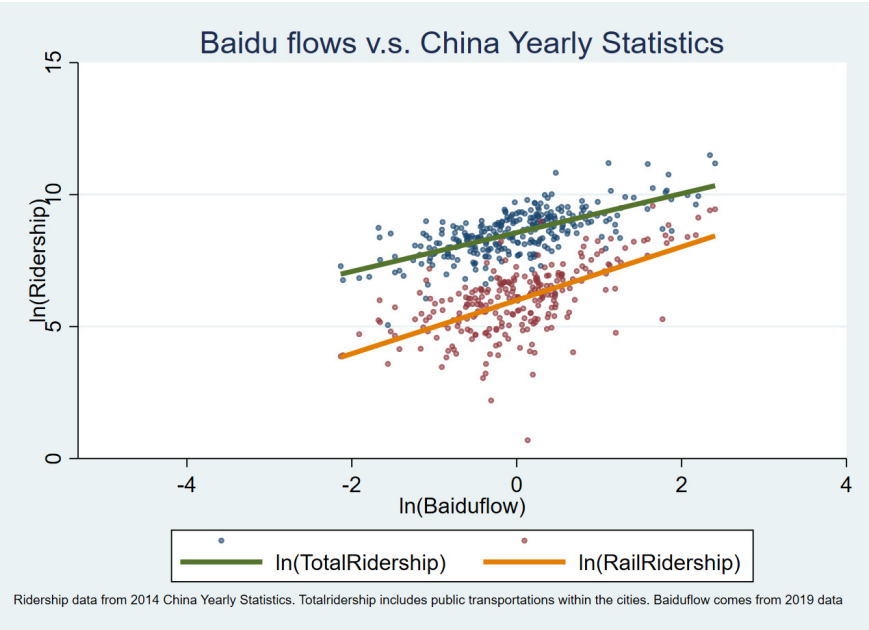
Note: Source: Up: Tencent LBS; Down: Baidu LBS.

Figure A.3: Tencent LBS Data Validation



Note: Comparison between Tencent GPS transportation ridership of Beijing with Beijing Yearly statistics from NBS. The aggregate total outflows in 2016 are normalized to one. Source: https://data.stats.gov.cn/.

Figure A.4: Baidu LBS data Validation



Note: Comparison between Baidu GPS travel flows with public transportation usage in China City Yearly Statistics. Source: https://data.stats.gov.cn/.

Figure A.5: City Activity Intensity: An example

Note: This graph shows the business activity intensity in Beijing over eight weeks. It shows robust cyclicality patterns and has lower business activity intensity in town during holidays. Source: Baidu LBS

Figure A.6: A typical HSR station

Note: This graph is to illustrate there are no goods loading areas in a typical HSR station. Therefore, HSR only affects travel costs but has no effect on trade costs.

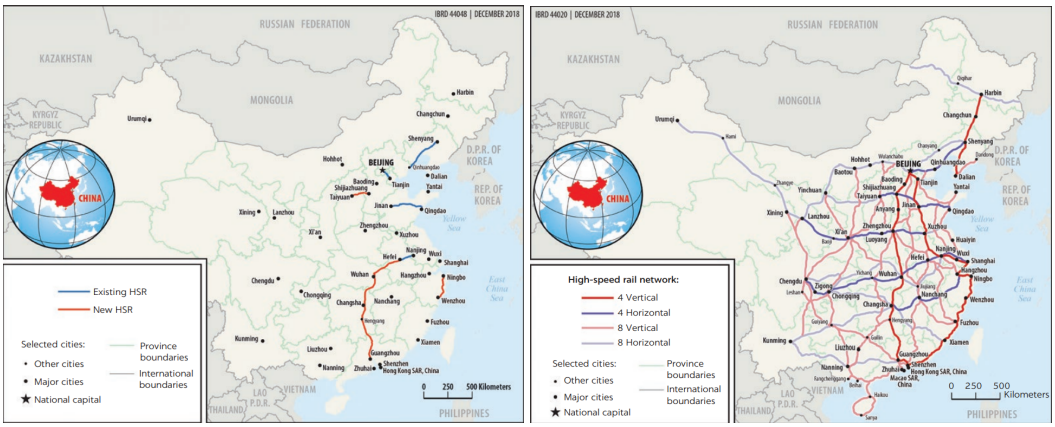


Figure A.7: The fast expansion of China HSR networks

Source: **?**

Table A.1: The effect of HSR connect on gross travel flows

	(1)	(2)	(3)
	dlngross	dlngross_weekends	dlngross_weekdays
	b/se	b/se	b/se
1.New	0.129***	0.142***	0.125***
	(0.049)	(0.054)	(0.047)
Const.	-0.488***	-0.510***	-0.480***
	(0.026)	(0.026)	(0.026)
N	325	325	325
R2	0.008	0.009	0.007

Note: Column (1) shows the change of average travel flows after HSR connection. Column (2) and (3) look at the change of travel flows during weekdays and weekends separately.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.2: The transportation demand by weekdays/ weekends

	(1) carshare b/se	(2) trainshare b/se	(3) airshare b/se
1.weekend	0.000	0.006	-0.007
	(0.003)	(0.007)	(0.007)
Const.	0.226***	0.271***	0.503***
	(0.005)	(0.010)	(0.010)
*Destination	Yes	Yes	Yes
N	310	310	310
R2	0.994	0.898	0.941

Note: Regression of riderships on weekends indicator. There are no systematic difference of transportation choice between weekends and weekdays

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.