# Quantifying the substitutability and complementarity between high-speed rail and air transport

Fangni Zhang[a,b], Daniel J. Graham[a,*], Mark Siu Chun Wong[a]

[a] *Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

[b] *Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom*

**Abstract**

This paper quantifies the substitution and complementary effects of high-speed rail (HSR) on air travel demand in terms of both route traffic and airport enplanement. Employing the difference-in-differences (DID) method, the first part of the analysis measures the effect of new HSR routes on parallel air route traffic with a focus on East Asian regions (Mainland China, Japan, South Korea, and Taiwan). The second part examines the effect of air-HSR integration on passenger enplanement at East Asian airports and compares with that in the Central European market. We find that in general the airport's access cost (reflected by the distance from central city) has a negative impact on the air traffic. The substitution effects of HSR are the most significant on short- and medium-haul (below 1000km) air routes while introducing HSR services has encouraged long distance (over 1000km) air travels in Mainland China. The complementary effect is investigated in the context of air-HSR integration, which has significantly positive impacts on airport enplanement at primary hub airports when fitted with on-site HSR links. The benefit is limited at secondary hubs and regional airports possibly by locations and HSR service frequencies.

**Keywords:** High-speed rail (HSR); air transport; air-HSR integration; substitution effect; complementary effect.

# 1 Introduction

Since the inauguration of the first high-speed railway (HSR) with a speed of 210km/h between Tokyo and Osaka in 1964, HSR has revitalised the railway industry while potentially causing decreased air transport patronage due to competition. In South Korea, there were 20-90% reductions in passengers on domestic air routes after HSR services were launched in 2004

*Corresponding author. Email: d.j.graham@imperial.ac.uk.

where most domestic routes are shorter than 450km (Park and Ha, 2006). In Japan, airlines were forced out of the Tokyo-Sendai, Tokyo-Morioka and Tokyo-Niigata markets by HSR (JR East, 2016). In Europe, air passengers between Frankfurt and Cologne (180km) dropped by two-thirds in three years after HSR entered in 2002, and the air route was eventually axed in 2009 (Clewlow et al., 2012). These qualitative observations imply that HSR has a substitution effect on air transport.

There have been attempts to quantitatively measure the substitution effect of HSR on air transport. For example, a series of passenger preference surveys were conducted by Park and Ha (2006), Román et al. (2007), Burge et al. (2011) and Behrens and Pels (2012), which showed that service frequency, travel distance, and access time are the key determinants of HSR's competitiveness against airlines. Others such as Jiménez and Betancor (2012), Castillo-Manzano et al. (2015), and Clewlow et al. (2014) developed econometric model to study airlines' responses to HSR and suggested that the magnitude of responses in the European market depends on the route distance, population and economic environment. Albalate et al. (2015) and Wan et al. (2016) analysed the changes in available seats provided on domestic air routes to capture airlines' reactions to HSR in Europe and East Asia respectively, and they both found negative impacts of HSR on the operating capacity of airlines. This body of work provided important insights on the competition between HSR and airline services. However, little attention has been paid to the ex-post impact of HSR on the realised travel demand in East Asia where HSR networks have experienced explosive growth over the last decade. Since realised demand is always less than the provided capacity, the estimates for the latter term cannot be adopted to the former otherwise there is a risk of underestimation. Moreover, a large number of HSR links in East Asia were opened in the 2010s. Given that existing studies have only covered the early 2010s, it is valuable to update our knowledge with latest data that are available.

In addition to the widely acknowledged substitution effect, it is important to note that some airports in Europe and Asia have witnessed cooperation and integration with HSR infrastructure (air-HSR integration). Table 1 summarizes the airports with immediate intercity HSR connections (data obtained from various HSR operators). By air-HSR integration, HSR services act as additional spokes of airlines to free up airport slots and enlarge airport catchment areas. Frankfurt and Paris CDG are among the successful pioneers. The traffic evolution of Frankfurt airport over the last 40 years evidenced that efficient access to intermodal transport facilities could improve the airport's competitiveness (Airports Commission, 2014).

Table 1: Summary of European and Asian airports integrated with HSR stations

| Continent | Country | Airport | IATA code | HSR inaugural year |
|---|---|---|---|---|
| Europe | France | Paris Charles de Gaulle | CDG | 1994 |
| Europe | France | Lyon Saint-Exupry | LYS | 1994 |
| Europe | Netherlands | Amsterdam Schiphol | AMS | 1996 |
| Europe | Germany | Frankfurt International | FRA | 1999 |
| Europe | Germany | Dusseldorf International | DUS | 2000 |
| Europe | Denmark | Copenhagen Kastrup | CPH | 2000 |
| Europe | United Kingdom | Birmingham International | BHX | 2002 |
| Europe | Germany | Leipzig Halle | LEJ | 2003 |
| Europe | Belgium | Brussels National | BRU | 2003 |
| Europe | Germany | Cologne Bonn | CGN | 2004 |
| Asia | China | Shanghai Hongqiao | SHA | 2010 |
| Asia | China | Changchun Longjia | CGQ | 2011 |
| Asia | China | Haikou Meilan | HAK | 2011 |
| Asia | China | Shijiazhuang Zhengding | SJW | 2013 |
| Europe | Austria | Vienna International | VIE | 2014 |
| Asia | South Korea | Seoul Incheon | ICN | 2014 |
| Asia | China | Chengdu Shuangliu | CTU | 2015 |
| Asia | China | Guiyang Longdongbao | KWE | 2015 |
| Asia | China | Lanzhou Zhongchuan | LHW | 2015 |
| Asia | China | Zhengzhou Xinzheng | CGO | 2015 |
| Asia | China | Sanya Phoenix | SYX | 2016 |

Through theoretical analysis and numerical simulations, Jiang and Zhang (2014, 2016), Takebayashi (2016), and Xia and Zhang (2016, 2017) modelled the cooperative behaviours of airlines and HSR operators, and examined the effects of cooperation in different scenarios. Jiang and Zhang (2016) suggested that the development of HSR would reform the airline network from the fully-connected to the hub-and-spoke structure. The transition would help airlines in protecting their market share on important trunk routes, while the HSR can provide a feeding service from hub airports. Takebayashi (2016) advocated the multiple-hub system, where secondary gateway airports are connected to HSR networks to reduce congestion at primary hubs.

Air-HSR integration is gaining popularity worldwide for its twofold merits – enhancing airport accessibility being the one and relieving airport traffic pressure the other. However, empirical studies of the integration effects are very limited. Albalate et al. (2015) studied the airline service frequency and available seats data and suggested that HSR can provide feeding services to long haul air services in hub airports, which implies that air-HSR integration can help improve intercity mobility and bring benefits to both passengers and operators. The impact of air-HSR integration on realised travel demand by air has not been examined. Given the availability of traffic data at the airport level, there is a good opportunity to rigorously examine complementarity between the two modes.

To these ends, this paper employs econometric methods to quantify the substitution and complementary effects of HSR on air transport based on realised traffic data, and to establish causal relationships between air passenger movement and HSR inventions of different forms. Specifically, the substitute role of HSR is mainly measured by the impact of new HSR routes on the passenger movement in parallel air routes using difference-in differences (DID) estimation. This part of analysis focuses on East Asian regions, particularly Japan, Korea, mainland China, and Taiwan, where an upward trend of HSR development has been observed recently. Air routes are categorised into groups according to their great circle distances to further investigate the effect of travel distance.[1] The results show that substitution effects are most significant in short- to medium-haul (below 1000km) markets. With many long-distance HSR corridors opening to traffic in mainland China, the impacts of HSR entries on air routes longer than 1500km are studied for the first time. It is observed that HSR services have a positive impact on airline traffic in the longer distance (over 1000km) markets of Mainland China, indicating that the introducing HSR into these markets can reinforce (rather than cutting) the travel demand for airline services.

The complementarity is investigated by examining the impact of air-HSR integration on airport passenger enplanement. Data from both East Asian regions (Mainland China, Japan, South Korea) and Europe are collected, where air-HSR integration has been practised during the sampling period. We find that while air-HSR integration has a positive and significant impact on airport enplanement at primary hub airports, it has much less impact on secondary hubs and regional airports, and even a negative impact in Europe. To our knowledge, this is the first empirical study to quantify complementarity between air transport and HSR explicitly in the context of air-HSR integration using econometric methods.

The remainder of this paper is organised as follows. Section 2 introduces the methodological approach and model setup. Section 3 describes the datasets visited to collect the data, the air routes and airports included in the sample, and the variables used in the analysis. Section 4 examines the effect of HSR entries on air route traffic and Section 5 looks into the effect of air-HSR integration on airport traffic. Section 6 concludes the paper and discusses

[1] Wan et al. (2016) studied airlines' responses to HSR in East Asia using a DID approach and focused on the the number of available seats (the capacity provided by airlines). In this paper we focus on passenger movements (the realised demand of passengers) and for the first time include the Taiwanese market, and a sizeable number of HSR routes that opened between 2012-2014, some of which cover distances over 1500km.

the policy implications and limitations of this study.

# 2 Methodology and model setup

## 2.1 The difference-in-differences (DID) method

In this paper we use difference-in-differences (DID) estimation to estimate the casual effects of HSR on air traffic. The DID method is based on comparing the observations from different individuals in the 'Treatment Group' versus the 'Control Group' in a natural experiment. It is a common econometric technique to assess the impact of policy interventions (Wooldridge, 2010, 2015). Estimation of treatment effects is based on data that take the the form of a random vector, $z_i = (y_i, d_i, x_i)$, $i = 1, ..., n$, where for the $i$-th unit of observation $y_i$ denotes a response, $d_i$ the treatment (or exposure) received, and $x_i$ a vector of pretreatment covariates. For treatment level $D = d$, where $d$ could take values in $\mathcal{D} \in \{1, 0\}$, or in $\mathcal{D} \equiv (d_0, d_1, ..., d_k)$, or in some bounded interval $\mathcal{D} \subseteq \mathbb{R}$; we assume the existence of a set of potential outcomes for unit $i$: $\mathcal{Y}_i = \{Y_i(d), d \in \mathcal{D}$ for $i = 1, ..., n\}$ and the full data for consideration of causal effects is then taken to be $(\mathcal{Y}_i, D_i, X_i)$. The target of the causal inference is the Average Treatment Effect (ATE) of the form $\tau(d) = \mathbb{E}\{Y_i(d)\} - \mathbb{E}\{Y_i(0)\}$.

The basic DID model compares outcomes $Y_{it}$, for units $i$, $i = (1, 2, ..., N)$ with binary treatment effect $D \in \{0, 1\}$ ($D = 1$ represents the treatment group and $D = 0$ the control group) in two time periods $t \in \{0, 1\}$ ($t = 0$ indicates the pretreatment period and $t = 1$ the post-treatment period) using

$$Y_{it} = \mu + \alpha \cdot D_i + \delta_t \cdot t + \tau_D \cdot (D_i \times t) + X_i^{'}\beta + \varepsilon_{it}, \tag{1}$$

where $\varepsilon_{it}$ is a potentially autoregressive error with mean zero in each time period. The ATE is captured by the parameter $\tau_D$, which is the sample counterpart to

$$\begin{aligned} &\tau_D = \mathbb{E}[Y_i(1)|X_i] - \mathbb{E}[Y_i(0)|X_i] \\ &= \{\mathbb{E}[Y_{i,1}|X_i, D_i = 1] - \mathbb{E}[Y_{i,0}|X_i, D_i = 1]\} - \{\mathbb{E}[Y_{i,1}|X_i, D_i = 0] - \mathbb{E}[Y_{i,0}|X_i, D_i = 0]\}, \end{aligned} \tag{2}$$

with least squares estimate

$$\hat{\tau}_D = \left(\overline{Y}_{11} - \overline{Y}_{10}\right) - \left(\overline{Y}_{01} - \overline{Y}_{00}\right), \tag{3}$$

where $\overline{Y}_{11}$ is the sample average outcome for treated units in year 1.

In this paper, a route $r$ is defined to be 'treated' if HSR service began operation on that route within the sampling period when modelling route traffic; and an airport $p$ is 'treated' if it is directly connected to intercity HSR networks when analysing the impact of air-HSR integration. Within the sampling period $t \in \{1, 2..., T\}$, time points of treatment vary across the individuals. To capture the yearly effect, the DID model can be reformulated by including a vector of year dummies:

$$Y_{it} = \mu + \alpha \cdot D_i + \sum_{t=1}^{T} \delta_t \cdot I_t + \tau_D \cdot (D_i \times I_t) + X_{it}^{'}\beta + \varepsilon_{it}, \tag{4}$$

where $I_t$ is an indicator variable for year $t$, and the interaction term $(D_i \times I_t)$ is the group-time treatment indicator which takes a value of 1 for groups and time periods that were subject to the treatment and 0 otherwise.

## 2.2 Pretreatment trend and propensity score analysis

The key assumption underlying the DID model is that the average outcomes for the treatment and control groups would have followed parallel paths over time in the absence of the treatment. If $Y_{it}(0)$ is the outcome that unit $i$ experiences in time $t$ in the absence of treatment, then for binary period $t \in \{0, 1\}$, DID requires the following assumption.

**Assumption 1.** *(Unconditional parallel trend). For identification of treatment effects in the basic DID model it is necessary that the average outcomes for the treatment and control groups would have followed parallel paths over time in the absence of the treatment,*

$$\mathbb{E}_i[Y_{i,1}(0) - Y_{i,0}(0)|X_i, D_i = 1] = \mathbb{E}[Y_{i,1}(0) - Y_{i,0}(0)|X_i, D_i = 0]. \tag{5}$$

The above assumption ensures that over the long run, the groups are comparable prior to treatment so that a difference in outcome reflects the effect of treatment. When it is violated, the treatment and control groups may differ in their pretreatment characteristics: subjects who receive the treatment may be systematically different from those who do not. Pretreatment differences may cause a difference in outcomes, rather than the treatment itself causing the difference (Adelson, 2013).

To deal with this, analytical tools are needed to adjust for these systematic differences between treatment and control groups with respect to a number of pretreatment characteristics. The propensity score can be used to model the relationship between pretreatment variables and treatment assignment as it represents the conditional probability of assignment to treatment based on measure pretreatment characteristics (Rosenbaum and Rubin, 1983). Statistically, the propensity score, $s(x)$, is the conditional probability of receiving the treatment given the observed pretreatment variables, $x$, given by

$$s(x) = prob(d_i = 1|x). \quad (6)$$

As defined above, $d_i$ is a binary variable indicating whether the $i$-th observation belongs to the treatment or control group, and $x$ denotes the pretreatment covariates. For subjects with the same propensity score, the joint distribution of the observed covariates is balanced between the treatment and control groups meaning that they have the equal probability of receiving treatment but accidentally appear in different groups (Rosenbaum and Rubin, 1983). The propensity score controls for systematic differences in background characteristics between treatment and control groups and reduces a number of pretreatment variables into a single composite indicator (Rubin, 1997).

Based on propensity scores, one-to-one matching (propensity score matching, PSM) or stratification can be applied to reconstruct a situation similar to random treatment assignment after the fact (Braitman and Rosenbaum, 2002). The rationales, models and practical methods for propensity score analysis and matching have been well established in, e.g., Rosenbaum and Rubin (1983), Holland (1986), Rubin (1986, 1997), Braitman and Rosenbaum (2002), Rudner and Peyton (2006), Adelson (2013), and Randolph and Falbe (2014). In a precedent study of airlines' responses to HSR-entry, Wan et al. (2016) employed the propensity score matching techniques prior to DID analysis. Similar methods are adopted here to reduce the selection bias.

## 2.3 Air route traffic models

The effect of HSR interventions on air route traffic is analysed using the model adopted from Eq. (4):

$$\begin{aligned} ROUTE_{rt} = \beta_0 + \beta_1 \cdot TREATED_r + \beta_2 \cdot YEAR_t + \beta_3 \cdot HSR_{rt} \\ + \beta_4 \cdot POP_{rt} + \beta_5 \cdot GDP_{rt} + \beta_6 \cdot ACCESS_{rt} + \varepsilon_{rt}, \quad (7) \end{aligned}$$

where $ROUTE_{rt}$ represents the total number of passengers that travelled on route $r$ in year $t$, and $TREATED_r$ is a dummy variable that is equal to one if HSR services were introduced to route $r$ during the sampling period, and zero otherwise; it is included in the model to capture the differences between observations in the treatment group and those in the control group before HSR interventions were introduced. $YEAR_t$ are year dummies that represent the year-specific fixed effects; they capture the changes in the dependent variable between the base year and year $t$ in absence of HSR entries. For example, the dummy variable for the year 2000 is equal to one for observations taken in 2000, and equates to zero in other years within the observation period. $HSR_{rt}$ (corresponding to the interaction term $(D_i \times I_t)$ in Eq. (4)) is the policy variable that indicates treatment status; it takes the value of one if HSR services are operating on route $r$ in year $t$, and zero otherwise. Therefore, $\beta_3$, the coefficient of the policy variable, measures the average impact of HSR entries on airline traffic at the route level, and is thus the coefficient of interest in the first part of the analysis.

The demographic and economic environment (population and GDP) are commonly acknowledged as the major determinants of air traffic and are widely employed as explanatory variables of air traffic in empirical studies (e.g., Brueckner, 1985; IATA, 2007; Chi and Baek, 2012; Albalate et al., 2015; Airbus, 2016). Since the aggregate GDP is normally highly correlated with population, this study thus considers the per capita GDP and the percentage of GDP that came from the service sector and the population of the concerned catchment areas as explanatory variables in order to eliminate multicollinearity. In addition, previous studies (e.g., Yao and Morikawa, 2005; Román et al., 2007; Adler et al., 2010) showed that passengers are concerned with the accessibility to airports when facing the alternative of travelling by rail. The impact of the airport accessibility ($ACCESS$) on travel demand is also of great interest in this study. $\varepsilon_{rt}$ is the error term.

It is suggested by literature (e.g., Albalate and Bel, 2012; Wan et al., 2016; Xia and Zhang, 2016) that HSR is more competitive than air transport in short-to-medium haul passenger markets, but its competitive edge diminishes with travel distance. To investigate the impact of travel distance, Eq. (7) is extended to incorporate distance classes into the model. Air routes are categorised into four classes according to their great circle distance (Swartz, 2017) which are represented by four dummy variables, $D1$ (up to 500km), $D2$ (501-1000km), $D3$ (1001-1500km) and $D4$ (over 1500km). In the extended model, $HSR_{rt}$ is replaced by interaction terms of distance dummies and policy variable as follows:

$$\begin{aligned} ROUTE_{rt} = \beta_0 + \beta_1 \cdot TREATED_r + \beta_2 \cdot YEAR_t \\ + \beta_3 \cdot (D1 \times HSR)_{rt} + \beta_4 \cdot (D2 \times HSR)_{rt} + \beta_5 \cdot (D3 \times HSR)_{rt} + \beta_6 \cdot (D4 \times HSR)_{rt} \\ + \beta_7 \cdot POP_{rt} + \beta_8 \cdot GDP_{rt} + \beta_9 \cdot ACCESS_{rt} + \varepsilon_{rt}, \quad (8) \end{aligned}$$

where the interaction term, for example, $(D2 \times HSR)_{rt}$ takes the value of one if air route $r$, with a great circle distance between 501km and 1000km, faces competition from HSR services in year $t$, and equals to zero in all other cases.

## 2.4 Airport traffic models

The effect of air-HSR integration on airport traffic is analysed using model:

$$AIRPORT_{pt} = \beta_0 + \beta_1 \cdot TREATED_p + \beta_2 \cdot YEAR_t + \beta_3 \cdot INTEGRT_{pt} \\ + \beta_4 \cdot POP_{pt} + \beta_5 \cdot GDP_{pt} + \beta_6 \cdot ACCESS_{pt} + \varepsilon_{pt}, \quad (9)$$

where $AIRPORT_{pt}$ represents the total number of passengers that used airport $p$ in year $t$, and $TREATED_p$ is a dummy variable that is equal to one if air-HSR integration was introduced to airport $p$ during the sampling period, and zero otherwise. $YEAR_t$ are year dummies that represent the year-specific fixed effects; they capture the changes in $AIRPORT_p$ between the base year and year $t$. $INTEGRT_{pt}$ is the policy variable that indicates the treatment status; it takes the value of one if air-HSR integration was introduced to airport $p$ in year $t$, and zero otherwise. Therefore, $\beta_3$, the coefficient of the policy variable, measures the average impact of air-HSR integration on airport travel demand, and thus is the coefficient of interest in this part of the analysis. Similar to the air route traffic models, the demographic indicator ($POP$), the economic indicator ($GDP$), and the airport accessibility measurement ($ACCESS$) are included in the model to describe background characteristics of airports. $\varepsilon_{pt}$ is the error term.

It is suggested in Albalate et al. (2015) that the impact of HSR entries on airline services (in terms of service frequency and seats provided) is more significant at major hubs than secondary hubs or regional airports. In order to investigate the impact of hub status on the air-HSR integration, Eq. (9) can be extended to incorporate hub status in the model. Airports are sorted into two classes which are represented by two dummy variables, $H1$ (primary hubs) and $H2$ (secondary hubs and regional airports). In the extended model, the policy variable $INTEGRT_{pt}$ is replaced by interaction terms of the hub dummies and the policy variable as follows:

$$AIRPORT_{pt} = \beta_0 + \beta_1 \cdot TREATED_p + \beta_2 \cdot YEAR_t \\ + \beta_3 \cdot (H1 \times INTEGRT)_{pt} + \beta_4 \cdot (H2 \times INTEGRT)_{pt} \\ + \beta_5 \cdot POP_{pt} + \beta_6 \cdot GDP_{pt} + \beta_7 \cdot ACCESS_{pt} + \varepsilon_{pt}, \quad (10)$$

where the interaction term, for example, $(H1 \times INTEGRT)$ takes the value of one if airport $p$ is a primary hub and has a dedicated HSR station onsite in year $t$, and equals to zero otherwise.

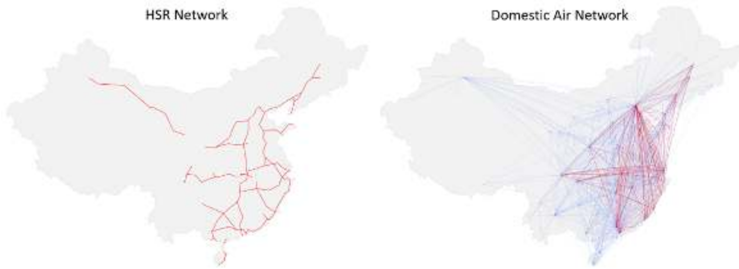
# 3 Data

## 3.1 Air route traffic data and explanatory variables

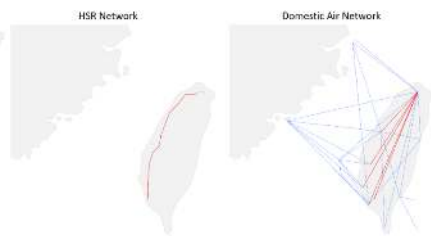
To measure the impact of the opening of HSR links on passenger movements on parallel air routes in East Asia, we constructed a regional dataset comprising both domestic air routes and HSR-related data from four East Asian economies, namely, Japan, South Korea, Taiwan and mainland China. Domestic air route traffic data are collected from the Ministry of Land, Infrastructure, Transport and Tourism (MLIT, 1989-2015 available) of Japan, Korea Airports Corporation (KAC, 1997-2016 available) of South Korea, the Civil Aeronautics Administration (CAA, 2000-2016 available) of Taiwan, and the Civil Aviation Administration of China (CAAC, 1998-2014 available). Since the common observation period is 2000-2014, observations before 2000 and after 2014 are removed from the database. HSR entry data are obtained from UIC (2017), and validated by articles from various media outlets. HSR entering service in the fourth quarter of the year are assumed to open the following year. For example, the Wuhan-Guangzhou HSR in mainland China opened to traffic on 26th December 2009, so the inaugural year of this link was taken as 2010, so that the impact of HSR is recorded in the air traffic count of 2010.

A route is defined to be 'treated' if direct HSR services began operations between the city pair within the sampling period, and routes that do not have HSR entries during the observation period belong to the 'control' group. It should be noted that routes are only considered to be treated if HSR provides direct connections that do not require interchanges, and operate above 200km/h for the majority of the routes. Figure 1 illustrates the domestic air routes and HSR networks in the concerned economies. The red lines represent the treated air routes that faced HSR entries between 2000 and 2014, and the blue lines represent air routes in the control groups including those where HSR services began revenue services before 2000.

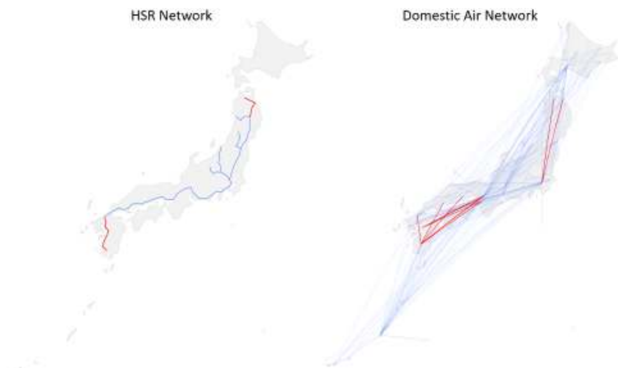
The full sample consists of a panel of 1178 routes for the observation period of 2000-2014. There are 245, 10, 9, and 5 treated routes in Mainland China, Japan, South Korea and Taiwan respectively. Table 2 presents descriptive statistics of annual passenger volume on air routes ($ROUTE$), which is the dependent variable in air route traffic models.



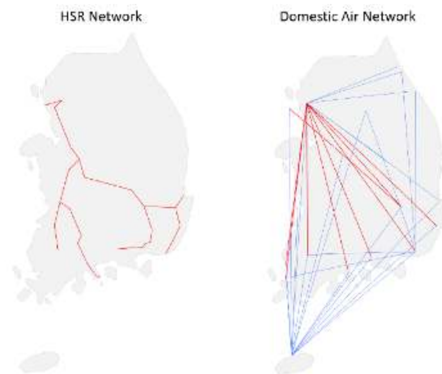

(a) Mainland China




(b) Taiwan




(c) Japan




(d) South Korea

Figure 1: Domestic air and HSR networks in (a) Mainland China, (b)Taiwan, (c)Japan, and (d)South Korea.

Table 2: Descriptive statistics of air route traffic

| Sample | Obs. | Mean | Standard deviation (SD) |
|---|---|---|---|
| Full sample | 10,293 | 366,215.4 | 697,343.3 |
| Full sample (Treated) | 2,529 | 501,707.5 | 719,807.7 |
| Full sample (Control) | 7,764 | 322,081.0 | 684,145.1 |
| Distance⩽500km | 2,688 | 300,303.4 | 760,466.4 |
| 501km<Distance⩽1000km | 3,633 | 347,301.0 | 673,911.6 |
| 1001km<Distance⩽1500km | 2,494 | 436,567.0 | 657,763.0 |
| Distance>1500km | 1,478 | 413,868.3 | 685,560.2 |
| (1) Mainland China | 6,307 | 374,758.2 | 515,113.9 |
| Mainland China (Treated) | 2,214 | 488,169.4 | 684,988.5 |
| Mainland China (Control) | 4,093 | 313,411.4 | 380,039.6 |
| (2) Japan | 3183 | 310,395.7 | 786,916.1 |
| Japan (Treated) | 121 | 386,588.9 | 347,165.8 |
| Japan (Control) | 3,062 | 307,384.8 | 799,219.4 |
| (3) South Korea | 353 | 855,820.9 | 1,803,146.0 |
| South Korea (Treated) | 119 | 833,750.8 | 1,166,610.0 |
| South Korea (Control) | 234 | 867,044.5 | 2,054,815.0 |
| (4) Taiwan | 450 | 257,247.6 | 509,163.5 |
| Taiwan (Treated) | 75 | 560,235.7 | 1,046,295.0 |
| Taiwan (Control) | 375 | 196,650.0 | 269,331.1 |

The population and economic data are collected from the World Bank, Statistics Japan, Statistics Korea, the Department of Household Registration of Taiwan, and the National Bureau of Statistics of China. The population data for South Korean and Taiwanese cities is updated on a yearly basis, while it is published every five years in Japan, and every ten years in mainland China. As a result, the population between census years are interpolated from census data in 1990, 1995, 2000, 2005, 2010 and 2015 for Japan, and that in 1990, 2000 and 2010 for mainland China. It is found that polynomial regressions fit the data better than cubic regressions, and polynomial regressions with higher degrees of freedom do not appear to improve the model fit significantly.

The airport access difficulty is measured by the shortest driving distance between the airport and the city centre of its catchment area and is collected from Google maps, Baidu maps, and various airport official websites. The great circle route distances are obtained from Swartz's (2017) online great circle mapper.

Since an air route involves two catchment areas associated with the origin and destination airports, the population variable of is defined by the total population in the two catchment areas (POP_TOT). The GDP variable includes two candidate covariates: average per capita GDP (GDP_PC), and the average percentage GDP contributed by service sectors (GDP_SERVICES) of the two catchment areas. Likewise, the airport access distance stands for the sum of access distance of the origin and destination airports (ACCESS_TOT). Table 3 presents the descriptive statistics of the explanatory variables for air route traffic. Table 4 summaries the distribution of great circle distances of the treated routes in the sample.

Table 3: Descriptive statistics for explanatory variables of air route traffic models

| Sample | Obs. | Mean (SD) | | | |
|---|---|---|---|---|---|
| | | POP_TOT ($\times 10^3$prs) | GDP_PC ($\times 10^3$\$) | GDP_SERVICES (%) | ACCESS_TOT (km) |
| Full sample | 10,293 | 9836.1 (6,992.8) | 16.3 (15.9) | 53.6 (14.4) | 54.5 (32.0) |
| Full sample (Treated) | 2,529 | 13809.7 (6,695.5) | 7.8 (8.5) | 46.8 (10.0) | 59.1 (19.8) |
| Full sample (Control) | 7,764 | 8,541.8 (6,588.7) | 19.1 (16.7) | 55.8 (14.9) | 53.1 (34.9) |
| (1) Mainland China | 6,307 | 11,065.9 (6,022.2) | 4.7 (2.7) | 43.0 (6.4) | 60.5 (22.9) |
| (2) Japan | 3,183 | 7,601.7 (7,577.4) | 38.8 (4.6) | 72.0 (1.5) | 50.3 (43.1) |
| (3) South Korea | 353 | 14264.8 (10,935.2) | 19.1 (5.3) | 59.5 (0.9) | 37.5 (25.6) |
| (4) Taiwan | 450 | 4930.2 (3,830.0) | 17.8 (2.9) | 65.9 (1.2) | 15.5 (5.5) |

Table 4: Summary of great circle distances of treated routes

| Great circle distance of treated routes | D1 (0-500km) | D2 (501-1000km) | D3 (1001-1500km) | D4 (>1500km) | Total |
|---|---|---|---|---|---|
| Full Sample | 54 | 125 | 64 | 26 | 269 |
| (1) Mainland China | 39 | 116 | 64 | 26 | 245 |
| (2) Japan | 1 | 9 | 0 | 0 | 10 |
| (3) South Korea | 9 | 0 | 0 | 0 | 9 |
| (4) Taiwan | 5 | 0 | 0 | 0 | 5 |

## 3.2 Airport traffic data and explanatory variables

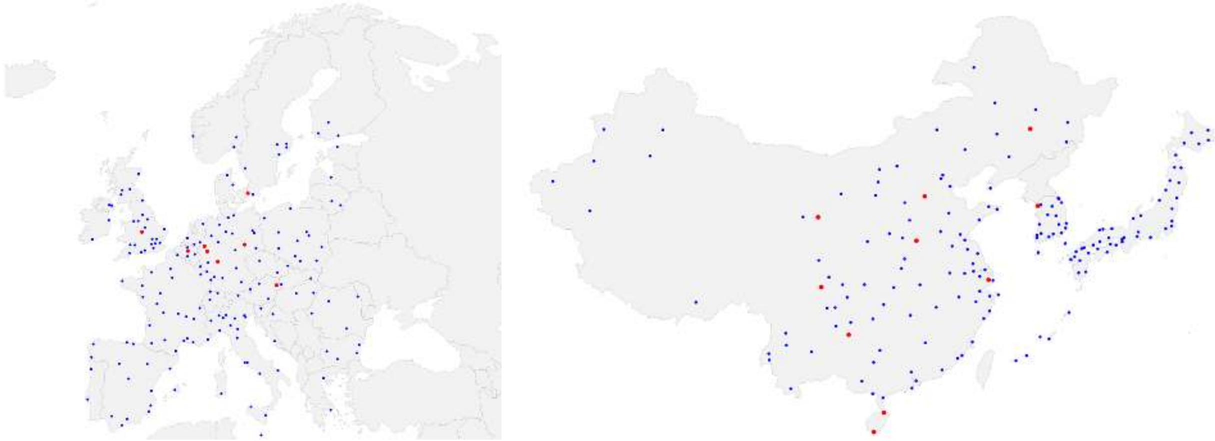
To further investigate the complementarity between HSR and air transport, the second part of this paper analyses the impact of air-HSR integration on air traffic at the airport level. The annual passenger enplanement data are collected from the MLIT for Japanese airports (2000-2016 available), the KAC for South Korean airports(1997-2016 available), and the CAAC for Chinese airports (2000-2016 available). For European airports (1997-2015 available), the Eurostat database is complemented by data from the Ministère de l'Environnement, de l'Energie et de la Mer (MEEM) of France, Statistics Belgium, Statistics Netherlands and Flughafenverband ADV of Germany. Note that many European airports have good rail connections to urban areas and some of them operate at or above 200km/h, such as the Flytoget service at Oslo Airport and the Arlanda Express at Stockholm. However, these train services merely operate within their respective metropolitan areas and fail to expand the airports' catchment areas to other urban agglomerations. Therefore, we consider an airport to be 'treated' (have dedicated HSR services) if HSR services are intercity and operate at or above 200km/h for most of their journeys. The inaugural years of air-HSR integrations are listed in Table 1. Where an HSR station opened in the fourth quarter of a particular year it is considered to take effect in the subsequent year.

Given that the air-HSR integration differs in many intrinsic aspects and inauguration years[2], separate analyses are conducted on different continents. Data from Central Europe are combined into one sample, and those from East Asia (particularly mainland China, South Korea and Japan) into another. The sampling period for Central European airports is 1997-2015, whereas that for East Asian airports is 2000-2016.[3] In Figure 2, we present the airports analysed in this study, in which red dots indicate airports that are treated and those in blue are in the control group.

In total, 180 airports (2602 observations) are included in the Central European sample, and 170 airports (2538 observations) in the East Asian sample. During the respective sampling periods, eight Central European airports and ten airports in East Asia were integrated into HSR networks.

[2]Most air-HSR integrations in Central European countries were introduced before 2005, while those in East Asia were not until the 2010s.

[3]Three treated airports in Europe which began HSR services before 1997, namely, Paris CDG, Lyon Saint-Exupéry and Amsterdam Schiphol, are excluded from the analysis.



(a) Central Europe (b) East Asia

Figure 2: Airports included in the study.

Table 5 summaries descriptive statistics of the dependent variable (passenger enplanement of airports per annum) and the explanatory variables involved in the airport traffic models. The data collection of explanatory variables for East Asian regions is described in Section 3.1. The economic and demographic data for European regions are collected from the Eurostat. Most of the population data for years before 2000 are missing from the Eurostat package, so the population of European metropolitan areas between 1995 and 2000 are estimated based on the average annual population growth rate between 2000 and 2010. The population (POP), the per capita GDP (GDP_PC), and the percentage contribution of service sector in GDP (GDP_SERVICES) in the airport catchment area, as well as the airport access distance (ACCESS), are used as explanatory variables.

It is noteworthy that most HSR services from airports are linked with HSR stations closer to cities (e.g., Amsterdam Schiphol, Frankfurt International), and there are typically multiple HSR stations in the catchment area of the airport in the 'Treatment Group'. Meanwhile, those in the 'Control Group' differ in their proximity to HSR services: some airports have overlapped catchment areas with HSR stations (in cities rather than on-site), while others do not. In the latter case, a difference in the airport traffic of 'Treatment Group' versus 'Control Group' is twofold – the airport traffic can be influenced not only by the air-HSR integration, but also by the availability of HSR service in the catchment area. This means that when analysing the effect of air-HSR integration based on the difference between 'Treatment Group' and the full 'Control Group', the effect of air-HSR integration may be contaminated by effect of introducing HSR in the broader area. To deal with this, airports in the whole 'Control Group' (referred to as the 'Full Control Group' thereafter) are categorised into two subgroups: those with HSR connections within the catchment area ('Control Group-A') and those without HSR services nearby (‘Control Group-B’). The airport enplanement of the ‘Treatment Group’ is analysed with reference to not only the ‘Full Control Group’ but also the ‘Control Group-A’ to anatomise the effects

Table 5: Descriptive statistics for dependent and explanatory variables in airport traffic models

| | Sample | Obs. | Mean (SD) | | | | |
|---|---|---|---|---|---|---|---|
| | | | AIRPORT (prs) | POP ($\times 10^3$prs) | GDP_PC ($\times 10^3$\$) | GDP_SERVICES (%) | ACCESS (km) |
| Central Europe | All | 2,950 | 5,755,804 (9,667,274) | 1,901.2 (2,593.5) | 33.4 (15.4) | 71.3 (5.8) | 18.0 (16.8) |
| | Treatment Group | 143 | 18,143,953 (14,784,654) | 1,963.4 (531.9) | 38.8 (9.9) | 71.2 (3.4) | 23.3 (23.4) |
| | Full Control Group | 2,807 | 5,124,702 (8,884,694) | 1,898.1 (2,656.0) | 33.1 (15.6) | 71.3 (5.9) | 17.7 (16.4) |
| | Control Group-A | 1,314 | 6,842,927 (11,643,559) | 2,606.9 (3,427.7) | 35.8 (11.4) | 73.1 (3.9) | 15.0 (12.2) |
| East Asia | All | 2538 | 5,356,017 (10,603,751) | 2,671.7 (4,304.1) | 16.8 (16.2) | 54.7 (13.0) | 23.4 (23.8) |
| | Treatment Group | 169 | 11,871,741 (12,249,761) | 5,417.5 (5,865.2) | 5.6 (5.8) | 45.6 (5.5) | 33.7 (17.5) |
| | Full Control Group | 2,369 | 4,891,197 (10,323,380) | 2,475.8 (4,102.5) | 17.6 (16.4) | 55.3 (13.1) | 22.7 (24.0) |
| | Control Group-A | 1,386 | 6,884,541 (12,624,306) | 3,533.6 (4,395.6) | 14.6 (15.3) | 53.0 (12.4) | 24.2 (16.3) |

In order to investigate the impact of hub status, airports in the two samples are respectively sorted into two groups: $H1$ (Primary hubs) and $H2$ (Secondary hubs and regional airports). The sample includes three hub airports in East Asia and one in Central Europe.

# 4 Effect of HSR entries on air route traffic

This section analyses the impact of HSR entries on air route traffic based on the air route traffic models introduced in Section 2.3. The original dataset constructed in Section 3.1 is firstly used to build up initial estimations of Eqs. (7) and (8) in order to envisage impacts of all covariates involved. It is further employed to analyse the pretreatment trends of yearly fixed effects in the treatment and control groups. Then the propensity score matching is carried out in each sample to ensure the unconditional parallel trend assumption (Assumption 1) of the DID model is satisfied. Finally, the air route traffic models are re-estimated based on post-matching samples upon which interpretations of results are drawn.

## 4.1 Initial regression and propensity score matching

Table 6 presents the initial regression results of air route traffic models using ordinary least square (OLS) method with fixed effects. It includes two sets of results, one is based on formulation of Eq. (7) where the effect of HSR entry is captured by a single HSR dummy variable and thus the coefficient associated with HSR represents the average impact across a particular sample; another set is the estimation of Eq. (8) which involves four interaction terms of HSR dummy and differentiated great circle distance classes, each representing the impact of HSR entry in that particular class.

The estimation of Eq. (7) based on the pooled sample (East Asian four regions) yields a significant and negative coefficient for HSR, which is consistent with the common belief that HSR has substitution effect on air route traffic. Region-wise analysis further shows that the substitution effect is the greatest in Taiwan, followed by South Korea and Japan. In contrast, the HSR has an overall positive impact on the airline traffic in Mainland China.

Recall that all treated routes in South Korea and Taiwan are shorter than 500km as summarised in Table 4. Given that previous studies have shown that HSR is most competitive against airlines in short-haul markets, it is not surprising that the HSR policy variable is strongly negative in Taiwan and South Korea, where all domestic routes are below 500km. However, most HSR routes in Japan that entered service within the sampling period range between 500km and 750km, and in Mainland China, most treated routes have great circle distances between 500km and 1000km, and there is a sizeable number of treated routes that traverse over 1000km. In view of the wide range of route distance in these two regions, a single dummy variable is not sufficient to capture the full picture. Therefore, further analysis on distance class-specific air routes is carried out on the pooled, Mainland Chinese, and Japanese samples based on Eq. (8).[4]

Estimations of Eq. (8) further show that across the three samples, the scale of the substitution effect is the greatest in the short distance class (D1×HSR, below 500km), followed by the median distance (D2×HSR, 501-1000km). For distance classes over 1000km (D3×HSR and D4×HSR) in Mainland China, the coefficients of HSR are positive. Given that the number of observations over 1000km constitutes one-third of treated observations

[4]Since all treated routes in Taiwan and South Korea are below 500km, the D1×HSR estimator of Eq. (8) would return the same values as those of HSR of Eq. (7) making this operation redundant.

Table 6: Initial estimation of air route traffic models Eqs. (7) and (8)

| | East Asia | | (1) Mainland China | | (2) Japan | | (3) South Korea | (4) Taiwan |
|---|---|---|---|---|---|---|---|---|
| | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (7) |
| POP_TOT | 46.001*** | 46.069*** | 44.795*** | 44.300*** | 41.285*** | 41.352*** | 102.327*** | 49.832*** |
| | (0.957) | (0.952) | (0.937) | (0.931) | (1.748) | (1.751) | (12.063) | (7.069) |
| GDP_PC | −10, 063.460 | −11, 096.490 | −496, 774.400 | −412, 237.000 | 6, 729.874 | 6, 677.050 | −82, 041.730 | −27, 341.420 |
| | (18, 444.040) | (18, 378.504) | (842, 912.100) | (839, 911.601) | (31, 355.430) | (31, 358.310) | (610, 464.400) | (35, 759.870) |
| GDP_SERVICES | 15, 305.230*** | 16, 958.640*** | 464, 229.000*** | 388, 897.300*** | −4, 886.427 | −4, 941.604 | 840, 302.800 | −13, 481.340 |
| | (2, 260.440) | (2, 254.803) | (72, 622.160) | (72, 385.730) | (13, 039.220) | (13, 040.640) | (4, 766, 537.000) | (89, 668.370) |
| ACCESS_TOT | −681.936*** | −749.650*** | 244.606 | 219.484 | −955.547*** | −958.252*** | −3, 261.801 | −4, 357.955 |
| | (200.339) | (199.174) | (229.135) | (227.159) | (305.801) | (305.856) | (3, 633.264) | (3, 793.044) |
| TREATED | −9, 074.755 | −7, 054.021 | 1, 646.538 | 1, 721.529 | 34, 099.630 | 33, 854.990 | −1, 328, 723.000*** | 590, 170.800*** |
| | (16, 781.880) | (16, 676.430) | (12, 086.490) | (11, 981.250) | (81, 332.470) | (81, 340.520) | (329, 582.300) | (92, 516.820) |
| HSR | −87, 694.490** | | 74, 737.370*** | | −309, 040.500** | | −768, 424.500** | −1, 041, 598.000*** |
| | (34, 951.360) | | (28, 827.510) | | (144, 576.600) | | (339, 989.500) | (106, 344.900) |
| D1×HSR | | −445, 709.800*** | | −254, 348.900*** | | −80, 712.710 | | |
| | | (53, 474.450) | | (66, 148.510) | | (373, 627.100) | | |
| D2×HSR | | −75, 090.300 | | −36, 847.820 | | −335, 878.700** | | |
| | | (48, 128.450) | | (35, 813.540) | | (150, 153.100) | | |
| D3×HSR | | 497, 400.500*** | | 442, 429.800*** | | | | |
| | | (75, 967.470) | | (51, 866.580) | | | | |
| D4×HSR | | 548, 781.200*** | | 486, 460.300*** | | | | |
| | | (133, 737.800) | | (90, 006.000) | | | | |
| Constant | −698, 244.000*** | −766, 456.300*** | −18, 218, 777.000*** | −15, 296, 658.000*** | 136, 546.300 | 142, 009.700 | −47, 373, 654.000 | 1, 528, 838.000 |
| | (97, 854.370) | (97, 579.200) | (2, 806, 812.000) | (2, 797, 776.000) | (1, 203, 963.000) | (1, 204, 098.000) | (266, 828, 861.000) | (6, 492, 392.000) |
| Observations | 10,293 | 10,293 | 6,307 | 6,307 | 3,183 | 3,183 | 353 | 450 |
| $R^2$ | 0.209 | 0.219 | 0.356 | 0.367 | 0.151 | 0.152 | 0.189 | 0.351 |
| Adjusted $R^2$ | 0.207 | 0.217 | 0.354 | 0.365 | 0.147 | 0.147 | 0.145 | 0.324 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

in the Mainland China sample, it can be inferred that the overall positive coefficient of HSR for Mainland China is primarily driven by the positive coefficients of distance classes over 1000km.

Among other explanatory variables involved in the regression (i.e., POP_TOT, GDP_PC, GDP_SERVICES, and ACCESS_TOT), population is a highly significant factor for air route traffic in all economies. The regression suggests a positive relationship between the population (POP_TOT) and air route traffic (ROUTE), with the strongest correlation in South Korea. For every 1000 increase in population in the catchment areas of the route, air route traffic increases by over 100 in South Korea, which is more than twice the amount of that in other East Asian economies. The impact of GDP_PC is not significant in all samples and appears with the wrong sign. Although the per capita GDP is widely cited as an explanatory variable in air traffic analysis, recent studies by Mukkala and Tervo (2013) and Airbus (2016) suggest that the impact of GDP on airline travel demand is not casually significant. In comparison, the percentage contribution of service sectors in the GDP (GDP_SERVICES) is more significant although the value of coefficient varies across the samples. The analysis on the pooled sample and the Mainland China sample reveals a positive and significant relationship between GDP_SERVICES and air route traffic. The relationship is insignificant for other three regions. The regression also suggests that the airport access distance (ACCESS_TOT) has a negative impact on air travel demand across three East Asian economies (Japan, South Korea, and Taiwan) although is less significant in the latter two regions. The impact of ACCESS_TOT is not significant for Mainland China and tends to be positive although small in scale.

As mentioned in Section 2.2, the DID approach makes the parallel-trend assumption which needs to be validated as a premise. Given that treatments are introduced in multiple years, the pretreatment trends in the outcome variable is examined as suggested in Roberts (2012). We thus plot in Figure 3 the pretreatment year-specific fixed effects, represented by year dummy estimates, to inspect the parallel trend assumption visually. Separate plots are made for the treatment groups and the control groups where the TREATED dummy variable is removed from the equations. Moreover, the post-treatment observations in the treatment groups are excluded from the sample, so the HSR policy variable is also removed from the regression model. By visual inspection, pretreatment yearly fixed effects in treatment groups do not closely follow the trend of those in control groups, implying that the parallel-trend assumption does not stand in original samples.

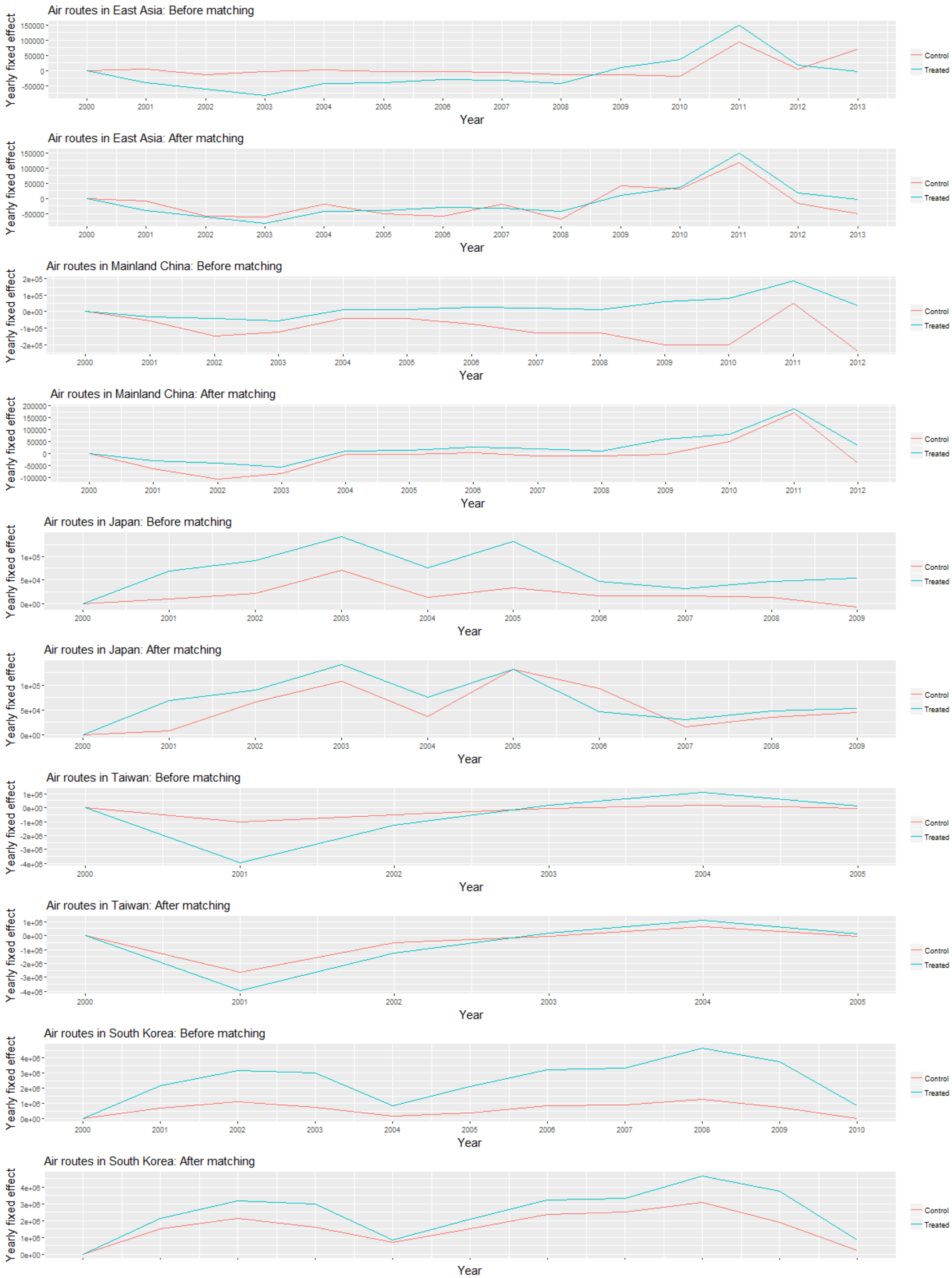



Figure 3: Yearly fixed effects of air route traffic models before and after matching.

To deal with this caveat, the propensity score matching (PSM) is implemented following the method of Randolph and Falbe (2014) which matches treated observations with those in the control group based on propensity scores. Shadish et al. (2008) noted that an important consideration in the PSM is which covariates to include in the propensity score calculation and it is suggested by Rubin (1997), Newgard et al. (2004) and Adelson (2013) that even the weakly predictive pretreatment variables should be included when constructing the propensity score as the biasing effects of omitting them may override the statistical efficiency gains of not including them. Therefore, the whole set of explanatory variables (i.e., POP_TOT, GDP_PC, GDP_SERVICES, and ACCESS_TOT) is used in the PSM. In addition, since the initial regression suggests that the impact of HSR differs across distance classes, the great circle distance (GC_DIST) of air routes is also included as a matching variable. Randolph and Falbe (2014) provided multiple matching methods including exact matching, nearest neighbor, optimal matching, and genetic matching. We compare the effectiveness among these methods and adopt the most effective ones – nearest neighbor and optimal matching.

Results showing the effectiveness of the PSM are relayed to the Appendix A, which demonstrate that the difference between the distributions of matching variables in control groups and that in treatment groups are reduced after PSM. Figure 3 shows that after matching there is more overlapping in yearly fixed effects of air route traffic between treatment and control groups in all samples across the pretreatment period.[5] Given that trends of pretreatment year-fixed effects in the two groups are generally consistent after matching, post-matching samples comply with the parallel-trend assumption and are used to re-estimate the DID models intending to refine results.

## 4.2 Post-matching estimation

Table 7 presents the regression results of air route traffic models based on post-matching samples. Since the per capita GDP (GDP_PC) resumes insignificant and continues producing wrong signs as in the initial regression, it is excluded the post-matching regression. The total population (POP_TOT) and average percentage contribution of service sectors in the GDP (GDP_SERVICES) of catchment areas of the route and the total airport access distance (ACCESS_TOT) are kept as explanatory variables. Comparing with the initial regression (Table 6), results are similar in terms of coefficients associated with HSR policy variables, positive effects of route-level population across all samples, and coefficients of regressors for the East Asia pooled sample. However, post-matching regression offers more sensible coefficients for other explanatory variables. The impact of ACCESS_TOT turns significant

[5]There is an upsurge in the yearly effect of Chinese airline traffic in the year of 2011. This reflects the traffic divergent effect of the high speed train crash happened in the July 2011, but it dissipated soon in 2012.

Table 7: Post-matching estimation of air route traffic models Eqs. (7) and (8)

| | East Asia | | (1) Mainland China | | (2) Japan | | (3) South Korea | (4) Taiwan |
|---|---|---|---|---|---|---|---|---|
| | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (8) | Eq. (7) | Eq. (7) |
| POP_TOT | 58.731*** | 58.374*** | 44.328*** | 44.157*** | 30.126*** | 30.383*** | 99.048*** | 488.043*** |
| | (1.823) | (1.814) | (1.652) | (1.635) | (2.670) | (2.720) | (19.117) | (87.901) |
| GDP_SERVICES | 3,708.993*** | 4,785.007*** | 12,164.570*** | 11,408.160*** | 27,128.840 | 26,730.760 | 284,996.500 | 345,240.200*** |
| | (1,091.415) | (1,093.088) | (1,467.784) | (1,455.247) | (27,218.070) | (27,273.530) | (350,395.100) | (119,741.900) |
| ACCESS_TOT | −737.229** | −894.497** | −941.000*** | −1,000.231*** | −3,399.079*** | −3,417.486*** | −3,258.715 | −28,099.770** |
| | (353.283) | (351.784) | (349.075) | (345.241) | (1,048.286) | (1,050.605) | (4,540.535) | (10,941.880) |
| TREATED | 17,729.690 | 21,026.840 | −50,649.080*** | −50,693.960*** | 160,845.900*** | 161,044.400*** | −1,283,417.000*** | −172,787.000 |
| | (26,853.570) | (26,690.540) | (14,421.490) | (14,259.610) | (52,690.140) | (52,778.020) | (409,920.700) | (249,801.100) |
| HSR | −177,536.700*** | | 47,504.790 | | −380,960.000*** | | −867,145.300** | −797,593.000*** |
| | (49,483.440) | | (30,913.560) | | (85,274.210) | | (437,924.200) | (237,205.300) |
| D1×HSR | | −504,988.400*** | | −278,328.300*** | | −294,627.500 | | |
| | | (72,327.220) | | (67,158.400) | | (186,902.800) | | |
| D2×HSR | | −147,887.800** | | −43,312.170 | | −390,356.800*** | | |
| | | (66,401.410) | | (37,305.180) | | (87,309.830) | | |
| D3×HSR | | 393,892.300*** | | 415,464.200*** | | | | |
| | | (103,928.100) | | (54,179.840) | | | | |
| D4×HSR | | 418,520.000** | | 449,837.800*** | | | | |
| | | (179,800.100) | | (92,259.770) | | | | |
| Constant | −422,553.500*** | −464,520.200*** | −637,097.900*** | −599,754.800*** | −1,676,615.000 | −1,651,181.000 | −16,392,336.000 | −25,706,907.000*** |
| | (85,817.000) | (85,465.560) | (66,354.180) | (65,781.570) | (1,932,511.000) | (1,936,302.000) | (20,500,300.000) | (7,851,099.000) |
| Observations | 5,058 | 5,058 | 4,428 | 4,428 | 242 | 242 | 238 | 150 |
| $R^2$ | 0.195 | 0.205 | 0.410 | 0.423 | 0.394 | 0.395 | 0.150 | 0.433 |
| Adjusted $R^2$ | 0.192 | 0.202 | 0.407 | 0.420 | 0.345 | 0.343 | 0.081 | 0.355 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

and negative for Mainland China whereas it was positive in the initial regression. The coefficient of GDP_SERVICES becomes significantly positive for Japan in contrast to the negative one yielded in the initial regression.

### 4.3 Summary and results interpretation

In summary, estimations of air route traffic models based on the East Asia pooled sample demonstrate that the substitution effect of HSR on aviation travel demand is mainly felt in short- and medium-haul markets (within 1000km), and the effect is the most significant in short-haul routes (below 500km). This echoes the airlines' response to HSR entries identified by Wan et al. (2016).

Region-wise analysis shows that HSR entries have significantly negative impacts on the air route passenger movements in South Korea and Taiwan, where all treated air routes are below 500km. The intervention of HSR reduces the average air route patronage by around 0.8 mllion. Results from Japan also reveal strongly negative impact of HSR on all treated routes, with the effects most pronounced on routes between 500km and 1000km. The negative impact on Japanese short-haul routes (within 500km) are statistically insignificant, but this can be explained by the fact that there is only one short-haul route in the treatment group during the sampling period.

Analysis of Mainland Chinese data reiterates that the HSR is most competitive on routes within 500km, and has moderately negative impacts on air route traffic for routes between 500km and 1000km. Positive coefficients associated longer distance routes further suggest that introducing HSR may encourage long distance air travels (over 1000km). One possible explanation is that most HSR links over 1000km were opened very recently in the 2010s. The origin and destination areas linked by such routes were relatively isolated before this. The opening of HSR links between these areas, by enhancing the travel mobility, might have initiated a number of social and economic activities which generated induced demand for travel. Part of the induced travel demand later spread onto other competitive transport modes such as the aviation service. As the result, the intervention of HSR reinforced instead of reducing the travel demand by air.

The analysis also establishes that the airport accessibility is a crucial factor in predicting the travel demand by the air mode. Specifically, we find that the air route traffic negatively correlates with the total access distance from original and destination airports to respective city centres. This finding indicates that facing the competition of HSR, improving airport accessibility is a sensible route of enhancing the mode share of aviation service.

In accordance with existing evidence, our results show that the catchment population is a significant and positive factor in predicting the air route travel demand. The positive impact of the size of service sectors is also noticeable, but it is less significant in Japan and

South Korea. This implies that GDP_SERVICES may have larger influences on air travel demand in developing economies like Mainland China (UNDP, 2017).

# 5 Effect of air-HSR integration on airport traffic

This section analyses the impact of air-HSR integration on airport passenger enplanement. We start by initial estimation of airport traffic models Eqs. (9) and Eq. (10) using the original datasets constructed in Section 3.2, and testing the unconditional parallel trend assumption (Assumption 1). The propensity score matching is then conducted and post-matching samples are employed to re-estimate the DID model and interpret results.

Recall that in Section 3.2 airports in the 'Full Control Group' are classified according to their proximity to HSR services. Those have overlapping catchment areas with HSR services are categorised into the 'Control Group-A' and those without into the 'Control Group-B'. In this section, the airport enplanement of the 'Treatment Group' is analysed with reference to not only the 'Full Control Group' but also the 'Control Group-A', intending to anatomise the effect of the availability of HSR services in the broad area from that of the availability of direct HSR connections at airports.

## 5.1 Initial regression and propensity score matching

Table 6 presents the initial regression results based on Eqs. (9) and (10) using the OLS method with fixed effects. When Eq. (9) prevails, the effect of air-HSR integration is captured by a single INTEGRT dummy variable and thus the coefficient associated with INTEGRT represents the average impact across a particular sample. When Eq. (10) is used, the regression involves two interaction terms of INTEGRT dummy and differentiated hub status classes, each representing the impact of air-HSR integration in that particular class.

For each sample, the DID estimation is carried out between the 'Treatment Group' and the 'Full Control Group' (columns entitled 'Treatment vs Full Control Group'), as well as between the 'Treatment Group' and the 'Control Group-A' (columns entitled 'Treatment vs Control Group-A').

Estimations of Eq. (9) based on Central European and East Asian samples both reveal positive coefficient for the policy variable INTEGRT, reflecting the complementary effect of HSR on aviation services in the presence of air-HSR integration. Except for the regression versus Control Group-A of East Asia, the impact is significant in all other samples. When looking into the hub status, estimations of Eq. (10) yield more positive and significant coefficients for the H1$\times$INTEGRT dummy than H2$\times$INTEGRT. This implies that the air-

Table 8: Initial estimation of airport traffic models Eqs. (9) and (10)

| | Central Europe | | | | East Asia | | | |
|---|---|---|---|---|---|---|---|---|
| | Treatment vs Full Control Group | | Treatment vs Control Group-A | | Treatment vs Full Control Group | | Treatment vs Control Group-A | |
| | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) |
| POP | 1,964.297*** | 1,917.370*** | 2,262.300*** | 2,155.431*** | 1,848.455*** | 1,842.165*** | 2,138.127*** | 2,146.802*** |
| | (59.505) | (56.155) | (89.188) | (83.143) | (31.561) | (32.232) | (41.470) | (42.750) |
| GDP_PC | 76,869.730*** | 76,417.680*** | 47,119.820 | 64,099.430** | −11,691.500 | −14,564.940 | −102,571.600 | −98,415.600 |
| | (11,114.950) | (10,479.130) | (32,340.040) | (30,059.860) | (52,199.670) | (52,285.820) | (74,545.420) | (74,717.820) |
| GDP_SERVICES | −151,524.100 | −129,515.000 | −557,337.100 | −471,538.900 | 67,248.300 | 70,735.910 | 159,949.400* | 154,603.500* |
| | (278,103.303) | (262,446.000) | (733,293.700) | (683,474.100) | (64,872.580) | (64,974.750) | (91,912.700) | (92,143.230) |
| ACCESS | −41,197.840*** | −30,927.440*** | −37,276.170* | −3,478.816 | −3,537.017 | −3,198.300 | −16,601.870 | −17,810.640 |
| | (9,038.236) | (8,538.058) | (20,893.680) | (19,535.190) | (5,822.899) | (5,833.620) | (11,807.470) | (11,896.580) |
| TREATED | 8,560,854.000*** | 8,597,985.000*** | 7,537,203.000*** | 7,449,363.000*** | 1,427,028.000** | 1,438,586.000** | 1,137,471.000 | 1,133,950.000 |
| | (1,248,668.000) | (1,177,239.000) | (1,613,053.000) | (1,498,288.000) | (612,266.900) | (612,393.900) | (694,406.200) | (694,486.700) |
| INTEGRT | 5,825,515.000*** | | 6,154,486.000*** | | 3,512,403.000*** | | 1,310,670.000 | |
| | (1,467,135.000) | | (1,851,780.000) | | (1,301,283.000) | | (1,462,512.000) | |
| H1×INTEGRT | | 37,062,701.000*** | | 36,622,531.000*** | | 5,071,685.000** | | 217,555.400 |
| | | (2,138,687.000) | | (2,648,714.000) | | (2,078,570.000) | | (2,339,209.000) |
| H2×INTEGRT | | −552,296.200 | | −198,108.400 | | 2,720,546.000* | | 2,082,807.000 |
| | | (1,422,736.000) | | (1,770,547.000) | | (1,539,771.000) | | (1,729,187.000) |
| Constant | 9,115,445.000*** | 7,475,027.000*** | 37,900,278.000*** | 31,090,618.000*** | −4,380,558.000* | −4,511,407.000* | −9,282,128.000** | −9,064,766.000** |
| | (2,074,122.000) | (1,957,346.000) | (5,331,663.000) | (4,972,710.000) | (2,651,984.000) | (2,655,509.000) | (3,783,532.000) | (3,792,799.000) |
| Observations | 2,950 | 2,950 | 1,457 | 1,457 | 2,538 | 2,538 | 1,555 | 1,555 |
| $R^2$ | 0.352 | 0.424 | 0.397 | 0.480 | 0.604 | 0.604 | 0.662 | 0.662 |
| Adjusted $R^2$ | 0.347 | 0.419 | 0.387 | 0.471 | 0.601 | 0.601 | 0.657 | 0.657 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

HSR integration is more likely to enhance airport enplanement if implemented at hubs. Comparing results from the two regions, the impact at hubs is stronger in Central Europe than in East Asia.

Inspecting other explanatory variables we find that the annual airport enplanement is strongly correlated with population size in both samples. The impacts of per capita GDP and GDP sector composition on air travel demand are distinct in the two samples. GDP_PC has a significantly positive impact in Europe but tends to be negative in the Asian sample. On the contrary, the coefficients of GDP_SERVICES are negative for the European sample and positive for East Asia. The initial regression also show that the airport access distance has a negative impact on the annual passenger volume, but is more significant at Central European airports than East Asian ones. The impact of these covariates will be re-examined after propensity score matching (PSM).

To verify the parallel-trend assumption, we plot in Figure 4 the pretreatment year-specific fixed effects represented by year dummy estimates, following the method adopted in Section 4.1. Separate plots are made for the 'Treatment versus Full Control Group' analysis, as well as the 'Treatment versus Control Group-A' analysis, and for Central European and East Asian samples respectively. By visual inspection, trends of pretreatment yearly fixed effects in treatment groups deviate from those in control groups. In order to reduce the selection bias introduced by pretreatment differences between treatment and control groups, PSM is carried out in each treatment and control group pair based on explanatory variables (POP_TOT, GDP_PC, GDP_SERVICES, and ACCESS_TOT). Results showing how PSM narrows the gaps between distributions of explanatory variables in the two groups are relayed to the Appendix B. Figure 4 demonstrates that after matching trends of pretreatment year-fixed effects in the two groups are generally consistent after matching.

## 5.2 Post-matching estimation

Post-matching samples are used to re-estimate airport traffic models. Results are presented in Table 9. Since the percentage contribution of service sectors in GDP (GDP_SERVICES) resumes insignificant and continues producing wrong signs as in the initial regression, it is excluded from the post-matching regression. The population (POP) and per capita GDP (GDP_PC) in the catchment area of the airport and the access distance (ACCESS) remain as explanatory variables.

In comparison with the initial regression (Table 8), the catchment population (POP) continues to be a significantly positive factor in predicting airport enplanement for all samples in Table 9, which is consistent with existing evidence. However, coefficients of all other variables are considerably more sensible in post-matching estimation than in the initial regression. For the per capita GDP (GDP_PC), the impact tends to be insignificant and negative for the

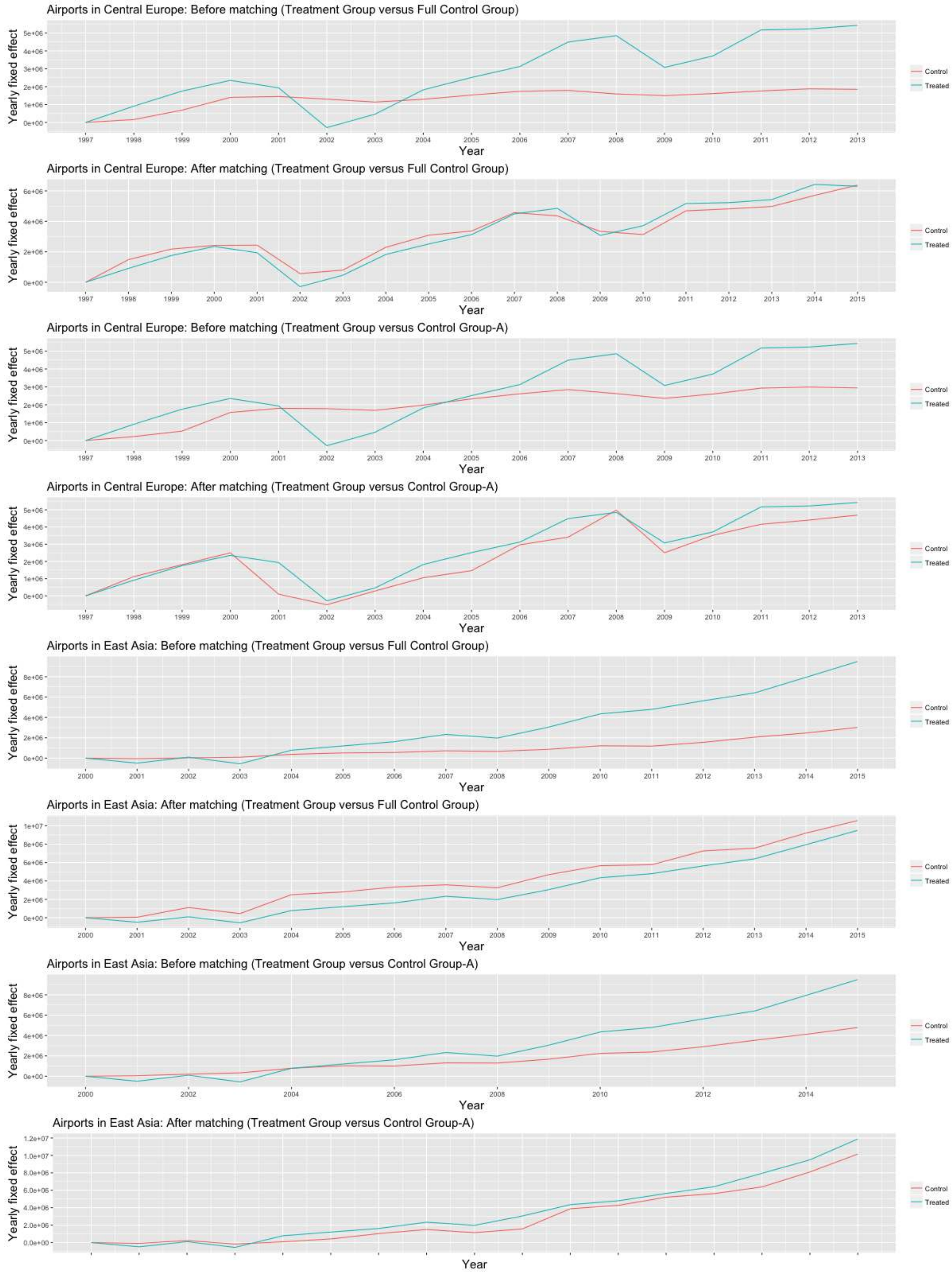



Figure 4: Yearly fixed effects of airport traffic models before and after matching.

Table 9: Post-matching estimation of airport traffic models Eqs. (9) and (10)

| | Central Europe | | | | East Asia | | | |
|---|---|---|---|---|---|---|---|---|
| | Treatment vs Full Control Group | | Treatment vs Control Group-A | | Treatment vs Full Control Group | | Treatment vs Control Group-A | |
| | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) | Eq. (9) | Eq. (10) |
| POP | 5,861.775*** | 5,709.527*** | 5,991.364*** | 5,725.809*** | 1,237.202*** | 969.595*** | 1,315.613*** | 1,074.738*** |
| | (296.180) | (193.222) | (298.638) | (199.117) | (72.242) | (80.591) | (72.515) | (81.758) |
| GDP_PC | 141,480.000** | 192,321.700*** | 132,432.400** | 176,792.200*** | 226,771.700*** | 256,424.700*** | 213,644.200*** | 235,209.600*** |
| | (68,804.790) | (44,928.710) | (54,314.100) | (36,198.690) | (56,114.170) | (53,229.170) | (58,144.130) | (55,712.670) |
| ACCESS | −140,131.000*** | −75,914.140*** | −132,309.600*** | −56,848.000** | −94,939.510*** | −86,705.680*** | −69,583.010*** | −62,484.060*** |
| | (33,701.880) | (22,229.610) | (35,015.140) | (23,649.580) | (17,569.690) | (16,652.210) | (16,615.800) | (15,933.590) |
| TREATED | −804,562.700 | 90,410.650 | −2,565,234.000 | −1,668,983.000 | 5,947,858.000*** | 6,881,144.000*** | 5,013,218.000*** | 5,744,233.000*** |
| | (2,222,073.000) | (1,449,150.000) | (2,301,056.000) | (1,530,898.000) | (717,676.100) | (694,338.000) | (668,312.300) | (652,220.200) |
| INTEGRT | 6,859,940.000*** | | 8,283,225.000*** | | 4,311,744.000*** | | 3,760,770.000*** | |
| | (2,337,029.000) | | (2,379,808.000) | | (1,120,417.000) | | (1,160,107.000) | |
| H1×INTEGRT | | 35,163,997.000*** | | 36,127,927.000*** | | 13,241,385.000*** | | 12,039,491.000*** |
| | | (2,138,419.000) | | (2,200,644.000) | | (1,779,098.000) | | (1,857,504.000) |
| H2×INTEGRT | | −218,492.200 | | 845,062.800 | | 945,041.600 | | 639,647.900 |
| | | (1,568,860.000) | | (1,634,350.000) | | (1,187,948.000) | | (1,243,090.000) |
| Constant | 2,673,274.000 | −681,538.000 | 4,017,019.000 | 945,291.800 | −1,388,030.000 | −1,577,340.000 | −1,834,655.000 | −1,868,594.000 |
| | (4,990,440.000) | (3,257,687.000) | (4,663,998.000) | (3,105,949.000) | (1,212,867.000) | (1,146,324.000) | (1,281,444.000) | (1,224,888.000) |
| Observations | 286 | 286 | 286 | 286 | 338 | 338 | 338 | 338 |
| $R^2$ | 0.637 | 0.846 | 0.626 | 0.835 | 0.799 | 0.821 | 0.782 | 0.802 |
| Adjusted $R^2$ | 0.605 | 0.832 | 0.594 | 0.820 | 0.785 | 0.808 | 0.768 | 0.788 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

East Asian sample in the initial regression. After PSM, the impact of GDP_PC is positive at high significance levels across all samples. Likewise, post-matching regression signifies that the airport access distance (ACCESS) is a significantly negative factor in airport passenger volume which reflects passengers' concerns about airport accessibility.

Coefficients of policy variables INTEGRT (based on Eq. (9)) and H1×INTEGRT (based on Eq. (10)) are positive with high level of significance in all samples. This demonstrates that the positive impact of air-HSR integration on airport enplanement is substantial in general, but is remarkably significant at hub airports. On the contrary, the effect is much less impressive at secondary hubs or regional airports suggested by the insignificant coefficients associated with H2×INTEGRT in all samples. Compared with the Central European sample, the effect of air-HSR integration is smaller in scale in East Asia. While the integration scheme can increase the average airport enplanemant by 6-8 million per annum in Central Europe, the increment is 3-4 million for East Asian airports. This can perhaps be explained by the fact that Europeans are pioneers of air-HSR integration, while the integration schemes have just begun in East Asia.

It is also of our great interest to compare results in the 'Treatment vs Full Control Group' columns and those in the 'Treatment vs Control Group-A' columns. In the former case, a difference in the airport traffic can be caused not only by the availability of direct HSR connections at airports, but also by the availability of HSR services in the whole catchment area. However in the latter case, HSR services are available in the catchment areas of airports in both treatment and control groups. The difference in airport enplanement thus entirely reflects the effect of air-HSR integration.

In line with this notion, results from Central Europe dictate that the positive effect of air-HSR integration would be underestimated if the 'Full Control Group' is employed as the reference point. Larger values of coefficients are obtained for policy variables (INTEGRT, H1×INTEGRT, and H2×INTEGRT) in the 'Treatment vs Control Group-A' columns than the 'Treatment vs Full Control Group' columns. Particularly for secondary hubs and regional airports, the coefficient of H2×INTEGRT is negative in the 'Treatment vs Full Control Group' column whereas it is positive in the 'Treatment vs Control Group-A' analysis. These differences demonstrate the existence of the counteracting substitution and complementary effects of HSR on aviation services in Europe. The complementary effect dominates at hubs while the substitution effect has influence over the other at secondary hubs and regional airports.

The East Asian analysis show that the positive effect of air-HSR integration on airport enplanement is larger in scale when comparing the 'Treatment Group' with the 'Full Control Group' than with the 'Control Group-A'. This means that the availability of HSR services in the catchment area reinforces the positive effect brought by the integration. At hub, the air-HSR integration accounts for more than 91% of the total effect (12,039,491 out of 13,241,385). The contribution of the integration at secondary hubs and regional airports is around 67% (639,647 out of 945,041.6).

## 5.3 Summary and results interpretation

In essence, estimations of airport traffic models find positive effects of air-HSR integration on airport enplanement which in turn demonstrates the complementary effect of HSR on aviation services.

The complementary effect is substantial at primary hub airports, but is much less at secondary hubs and regional airports. Since primary hub airports commonly have capacity issues, HSR services linked to hubs can serve as substitutes to short-haul flights and complements to long-haul flights. Given that widebody aircrafts mostly serving long-haul routes usually carry more passengers than single aisle aircrafts serving short-haul routes, the overall passenger enplanement at primary hub airports may increase due to integration. This implies that connecting to the HSR network is highly beneficial to primary hub airports. For secondary hubs and regional airports, the major benefit of air-HSR cooperation is to capture the spillover effects of primary hubs and to expand catchment areas. This echoes with the theoretical prediction of Jiang and Zhang (2016) that the development of HSR networks would reform airlines' network from the fully-connected to the hub-and-spoke structure. With air-HSR integration at hubs, HSR can provide feeding services for hubs while might compete with air mode for non-hubs.

Region-wise analysis reiterates the benefit of air-HSR integration at European hubs. The availability of HSR in the catchment area and the availability of air-HSR integration simultaneously influence the airport enplanement but along contrary directions. The former reflects the substitution effect which dominates at secondary hubs and regional airports. The latter constitutes the complementary effect which has overwhelming influence at hubs.

Geographical locations of secondary hubs and regional airports may have played a vital role in limiting the success of air-HSR integration. Regional airports, such as Brussels National (BRU) and Cologne Bonn (CGN), serve metropolitan areas that lie within the expanded catchment areas of primary hub airports that have air-HSR intermodal connections. For example, travellers in the Belgian capital may choose to take the TGV to Paris CDG or the Thalys to Amsterdam Schiphol for their connecting flights, rather than flying from Brussels National. Similarly, Cologne is sandwiched between Frankfurt and Dusseldorf. From Deutsche Bahn's (2017) timetables, air-HSR services put Cologne a comfortable 33 minutes away from Dusseldorf Airport and 49 mins away from Frankfurt Airport, leading to serious overlapping of airport catchment areas. Realising that it could not compete with its bigger neighbours for premier traffic, Cologne Bonn Airport (2016) had re-posited itself as a low-cost airport. Therefore, one of the implications from European practice of air-HSR integration is that building HSR stations at primary hubs will help in consolidating their hub status, but it may have negative impact on nearby regional airports.

The air-HSR intermodality also brings significant benefits to East Asian airports, but to a lesser extent than the European case. While the integration scheme can increase the average airport enplanemant by 6-8 million per annum at European airports, the increment is 3-4 million in East Asia. This can perhaps be explained by the fact that Europeans are pioneers of air-HSR integration, while it is emerging in East Asia. This strengthens the argument that more coordination between transport operators may be required to unlock the full potential of the concept in East Asia, especially at regional airports. The underutilization of HSR stations at airports, evidenced by the fact that there are only 13 daily HSR services at Guiyang Airport in mainland China (Ministry of Transport, 2017), may partially account for the incumbent benefit.

The airport traffic analysis re-establishes the strong influences of the population, GDP, and the airport accessibility on the travel demand. Both the former two factors have positive impact on airport enplanement, but the scale of effect is divergent in two regions. An increase of 1000 in the population in a city is predicted to induce 5800 more enplanements at a European airport and roughly 1000 in East Asia. The effect of per capita GDP in driving airport traffic of East Asia is two time of that in the Central European sample. The negative effect of the airport access distance is also significant in both samples in comparable sizes.

# 6 Conclusions

In this paper, substitution and complementary effects of high-speed railways (HSR) on the aviation industry are analysed in two contexts: the effect on competitive airline traffic, and the effect on airport enplanement through air-HSR integration.

The substitute role of HSR is mainly measured by the impact of new HSR routes on passenger movements in parallel air routes, with a focus on East Asian markets, where there is has been a massive development in HSR networks in recent years. Our results indicate that airlines face intense competition from HSR on routes less than 500km (short-haul market). In this distance class, the intervention of HSR reduces the annual average air route patronage by around 0.8 mllion in South Korea and Taiwan, and around 0.3 million in Mainland China and Japan.

The substitution effect is most significant in medium-haul routes (between 500km and 1000km) in Japan, but is moderate in the mainland China market. This implies that the medium-haul market is still a battleground for competing transport modes in China. To our knowledge, our paper is the first to include a sizeable number of HSR routes that opened after

2012, as well as the Taiwanese market. Moreover, with many long-distance HSR corridors opening to traffic in mainland China between 2012 and 2014, the impacts of HSR entries on air routes longer than 1500km are studied for the first time. We find that HSR services have a positive impact on airline traffic in long distance markets (over 1500km), indicating that introducing HSR may encourage long distance air travels.

While the existing literature usually focuses on competitive behaviour alone, this paper further investigate complementarity between the two modes by examining the effect of air-HSR integration. Central European and East Asian airports are included in the study. Air-HSR integration is an idea originating from European major hub airports, so it is not surprising that the concept is highly successful at primary hubs in this continent. The benefits at secondary hubs and regional airports are much less. One possible reason for this is underutilisation of the facilities. In East Asia, transport planners have embraced air-HSR integration in recent years. Likewise, our results indicate that integration boosts passenger enplanement at primary hubs in East Asia, and regional airports have also benefited, but to a lesser degree. To our knowledge, this is the first empirical study to quantify complementarity between air transport and HSR in the context of air-HSR integration using econometric methods.

In addition to intervention of HSR, this study identifies that the airport accessibility is also a crucial factor in the travel demand by air mode. Specifically, we find that both air route and airport traffic negatively correlates with the access distance from city centres to airport infrastructures. This finding indicates that facing the competition of HSR, improving airport accessibility is a sensible route of enhancing the mode share of aviation service. Besides, our results reiterate the significance of population and GDP in predicting the aggregate travel demand by air.

Taking a holistic approach to analyse interaction between HSR and aviation, our analyses yield important policy implications. First, airlines may wish to consider codeshare agreements with HSR operators to substitute short-haul flights with HSR services, as it is difficult for airlines to compete directly with HSR on routes shorter than 500km. Second, air-HSR integration is a likely successful business model for primary hubs, so that airports suffering from capacity constraints, such as London Heathrow (LHR) and Beijing Capital (PEK), should study the feasibility of this idea. Third, secondary hubs and regional airports should exercise more caution when considering air-HSR inter-modality, as the potential benefits may be much less than that experienced by primary hubs.

In this study, the air route analysis focuses on East Asian regions and those in Europe are not included due to limited availability of data. Future works are expected to include the European market, especially long distance routes over 1000km. Given that there exist intrinsic differences between the intermodal services provided by East Asian integration and that in Europe, separate studies are carried out for the two samples. Future study could look into effects of specific intermodality characteristics, for example, the transfer distance and the availability of integrated ticketing/baggage handling. Moreover, given that a number of HSR stations were opened after 2014 at East Asian airports which are not covered in the current sampling period, further research on the complementary effects of HSR on airport-level traffic in East Asia are called for when more data become available in the future.

## Acknowledgement

The work described in this paper was financially supported by the EPSRC grant: Airport Capacity Consequences Leveraging Aviation Integrated Modelling.

# Appendix A. Propensity score matching results of air route traffic

| | | *Matched data* | | | | |
|---|---|---|---|---|---|---|
| | | Means Treated | Means Control | SD Control | Mean Diff | Percentage reduction of Mean Diff. (%) |
| East Asia | GC_DIST | 857.7803 | 865.3630 | 445.8174 | −7.5827 | 89.2649 |
| | POP_TOT | 13809.6806 | 11621.6338 | 6270.6015 | 2188.0468 | 58.4647 |
| | GDP_PC | 6.3532 | 5.6401 | 8.0248 | 0.7131 | 94.2275 |
| | GDP_SERVICES | 46.0686 | 45.2080 | 6.9381 | 0.8606 | 91.5865 |
| | ACCESS_TOT | 59.1042 | 60.9286 | 38.7456 | −1.8244 | 69.7853 |
| Mainland China | GC_DIST | 923.6762 | 921.5384 | 443.8415 | 2.1378 | 99.0237 |
| | POP_TOT | 13458.7349 | 11705.2556 | 5811.8854 | 1753.4792 | 52.4433 |
| | GDP_PC | 3.4856 | 3.4453 | 2.0420 | 0.0403 | 82.5104 |
| | GDP_SERVICES | 43.2462 | 43.2037 | 1.9141 | 0.0425 | 79.4305 |
| | ACCESS_TOT | 62.2060 | 62.3394 | 28.2385 | −0.1334 | 95.0394 |
| Japan | GC_DIST | 590.2066 | 658.3636 | 307.5331 | −68.1570 | 64.3407 |
| | POP_TOT | 10877.1902 | 12683.7156 | 8622.0619 | −1806.5253 | 46.9445 |
| | GDP_PC | 38.8994 | 38.6623 | 4.5695 | 0.2370 | −269.4367 |
| | GDP_SERVICES | 72.0779 | 72.0698 | 1.4699 | 0.0082 | 75.6998 |
| | ACCESS_TOT | 48.4603 | 49.6397 | 65.2775 | −1.1793 | 36.7440 |
| South Korea | GC_DIST | 292.2192 | 286.7027 | 103.1387 | 5.5165 | 30.1125 |
| | POP_TOT | 25391.6052 | 14396.5574 | 9135.2761 | 10995.0478 | 34.4959 |
| | GDP_PC | 19.5137 | 18.9784 | 5.4358 | 0.5354 | 10.4029 |
| | GDP_SERVICES | 59.5720 | 59.5228 | 0.9067 | 0.0492 | 34.2698 |
| | ACCESS_TOT | 38.5269 | 39.0672 | 31.5069 | −0.5403 | 66.5322 |
| Taiwan | GC_DIST | 241.5780 | 191.9063 | 98.7635 | 49.6717 | 15.1148 |
| | POP_TOT | 10524.0295 | 6123.2153 | 3820.7091 | 4400.8143 | 34.4399 |
| | GDP_PC | 17.6165 | 17.8798 | 2.8999 | −0.2633 | −17.6974 |
| | GDP_SERVICES | 65.9973 | 65.9425 | 1.1487 | 0.0548 | 29.0401 |
| | ACCESS_TOT | 17.3600 | 15.6818 | 5.8254 | 1.6782 | 25.1507 |

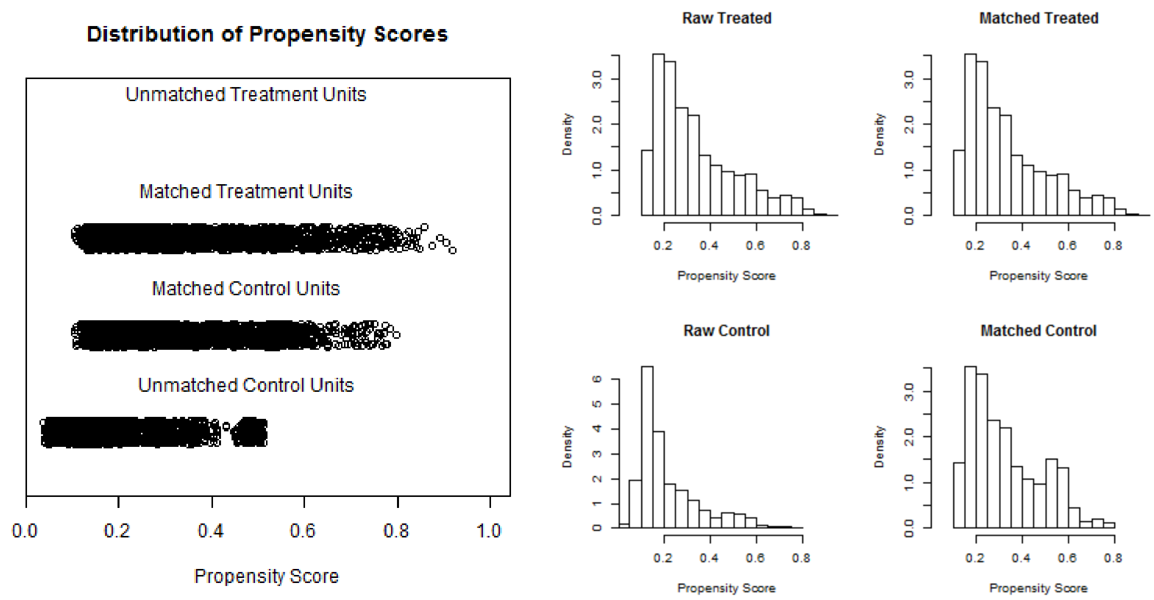



Figure 5: Propensity score matching results: East Asia.

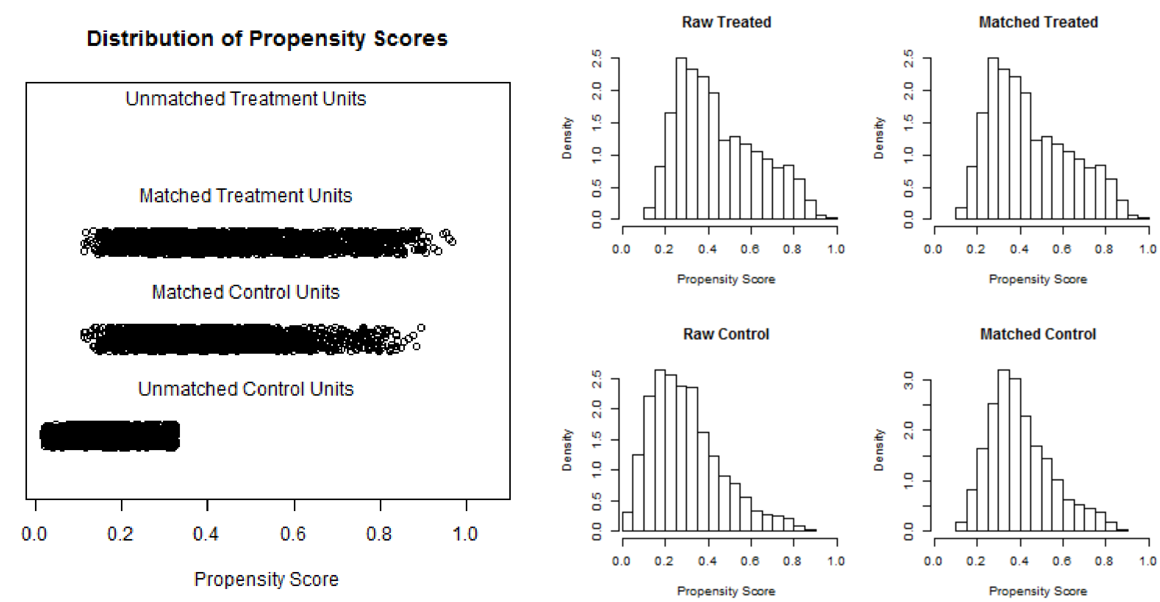



Figure 6: Propensity score matching results: Mainland China.

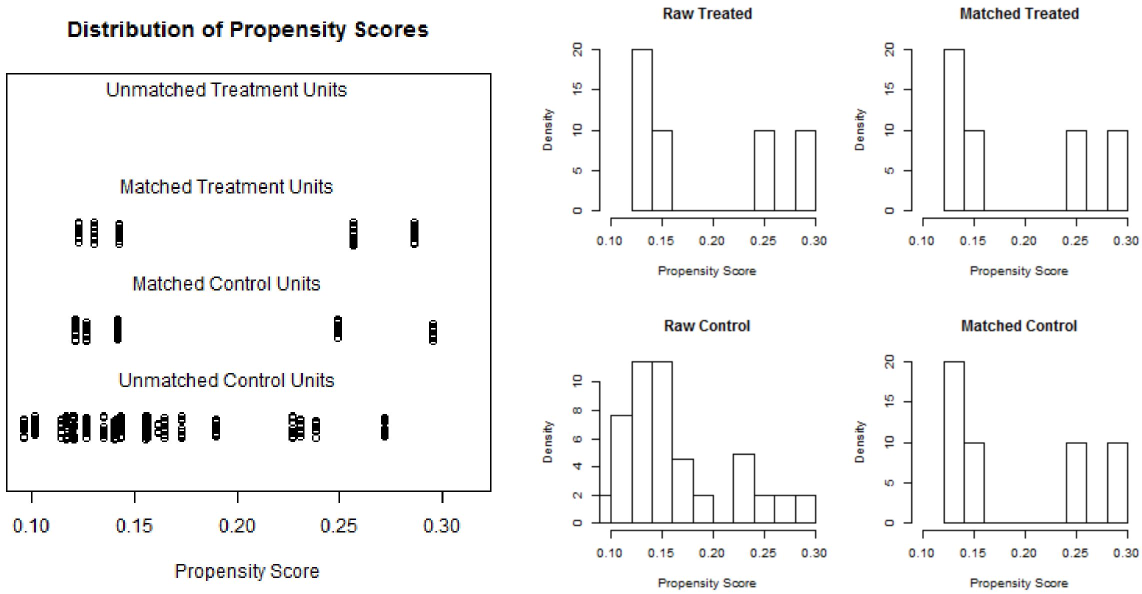



Figure 7: Propensity score matching results: Japan.

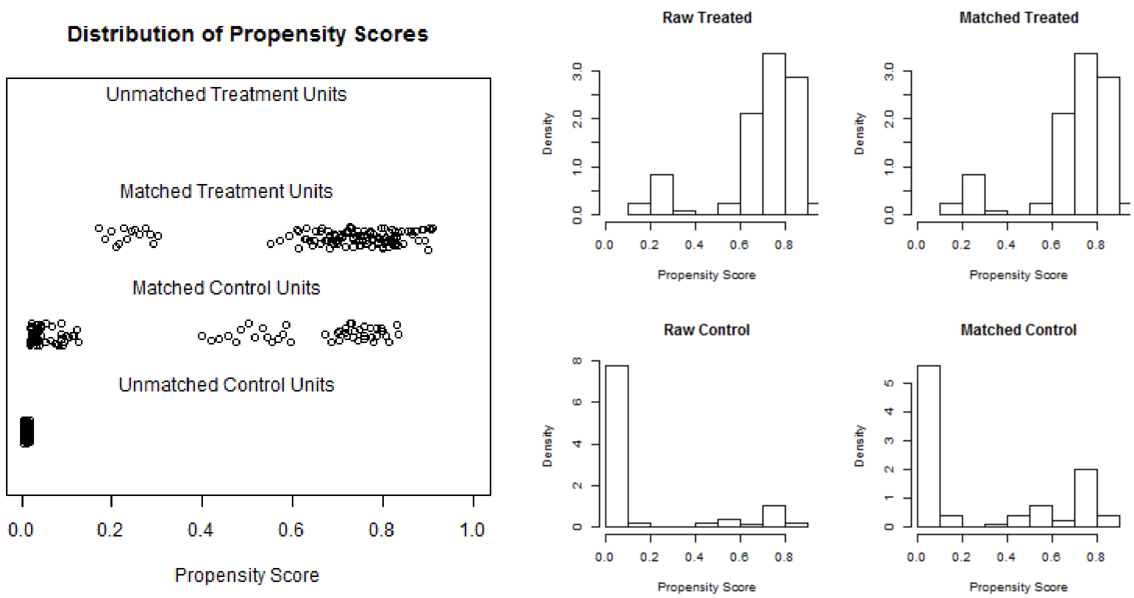



Figure 8: Propensity score matching results: South Korea.

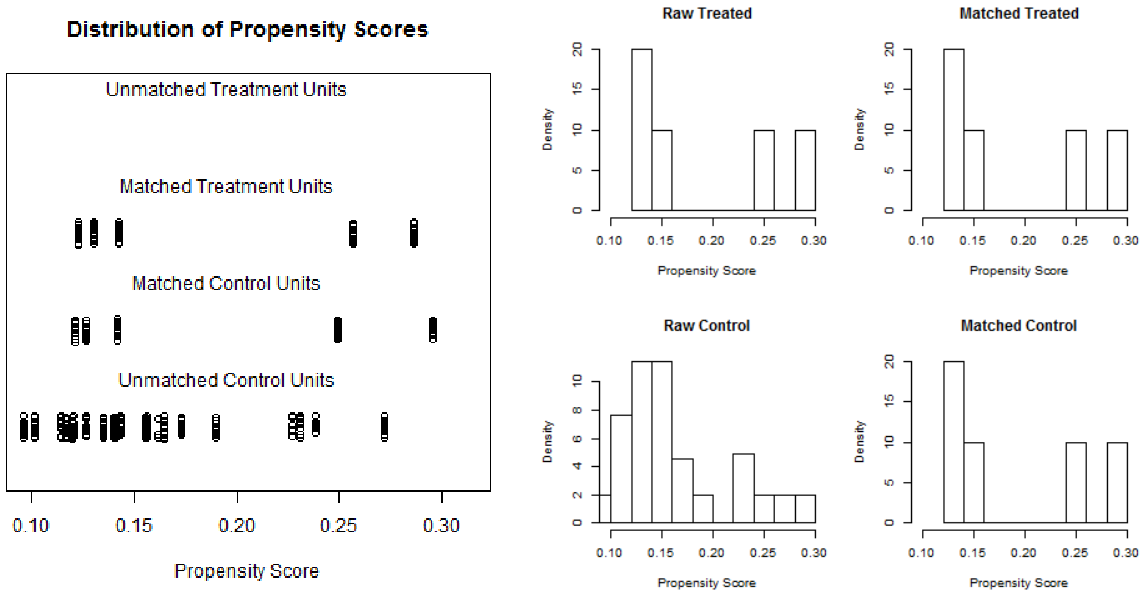



Figure 9: Propensity score matching results: Taiwan.

# Appendix B. Propensity score matching results of airport traffic

| | | *Matched data* | | | | |
|---|---|---|---|---|---|---|
| | | Means Treated | Means Control | SD Control | Mean Diff. | Percentage reduction of Mean Diff. (%) |
| **Central Europe** (Treatment Group versus Full Control Group) | POP | 1963.3871 | 1982.7097 | 2093.1958 | −19.3226 | 96.9972 |
| | GDP_PC | 38.8052 | 38.6799 | 15.4957 | 0.1253 | 95.8981 |
| | GDP_SERVICES | 71.1716 | 70.8835 | 3.0400 | 0.2880 | 84.9247 |
| | ACCESS | 23.3427 | 25.0545 | 20.6019 | −1.7119 | 79.5233 |
| **Central Europe** (Treatment Group versus Control Group-A) | POP | 1963.3871 | 1764.2378 | 2287.1091 | 199.1493 | −204.9367 |
| | GDP_PC | 38.8052 | 39.0578 | 15.3344 | −0.2526 | 95.5490 |
| | GDP_SERVICES | 71.1716 | 70.6086 | 6.1529 | 0.5630 | −279.5943 |
| | ACCESS | 23.3427 | 22.0406 | 20.8601 | 1.3021 | 76.9071 |
| **East Asia** (Treatment Group versus Full Control Group) | POP | 5417.4951 | 4160.5821 | 5621.1833 | 1256.9130 | 57.2724 |
| | GDP_PC | 5.5830 | 3.9420 | 4.4884 | 1.6410 | 86.3005 |
| | GDP_SERVICES | 45.5705 | 43.7765 | 3.9759 | 1.7941 | 81.6246 |
| | ACCESS | 33.6663 | 33.2024 | 27.0348 | 0.4639 | 95.7832 |
| **East Asia** (Treatment Group versus Control Group-A) | POP | 5417.4951 | 4519.8175 | 4487.4862 | 897.6775 | 52.3495 |
| | GDP_PC | 5.5830 | 3.8503 | 3.0222 | 1.7327 | 80.8184 |
| | GDP_SERVICES | 45.5705 | 43.8347 | 3.5163 | 1.7359 | 76.5571 |
| | ACCESS | 33.6663 | 31.8065 | 13.4333 | 1.8598 | 80.4228 |

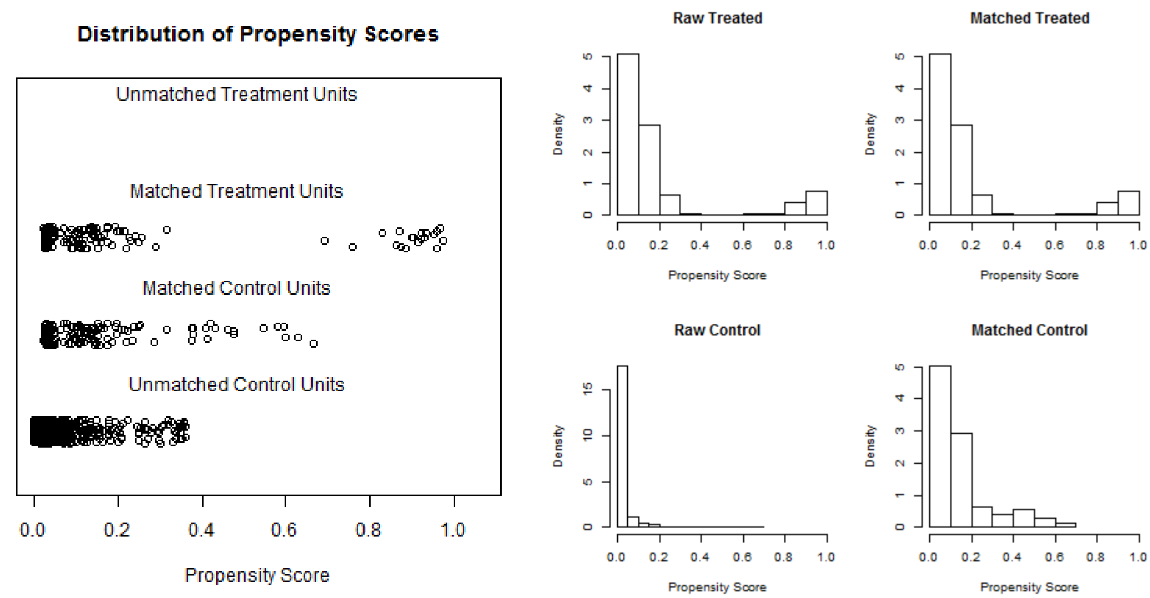



Figure 10: Propensity score matching results: Central Europe (Treatment Group versus Full Control Group).

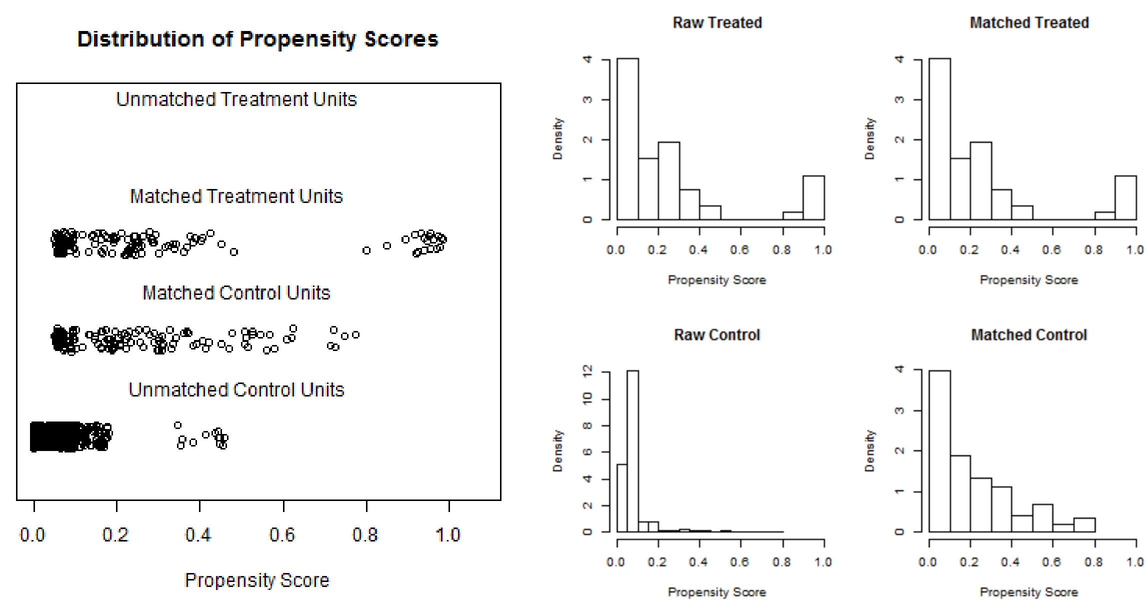



Figure 11: Propensity score matching results: Central Europe (Treatment Group versus Control Group-A).

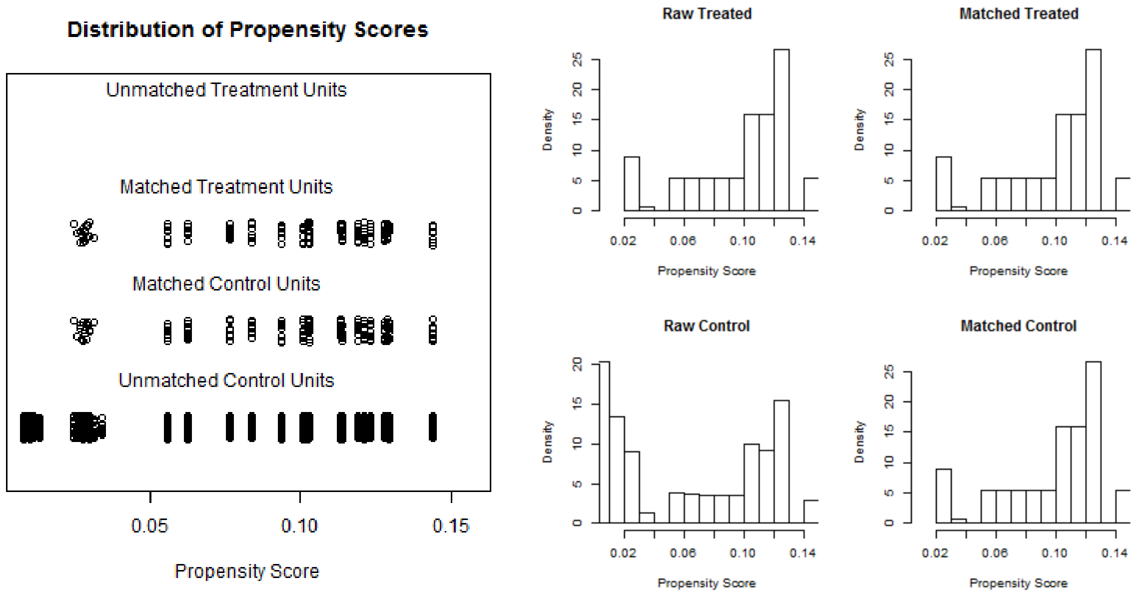



Figure 12: Propensity score matching results: East Asia (Treatment Group versus Full Control Group).

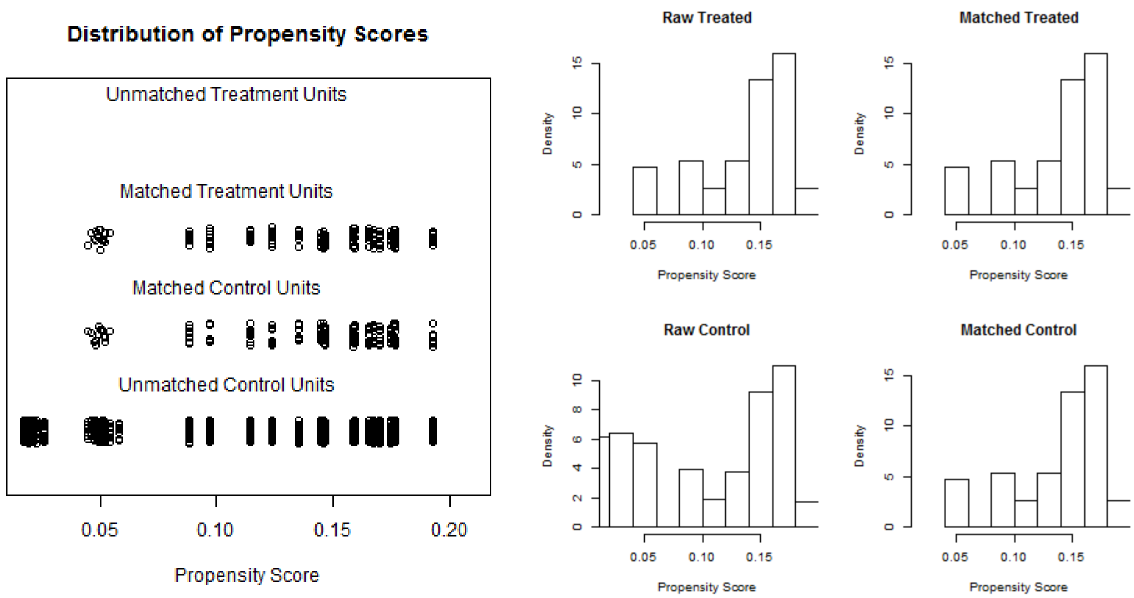



Figure 13: Propensity score matching results: East Asia (Treatment Group versus Control Group-A).

# References


Adelson, J. L. (2013). Educational research with real-world data: Reducing selection bias with propensity scores. *Practical Assessment, Research & Evaluation*, 18(15):1–11.

Adler, N., Pels, E., and Nash, C. (2010). High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. *Transportation Research Part B: Methodological*, 44(7):812–833.

Airbus (2016). 2016 global market forecast. *Airbus*. Available at <http://www.airbus.com/company/market/global-market-forecast-2016-2035/?eID=maglisting_push&tx_maglisting_pi1%5BdocID%5D=111117>. [Accessed 25th May 2017].

Airports Commission (2014). Local economy: literature review. *Airports Commission, UK*. Available at <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371811/3-local-economy--local-economy-literature-review.pdf>.

Albalate, D. and Bel, G. (2012). High-speed rail: Lessons for policy makers from experiences abroad. *Public Administration Review*, 72(3):336–349.

Albalate, D., Bel, G., and Fageda, X. (2015). Competition and cooperation between high-speed rail and air transportation services in europe. *Journal of transport geography*, 42:166–174.

Behrens, C. and Pels, E. (2012). Intermodal competition in the london–paris passenger market: High-speed rail and air transport. *Journal of Urban Economics*, 71(3):278–288.

Braitman, L. E. and Rosenbaum, P. R. (2002). Rare outcomes, common treatments: analytic strategies using propensity scores. *Annals of internal medicine*, 137(8):693–695.

Brueckner, J. K. (1985). A note on the determinants of metropolitan airline traffic. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, 12(2):175–184.

Burge, P., Kim, C. W., and Rohr, C. (2011). Modelling demand for long-distance travel in great britain: Stated preference surveys to support the modelling of demand for high-speed rail. *Santa Monica, CA, RAND Corporation*, Report number: TR899.

Castillo-Manzano, J. I., Pozo-Barajas, R., and Trapero, J. R. (2015). Measuring the substitution effects between high speed rail and air transport in spain. *Journal of Transport Geography*, 43:59–65.

Chi, J. and Baek, J. (2012). A dynamic demand analysis of the united states air-passenger service. *Transportation Research Part E: Logistics and Transportation Review*, 48(4):755–761.

Clewlow, R., Sussman, J., and Balakrishnan, H. (2012). Interaction of high-speed rail and aviation: exploring air-rail connectivity. *Transportation Research Record: Journal of the Transportation Research Board*, (2266):1–10.

Clewlow, R. R., Sussman, J. M., and Balakrishnan, H. (2014). The impact of high-speed rail and low-cost carriers on european air passenger traffic. *Transport Policy*, 33:136–143.

Cologne Bonn Airport (2016). Annual report 2016. Available at <https://www.koeln-bonn-airport.de/uploads/tx_download/KBA_GeBer2016_engl_WEB100dpi_2017_04_12.pdf>. [Accessed 15th July 2017].

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

IATA (2007). Estimation air travel demand elasticities. *International Air Transport Association (IATA)*. Available at <https://www.iata.org/whatwedo/Documents/economics/Intervistas_Elasticity_Study_2007.pdf>.

Jiang, C. and Zhang, A. (2014). Effects of high-speed rail and airline cooperation under hub airport capacity constraint. *Transportation Research Part B: Methodological*, 60:33–49.

Jiang, C. and Zhang, A. (2016). Airline network choice and market coverage under high-speed rail competition. *Transportation Research Part A: Policy and Practice*, 92:248–260.

Jiménez, J. L. and Betancor, O. (2012). When trains go faster than planes: the strategic reaction of airlines in spain. *Transport Policy*, 23:34–41.

JR East (2016). Initiatives responding to change-annual report 2016. Available at <https://www.jreast.co.jp/e/investor/ar/2016/pdf/ar_2016-all.pdf>. [Accessed 15th July 2017].

Ministry of Transport (2017). Railway timetables. Available at <https://kyfw.12306.cn/otn/>. [Accessed 15th July 2017].

Mukkala, K. and Tervo, H. (2013). Air transportation and regional growth: which way does the causality run? *Environment and Planning A*, 45(6):1508–1520.

Newgard, C. D., Hedges, J. R., Arthur, M., and Mullins, R. J. (2004). Advanced statistics: the propensity scorea method for estimating treatment effect in observational research. *Academic Emergency Medicine*, 11(9):953–961.

Park, Y. and Ha, H.-K. (2006). Analysis of the impact of high-speed railroad service on air transport demand. *Transportation Research Part E: Logistics and Transportation Review*, 42(2):95–104.

Randolph, J. J. and Falbe, K. (2014). A step-by-step guide to propensity score matching in r. *Practical Assessment, Research & Evaluation*, 19(18):1–6.

Roberts, M. R. (2012). Differences in differences-empirical methods. Available at <http://finance.wharton.upenn.edu/~mrrobert/resources/Teaching/CorpFinPhD/Dif-In-Dif-Slides.pdf>. [Accessed 2nd July 2017].

Román, C., Espino, R., and Martin, J. C. (2007). Competition of high-speed train with air transport: The case of madrid–barcelona. *Journal of Air Transport Management*, 13(5):277–284.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81(396):961–962.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine*, 127(8_Part_2):757–763.

Rudner, L. M. and Peyton, J. (2006). Consider propensity scores to compare treatments. *Practical Assessment Research & Evaluation*, 11(9):1–9.

Shadish, W. R., Clark, M. H., and Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? a randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, 103(484):1334–1344.

Swartz, K. L. (2017). Great circle mapper. *ResearchGate*. Available at <http://www.gcmap.com/>. [Accessed 25th June 2017].

Takebayashi, M. (2016). How could the collaboration between airport and high speed rail affect the market? *Transportation Research Part A: Policy and Practice*, 92:277–286.

UIC (2017). High speed lines in the world. Available at <http://www.uic.org/IMG/pdf/20170201_high_speed_lines_in_the_world.pdf>. [Accessed 20th April 2017].

UNDP (2017). Human development index and its components. Available at <http://hdr.undp.org/en/composite/hdi>. [Accessed 31th May 2017].

Wan, Y., Ha, H.-K., Yoshida, Y., and Zhang, A. (2016). Airlines reaction to high-speed rail entries: Empirical study of the northeast asian market. *Transportation Research Part A: Policy and Practice*, 94:532–557.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

Xia, W. and Zhang, A. (2016). High-speed rail and air transport competition and cooperation: A vertical differentiation approach. *Transportation Research Part B: Methodological*, 94:456–481.

Xia, W. and Zhang, A. (2017). Air and high-speed rail transport integration on profits and welfare: Effects of air-rail connecting time. *Journal of Air Transport Management*, 65:181–190.

Yao, E. and Morikawa, T. (2005). A study of on integrated intercity travel demand model. *Transportation Research Part A: Policy and Practice*, 39(4):367–381.