The Welfare Effects of Passenger Transportation Infrastructure: Evidence from China

Panle Jia Barwick  Dave Donaldson  Shanjun Li  Yatang Lin
Cornell and NBER  MIT and NBER  Cornell and NBER  HKUST

ABFER 6th Annual Conference, May, 2018
Motivation

- Enormous public resources invested in passenger transportation infrastructure across the globe:
  - Airports, railways, highways, subways
  - New high-speed railway projects under discussion in UK, US, India

- Limited research on the importance of these large-scale projects:
  - Reduced form evidence on passenger transportation largely within city (e.g., subways)
  - Quantitative welfare evaluation mostly focuses on the flow of goods
  - Substantial data challenges in obtaining ideal data on bilateral passenger flows across the entire network
This paper

- Study China’s high-speed railway system (HSR), one of the largest passenger infrastructure projects of the world
  - Total length exceeding 15,000 miles (25,000 km) in 2017, connecting cities from 29 provinces out of 33
  - Total investment of $300 billion from 2011 to 2015

- Draw on new data on universe of debit/credit card transactions (40 Trillion Yuan in 2015) to measure:
  - City-to-city passenger flows
  - City-to-city transactions

- Develop a quantitative model for evaluating the welfare implications of passenger transportation infrastructure improvements
  - Aggregate consumer welfare gain of the HSR network
  - Distributional impacts
Related literature

- Transportation infrastructure and development

- Quantitative evaluation of infrastructure projects:

- Evaluation of the HSR system
Roadmap

- Background and data construction
- CES model on demand for travel and goods
- Model estimation
- Ongoing: Random coefficient logit framework to capture passenger heterogeneity
Background

- China’s HSR expansion
  - In 2003, the first line opened connecting Qinhuangdao and Shenyang
  - By 2017, over 20,000 km routes in service, with 7 bn cumulative number of trips
  - Operation speed: 250 km/h - 350 km/h, versus up to 120 km/h for regular railway

- Ministry of Railway plan (2008)
  - Main network: four horizontal and four vertical lines
  - Connect major cities with more efficient means of transportation
  - Environmental and national security concerns

- Planning and Financing
  - Centrally planned and managed by Ministry of Railway (later China Railway Corporation), mostly funded by MOR
  - Local government helps with compensation for land use; limited private investment
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Expansion of HSR from 2003-2016
Data at the City-pair and Month Level

- **Transport network and cost features:**
  - HSR and traditional train: routes, schedule, and fares
  - Air travel: routes, schedule, and number of seats; fares for a small subsample
  - Road: highway distance and travel time for all city pairs in 2017

- **Bilateral passenger flows and transaction values constructed from Unionpay card (credit and debit card) transactions**

- **Consumption goods price indices constructed from Unionpay data**
Coverage of the Bank Card Data 2011-2017

- In 2015: 2.7 bn cards, 48% of retail sales of consumer goods, 40 tn RMB worth of transactions (China’s GDP is 69 tn RMB)
- Number of active bank cards per capita similar across cities

**Figure:** Number of Active Cards per Capita, 2015

- About 20% of the transactions are made out-of-town, increasing over time

**Figure:** Share of out-of-town spending: by the size of resident city
## Summary Statistics

Bilateral transaction and trip flows

<table>
<thead>
<tr>
<th></th>
<th>Exclude Own city Flows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
</tr>
<tr>
<td>Transaction value (Y)</td>
<td>1,935,262</td>
</tr>
<tr>
<td>Transaction count (N)</td>
<td>1,936,603</td>
</tr>
<tr>
<td>Number of trips (T)</td>
<td>1,783,886</td>
</tr>
</tbody>
</table>
Motivating Evidence

- Question: how does direct HSR connection change cross-city travel and consumption behavior?

\[
\ln(y_{ijt}) = \beta \text{connect}_{ijt} + \alpha_{ij} + \eta_{it} + \gamma_{jt} + \epsilon_{ijt}
\]  

- \(y_{ijt}\): number of trips or total transaction value made by residents in city \(i\) to city \(j\)
- \(\text{Connect}_{ijt} = 1\) if city \(i\) and city \(j\) are connected by HSR at month \(t\)
- \(\alpha_{ij}\) city pair FE; \(\eta_{it}\) and \(\gamma_{jt}\) are origin/destination*month FE.
## Motivating Evidence

<table>
<thead>
<tr>
<th></th>
<th>$\ln(\text{Trips})$</th>
<th>$\ln(\text{Value})$</th>
<th>$\ln(\text{Trips})$</th>
<th>$\ln(\text{Value})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSR connection</td>
<td>0.37***</td>
<td>0.37***</td>
<td>0.35***</td>
<td>0.28***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,214,670</td>
<td>2,214,670</td>
<td>2,214,597</td>
<td>2,214,597</td>
</tr>
<tr>
<td>Pair FE, Month FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Origin*month FE</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Destination*month FE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.83</td>
<td>0.58</td>
<td>0.86</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Model: Setup

- Representative agents from city $i$ with fixed income $X_i$ make travelling decisions in three steps (3-layer nested CES)
  - Inner nest: transportation mode choices
  - Middle nest: trips and consumption goods
  - Outer nest: decision about travelling across all destinations
Model of Demand for Travel and Goods

- Basic idea draws on logic of revealed preference: use changes in ridership and consumption associated with changes in travel cost to back out consumers’ willingness to pay for HSR

- Outer layer:

\[ U_{it} = \left( Q_{ijt} \phi_{ijt} \right)^{\frac{\sigma - 1}{\sigma}} \]

- Middle layer:

\[ Q_{ijt} = \left[ (q_{ijt})^{\frac{\delta - 1}{\delta}} + (T_{ijt} \epsilon_{ijt})^{\frac{\delta - 1}{\delta}} \right]^{\frac{\delta}{\delta - 1}} \]

- Inner layer:

\[ T_{ijt} = \left[ (t_{ij1t} \eta_{ijt})^{\frac{\rho - 1}{\rho}} + (t_{ij2t})^{\frac{\rho - 1}{\rho}} \right]^{\frac{\rho}{\rho - 1}} \]
Model: Inner Nest

- Holding constant spending and the number of trips made from $i$ to each destination $j$ at month $t$, agents decide on transportation

$$T_{ijt} = \left[ \left( t_{ij1t} \eta_{ijt} \right)^{\frac{\rho-1}{\rho}} + \left( t_{ij2t} \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}}$$

- Two modes: air and other (including HSR, traditional trains, and highway)
- $\eta_{ijt}$ idiosyncratic demand shifter for air travel
- Travel cost for air $c_{ij1t}$: a function of both travel time and fare cost
- Travel cost for non-air mode $c_{ij2t}$: the minimum travel cost among HSR, traditional trains, and highway
**Preferences: Inner Nest**

- Intermodal choice can be used to identify $\rho$:

$$\frac{t_{ij1t}}{t_{ij2t}} = \left(\frac{c_{ij1t}}{c_{ij2t}}\right)^{-\rho} (\eta_{ijt})^{\rho-1}$$

- Once $\rho$ (and residual $\eta_{ijt}$) estimated, obtain travel cost index $c_{ijt}$ across all transportation modes as follows:

$$c_{ijt} = \left[\left(\frac{c_{ij1t}}{\eta_{ijt}}\right)^{1-\rho} + (c_{ij2t})^{1-\rho}\right]^{\frac{1}{1-\rho}}$$
Preferences: Middle Nest

- Holding constant total consumption quantity at each destination $j$ as $Q_{ijt}$, agents allocate it across goods consumption $q_{ijt}$ and trips $T_{ijt}$

$$Q_{ijt} = \left[ (q_{ijt})^{\frac{\delta-1}{\delta}} + (T_{ijt}\epsilon_{ijt})^{\frac{\delta-1}{\delta}} \right]^{\frac{\delta}{\delta-1}}$$

- Intuition: consumers derive utility from access to consumption goods $(q)$ in city $j$, as well as free local amenity (tourist attractions, visiting family/friends etc.), which is a function of trips made $(T)$

- Reductions in travel cost might induce consumers to make more frequent trips, but spend less per trip

- $\epsilon_{ijt}$: idiosyncratic demand shifter between goods consumption and trips
Preferences: Outer Nest

- Consumers from city $i$ allocate total spending across all destination cities to maximize utility, subject to the budget constraint

$$U_{it} = \left( Q_{ijt} \phi_{ijt}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

$$Q_{ijt} = \frac{P_{ijt}^{-\sigma}}{P_{it}^{1-\sigma}} X_{it} \phi_{ijt}^{\sigma-1}$$

$$P_{ijt} = \left[ (p_{jt})^{1-\delta} + (c_{ijt}/\epsilon_{ijt})^{1-\delta} \right]^{\frac{1}{1-\delta}}$$

$$P_{it} = \left[ \sum_{j=1}^{J} (P_{ijt}/\phi_{ijt})^{1-\sigma} \right]^{\frac{1}{1-\sigma}}$$

- $Q_{ijt}$: total consumption quantity in city $j$; $X_{it}$: total spending for consumers from city $i$; $\phi_{ijt}$: taste shocks across destination cities
Welfare Impact

\[ U_{it} = \frac{X_{it}}{P_{it}} \]

- We focus on the “consumer” benefits of HSR, which appear purely through \( P_{it} \). Further, impact of HSR on \( P_{it} \) comes purely through impact of HSR on \( c_{ijt} \)

- So far, following effects omitted:
  - HSR changes the price of consumption goods in cities
  - \( \Rightarrow \) No business stealing effects in this model: cities will not be worse off after the HSR connection
  - HSR changes incomes (and hence \( X_{it} \))
Estimation of the demand system

- Estimation in three steps

- Step 1: Inner layer:

\[
\ln\left(\frac{t_{ij1t}}{t_{ij2t}}\right) = -\rho \ln\left(\frac{c_{ij1t}}{c_{ij2t}}\right) + \alpha_{ij} + \beta_it + \gamma_{jt} + \tilde{\eta}_{ijt}
\]

- Obtain \(\rho\) and \(\eta_{ijt}\) from estimating the equation above, using data on travel cost and frequency (air travels measured by seat capacity) for different modes of transportation

- Construct \(c_{ijt}\) from \(c_{ij1t}, c_{ij2t}, \rho,\) and \(\eta_{ijt}\)

- Construct \(T_{ijt}\) from \(t_{ij1t}, t_{ij1t}, \rho,\) and \(\eta_{ijt}\)

- Instrument \(\ln\left(\frac{c_{ij1t}}{c_{ij2t}}\right)\) with HSR indirect connection dummy
Instrument

- Instrument $\ln\left(\frac{c_{ij1t}}{c_{ij2t}}\right)$ with HSR indirect connection dummy to take care of:
  - Measurement errors in travel cost
  - Fare price endogeneity
  - Endogeneity in the availability of new transportation modes

- Idea of Instrument
  - China’s HSR network is quite intensive with four main horizontal lines and four main vertical lines
  - When a (segment of) a horizontal line gets joined with a vertical line, non-nodal cities from both lines get ”indirectly connected”, which are less likely to be planned in advance
Yichun and Shaoguan is considered to be indirectly connected after both Changsha-Nanchang and Changsha-Guangzhou lines are in operation.
# Reduced form evidence

<table>
<thead>
<tr>
<th></th>
<th>(\ln(\text{Trips}))</th>
<th>(\ln(\text{Value}))</th>
<th>(\ln(\text{Trips}))</th>
<th>(\ln(\text{Value}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect connection</td>
<td>0.171***</td>
<td>0.194***</td>
<td>0.148***</td>
<td>0.117***</td>
</tr>
<tr>
<td></td>
<td>(0.0201)</td>
<td>(0.0368)</td>
<td>(0.0207)</td>
<td>(0.0364)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,147,311</td>
<td>2,147,311</td>
<td>2,147,241</td>
<td>2,147,241</td>
</tr>
<tr>
<td>Pair FE, Month FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Origin*month FE</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Destination*month FE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.799</td>
<td>0.549</td>
<td>0.835</td>
<td>0.595</td>
</tr>
</tbody>
</table>
Estimation of the demand system

Constructing $c_{ij1t}$ and $c_{ij2t}$

- Travel cost for different modes of transportation
  - $c_{ijkt} = farecost_{ijkt} + ValueforTravelTime(VTT) \times traveltime_{ijkt}$; VTT assumed to be $1/3$ of hourly wage
  - Air: distance and duration of all flights from 2010 to 2017, price data available for a small subset
  - HSR and traditional trains: railway timetable data that report duration and ticket price for all train schedules from 2008 to 2016
  - Road: calculate duration and distance of travel by road for any city pairs using OpenStreetMap
  - $c_{ij2t}$: the minimum of travel cost across HSR, traditional trains, and road
Estimation of the demand system

Constructing passenger flows and consumption

- Bilateral passenger transportation ridership on air \((t_{ij1t})\) and the rest \((t_{ij2t})\)
  - Air: total seats of all flights serving each city pair
  - Total number of trips made by card holders from city \(i\) in city \(j\): constructed using UnionPay data

- Bilateral consumption \(q_{ijt}\) and destination city price index \(p_{jt}\)
  - Assume the distribution of quantity purchased per transaction constant over time, and use the average value per transaction as a proxy for \(p_{jt}\) (alternative approaches)
Estimation of the demand system

Parameter estimation

- Step 2: Middle layer:

\[
\ln\left(\frac{T_{ijt}}{q_{ijt}}\right) = -\delta \ln\left(\frac{c_{ijt}}{p_{jt}}\right) + \alpha_{ij} + \beta_{it} + \gamma_{jt} + (\delta - 1) \ln(\tilde{\epsilon}_{ijt})
\]

- Plug in \( T_{ijt} \) and \( c_{ijt} \) from the inner nest
- Obtain \( \delta \) and \( \epsilon_{ijt} \)
- Construct \( P_{ijt} = \left[ (p_{jt})^{1-\delta} + \left(\frac{c_{ijt}}{\epsilon_{ijt}}\right)^{1-\delta} \right]^{\frac{1}{1-\delta}} \)
- Instrument \( \ln\left(\frac{c_{ijt}}{p_{jt}}\right) \) with HSR indirect connection dummy
- In practice, inclusion of \( \gamma_{jt} \) means that regressor is effectively just \( \ln(c_{ijt}) \).
Estimation of the demand system

Parameter estimation

Step 3: Outer layer:

\[ \ln(X_{ijt}) = (1 - \sigma) \ln(P_{ijt}) - (1 - \sigma) \ln(P_{it}) + \ln(X_{it}) + (\sigma - 1) \ln(\phi_{ijt}) \]

- \( X_{ijt} = p_{jt} q_{ijt} + c_{ijt} T_{ijt} \): total spending by consumers from city \( i \) in destination city \( j \)
- Plug in \( P_{ijt} = \left( (p_{jt})^{1-\delta} + (c_{ijt}/\epsilon_{ijt})^{1-\delta} \right)^{\frac{1}{1-\delta}} \) from the previous step
- Add city-pair FE and origin/destination*monthFE, absorbing \( \ln(P_{it}) \) and \( \ln(X_{it}) \)
- Obtain \( \sigma \) and \( \phi_{ijt} \) to construct final city-level price index \( P_{it} \)
- Instrument \( \ln(P_{ijt}) \) with HSR indirect connection dummy
Estimation of the demand system: Results

<table>
<thead>
<tr>
<th>Columns Variables</th>
<th>(1) (\ln(\text{air/non-air}))</th>
<th>(2) (\ln(\text{trip/consumption}))</th>
<th>(3) (\ln(X_{ijt}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation steps</td>
<td>Inner layer</td>
<td>Middle Layer</td>
<td>Outer layer</td>
</tr>
<tr>
<td>(\ln(\text{cost air/cost non-air}))</td>
<td>-2.33** (0.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\ln(\text{travel cost}))</td>
<td></td>
<td>-0.07 (0.20)</td>
<td></td>
</tr>
<tr>
<td>(\ln(P_{ijt}))</td>
<td></td>
<td></td>
<td>-1.92* (1.15)</td>
</tr>
</tbody>
</table>

Model interpretation: \(-\rho\), \(-\delta\), \(1 - \sigma\)

Estimation method: IV with connect dummy

Observations: 81,807 | 1,927,482 | 1,927,482
R-squared: 0.69 | 0.18 | -1.89
Results

- A direct HSR connection leads to 13% drop in bilateral travel costs
- Trips and spending in destination city very closely complementary, with an elasticity of substitution around 0.07
- Substitution elasticity between different cities around 2.9
- Removing the whole HSR network increases $P_{it}$ by 2.8% on average
  - Our model did not take into consideration of utility from local consumption. Accounting for it would mean the total effects on aggregate welfare to be around 0.2*2.8% given the share of out-of-city spending
Limitations of the current framework:

- Limited substitution patterns across different transportation modes
- Choice over transportation mode is multi-dimensional: fare cost, time, frequency, delays etc.
- Passenger heterogeneity: different groups of people have different valuation over these characteristics (income; business vs. personal trips)
- Distributional consequences

Extend the current framework to allow for:

- Multiple transportation mode characteristics
- Heterogeneity across income distribution
Random Coefficient Mixed Logit Framework

- Nested logit: consumers choose destination city first, then transportation modes
- The utility of consumer $i$ travelling to city $k$ by travel mode $j$ is defined as

$$U_{ijt} = x_{ikt} \beta_i + x_{ijt} \eta_i - \alpha_i p_{ijt} + \nu_{ikt} + \xi_{ijt} + \mu_{ijt}(\lambda), j \in C_{ikt} \tag{2}$$

- $x_{ikt}$ is a vector of destination city characteristics, such as city GDP, population, tourist attractions, etc.
- $x_{ijt}$ is a vector of transportation mode characteristics (duration, frequency of flights etc.),
- $\nu_{ikt}$ is the unobserved (to researchers) characteristic/amenities of city $k$ to residents from city $i$.
- $\xi_{ijt}$ is the unobserved (to researchers) characteristic of travel mode $j$ that deviates from the nest average
- $\mu_{ijt}$ is a nested logit random taste shock (Type I extreme-value distribution)
Random Coefficient Mixed Logit Framework

\[ U_{ijt} = x_{kt}\beta_i + x_{ijt}\eta_i - \alpha_i p_{ijt} + v_{ikt} + \xi_{ijt} + \mu_{ijt}(\lambda), j \in C_{ikt} \] (3)

- Passenger Heterogeneity: We assume \( \beta_i \) and \( \eta_i \) to be functions of observed and unobserved household demographics:
  \[ \beta_i = \beta + \alpha_h z_{ih} + \epsilon_i \] (4)

- We randomly draw individuals from city income distributions and pick parameters to minimize the distance between simulated market shares and
  - The share of origin \( i \) passengers travelling to different destination \( k \)
  - The share of passengers travelling by air for each city pair per month
Conclusion

- Goal of paper: combine various novel datasets to evaluate the impacts of HSR in China
  - Bilateral consumption and travel patterns in China using card transaction information
  - Evolution of transportation network and travel cost of various modes of transportation over time
- Framework for assessing welfare impact of passenger transportation infrastructure improvements via a “revealed-preference”-like approach
### Estimation of the demand system: Results

<table>
<thead>
<tr>
<th>Columns Variables</th>
<th>(1) ( \ln(\text{air/non-air}) )</th>
<th>(2) ( \ln(\text{trip/consumption}) )</th>
<th>(3) ( \ln(X_{ijt}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation steps</td>
<td>Inner layer</td>
<td>Middle Layer</td>
<td>Outer layer</td>
</tr>
<tr>
<td>( \ln(\text{cost air/cost non-air}) )</td>
<td>-3.86***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln(\text{travel cost}) )</td>
<td></td>
<td>-0.67***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td>( \ln(P_{ijt}) )</td>
<td></td>
<td></td>
<td>-1.46***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.35)</td>
</tr>
<tr>
<td>Model interpretation</td>
<td>(-\rho)</td>
<td>(-\delta)</td>
<td>(1 - \sigma)</td>
</tr>
<tr>
<td>Estimation method</td>
<td>IV with indirectconnect dummy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>82,027</td>
<td>2,000,336</td>
<td>2,000,336</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.90</td>
<td>0.24</td>
<td>-6.64</td>
</tr>
</tbody>
</table>