https://doi.org/10.1016/j.tra.2016.10.014

Airlines' Reaction to High-Speed Rail Entries: Empirical Study of the Northeast Asian Market

Yulai Wan,[a, *] Hun-Koo Ha,[b] Yuichiro Yoshida,[c] and Anming Zhang[d, e]

[a] Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hong Kong. sarah.wan@polyu.edu.hk

[b] Asia-Pacific School of Logistics, Inha University, Incheon, South Korea. hkha@inha.ac.kr

[c] Graduate School for International Development and Cooperation, Hiroshima University, Japan. yuichiro@hiroshima-u.ac.jp

[d] Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T 1Z2, Canada. anming.zhang@sauder.ubc.ca

[e] China Academy of Financial Research, Shanghai Jiao Tong University, China

* Corresponding author

April 2016; revised: August 2016

Abstract: We investigate the impact of the commencement of high-speed rail (HSR) services on airlines' domestic available seats on affected routes in China, Japan, and South Korea. The study is based on a dataset covering the 1994-2012 period. We use the propensity score matching method to pair HSR affected routes with routes without HSR services. The difference-in-difference approach is used to estimate the impact of HSR entry. We find that HSR entries may, on average, lead to a more significant drop in airlines' seat capacity in China than in Japan and Korea given similar HSR service speed. In China, HSR services with a maximum speed about 200km/h can produce strong negative impacts on medium-haul air routes but induce more air seat capacity on long-haul routes. HSR services with a maximum speed of 300km/h have little extra impact on medium-haul routes but a strong negative impact on long-haul routes. Finally, although HSR has a strong negative impact in Japan's short-haul and medium-haul air markets, little impact is observed in its long-haul markets.

Keywords: Difference-in-difference estimator; High-speed rail; Airline available seats; Northeast Asia; Propensity score matching.

Acknowledgement: We are very grateful to three anonymous referees and Kun Wang for constructive comments that have led a significant improvement of the paper and Shuli Liu for excellent assistance during the revision. We also thank seminar participants at Asia-Pacific School of Logistics, Inha University, and the Conference on Logistics and Maritime Studies on "One Belt, One Road" at the Hong Kong Polytechnic University for helpful comments. Financial support from Hong Kong Polytechnic University (code: G-UC62) is gratefully acknowledged.

1. Introduction

Starting from the first experimental high-speed rail (HSR), Qinhuangdao-Shenyang line, China has been expanding its HSR network and has already achieved an extensive track length of 8,358 km by the end of 2010. The Chinese HSR system has grown into the largest HSR network in the world within a short period of time, and according to the plan the network will continue to grow during the next decade. In the other two major Northeast Asian countries, Japan and South Korea, HSR also plays a significant role in domestic inter-city passenger transport and the HSR networks of both countries are still expanding. Such ambitious plans for the development of HSR have important implications for the domestic aviation market. Intensified competition between HSR and airlines on certain routes has been recorded, which may lead to various changes in airlines' route selections and service levels on the affected routes.

The impacts of HSR on air transport have received substantial, and increasing, attention since they have major policy implications on many aspects, such as climate change mitigation (e.g., Givoni, 2007; Ha et al., 2011; D'Alfonso et al., 2015, 2016), passenger welfare and social welfare (e.g., Janic, 1993; Adler et al., 2010; Rothengatter, 2011; Yang and Zhang, 2012; Roman and Martin, 2014; Álvarez-SanJaime et al., 2015), airline entry barriers and market power (e.g., Kappes and Merkert, 2013; Zhang et al., 2014), competitive and cooperative behavior of rail operators and airlines (e.g., Jiang and Zhang, 2014; Albalate et al., 2015; Xia and Zhang, 2016a and 2016b), airport and HSR infrastructure investments (e.g., Goldman Sachs, 2010; Ollivier et al., 2014) and so on. Dobruszkes et al. (2014) provided a comprehensive review of the existing literature that investigates the ex-post intermodal impacts of HSR entry. The literature consists of both theoretical modeling and empirical studies. Most of the empirical studies do not apply econometric methods and hence rely heavily on either comparing traffic volumes or market shares before and after the HSR operation, or surveying passengers about their modal choice decisions. Studies based on econometric methods are mainly related to various European markets (see Table 1 for a list of related studies), while studies on the Northeast Asian markets, especially the Chinese markets, are rare. Fu et al. (2012) provided some descriptive analysis on the impact of HSR entry in Chinese air transport markets. Zhang and Zhang (2016) used gravity models to examine the determinants of air passenger flows in China with the HSR presence as one of the explanatory variables. Earlier, Park and Ha (2006) conducted a survey on passengers' stated preference over air and HSR to predict the impact of the entry of HSR service in South Korea.

[Insert Table 1 here.]

Furthermore, existing empirical studies fall into three major streams as listed in Table 1. The first stream applies the discrete choice models, i.e. various versions of logit models, to analyze factors influencing passengers' choice between air, HSR and sometimes other modes of transport. The estimated models are then often used to predict the market share split between air and HSR. The second stream aims at quantifying how HSR service levels, such as rail travel time (Clewlow et al, 2013; Dobruszkes et al, 2014), HSR frequency (Dobruszkes et al, 2014) and HSR passenger number (Castillo-Manzano et al, 2015), relate to changes in airlines' seats, frequency or aircraft size. None of these two streams is able to measure the ex-post impact of HSR entry into a market traditionally served by airlines by comparing with the situation where no HSR service is available. The third stream addresses this issue by adding an HSR dummy variable which indicates the entry and existence of HSR services and it is most relevant to our study. However, existing studies in this stream ignore a few important issues. First, there might be some inherent differences between the HSR affected routes and unaffected routes, and hence on average these two groups of routes differ in traffic volumes. Second, the impact of HSR entry could be dynamic. For example, there might be a time lag between the entry of HSR and the impact on airline traffic. Third, traffic of both the HSR affected routes and unaffected routes can change overtime, owing to certain common factors, and hence it might be better to control for this kind of traffic change, so as to compare the traffic change of treated routes with the hypothetical situation where HSR services were not started. Finally, it would be better to measure the impacts based on benchmarking with the unaffected routes that are similar to the routes affected by HSR services.

Most of the papers in the literature recognize adverse impacts of HSR on airlines' traffic, operations and market shares, which can last for two to five years (Campos and Gagnepain, 2009), but the amount of impacts varies across routes depending on many factors. Some papers predict that HSR will mainly compete with air on routes over 500km (e.g. Martin and Nombela, 2007; Armstrong and Preston, 2011) while others predicted that HSR is competitive to air for routes within three hours (e.g. González-Savignat, 2004). Bilotkach et al. (2010) concluded that HSR entry may impose a competition pressure on airlines to raise quality by increasing frequency but it has no impact on routes less than 550km. However, Jimenez and Betancor (2012) found that

HSR entry led to an average 17% reduction in air flights for a sample of Spanish routes (all less than 500km). Even for routes in the same country and of similar distances, the results are not always the same. For example, De Rus and Inglada's (1997) ex-post cost benefit analysis finds the 471km Madrid-Seville high speed train (HST) caused an almost 50% air passenger drop within four years of HST entry while Jimenez and Betancor's (2012) regression analysis suggests no impact of HSR entry on the 483km Madrid-Barcelona route. Thus, it seems that the literature has not reached a consensus yet on measuring the impact of HSR even for the European market.

In this paper, we employ econometric methods to examine the route-based impact of HSR entry on air traffic, in particular, the air carriers' domestic available seats, by focusing on the Northeast Asia markets, in particular Japan, South Korea and mainland China. We also compare the impacts of HSR entry in China with those in Japan and Korea noting that Japan has the world's first modern HSR service – the route between Tokyo and Osaka with a maximum speed of 210 km/hour – in 1964, while the Chinese market plays a significant role during our sampling period. Moreover, in order to address the issues mentioned above, we apply the difference-in-difference (D-in-D) estimator with propensity score matching (PSM) approach. To our knowledge, this is the first paper using the D-in-D approach to measure the impact of HSR entry on air transport.[1] We find differentiated HSR impacts in the three Northeast Asian countries. In China, the strongest impact is on short-haul air routes with distance less than 500km, while in Japan the strongest impact seems to be on medium-haul routes (between 500km and 800km). Overall, the HSR entry tends to have stronger impacts in China than in Japan and South Korea, especially in short-haul and long-haul markets. The HSR entry impact is also relevant to the speed of HSR services. In China, HSR services with a maximum speed about 200km/h can cause significant reduction in airline seat capacity on medium-haul routes but induce more seat capacity on long-haul routes. The entry of higher speed HSR services (with a maximum speed about 300km/h) do not lead to further reduction in airline seat capacity on medium-haul routes but will cause strong negative impact on long-haul routes. However, little impact is observed in Japan's long-haul markets even though its HSR services have a similar speed as those faster ones in China. We also observe

[1] The D-in-D approach requires the estimation with panel data and two-way fixed effects, i.e. time and route fixed effects. To our best knowledge, either one or two of the fixed effects are missing in the existing studies when panel data is used (see Table 1 for details). Moreover, PSM has never been applied in the literature on measuring HSR impacts.

different outcomes before and after propensity score matching, suggesting that an appropriate sampling process is necessary to quantify the impact of HSR.

The rest of the paper is organized as follows. Section 2 summarizes the datasets and airline routes included in our study and describes the variables used in the analysis. Section 3 describes our econometric methods and presents the preliminary results using the D-in-D method before PSM is carried out. Section 4 reports the major results after the matching is conducted and the new control group is constructed. Section 5 provides further discussion and interpretation on the differentiated findings in the three countries. Finally, Section 6 provides concluding remarks.

2. Data and variable construction

2.1 Airline routes and HSR services included in the sample

Our empirical study is based on a number of domestic routes originated from ten major airports in Northeast Asia: namely, Beijing Capital, Shanghai Hongqiao, Shanghai Pudong, and Guangzhou airports in China; Tokyo Narita, Tokyo Haneda, Osaka Itami and Osaka Kansai airports in Japan; and Gimpo and Incheon airports in South Korea.[2] Shanghai Hongqiao Airport was Shanghai's only airport until 1999 when Shanghai Pudong Airport commenced operation. Shanghai Hongqiao operated only domestic flights for a long time after the opening of Shanghai Pudong, but it resumed some regional and international flights (mainly to/from Japan and South Korea) in 2007. Given that both airports are operated by the same airport authority, we combined these two airports into one airport entity. Similarly, Gimpo Airport used to be the only airport in the Seoul area until Incheon Airport opened in 2001 and hence we aggregated Incheon and Gimpo airports into one airport. As a result, there are in total eight *de facto* origin airports in our data set (Table 2).

[Insert Table 2 here.]

We include all the domestic airline routes originated from these eight airports in the study based on the OAG (Official Airline Guides) scheduling data, and the sample consists of a panel of 19 years' annual data from 1994 to 2012. Here, we define a route as a non-stop city-pair air travel market. Each observation consists of a route-year pair. We consider HSR affected routes, i.e. those

[2] These are all major airports in their respective countries, which might produce biased results regarding the impact of HSR entry. That is, there could be different HSR impacts on routes serving smaller airports than our sample airports. We discuss the issue further in both Sections 5 and 6 (concluding remarks).

that saw a direct entry of HSR service, as routes belonging to the "treated" group, while those not facing HSR entry within the sampling period as routes belonging to the "untreated" (or control) group.

In total there are 503 domestic airline routes and 9,557 observations throughout the sampling period. In terms of HSR services, we are only interested in those that are operating in parallel with one of the 503 domestic airlines routes, i.e. HSR services that offer an alternative to the domestic airline routes included in this study. The entry dates of HSR services between two airports are collected from various sources, such as news media, newsletters and announcements from government agencies. Up to the end of the sampling period, 119 of the 503 air routes have an alternative HSR service. However, since we would like to analyze the impact of HSR entry by comparing the airlines' available seats before and after the HSR entry, we only take into account HSR services that commenced after 1994 and before the end of 2012. Thus, 36 Japanese routes which started HSR services before 1994 are excluded. Given that the analysis is based on annual data while entry of HSR may close to the end of a year and hence has very limited impacts on the air traffic in the year of entry, we consider the "effective" entry year as one year after the actual entry if HSR enters in the fourth quarter of the year. In China, 15 HSR entries occurred in December 2012, so these routes are considered as untreated. Thus, among these 119 treated routes, only 68 satisfy this requirement and are kept in the sample as treated routes (Table 3). As a result, only 467 airline routes are kept in the analysis, 68 treated routes and 399 untreated routes, leading to a total sample size of 8873 observations.

[Insert Table 3 here.]

As shown in Table 3, China has the most HSR entries during the sampling period, followed by Japan and then Korea. All the Korean HSR entries are in the short-haul markets while in China and Japan the HSR affected routes are relatively evenly distributed across the three distance categories. Appendix A shows the development of relevant HSR network across time in the three countries based on actual entry years. HSR lines which are not parallel with air routes starting from the airport cities listed in Table 2 are excluded from the graphs.

2.2 Variables and data sources

The dependent variable of our D-in-D analysis is the airlines' seats available (Y) on each route. This data is retrieved from OAG. The top part of Table 4 shows the average Y and frequency (F) for the full sample as well as subsamples, including the treated and untreated routes. The bottom part represents the difference in Y and F between the control and treatment groups and the corresponding standard error of t-test. The routes in the treatment group on average have larger available seats and frequency than those in the control group, while this difference becomes smaller after the propensity score matching with the nearest neighbor approach.

[Insert Table 4 here.]

Airline route-based explanatory variables used in the D-in-D analysis include total population of the route (rPOP), per capita real GDP of the route (rGDP_POP) and low cost carrier dummy (LCC). The route total population is the sum of the two endpoints' population of each route. The per capita real GDP of a route is calculated as the total real GDP of the two endpoints divided by their total population, which is further converted to millions of year 2000 US dollars. The nominal GDP data and corresponding deflator data come from the National Bureau of Statistics of China, Cabinet Office of Japan and Statistics Korea. As each country has its own administrative division which might differ from other countries, we use prefectural city level GDP and population data for China (Beijing and Shanghai are provincial level metropolitans and hence the provincial level data are used for these two cities), the prefectural city level data for Japan and provincial level data for South Korea. Both population and GDP are calculated based on the endpoints' catchment. The origin airport cities' catchments are defined in Table 2 and the destination airport cities' catchments are defined in different ways. In general, as cities in China tend to scatter apart from each other, we define the prefectural level city containing the airport as the catchment of each destination airport. In Japan and Korea, a destination airport's catchment consists of not only the province or prefectural level city containing the airport, but also some surrounding area.

Table 5 presents the summary statistics of major variables. In general, routes in China have slightly smaller average population but much smaller average per capital GDP and available seats (Table 4) than those in Japan and South Korea. Table 5 also lists the summary statistics of air route distance (Distkm) and road distance (droad) used in PSM. The former is defined as the greater circle distance between the origin and destination airports of a particular route and the latter is the shortest road distance between the origin and destination airports found via Google map. The destination airport catchments' population (POP) is used in PSM for Japan.

[Insert Table 5 here.]

3. Econometric methodology

3.1 D-in-D approach

We quantify the impact of HSR entry by comparing the seat capacity of the treated and control routes before and after the treatment (i.e. HSR entry) with the D-in-D method. Similar to the common D-in-D approach with multiple time periods, following Bakis et al. (2015), we estimate the following specification for each country:

$$Y_{rt} = \beta_0 + \beta_1 HSR_{rt} + \beta_2 rGDP\_POP_{rt} + \beta_3 rPOP_{rt} + \beta_4 LCC_{rt} + Year_t + route_r + \varepsilon_{rt} \ , \qquad (1)$$

where Y_{rt} is the airlines' total seats available on route *r* in year *t*. HSR_{rt} is the policy variable measuring the difference in seats available before and after the treatment. Hence it equals one if parallel operation of HSR exists on route *r* in year *t*; otherwise, it equals zero. That is, if a route encounters HSR entry in 2007, its HSR_{rt} equals one for all the periods from 2007 to 2012 and equals zero for all the remaining years. Thus, the coefficient (β_1) of HSR_{rt} measures the average impact of HSR service on airlines' route-level seat capacity and hence is the main focus of this paper. Variables which control for the route characteristics include the route-level per capita real GDP ($rGDP\_POP_{rt}$), the route-level population ($rPOP_{rt}$) and the low cost carrier dummy (LCC_{rt}) which equals one if low cost carriers operate on the route in period *t*. $Year_t$ controls for the year-specific fixed effect throughout the sampling period. $route_r$ captures the route-specific fixed effects which vary across routes, but tends to be constant over time. Since all time-invariant features can be captured by the route-level fixed effect, time-invariant route characteristics used in the literature, such as route distance, hub airports and treatment, are excluded. Finally, ε_{rt} is the error term.

The most crucial assumption in the D-in-D approach is that the treatment and control groups should have a common trend before the treatment. This condition is difficult to verify especially because in our sample the HSR entry years vary across treated routes. Applying Galiani et al.'s (2005) approach, we estimate equation (1) by removing policy variable HSR_{rt}, separating the yearly dummies for treated and untreated routes, and including only pretreatment observations in the estimation. That is, treated observations taken on or after the HSR entry years are removed. If there are no statistically significant difference in the coefficients of yearly dummies between treated and untreated groups, we consider that the common trend requirement is satisfied. Figure 1 plots yearly fixed effects of treated and control groups when equation (1) is fitted and Appendix B reports the statistical tests for the difference between these two groups. In general, we cannot reject the null hypothesis that treatment and control groups have the same yearly fixed effects for China both before and after the matching. In the case of Japan, although the fixed effects are different for certain years before matching, we observe no statistically significant difference after matching. Thus, the common trend assumption is hold in general for China and Japan, especially after the adoption of PSM. The case of South Korea is a bit complicated. The treated routes have a substantially higher yearly fixed effects before 2003 than untreated routes, suggesting a violation of common trend condition. However, no difference between treated and control groups is observed after all observations prior to 2003 are removed. Thus, the common trend requirement may still be satisfied if we only take into account the periods after 2003, which is confirmed by statistical tests in Appendix B. Thus, our regression analysis for Korea only covers periods from 2003 to 2012.

[Insert Figure 1 here.]

The initial D-in-D regression results before PSM is presented in Table 6.[3] One potential problem is heteroscedasticity due to the high diversity among sampled routes, thus we report the White heteroscedasticity consistent robust standard errors. As the sampling period covers 19 years, autocorrelation may occur since the modified BFN Durbin Watson statistics are far below 2 for all the three countries. Following Galiani et al. (2005), this issue is addressed by reporting robust standard errors clustered at route levels which takes into account both cross-sectional heteroskedasticity and serial correlation. The model is fitted by Ordinary-Least Square (OLS) method with fixed effects.[4] For China and Japan, Table 6 shows two sets of results, one using only

[3] We also tried specifications with semi-log and double-log forms, but find the presented no-log form the best for a few reasons. First, although we are not able to reject the null hypothesis that there is no difference in yearly fixed effects between treatment and control groups after matching with the other two specification forms, the common trend requirement seems to be better satisfied without log transformation from the fixed effect plots (see Appendix C). Second, the modified Wald tests for the cross-sectional heteroskedasticity of fixed effect panel model suggest that the group wise heteroscedasticity problem is stronger in double-log and semi-log forms than the no-log form, especially after adopting PSM (Appendix D). Third, log transformation tends to produce unreasonable coefficients of per capita GDP and population.

[4] Although the random effect model can be more efficient, the Hausman test rejects the null hypothesis at the 0.01 level of

one HSR dummy to capture the average impact across routes and the other using three interaction terms of HSR dummy and air route distance groups to capture differentiated impacts on distance groups. D1 stands for short-haul routes (less than or equal to 500km), D2 stands for medium-haul routes (over 500km but no longer than 800km) and D3 stands for long-haul routes (more than 800km). In Japan, the per capita GDP data are missing in 2011 and 2012 during which HSR entered most of the treated routes in our sample. As the coefficient of per capita GDP is statistically insignificant and removing this variable leads to limited changes in coefficients of other non-policy related variables, we exclude per capita GDP from the model.

[Insert Table 6 here.]

When pooling all air route distances together, the coefficients of the policy variable are negative for both Japan and Korea, but they are statistically insignificant in the case of China. However, the entry of HSR has a strong negative impact on the short-haul routes and a strong positive impact on the long-haul routes in China but in Japan we observe negative and statistically significant impacts in short-haul and medium-haul markets. The existence of low cost carriers and an increase in population are associated with higher seat capacity in all the countries. In China, an increase in per capita GDP is associated with more seat capacity but it is negatively associated with air operation in Korea.

In addition to the common trend condition, the D-in-D regression approach is suitable for measuring the impact of external shocks if the following two conditions are satisfied: (1) HSR chooses the routes to enter independently and randomly in the sense that only observable characteristics affect HSR's route selection, and (2) the treatment and control groups should satisfy the common support requirement, i.e. the domains of treatment and control groups should be sufficiently overlapped. The first condition is satisfied with equation (1), though rail networks are rarely formed in a random way. It is true that railway companies usually select railway routes carefully before constructing the entire network, as HSR construction is quite expensive.[5] Therefore, to justify the high construction cost, HSR tends to serve high density markets, in

significance that there is no systematic difference between the coefficients estimated by fixed effect and random effect models. Given that fixed effect models tend to produce consistent estimators while it may not be the case for random effect models, we use fixed effect models in the paper. Due to the space limitation, we do not provide the Hausman test results in the paper, but they are available from the authors upon request.

[5] According to Ollivier et al (2014), by the end of 2013, the average construction cost of the 250km/h railways in operation ranges from RMB 70 million per km to RMB 169 million per km, while the 350km/h railways are about RMB 94-183 million per km.

particular, markets with a high level of population and economic activities and hence a high level of travel demand. Since the major factors that influence the entry of HSR are controlled in the D-in-D regression by adding GDP per capita and population as explanatory variables in equation (1), the selection bias is not an issue in this paper. However, the second condition, i.e. the common support assumption, could be violated due to HSR's tendency of entering high density routes which may have quite different characteristics than markets with low travel demand and this is the reason why PSM is employed to match treated routes with similar control routes.

3.2 Propensity score matching

One widely used approach to match observations in the control group with those in the treatment group is to calculate the propensity scores and match observations from the two groups based on these scores. Only those untreated routes which are matched with a particular treated route are kept for the estimation of HSR impact in the D-in-D regression.

A common way to obtain the propensity scores is to fit a binary choice model, such as probit and logit models. In particular, we use the probit or logit model to estimate the conditional probability (propensity score) that a particular route *r* is assigned into the treatment group given a set of observed route-related characteristics or factors. That is,

$$p(x) = \Pr(treat_r = 1 \mid x), \tag{2}$$

where $treat_r$ is a binary variable, indicating if the route is in the treatment group or the control group. The vector of characteristics, *x*, affects the likelihood that a route is assigned into the treatment group. In this paper, we conduct the matching for routes in China and routes in Japan separately. That is, treated routes in China must be matched with untreated routes in the same country and similarly each treated route in Japan will be matched with untreated routes in Japan. Thus, an independent binary model is fitted for Chinese and Japanese routes respectively. As the numbers of treated and untreated routes in Korea are 4 and 12 respectively, propensity score matching cannot be performed for Korea. Thus, in this paper, D-in-D analysis after matching is not available in the case of Korea.

The matching process requires the probability of HSR entry not affected by the status of the treatment, i.e. to satisfy the unconfoundedness assumption (Bakis et al, 2015; Lechner, 2011). In the present paper, this assumption would be violated if we conducted the matching based on years after the entry of HSR services, because in addition to potentially substituting to airline services, HSR may at the same time induce first-time travelers who would otherwise have not traveled by air and extra trips of current travelers (Goldman Sachs, 2010). The increased mobility could in turn affect the endpoint cities' population and per capita GDP,[6] two factors which are likely to associate with the chance of HSR entry. Therefore, we match routes in China with the data taken in 2005 since GDP or population data are not available for many sampled endpoint cities in China before 2005 and most importantly the first HSR entry in our Chinese route sample happened in 2007 which is two years after 2005. Thus, we believe that using this year's data to match the Chinese routes would minimize the chance of violating the unconfoundedness assumption. Similarly, routes in Japan are matched with the 1994 data, since 1994 is the first sampling year and it is three years before the first sampled HSR entry happened in 1997 in Japan.

To fit the binary model, we consider a few characteristics as *x* regressors which influence the probability of HSR entry. In the case of China, the *x* regressors include the natural log of total population of the city pair route (lnrPOP), the natural log of per capita real GDP of the route (lnrGDP_POP), the natural log of air route distance (lnDistkm), the natural log of road distance (lndroad) and origin airport dummies. In the case of Japan, neither lnrPOP nor lnrGDP_POP is statistically significant in the probit or logit model and hence they are not good predictors of the probability of offering HSR services. Rather, the natural log of destinations' population (lnPOP) is a better predictor and hence for routes in Japan, we replace lnrPOP and lnrGDP_POP with lnPOP. The origin airport dummies are excluded for Japan, because none of them are statistically significant. Nine untreated routes in Japan involve island destinations which are not reachable via ground transport and hence they are removed. The Lagrange Multiplier test for normality is conducted after fitting the probit model. The test rejects the normality assumption for Japan and therefore a logit model is fitted for Japan. Table 7 presents the probit regression results for China and the logit regression results for Japan. In China, the probability of encountering HSR entry increases in per capita real GDP and population of the route but decreases in the route distances. However, in Japan, the probability of HSR entry increases in the destination airport catchment's population as well as the route distance.

---

[6] Levinson (2012) summarized mixed findings on the HSR's economic development effects. For example, in Japan, metropolitan population growth was found correlated with the HSR entries but the causality was unclear. There is also evidence showing that HSR entry did not affect firms' location choice in the Netherlands.

[Insert Table 7 here.]

With the estimation results in Table 7, the propensity scores are calculated by predicting the probability of having $treat_r$ equal one with the fitted models, probit for China and logit for Japan. Then, three matching methods, the nearest neighbor, caliper and radius, are applied and both with replacement and without replacement approaches are used. When caliper and radius matching are conducted, we tried different tolerance levels to search for a better matching outcome. Treated routes outside of the range of common support are excluded from matching. In general, the nearest neighbor with no replacement method produces less effective matching than the other methods (see Appendix E for an example) as the bias across covariates keeps to be high after matching and the matching does not lead to similar propensity score distributions between treated and untreated routes. Thus, the matching results of the nearest neighbor with no replacement approach is not used for further D-in-D analysis. No matching method is the best for all the cases and hence we only present one good matching outcome for each country in this section as examples. Figure 2 shows the common support and compares standardized percentage bias for covariates and the propensity score distributions before and after the radius matching for China. Figure 3 shows the outcomes of caliper matching (with replacement) for Japan. Due to the space limit, the outcomes of the other matching methods are available from the authors upon request.

[Insert Figure 2 here.]

[Insert Figure 3 here.]

The matching outcome should satisfy the balancing property in the sense that the treated and matched routes have similar distribution of observable characteristics used to predict the propensity scores. After matching, the differences in the covariates between treatment and control groups has been substantially reduced and become statistically insignificant. Thus, the balancing property has been largely satisfied. The propensity score distributions of treated and untreated groups become more overlapped after matching, suggesting that the common support assumption is satisfied.

4. D-in-D estimation after matching

Tables 8 and 9 present the D-in-D regression results for China and Japan respectively based on various matching methods using lnDistkm as an independent variable when calculating the propensity scores. Using lndroad generates similar results which are not presented due to space limit but are available from the authors upon request. The results are in general consistent with those obtained before matching in terms of signs and statistical significance of the coefficients, but the magnitudes differ. An increase in route-level population by 1 person would raise annual seat capacity by 0.07-0.09 seats in China but only 0.01-0.04 seats in Japan. Route level per capita real GDP also has strong impact on Chinese route seat capacity. After matching, the impact of low cost carriers is reduced in Japan but increased in China.

[Insert Table 8 here.]

[Insert Table 9 here.]

Similar to the results presented in Table 6, different HSR impacts in China and Japan are observed. In particular, there is a statistically insignificant overall impact of HSR entry on airline seat capacity in China but a strong negative impact on short-haul routes, a strong positive impact on long-haul routes and statistically insignificant impact on medium-haul routes (Table 8). However, the magnitudes of the impacts on these two distance groups have reduced after PSM. According to our point estimation based on various fitted models, the entry of HSR would lead to an average 83% reduction in airline seat capacity on short-haul routes and an average 28% increase on long-haul routes, while the change in medium-haul routes is only about -4%. To better understand whether the change in seat capacity is mainly driven by the change in frequency or aircraft size, we conduct a similar D-in-D analysis by replacing the dependent variable (Y) with frequency (F), which is presented in column (9) of Table 8. We find that the coefficients of the policy variables are consistent with those presented in columns (5)-(8) and on average there are about 78% frequency drop in short-haul routes and 25% frequency increase in long-haul routes. This is partly consistent to Bilotkach et al.'s (2010) finding based on European air routes that airlines may increase frequency to compete with HSR long-haul routes by offering a higher "service quality", but in our study we also find a drop in frequency in short-haul routes which is not observed by Bilotkach et al. (2010). Given that the percentage changes in frequency are close to the changes in seat capacity, it is possible that there is limited change in aircraft size,[7] which is consistent with Givoni and Rietveld's (2009) finding. Thus, the increase in seat capacity in long-

[7] We are not able to conduct a separate analysis regarding the impact on aircraft size, since quite a few routes in our sample have zero operation in certain years and hence the appropriate aircraft size data are not available for these routes, while there is no good reason to drop these routes.

haul routes is most likely contributed by the frequency competition by airlines when facing the entry of HSR. In Japan, HSR entry has a strong overall negative impact on airline seat capacity (Table 9), but we observe a relatively small seat capacity reduction in short-haul routes (about 28% on average), a large reduction in medium-haul routes (about 79%) and statistically insignificant change in long-haul routes in Japan. We also find that the impacts on seat capacity and frequency are consistent.

We also investigate the impact of HSR entry over the time by adding leads and lags of the treatment. In particular, variable HSR00 indicates the year of HSR entry for a particular route, HSRk0 indicates k years before the HSR entry (leads) and HSR0k indicates k years after the HSR entry into a particular route (lags). We consider up to five years before and after HSR entries and HSR06 indicates all observations more than 5 years after HSR entries. Figure 4 plots the coefficients of HSRk0's, HSR0k's and HSR00 as well as their respective 95% confidence intervals and hence tells the impact of HSR entry over time. When all distance groups are pooled together, all the dummy variables indicating years before and after the HSR entry are statistically insignificant in China, though there seem to be slightly negative impact since two years after the HSR entry. In Japan, although a statistically significant negative impacts are observed two years after the entry, the magnitudes are small for the following years. However, stronger negative impacts in Korea are observed for five years after the entry, mainly due to the fact that Korean routes are all short-haul. Considering the higher average seat capacity in Korea than in China and Japan, the large magnitude in Korea's seat capacity reduction can be translated into only about 56% drop, which is much lower than the percentage change in Chinese short-haul routes.

[Insert Figure 4 here.]

As mentioned earlier, the impact of HSR entry differs across air route distances, so we further analyze the impact over time by distance groups by adding leads and legs only for routes belonging to a particular distance group while keeping the distance group-HSR entry interaction variables of the other two distance groups. For example, to analyze the impact on short-haul routes overtime, we include HSRk0's, HSR0k's and HSR00 only for short-haul routes and D2*HSR and D3*HSR for medium-haul and long-haul routes respectively. Figure 5 shows the HSR-entry impacts on Chinese routes over time by different distance categories, and it reveals some interesting findings. First, the strong negative impact occurs on short-haul routes. For this category of routes, the seat capacity reduction is observed since the HSR entry year and the capacity reduction continues as the time passes by, since the coefficients of HSR lags are all statistically significant and become increasingly more negative over time. Second, the impact on routes in the medium-haul category is not statistically significant. However, there might be a negative impact in the third year after the HSR entry year, as the coefficient of HSR02 is quite negative though statistically insignificant. Similarly, the impact on routes in the long-haul category is in general not statistically significant except in the fifth year after the HSR entry, but there seems to be an upward pressure on seat capacity over time.

[Insert Figure 5 here.]

The above differentiated impacts across route distances are not exactly the same as in Japan (Figure 6).[8] In particular, both short-haul and medium-haul routes in Japan have encountered capacity reduction after HSR entries. The short-haul routes started to reduce capacity in the second year after the HSR entry while the medium-haul routes seem to start substantial capacity cut from the HSR entry year. However, no statistically significant impact on airline seat capacity is found on long-haul routes.

[Insert Figure 6 here.]

5. Discussion and interpretation of the results

Further analysis reveals that both route distance and HSR speed would cause different impacts of HSR entry between China and Japan in medium-haul and long-haul markets. In fact, China started train services with rolling stocks which are claimed to have a maximum speed of more than 200km/h during the country's sixth railway speedup campaign in 2007. These rolling stocks have a designation of CRH (for "China Railway High-speed") and the corresponding train services have a letter "D" in front of the service codes so that they can be differentiated from other conventional train services. Although these D services are not considered as high-speed rail in China, they represent a significant speed increase and satisfy the definition of high-speed rail services in many other regions of the world. D services used to be the only services using CRH trains until the

[8] Note that we have only two years' observations after HSR entries on medium-haul and long-haul routes. This is because all treated long-haul routes encountered HSR entry in 2011 and medium-haul routes starting HSR operations before 2011 are not on the common support during the PSM and are excluded from D-in-D analysis.

services with a maximum speed of 300km/h started since 2008, which included some C services (inter-city rail) and all the G services (high-speed rail) according to the pre-2015 classification of train services.[9] Thus, our study considers both the 200km/h and 300km/h services as HSR, and we simply call the former as D-train and the latter G-train. Appendix F summarizes the operating speed of sampled HSR services in China, Japan, and Korea. The operating speed is estimated by the ratio of travel distance and scheduled travel time. Note that the average operating speed of HSR in Japan is comparable with that of G-train services in China, while the average operating speed of HSR in Korea is comparable with that of D-train services in China. Thus, to have a better cross-country comparison, it is necessary to separate the impacts of D-train and G-train.

While keeping control variables such as per capita GDP, population and LCC, we separate the D-train effect for medium-haul routes in China by including leads and legs of policy variables for medium-haul treated air routes entered by D-train only and adding three interaction variables, D1*HSR, D3*HSR, and D2*GHSR, where GHSR equals one if G-train services are available on the route. These variables are used to control for the HSR impact on short-haul and long haul routes as well as the G-train effect on medium-haul routes. Similarly, we can also separate the G-train effect over time for medium-haul routes by including leads and legs of the policy variable for observations of medium-haul treated routes when G-train services are available and adding interaction variables, D1*HSR, D3*HSR, and D2*ODHSR, where ODHSR indicates the observations of affected routes when only D-train services are available. Thus, D2*ODHSR is used to control for the D-train effect on medium-haul routes. The same idea is applied to analyze the G-train and D-train impacts on long-haul routes.

Figure 7 presents the estimated G-train and D-train impacts based on caliper matching with no replacement. For medium-haul routes, airline seat capacity starts to drop one year after the entry of D-train service and there seems no significant impact after the entry of G-train. Since in our sample seven out of the eight G-train affected medium-haul routes were first affected by the D-train services and then G-train, the entry of D-train is effective enough to compete with air by cutting travel time and diverting discretionary passengers to rail. However, the later entry of G-train is less effective in terms of further diverting the remaining air passengers to rail because the

[9] The definition of high-speed rail in China has been changed several times and the final version was set in later 2014 and started to be in effect in early 2015. Since then, many CRH train services are reclassified by changing the letters in front of the service codes. Thus, given that our data were collected in 2013, the classification mentioned in the paper may not match the classification observed nowadays.

small improvement in G-train's travel time compared with D-train (owning to the shorter travel distances) may not be enough to justify the higher prices charged by the faster service. As a result, the entry of G-train (after the D-train entry) seems to have little extra impact on air traffic in the medium-haul markets. For long-haul routes, due to the slower speed and longer travel distances, D-train does not have much competitive advantage over air and hence, for airlines the best strategy is to compete head-on with rail by improving frequency. Thus, we find an increase in airline seat capacity right after the entry of D-train in the long-haul markets. However, after the entry of G-train, the substantial improvement in speed makes HSR more attractive and hence we observe a significant reduction in airline seat capacity on long-haul routes one year after the entry of G-train.

[Insert Figure 7 here.]

The above analysis suggests that similar to the case of Japan, HSR does have negative impacts on medium-haul routes in China. In particular, the entry of D-train in China reduces, on average, 169 thousand seats per year continuously for four years after the entry (Figure 7), while the entry of HSR in Japan leads to an average reduction of 325 thousand seats per year and our data only allows us to observe the impact for two years. Since the average medium-haul seat capacity in Japan is about 1.7 times of that in China after matching, the percentage drop in seat capacity should be similar for China and Japan in medium-haul markets. However, in long-haul markets, HSR services with comparable speeds tend to have stronger negative impacts in China while little impacts in Japan.

With respect to short-haul routes, we find significant and long-lasting impacts of HSR entry which can extend to five years after the entry in all the three countries. This is consistent to Campos and Gagnepain's (2009) statement as well as the finding by González-Savignat (2004). Nevertheless, Chinese air routes have the most substantial percentage loss in seat capacity, followed by Korea and then by Japan. This might be caused by the different air route distance distributions of the sampled treated routes (Appendix G). China has a wider range of treated route distances, in which 10 of the 17 short-haul routes are shorter than 300km, while in Korea only one of the four treated routes are shorter than 300km and in Japan all short-haul routes are longer than 400km. Thus, it is not surprising that the entry of HSR has on average a stronger impact in Chinese short-haul air markets. However, after removing routes shorter than 300km from the sample of Chinese routes, although the impact of HSR entry on short-haul routes reduces to a 67% seat capacity drop on average,[10] it is still stronger than those in Japan and Korea in terms of percentage seat capacity reduction.

Thus, except the medium-haul markets, for given comparable HSR speeds airlines in China in general tend to adjust (reduce) seat capacity more than airlines in Japan and Korea when HSR entry occurs. This might be explained by a few differences between the Chinese and Japanese/Korean air travel markets. First, the Chinese domestic market is still in the fast growing and developing stage. Comparing with Japan and Korea, China's GDP per capita is still quite low (see Table 5) but is growing rapidly, and hence there is much more opportunity to develop new air travel markets in China than in Japan or Korea.

Second, the domestic air travel market regulations are quite different between China and its two Northeast Asian neighbors (e.g., Ha et al., 2010, 2013). The Chinese government has much more control on the entry into the air transport business during the sampling period. Although China's big three airlines (Air China, China Eastern, and China Southern) do compete with each other on certain routes, they are still dominating the air transport market and given that they are all state-owned while low cost carriers only account for roughly 7% of the market share in China in 2015, the level of competition in the passenger air transport market is relatively limited. Thus, there are still new or existing markets which can be developed further. In fact, after the Civil Aviation Authority of China (CAAC) announced to again allow entry of private and/or low cost airlines in 2013,[11] a large number of potential entrants submitted application, suggesting that the market is far from saturated. However, compared with the air travel market in China, domestic air travel markets in Japan and Korea have been more competitive, since in both countries the airline industry has been liberalized and low cost carriers have already played a significant role in domestic operations. (For example, 48% of South Korea's domestic traffic was operated by low

---

[10] The impacts on medium-haul and long-haul routes remain almost the same as before even if routes below 300km are removed.

[11] China's private airlines emerged in 2005 with more than 10 being licensed in that year following the release of the "Regulation on Domestic Investment on Civil Aviation" by the Chinese government in 2004, which allowed private sector participation in the civil aviation industry, including setting up new airlines. As a result, three private carriers (Okay Airways in Tianjin, United Eagle Airlines in Chengdu and Spring Airlines in Shanghai) launched their maiden flights in 2005 (Lei and O'Connell, 2011). Spring Airlines positioned itself as an LCC and believed that this model would help them secure a slice of the market dominated by their state-owned counterparts (Zhang and Lu, 2013). In 2006 two other private carriers (Juneyao in Shanghai and East Star Airlines in Wuhan) began operating. By 2007 some 20 new private airlines had been approved in China. The expansion in the number of new airlines led to the civil aviation authority's decision to suspend approval of new domestic entrants until 2010. Given safety concerns following the crash of an aircraft of a local airline in 2010, the government extended this policy until 2013. Subsequently, another wave of private airlines emerged in 2013 and 2014: nine have been approved and more are in the application process (Zhang and Zhang, 2016). Finally, the two aspects, high traffic growth and restrictive regulations, have also been discussed in detail in Hu and Zhang (2016). Hu and Zhang further explore their implications for carriers' aircraft acquisition in terms of plane types.

cost carriers in 2015.) Although developing new markets is still possible in Japan as for the island markets or routes reaching the far northern part of Japan, given that airlines in Japan are all privately owned and hence they are more profit-oriented (the top-4 airlines in China are all state-owned), they possess higher tendency to stick to trunk routes rather than serve less profitable island routes or other marginal routes. The above two differences can also be seen from the OAG data that airlines in China had experienced frequent route entry and exit during the sampling period, suggesting that they are still exploring opportunities in various new markets. However, airlines in Japan and South Korea had not switched services among different domestic routes very frequently, suggesting that the airlines in these two countries had a limited opportunity to develop new domestic routes. In addition, for the period China's international air travel had grown much faster, driven mainly by the explosion of outbound tourism, than Japanese or Korean international travel. Therefore, after encountering HSR entries, airlines in China have larger flexibility to switch to other domestic or international routes, while airlines in Japan and Korea are relatively less responsive in terms of redeploying their capacity to other routes.

Third, compared with major airports in Japan and Korea, those in China, especially airports in Beijing, Shanghai and Guangzhou, have been experiencing prolonged delays for many years and the on-time performance can drop even below 40% during the peak seasons. This high level of flight delays primarily due to air traffic congestion can reduce passengers' net utility of taking a flight and make alternative HSR services more attractive.

Before closing this section, it is worth discussing the issue indicated in footnote 2: all our sample airports are major airports in their respective countries, of which Narita, Incheon, Beijing, and Shanghai Pudong might be considered as international "hub" airports. For a hub, HSR may provide feeding services to airlines for long-haul domestic or international flights while competing with airlines at the same time. Thus, the entry of HSR may benefit some "untreated" air routes which are difficult to be identified technically. Thus, there might be a certain level of overestimation on the impact of HSR in our study. However, we believe the amount of overestimation is limited especially for China. First, although origin airports included in the study operate large traffic volumes, they are traditionally not good at handling connecting flights, especially in China. According to the latest statistics we have, more than 10% of the passenger traffic in Shanghai Pudong are connecting passengers in 2015 while this number is only 7% and 6% in Beijing and Guangzhou a few years ago. The two largest connecting airports in our sample, Narita and Incheon, have only about 20% of the passenger traffic being connecting. Thus, even if all the original air-air connections were replaced by HSR-air connections, it would not constitute a significant share of the traffic. Second, HSR feeding may benefit the international routes the most as international flights tend to be concentrated in a few gateway airports but many medium/small sized airports provide quite a few domestic flights. Thus, HSR linkage makes international flights in gateway airports more accessible to people living far away. Since all international flights are excluded from our study, this issue should have negligible impact on our study. Third, the effectiveness of using HSR to provide feeding services depends on many factors, especially the convenience of transferring passengers between HSR stations and airports. During our sampling period, in the case of China, only the Hongqiao Airport in Shanghai has an HSR station nearby and hence an excellent access to HSR since July 2010. Shanghai Pudong Airport is 60km away from the HSR station, Beijing Airport is about 40km away and Guangzhou Airport 50km away. Considering the crowded subways and congested roads in these three cities and the lack of railway services linking these airports and HSR stations, air-HSR connection can be somewhat difficult in these three cities than the other major airport cities in Japan and South Korea in our sample.

6. Concluding remarks

We have empirically investigated the impact of HSR entry on airlines' adjustment in route-level seat capacity. In particular, the D-in-D approach was applied to distinguish the treated routes that encountered HSR entries during the sampling period, from the untreated (control) routes that had never encountered HSR entries, while taking account of the difference across time periods as well as the difference before and after the treatment (i.e. HSR entry). The propensity score matching was applied to pair treated routes with similar untreated routes which form the new control group for further D-in-D regressions.

The literature has usually identified negative impacts of HSR entry in various countries, such as European countries and Japan, in the sense that for short-haul air routes (e.g. less than 700km or 800km) HSR is likely to have a major advantage over airlines and would capture the majority of the market share after commencing the service. However, to our knowledge no existing study has applied the D-in-D approach which seems to produce different results as shown in the present paper: HSR's impacts could vary across countries, HSR service speed as well as distance. Overall, with similar HSR speeds, the entry of HSR tends to have more severe impacts in China. In particular, in short-haul (less than 500km) routes airlines begin to divert their seat capacity to other routes right after the entry of HSR in China but such significant reduction would not start in South Korea and Japan until one or two years after the entry, leading to lower percentage seat capacity reduction in Korea and Japan. In Japan, the impact of HSR on airlines' seat capacity is negative for medium-haul (between 500km and 800km) routes but statistically insignificant for long-haul routes. In China, however, it seems that the impact is related to both the HSR speed and route distance. That is, in medium-haul markets, seat capacity reduction is observed after the entry of HSR services with lower speed (comparable to the speed of Korean HSR) and little extra impact is found after the entry of higher speed HSR service (comparable to the speed of Japanese HSR). In long-haul (more than 800km) markets, the entry of the lower speed HSR induces more seat capacity due possibly to airlines' strategy to compete with improved frequency, but substantial seat capacity cut is observed after the entry of higher speed HSR services. Both phenomena are not observed in long-haul markets in Japan. Thus, the argument by Martin and Nombela (2007) and Armstrong and Preston (2011), that airlines and HSR mainly compete in the routes over 500km, does not seem to always hold in the case of Northeast Asian markets.

Our results call for research to further identify the underlying reasons for the differentiated impacts in China and the other two Northeast Asian countries. Differences in per capita income levels, domestic airline regulations and airport congestion could be part of the reasons as discussed in Section 5 of the paper. Other possibilities that deserve an in-depth study in the future could be the geographic differences and hence HSR network differences of the three countries. Second, airfare is omitted in our models due to the unavailability of the appropriate data. We acknowledge that the presence of HSR could also put a downward pressure on airfare which in turn will lead to an increase in passenger traffic. While the inclusion of HSR, low cost carrier dummy and other route-level variables could partly remedy the omission of the airfare variable, it is important to quantitatively control for the impact of price competition once appropriate data is available in the future. Third, due again to data constraints, air routes included in this study are mainly originated from major hub airports, it would be interesting to see if air routes serving medium- and small-size airports will encounter similar impacts.

Fourth, according to the latest plan, HSR in China would consist of eight vertical (roughly the North-South direction) corridors and eight horizontals (roughly the East-West direction) corridors together with regional lines ("branches") and hence the Chinese HSR network looks like a grid. The first part of such a grid structure, with four vertical lines and four horizontal lines, had been largely formed by the end of 2015. However, HSR networks in Japan and especially in Korea more look like a line or star rather than a grid. We have little knowledge on how different HSR network structures would affect airlines' reactions to HSR entries. In the short run, airlines could simply divert some capacity to routes with no HSR competition, such as international routes or low density domestic routes where constructing HSR routes is physically, politically or economically infeasible, and cover more fringe routes (Jiang and Zhang, 2016). In the long run, however, the entire network of an airline could be reconfigured as a result. For example, as predicted by Jiang and Zhang's (2016) theoretical model, airlines may have a greater incentive to move towards hub-and-spoke network structure as a long-run response to HSR entries. Finally, in addition to competition, air and HSR can also be complements. For example, in Japan after the entry of Hokkaido Shinkansen, the local government of Hokkaido planned to boost the local aviation market via the HSR connection at the Hakadate HSR station. Such complementarity between air and HSR should be considered in the future work.

References

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.

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.

Álvarez-SanJaime, Ó., Cantos-Sánchez, P., Moner-Colonques, R. and Sempere-Monerris, J.J. (2015), "A model of internal and external competition in a high-speed rail line", *Economics of Transportation*, 4 (3), 178-187.

Armstrong, J. and Preston, J. (2011), "Alternative railway futures: growth and/or specialisation?", *Journal of Transport Geography*, 19 (6), 1570–1579.

Bakis, O., Inci, E. and Senturk, R.O. (2015), "Unbundling curbside parking costs from housing prices and rents", Working Paper, Sabanci University, Turkey.

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.

Bilotkach, V., Fageda, X. and Flores-Fillol, R. (2010), "Scheduled service versus personal transportation: The role of distance", *Regional Science and Urban Economics*, 40 (1), 60–72.

Campos, J. and Gagnepain, P. (2009), "Measuring the intermodal effects of high speed rail", in: de Rus, G. (ed.), *Economic Analysis of High Speed Rail in Europe*, Bilbao, Fundación BBVA, pp. 71–88.

Cascetta, E., Papola, A., Pagliara, F. and Marzano, V. (2011), "Analysis of mobility impacts of the high speed Rome-Naples rail link using within day dynamic mode service choice models", *Journal of Transport Geography*, 19 (4), 635–643.

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.

Clever, R. and Hansen, M. (2008), "Interaction of air and high-speed rail in Japan", *Transportation Research Record: Journal of the Transportation Research Board*, 2043, 1–12.

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.

D'Alfonso, T., Jiang, C. and Bracaglia, V. (2015), "Would competition between air transport and high-speed rail benefit environment and social welfare?", *Transportation Research Part B: Methodological*, 74, 118-137.

D'Alfonso, T., Jiang, C. and Bracaglia, V. (2016), "Air transport and high-speed rail competition: Environmental implications and mitigation strategies", *Transportation Research Part A: Policy and Practice*, forthcoming.

De Rus, G. and Inglada, V. (1997), "Cost–benefit analysis of the high-speed train in Spain", *The Annals of Regional Science*, 31 (2), 175–188.

Dobruszkes, F., Dehon, C. and Givoni, M. (2014), "Does European high-speed rail affect the current level of air services? An EU-wide analysis", *Transportation Research Part A: Policy and Practice*, 69, 461-475.

Fu, X., Zhang, A. and Lei, Z. (2012), "Will China's airline industry survive the entry of high speed rail?" *Research in Transportation Economics*, 35, 13-25.

Galiani, S., Gertler, P. and Schargrodsky, E. (2005), "Water for life: The impact of the privatization of water services on child mortality", Journal of Political Economy, 113(1), 83-120.

Givoni, M. (2007), "Environmental benefits from mode substitution: Comparison of the environmental impact from aircraft and high-speed train operations", *International Journal of Sustainable Transportation*, 1 (4), 209-230.

Givoni, M. and Rietveld, P. (2009), "Airline's choice of aircraft size – explanations and implications", *Transportation Research Part A: Policy and Practice*, 43 (5), 500–510.

Goldman Sachs (2010), "High-speed rail less of a threat than you may think", Global Investment Research, Oct 28, 2010, The Goldman Sachs Group.

González-Savignat, M., (2004), "Competition in air transport: The case of the high speed train", *Journal of Transport Economics and Policy*, 38 (1), 77–108.

Ha, H.K., Wan, Y., Yoshida, Y. and Zhang, A. (2013), "Airline market structure and airport efficiency: Evidence from Northeast Asian airports", Journal of Air Transport Management, 33, 32-42.

Ha, H.K., Yoshida, Y. and Zhang, A. (2010), "Comparative analysis of efficiency for major Northeast Asian airports", Transportation Journal, 49 (4), 9-23.

Ha, H.K., Yoshida, Y. and Zhang, A. (2011), "Social efficiency benchmarking of Japanese domestic transport services: A comparison of rail and air", *Transportation Research, Part D: Transport and Environment*, 16, 554-561.

Hu, Q. and Zhang, A. (2016), "Why do emerging economies use less regional jets? A real option analysis of traffic growth and regulatory restriction", Unpublished manuscript.

Janic, M. (1993), "A model of competition between high speed rail and air transport", *Transportation Planning and Technology*, 17 (1), 1-23.

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*, forthcoming.

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.

Kappes, J.W. and Merkert, R. (2013), “Barriers to entry into European aviation markets revisited: A review and analysis of managerial perceptions”, *Transportation Research Part E: Logistics and Transportation Review*, 57, 58-69.

Lei, Z. and O’Connell, J.F. (2011), “The evolving of Chinese aviation policies and impact of a deregulating environment on Chinese carriers”, *Journal of Transport Geography*, 19, 829-839.

Levinson, D.M. (2012), “Accessibility impacts of high-speed rail”, *Journal of Transport Geography*, 22, 288-291.

Lechner, M. (2011), “The estimation of causal effects by difference-in-difference methods”, Discussion Paper no. 2010-28, Department of Economics, University of St. Gallen, Switzerland.

Martin, J.C. and Nombela, G. (2007), “Microeconomic impacts of investments in high speed trains in Spain”, *The Annals of Regional Science*, 41 (3), 715–733.

Ollivier, G., Sondhi, J. and Zhou, N. (2014), “High-speed railways in China: A look at construction costs”, China Transport Topics No. 9, World Bank Office, Beijing.

Pagliara, F., Vassallo, J.M., and Roman, C. (2012), “High-speed rail versus air transportation: Case study of Madrid-Barcelona, Spain”, *Transportation Research Record: Journal of the Transportation Research Board*, 2289, 10–17.

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, 95-104.

Roman, 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, 277-284.

Roman, C. and Martin, J.C. (2014), “Integration of HSR and air transport: Understanding passengers’ preferences”, *Transportation Research Part E: Logistics and Transportation Review*, 71, 129-141.

Rothengatter, W. (2011), “Competition between airlines and high-speed rail”, In: R. Macario and E. van de Voorde (eds.), *Critical Issues in Air Transport Economics and Business*, Oxford, UK: Routledge.

Steer Davies Gleave (2006), “Air and rail competition and complementarity”, Final report for the European Commission (DG TREN).

Xia, W. and Zhang, A. (2016a), “High-speed rail and air transport competition and cooperation: A vertical integration approach”, Working Paper, Sauder School of Business, University of British Columbia.

Xia, W. and Zhang, A. (2016b), “Effects of air and high-speed rail transport integration on profits and welfare: The case of air-rail connecting time”, Working Paper, Sauder School of Business, University of British Columbia.

Yang, H. and Zhang, A., (2012), “Effects of high-speed rail and air transport competition on prices, profits and welfare”, *Transportation Research Part B: Methodological*, 46 (10), 1322–1333.

Zhang, Q., Yang, H., Wang, Q. and Zhang, A. (2014), “Market power and its determinants in the Chinese airline industry”, *Transportation Research Part A: Policy and Practice*, 64, 1-13.

Zhang, Y. and Lu, Z. (2013), “Low cost carriers in China and its contribution to passenger traffic flow”, *Journal of China Tourism Research*, 9(2), 207-217.

Zhang, Y. and Zhang, A. (2016), “Determinants of air passenger flows in China and gravity model: Deregulation, LCC and high-speed rail”, *Journal of Transport Economics and Policy*, 50 (3), 287-303.

Tables

Table 1 Related empirical studies with econometric methods

Paper	Market	Data	Method	Variables	Main results
Demand/market share estimation with logit models					
Behrens and Pels (2012)	London-Paris	Trip-based cross-sectional survey data, 2003-2009	Multinomial and mixed logit models (five airport-airline pairs vs. HSR)	Period, fare type, average fare, road distance to the port, on-time performance, total travel time, weekly frequency	Observed stronger competition in the market than in other markets. The degree and pattern of intermodal competition depends on trip purpose. The improvement in HSR makes passengers less sensitive to HSR travel time but more sensitive to air travel time.
Campos and Gagnepain (2009)	Paris-Amsterdam	Price and market share, 2005	System of demand, pricing and cost functions (Traditional air, low-cost air, HSR, conventional train)	Price, market share	No econometric analysis due to lack of data. Simulation was conducted for the demand and pricing functions only. HSR price change has limited impact on airline demand.
Cascetta et al. (2011)	Rome-Naples	Survey, March 2008	Nested logit mode choice model (HSR, conventional trains vs. cars)	Schedule delays, travel cost, travel time, travel frequency, access/egress time	Car users are generally inelastic to HSR travel time and cost
Clever and Hansen (2008)	82 airport pairs and 1260 HSR station pairs in Japan	Intercity travel survey, 1995	Nested logit model, mode choice followed by terminal pair choice nested in each mode	Access/egress distance, line-haul time, fare, frequency, transfers	Air only competes in markets with medium access/egress distance
Pagliara et al. (2012)	Madrid-Barcelona	Survey, February and March 2010	Multinomial and mixed logit models (HSR vs. air)	Travel cost, frequency, check-in efficiency, parking capacity, access time/cost	Market share taken by HSR was lower than expected. Price and frequency are the most important determinants in competition
Park and Ha (2006)	Seoul-Daegu	Survey	Logit model (air vs. HSR)	Access and egress time, fare, frequency	Large market share drop in air was predicted.
Roman et al. (2007)	Madrid-Barcelona	Survey	Nested logit model (car, bus, train, air vs. HSR)	Travel time, travel cost, headway, access/egress time, waiting time, delay,	HSR is predicted to obtain limited market share (less than 35%) when competing with air.
Steer Davies Gleave (2006)	15 routes in Europe (incl. 8 HSR routes)	Cross-sectional	Logit model used to predict passenger choice between rail and air	Market share, scheduled journey time, frequency, check-in time, price, access time and cost, reliability, airport links, price variability, service quality	The fitted logit model is used as part of the simulation to predict future market share. Mixed results on rail market shares based on various scenarios
Impacts of HSR service levels					
Castillo-Manzano et al. (2015)	Total air and HSR traffic from the Madrid Barajas Airport	Time series monthly data, January 1996 – December 2012	Estimate the substitution effect between HSR and air with dynamic linear regression	DV: air passenger number in Madrid Barajas airport IV: passenger number in the national HSR network, one-period lagged air operation, unemployment rate, snowfall dummy, population, difference between business days and weekends, seasonal dummies	The rate of substitution dropped when new HSR lines with low population opened. Only 13.9% of the HSR passenger demand came from the air mode. There is no evidence that HSR has a strong network effect to attract air passengers.
Clewlow et al. (2013)	90 routes in Europe	Route-level panel data,	HSR impact is captured by rail travel time, no route	DV: air passenger number IV: jet fuel price, GDP, population,	Lower HSR travel time is associated with lower air passenger traffic volume

		1995-2009	fixed effect, no time fixed effect	population density, rail travel time, airline hub, LCC dummy	
Dobruszkes et al. (2014)	161 city-pairs with HSR services in Europe	Route-level cross-sectional data, January 2012	HSR impact is captured by HSR travel time and frequency	DV: airline seats, flight frequencies IV: population, share of low-cost air service, airline hubs, HSR travel time, air-HSR integration, HSR calling of both central and peripheral rail stations, country dummies	Lower HSR travel time has strong association with fewer airline frequencies and seats. Higher HSR frequency has limited impact on airline seats. Other HSR related variables are not statistically significant.
Impacts of HSR entry					
Bilotkach et al. (2010)	887 airline routes in Europe	Route-level panel data, May 2006 – April 2007	HSR impact is captured by HSR dummy which equals to 1 if HSR service exists, no route fixed effect, no time fixed effect	DV: individual airlines' frequencies IV: air route distance, airport access distance, road quality, population, GDP per capita, tourism, frequency HHI, LCC dummy, hub airline dummy, number of destinations, HSR dummy	HSR may have a positive pressure on airline flight frequencies based on the entire sample, but no statistically significant impact was found in short-haul routes subsamples (less than 550km).
Givoni and Rietveld (2009)	549 routes worldwide	Route-level cross-sectional data, 2003	HSR impact is captured by a less-than-three-hour HSR dummy variable	DV: aircraft size IV: route density, route distance, HHI, LCC dummy, continent dummies, number of runways, hub airport, slot controlled airport	No significant impact of HSR on aircraft size was found.
Jiménez and Betancor (2012)	9 routes from Madrid (incl. 4 HSR routes)	Route-level panel data, January 1999 – December 2009	Impact of HSR entry captured by HSR entry dummy, route fixed effect captured by destination airport dummies, no time fixed effect	DV: airline monthly frequencies, total (air+rail) passengers, air market share IV: lagged number of air passengers, number of rail passengers, air distance, GDP per capita, tourism per capita, percentage of international routes, time trend	HSR entry leads to reduction of 17% in flight frequencies on average, except for the Madrid-Barcelona route, and the decline of air market shares was observed.
Zhang et al. (2014)	93 routes in China	Route-level panel data, 1st quarter of 2010 – 4th quarter of 2011	HSR impact is captured by HSR dummy equal to 1 if parallel HSR service exists for the same route, no route fixed effect, no time fixed effect	DV: airline Lerner index, yield IV: route distance, number of air passengers, population, per capita income, tourism cities, LCC dummy, HSR dummy, GDP growth, seasonality	The existence of HSR service has strong negative impact on both airline market power and average airline yield.

Table 2 Origin airports in the sample and their corresponding catchment area

Code	Airport/City	Catchment
PEK	Beijing Capital Airport	Beijing
CAN	Guangzhou Baiyun Airport	Guangzhou
ITM	Osaka Itami Airport	Osaka, Kyoto, Nara, Shiga, Wakayama, and Hyogo (up to 2005)
KIX	Osaka Kansai Airport	
HND	Tokyo Haneda Airport	Included prefectures: Tokyo, Kanagawa, Chiba, Saitama, Gunma, Tochigi, Ibaraki, Yamanashi
NRT	Tokyo Narita Airport	
GMIC	Incheon (Gimpo + Incheon)	Seoul, Gyeunggi and Incheon
SHPV	Shanghai (Hongqiao + Pudong)	Shanghai

Table 3 Number of treated airline routes included in the sample

Treated airline route distance	"Effective" HSR entry year	Country			**Total**
		China	Japan	Korea	
Less than or equal to 500km	1997	0	1	0	**1**
	2004	0	0	1	**1**
	2007	9	0	0	**9**
	2008	1	0	0	**1**
	2009	4	0	0	**4**
	2011	3	2	2	**7**
	2012	0	0	1	**1**
	Total	**17**	**3**	**4**	**24**
More than 500km but less than or equal to 800km	1997	0	2	0	**2**
	2007	5	0	0	**5**
	2008	1	0	0	**1**
	2009	3	0	0	**3**
	2010	2	0	0	**2**
	2011	2	3	0	**5**
	2012	1	0	0	**1**
	Total	**14**	**5**	**0**	**19**
More than 800km	2007	4	0	0	**4**
	2008	1	0	0	**1**
	2009	5	0	0	**5**
	2010	4	0	0	**4**
	2011	3	4	0	**7**
	2012	4	0	0	**4**
	Total	**21**	**4**	**0**	**25**
All route distances	**Grand Total**	**52**	**12**	**4**	**68**

Table 4 Summary statistics for *Y* and *F* and the difference between treated and untreated groups

	Samples	Obs	Y ('000 seats) Mean or difference	Y ('000 seats) Std. Dev. or Std. Err.	F (flights) Mean or difference	F (flights) Std. Dev. or Std. Err.
Summary statistics	Full sample	8873	194.2	501.7	1021.2	2077.8
	China	6422	130.4	298.4	815.8	1600.0
	Japan	2147	318.1	724.9	1260.1	2106.0
	Korea	304	666.9	1158.0	3672.7	5691.4
	Distance ≤ 500km	1805	225.6	558.2	1281.8	2771.3
	500km < Distance ≤ 800km	1615	183.4	272.3	989.5	1294.5
	Distance > 800km	5453	187.0	533.0	944.3	1989.8
	Treatment	1615	383.9	644.9	2141.2	2951.3
	Control	7315	163.3	480.8	827.0	1873.8
	Matched treatment	1216	301.9	496.2	1667.2	2252.7
	Matched control	1216	203.7	344.9	1107.7	1764.8
Difference (control-treatment)	Full sample	8873	-188.7	15.0	-1109.5	61.4
	Distance ≤ 500km	1805	-95.1	30.2	-486.7	149.7
	500km < Distance ≤ 800km	1615	-122.7	16.0	-898.7	74.0
	Distance > 800km	5453	-310.1	25.3	-1697.4	92.8
	Matched	2432	-98.1	17.3	-559.5	82.1
	Distance ≤ 500km (matched)	817	-15.1	13.5	17.9	73.7
	500km < Distance ≤ 800km (matched)	646	-157.6	19.0	-882.7	110.7
	Distance > 800km (matched)	969	-125.5	37.7	-791.2	169.8

Note: Korean routes are not matched due to insufficient number of routes.

Table 5 Summary statistics for other variables

	China Obs	China Mean (Std. Dev.)	Japan Obs	Japan Mean (Std. Dev.)	Korea Obs	Korea Mean (Std. Dev.)	Treatment Obs	Treatment Mean (Std. Dev.)	Control Obs	Control Mean (Std. Dev.)
rPOP	5187	19743.3	2147	27788.5	304	25840.5	1182	23361.8	6456	22043.4
('000 persons)		(7079.7)		(12485.2)		(2235.60		(7257.4)		(9913.6)
rGDP_POP	5184	5.6	1921	40.0	288	13.6	1154	11.8	6239	15.4
('000 USD)		(3.3)		(5.1)		(2.7)		(13.3)		(15.8)
LCC	6422	0.02	2147	0.04	303	0.05	1292	0.05	7580	0.02
		(0.15)		(0.20)		(0.22)		(0.22)		(0.15)
Distkm	5187	5549.4	2147	1000.5	304	2848.3	1182	5435.3	6456	3930.4
(km)		(4691.3)		(1425.7)		(1583.7)		(3219.8)		(4602.3)
droad	6422	1171.1	2147	727.0	304	261.7	1292	664.5	7581	1095.2
(km)		(622.7)		(421.7)		(78.1)		(330.1)		(632.4)
POP	6422	1482.4	1976	953.8	304	305.9	1292	826.3	7410	1407.6
('000 persons)		(814.9)		(511.0)		(86.5)		(389.8)		(818.3)

Note: The brackets show standard deviation. The population or real GDP data are not available for some of the observations.

Table 6 Initial regression results base on equation (1) without matching

DV=Y	(1) China	(2) China	(3) Japan	(4) Japan	(5) Korea (Year≥2003)
HSR	17.662		-85.451		-331.698
	(18.836)		(25.487) ***		(163.973) **
	[40.061]		[40.044] **		[177.940] ***
D1*HSR		-176.578		-70.178	
		(16.561) ***		(25.392) ***	
		[38.530] ***		[27.700] **	
D2*HSR		20.763		-117.298	
		(21.463)		(41.429) ***	
		[44.863]		[68.736] *	
D3*HSR		235.335		-51.216	
		(43.178) ***		(48.325)	
		[85.421] ***		[77.140]	
rGDP_POP	69.205	67.835			-467.531
	(6.205) ***	(5.826) ***			(139.986) ***
	[19.496] ***	[18.283] ***			[302.409] ***
rPOP	0.067	0.066	0.054	0.054	0.914
	(0.006) ***	(0.005) ***	(0.005) ***	(0.005) ***	(0.447)
	[0.018] ***	[0.017] ***	[0.012] ***	[0.012] ***	[0.859]
LCC	209.610	195.818	282.879	281.998	63.954
	(23.999) ***	(23.233) ***	(41.793) ***	(42.091) ***	(205.228)
	[41.055] ***	[39.952] ***	[114.167] ***	[115.169] **	[374.427]
Constant	-1453.825	-1441.258	-1197.314	-1190.050	-17977.060
	(205.859) ***	(196.061) ***	(132.211) ***	(132.085)	(11262.050)
	[420.068] ***	[394.635] ***	[341.540] ***	[341.525]	[21013.868]
N	5184	5184	2147	2147	143
R-squared	0.8665	0.8748	0.9762	0.9762	0.9719

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1
The robust standard errors and robust standard errors clustered at route level are shown in parentheses and brackets, respectively. Year and route dummies are not presented to save space.

Table 7 Binary regression results (China and Japan)

	China (probit)		Japan (logit)	
DV=treat	Air distance	Road distance	Air distance	Road distance
lnrGDP_POP	5.5280 ***	5.5846 ***		
	(1.2421)	(1.1327)		
lnrPOP	4.8020 ***	4.8287 ***	1.3697 ***	1.3772 ***
	(0.8931)	(0.8254)	(0.3686)	(0.3613)
lnDistkm	-1.0395 ***		0.6053 *	
	(0.2185)		(0.3205)	
lndroad		-1.1404 ***		0.6732 *
		(0.1747)		(0.3559)
CAN	1.6941 ***	1.7019 ***		
	(0.5347)	(0.4674)		
PEK	1.2713 ***	1.2595 ***		
	(0.3618)	(0.3100)		
Constant	-51.2925 ***	-50.7168 ***	-15.4367 ***	-16.1476 ***
	(10.3112)	(9.7509)	(2.6381)	(2.7228)
N	339	339	104	104
Wald chi2	56.36	62.39	23.59	25.17
Pseudo R2	0.3121	0.3287	0.1631	0.1656
Ho: Normality (Prob > chi2)	0.2710	0.5603		

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1
The robust standard errors are shown in the parentheses.

Table 8 Regression results base on equation (1) after matching (China, using air distance in PSM)

China	(1) DV=Y (NR)	(2) DV=Y (CR)	(3) DV=Y (CN)	(4) DV=Y (RA)	(5) DV=Y (NR)	(6) DV=Y (CR)	(7) DV=Y (CN)	(8) DV=Y (RA)	(9) DV=F (CN)
HSR	-9.426	-6.691	-35.702	-13.471					
	(17.102)	(17.411)	(19.641) *	(18.338)					
	[41.533]	[41.678]	[46.057]	[38.840]					
D1*HSR					-148.345	-146.165	-169.883	-160.121	-991.537
					(16.176) ***	(17.087) ***	(22.050) ***	(20.534) ***	(129.675) ***
					[36.299] ***	[37.215] ***	[46.994] ***	[37.555] ***	[260.901] ***
D2*HSR					-7.966	-10.766	-38.445	-23.034	-137.899
					(22.688)	(22.685)	(25.130)	(22.899)	(164.209)
					[50.274]	[50.134]	[53.291]	[45.792]	[341.710]
D3*HSR					122.506	118.494	56.891	121.371	522.217
					(30.149) ***	(30.207) ***	(31.778) *	(33.567) ***	(186.976) ***
					[66.486] *	[66.332] *	[69.031]	[68.235] *	[383.625]
rGDP_POP	91.696	92.501	115.369	114.916	85.534	86.379	111.082	111.668	603.964
	(11.649) ***	(11.675) ***	(14.762) ***	(9.759) ***	(11.209) ***	(11.230) ***	(14.770) ***	(9.101) ***	(85.383) ***
	[37.169] **	[37.240] **	[46.656] **	[27.325] ***	[36.534] **	[36.609] **	[46.710] **	[25.987] ***	[264.681] **
rPOP	0.072	0.074	0.077	0.090	0.069	0.071	0.077	0.088	0.334
	(0.009) ***	(0.009) ***	(0.011) ***	(0.007) ***	(0.009) ***	(0.009) ***	(0.011) ***	(0.007) ***	(0.061) ***
	[0.030] **	[0.030] **	[0.036] **	[0.020] ***	[0.030] **	[0.030] **	[0.036] **	[0.020] ***	[0.192] *
LCC	251.974	244.848	278.801	240.535	240.003	235.455	267.409	234.681	1761.068
	(29.243) ***	(29.126) ***	(37.201) ***	(28.426) ***	(29.320)	(29.201)	(37.729)	(28.307)	(226.051)
	[64.881] ***	[64.657] ***	[65.165] ***	[61.676] ***	[63.844]	[63.445]	[64.603]	[57.489]	[391.208]
Constant	-1559.801	-1590.759	-1833.073	-1927.546	-1498.472	-1527.129	-1822.167	-1898.958	-8212.940
	(162.203) ***	(162.757) ***	(205.475) ***	(125.354) ***	(164.884)	(165.427)	(205.297)	(123.240)	(1091.034)
	[513.675] ***	[514.442] ***	[596.449] ***	[339.137] ***	[523.172]	[523.909]	[595.653]	[333.955]	[3170.168]
N	1580	1550	1285	4811	1580	1550	1285	4811	1285
R-squared	0.8811	0.8824	0.8791	0.8811	0.8901	0.8907	0.8836	0.8901	0.8664

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1

The robust standard errors and robust standard errors clustered at route level are shown in parentheses and brackets, respectively.

Year and route dummies are not presented to save space.

NR=nearest neighbor with replacement; CR=caliper with replacement; CN=caliper without replacement; RA=radius matching

Table 9 Regression results base on equation (1) after matching (Japan, using air distance in PSM)

Japan	(1) DV=Y (NR)	(2) DV=Y (CR)	(3) DV=Y (CN)	(4) DV=Y (RA)	(5) DV=Y (NR)	(6) DV=Y (CR)	(7) DV=Y (CN)	(8) DV=Y (RA)	(9) DV=F (RA)
HSR	-70.941	-86.394	-70.681	-102.520					
	(24.046) ***	(27.855) ***	(27.135) ***	(39.832) **					
	[37.556] *	[46.305] *	[40.495] *	[59.128]					
D1*HSR					-66.194	-77.348	-45.323	-76.358	-204.081
					(26.247) **	(24.960) ***	(34.065)	(29.506) **	(185.286)
					[35.610] *	[34.403] **	[46.042]	[36.850] *	[253.552]
D2*HSR					-117.088	-131.389	-104.248	-305.836	-1278.454
					(37.944) ***	(47.928) ***	(39.101) ***	(46.587) ***	(381.632) ***
					[62.172] *	[80.602]	[62.000]	[62.533] ***	[611.489] *
D3*HSR					2.139	26.406	-20.187	32.836	-166.274
					(36.038)	(20.247)	(47.490)	(30.286)	(200.940)
					[54.492]	[33.204]	[72.841]	[37.700]	[303.660]
rPOP	0.034	0.013	0.047	0.019	0.031	0.015	0.044	0.030	0.145
	(0.009) ***	(0.009)	(0.009) ***	(0.014)	(0.009) ***	(0.009) *	(0.009) ***	(0.013) **	(0.067) **
	[0.021]	[0.017]	[0.020] **	[0.018]	[0.020]	[0.017]	[0.020] **	[0.015] *	[0.127]
LCC	197.741	78.206	203.902	102.732	190.290	87.937	196.989	99.945	948.923
	(28.282) ***	(36.934) **	(30.307) ***	(52.326) *	(27.827) ***	(35.854) **	(29.796) ***	(45.063) **	(581.675)
	[32.278] ***	[29.661] **	[30.899] ***	[36.855]	[29.568] ***	[31.408] **	[30.477] ***	[36.498] **	[562.723]
Constant	-596.302	-106.145	-974.611	-146.904 **	-537.316	-157.447	-896.627	-399.108	-2297.946
	(220.696) ***	(207.744)	(252.438) ***	(298.327)	(216.105)	(201.349)	(250.380)	(280.555)	(1481.092)
	[525.178]	[379.458]	[561.317] *	[384.202]	[496.351]	[369.467]	[543.744]	[312.251]	[2730.754]
N	456	342	380	266	456	342	380	266	266
R-squared	0.9626	0.9376	0.9722	0.9140	0.9633	0.9397	0.9726	0.9229	0.8934

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1

The robust standard errors and robust standard errors clustered at route level are shown in parentheses and brackets, respectively.

Year and route dummies are not presented to save space.

NR=nearest neighbor with replacement; CR=caliper with replacement; CN=caliper without replacement; RA=radius matching

Figures

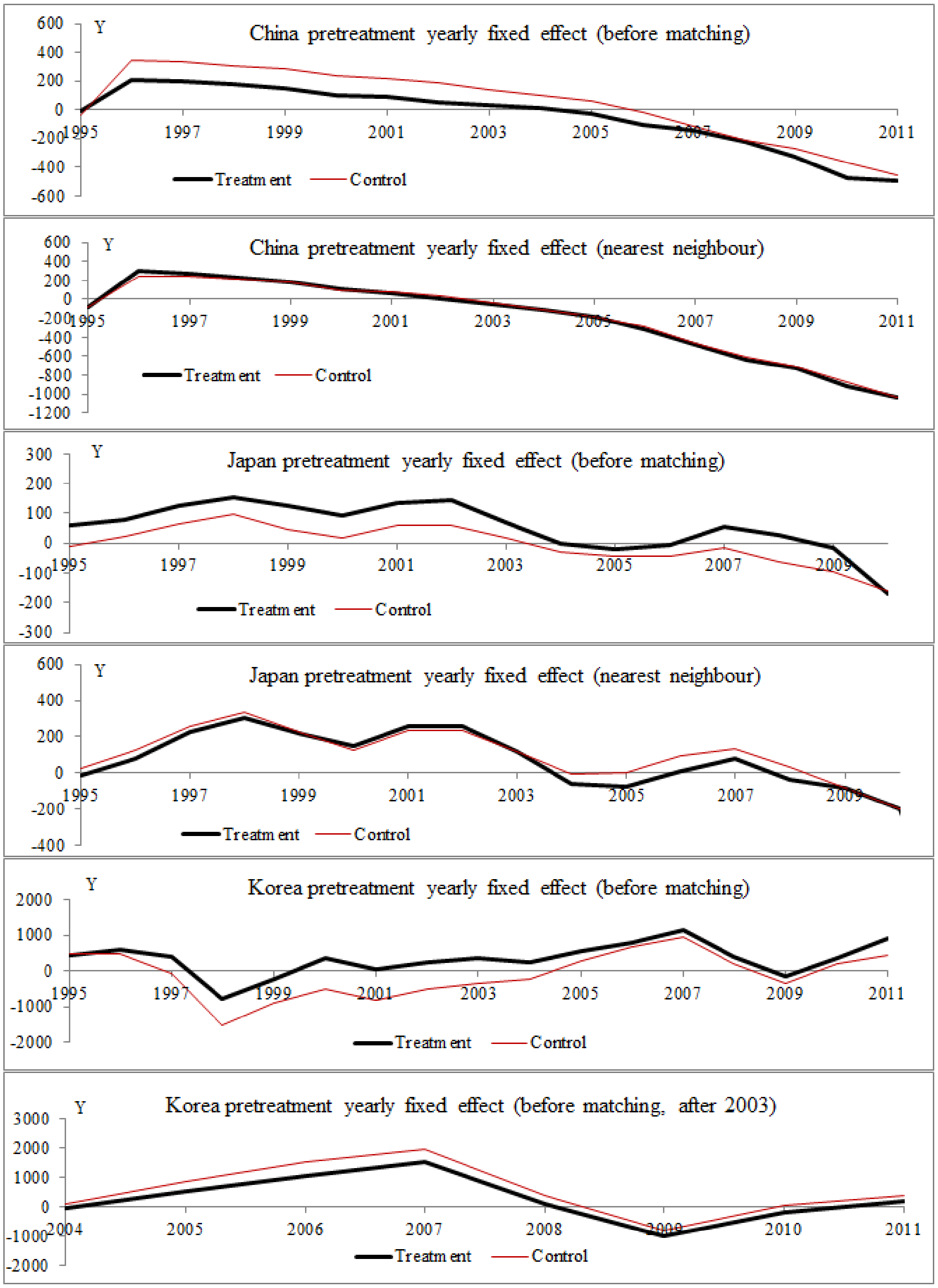


Figure 1 Yearly fixed effect prior to HSR entry on Y

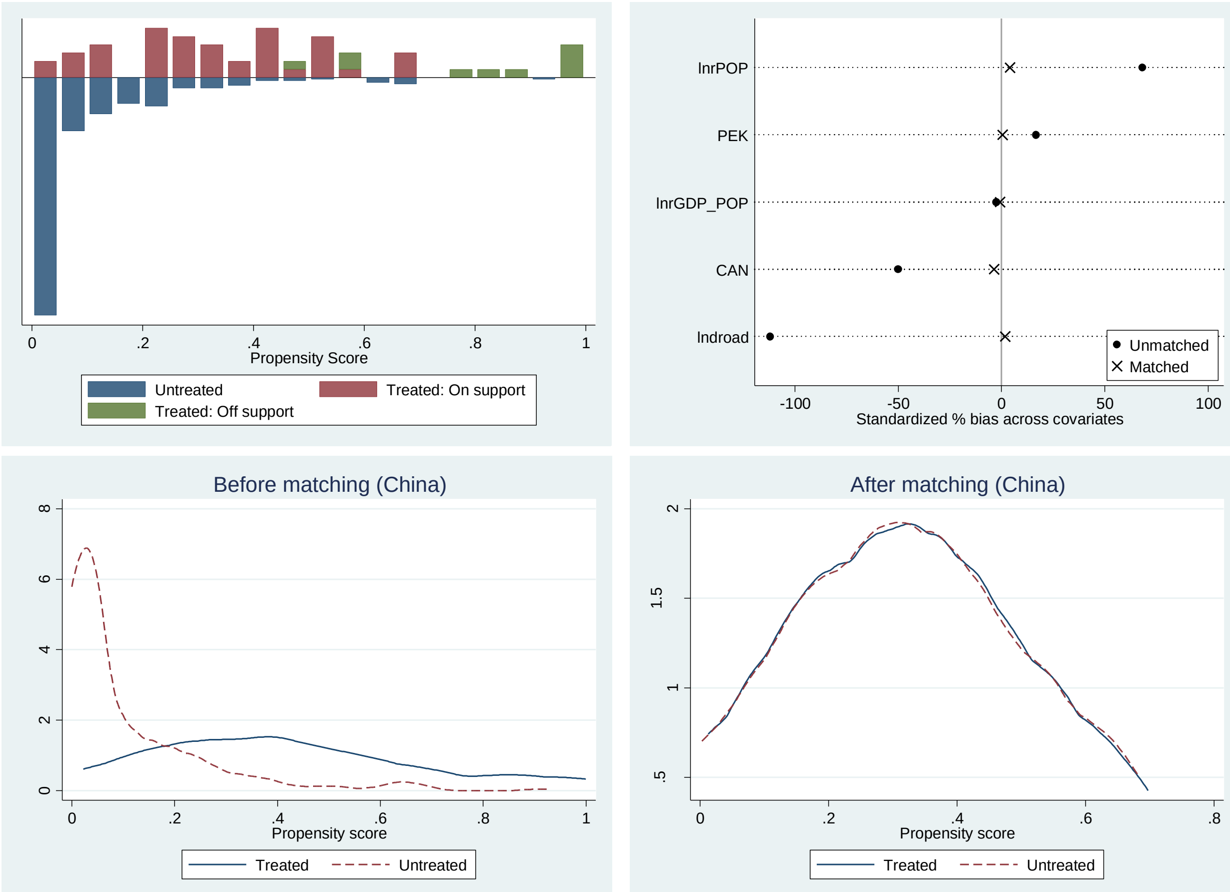


Figure 2 Matching outcomes (China, radius, road distance)

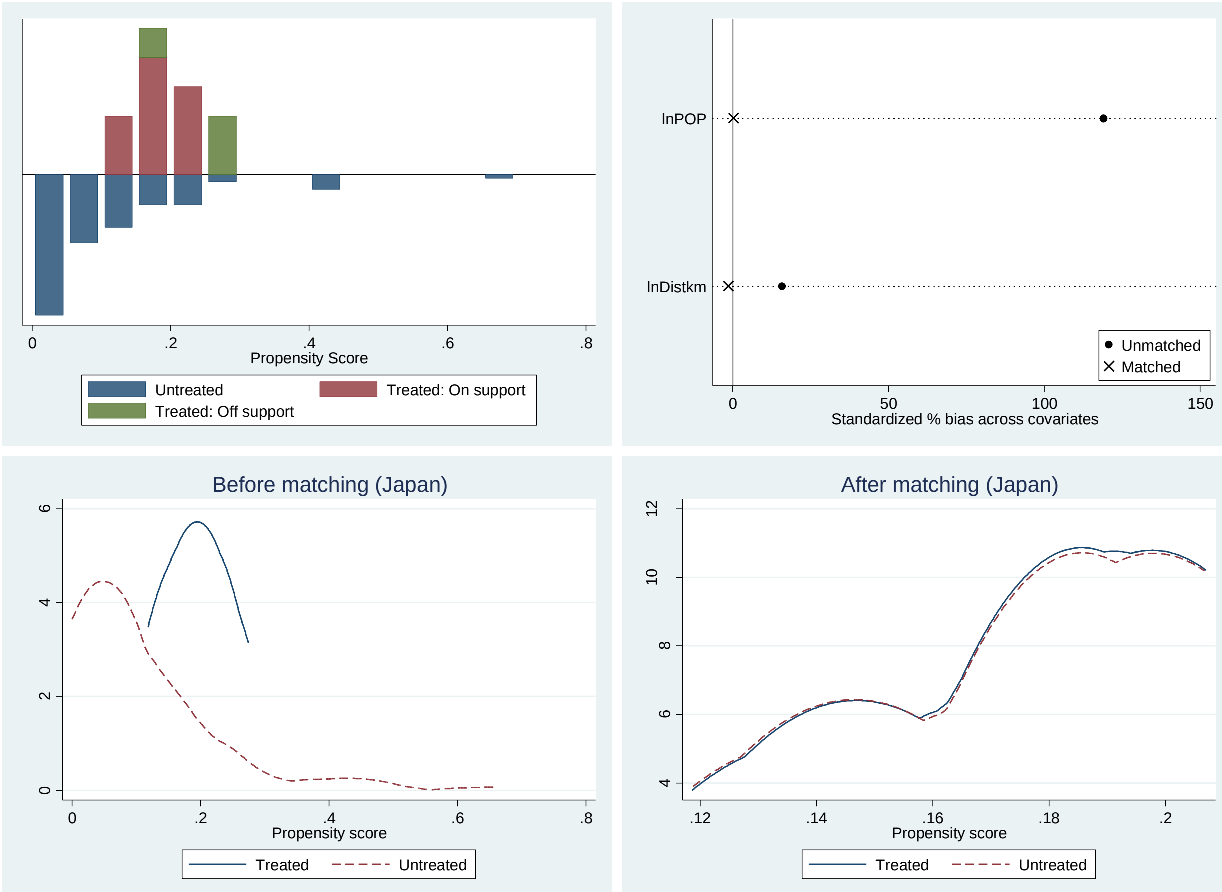


Figure 3 Matching outcomes (Japan, caliper with replacement, air distance)

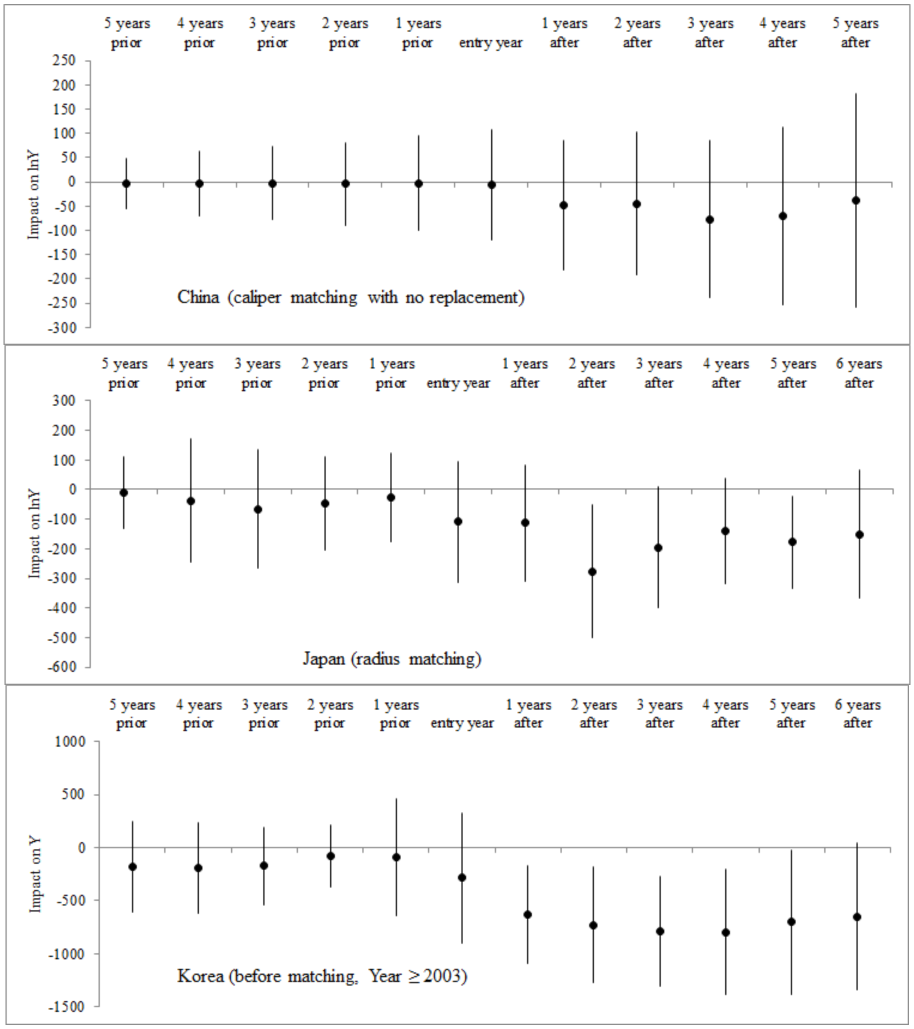


Figure 4 Impact of HSR entry over time (pooling all distance groups)
(The figure shows the coefficients and 95% confidence intervals.)

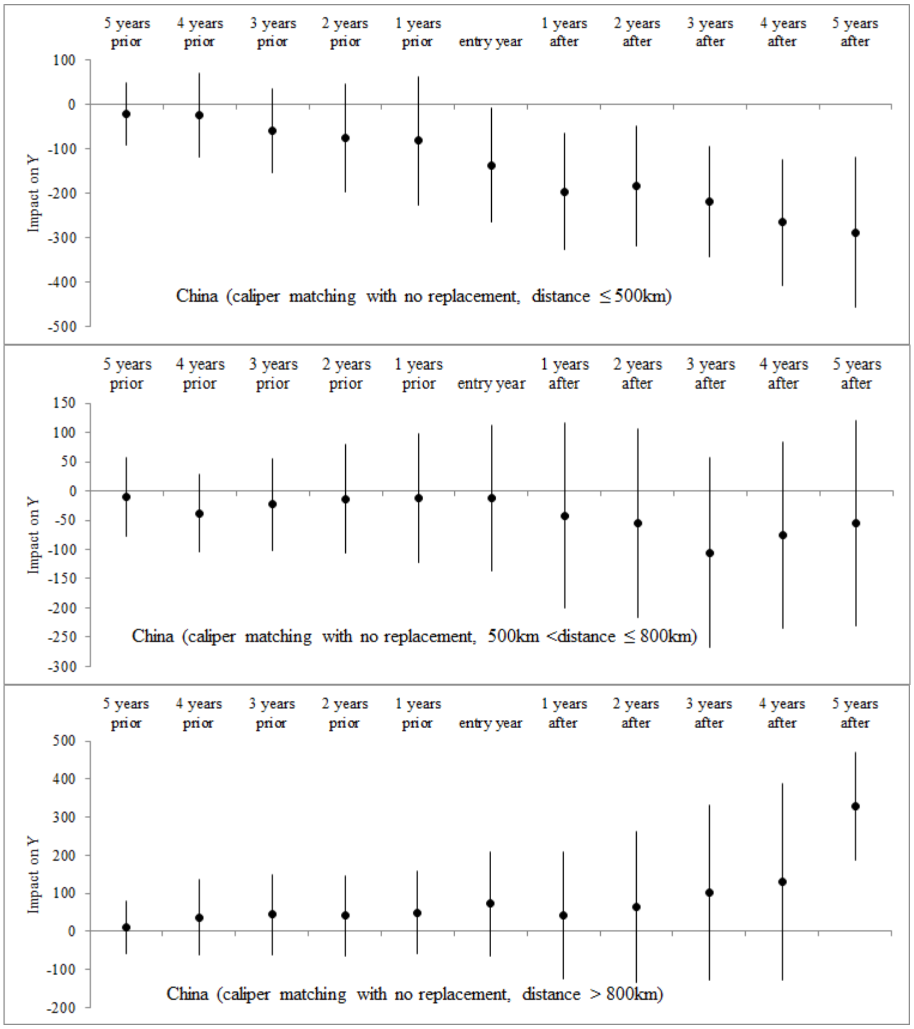


Figure 5 Impact of HSR entry over time by distance groups (China)
(The figure shows the coefficients and 95% confidence intervals.)

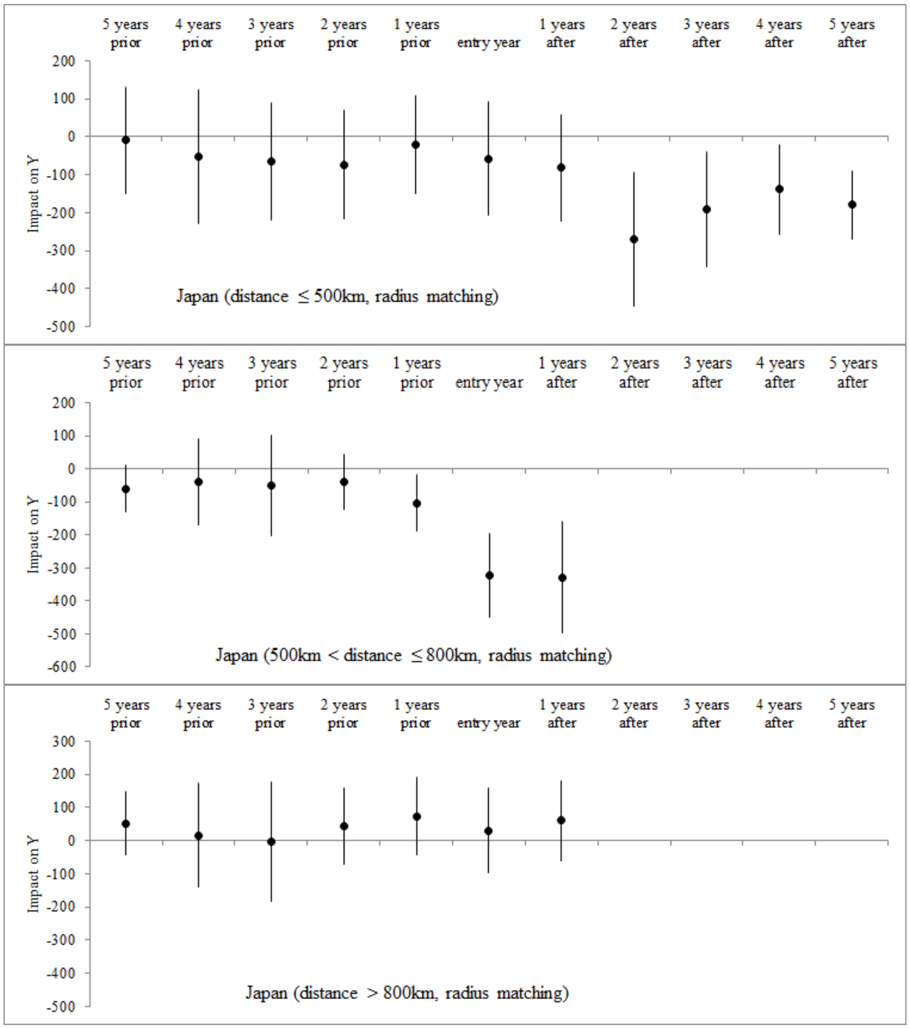


Figure 6 Impact of HSR entry over time by distance groups (Japan)
(The figure shows the coefficients and 95% confidence intervals.)

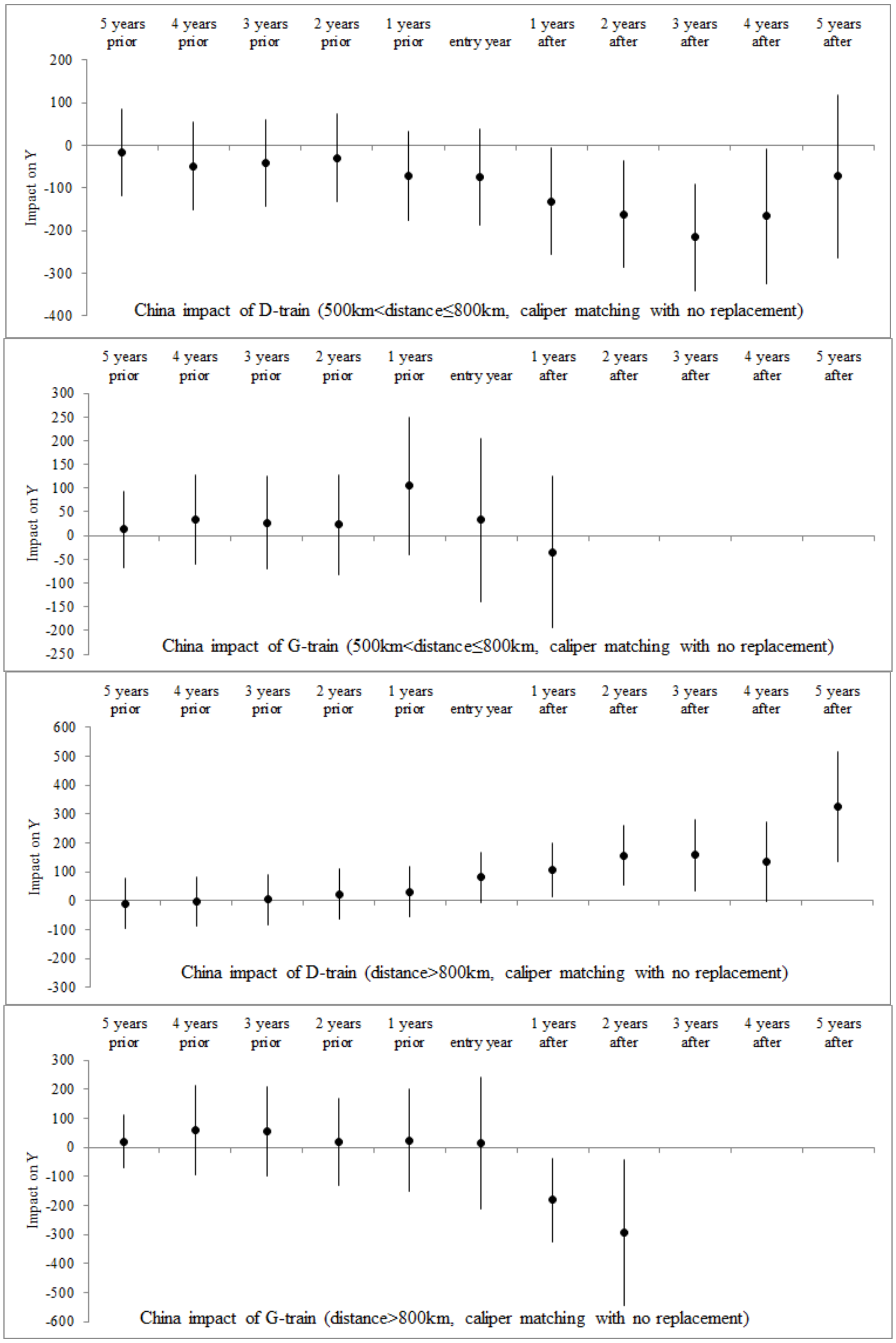
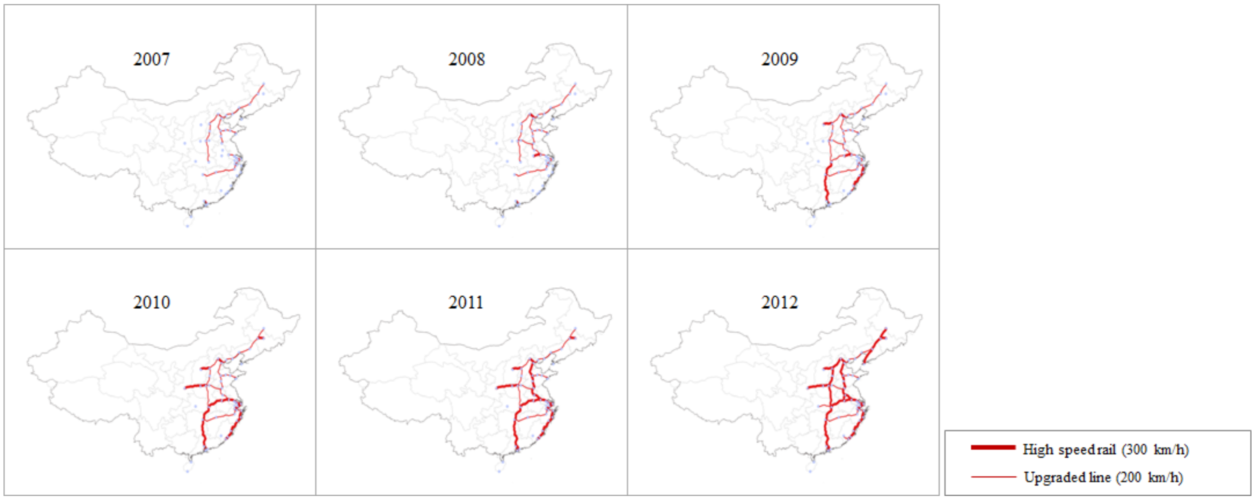


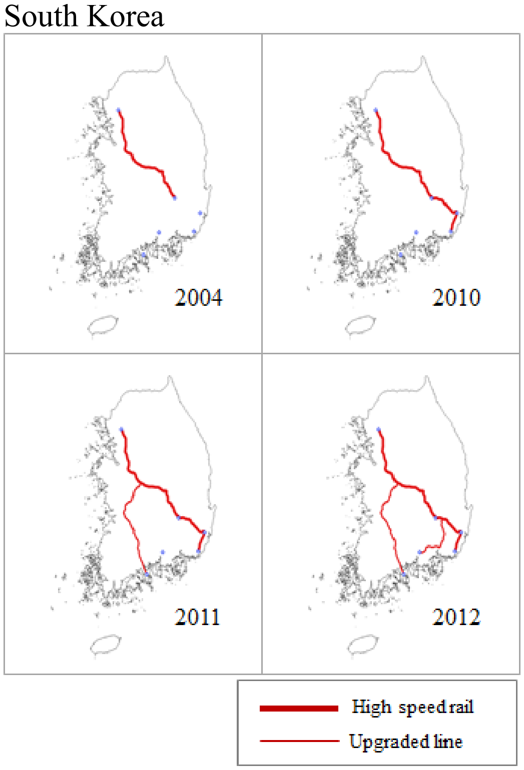
Figure 7 Impacts of China's D-train and G-train services overtime (medium-haul and long-haul) (The figure shows the coefficients and 95% confidence intervals.)

Appendix A. HSR expansion by countries based on actual entry years
(only HSR services parallel to sampled air routes are included)

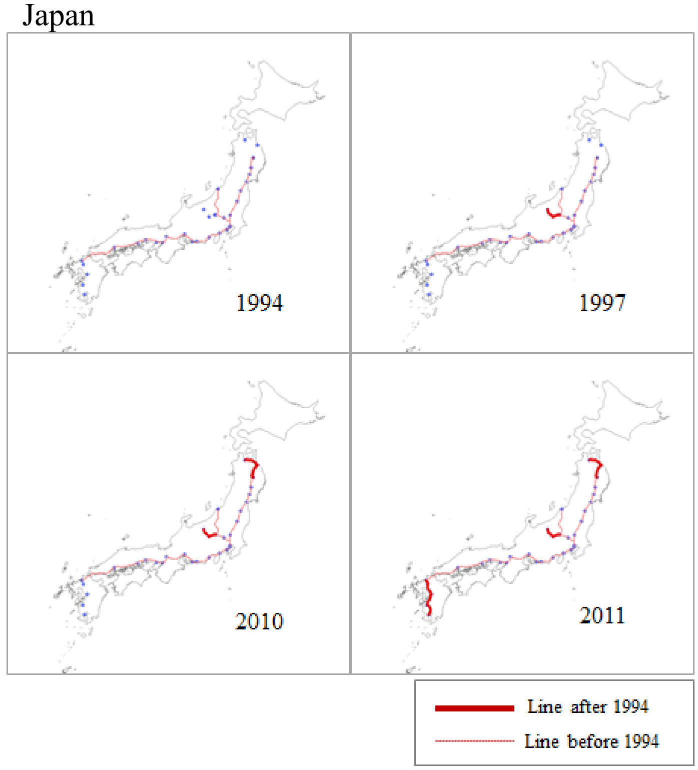
China



South Korea



Japan



Appendix B. Tests for the difference of pretreatment yearly fixed effects

Ho: Difference (treatment – control) = 0

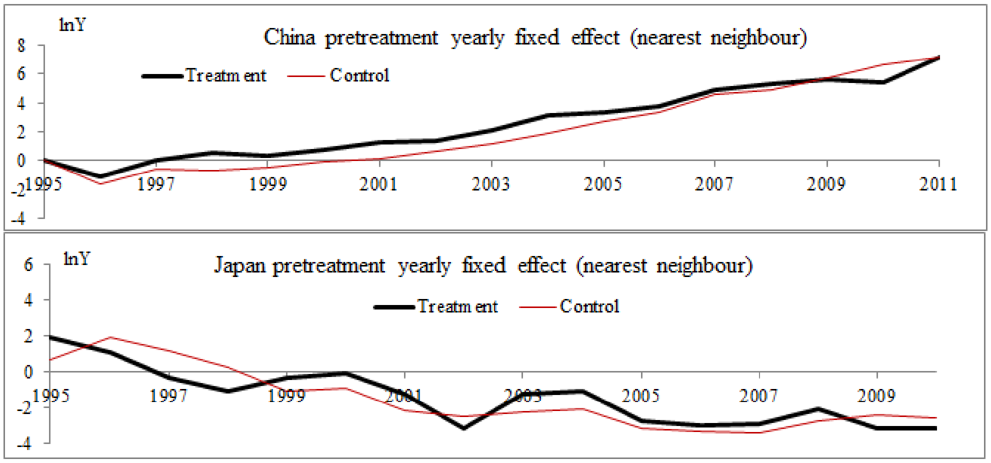
	China				Japan			
DV=Y	Before matching		After matching (nearest neighbor)		Before matching		After matching (nearest neighbor)	
	Diff.	Std. Err.	Diff.	Std. Err.	Diff.	Std. Err.	Diff.	Std. Err.
Year1995	26.9147	44.8367	52.5730	88.5086	69.9485	26.6211 **	-41.9731	99.4290
Year1996	-132.8704	340.9816	32.2099	85.5528	54.7230	26.0636 **	-30.8570	84.2454
Year1997	-128.3707	347.3164	20.4313	83.4262	59.5734	20.3341 ***	-32.2196	80.5773
Year1998	-122.5906	350.3925	4.4668	81.2458	58.0307	26.8565 **	-48.5891	94.5704
Year1999	-132.5013	351.0756	15.5740	78.8220	77.3061	35.0387 **	-28.9858	131.0022
Year2000	-136.2927	352.3848	-13.3575	78.2808	73.5205	30.5098 **	-25.1036	119.5831
Year2001	-130.3384	357.2039	-24.6854	77.4667	75.1034	31.3878 **	-7.6139	96.5128
Year2002	-129.9661	361.7264	-9.7925	77.9464	81.6039	33.0150 **	22.5789	84.0638
Year2003	-108.5295	358.6174	-4.6374	74.4705	49.3211	33.1747	20.6751	81.6989
Year2004	-92.6336	363.8431	-4.3423	73.2905	28.2121	45.9646	27.1001	56.2820
Year2005	-85.7912	364.1741	-26.1787	68.7653	26.4595	46.4985	-4.7880	75.2303
Year2006	-85.6168	371.6825	-7.2074	63.5464	35.1263	41.9432	-55.0278	109.1286
Year2007	-35.5588	383.8564	-23.2368	59.3132	71.2335	41.8685 *	-80.0077	149.1936
Year2008	-11.5438	395.3139	-21.3360	56.2087	88.0069	40.1203 **	-87.1415	167.5240
Year2009	-54.4821	362.7584	-33.6511	47.9629	78.3656	39.5567 *	-52.6509	130.2723
Year2010	-104.9703	361.1602	52.5730	88.5086	-11.6595	52.5105	-72.4648	121.8367
Year2011	-38.4983	367.2456	32.2099	85.5528				
Year2012								

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1
Certain yearly dummies are omitted as there are no treated observations as HSR has started operations on all treated routes during these years.

	Korea (before matching)			
DV=Y	All year		Year ≥ 2003	
	Diff.	Std. Err.	Diff.	Std. Err.
Year1995	-36.7558	96.9436		
Year1996	96.0129	80.6788		
Year1997	493.6742	291.2218		
Year1998	725.5407	397.9757 *		
Year1999	687.6638	370.1179 *		
Year2000	865.0525	464.8908 *		
Year2001	871.1136	475.8601 *		
Year2002	781.3332	382.0150 *		
Year2003	697.0194	307.8658 **		
Year2004	452.8792	288.6824	-178.1184	234.9040
Year2005	304.9958	291.1070	-342.6262	403.0089
Year2006	150.1634	343.4772	-487.4733	487.1755
Year2007	171.0980	341.4904	-429.6287	509.6715
Year2008	187.8965	299.0047	-303.1139	495.7122
Year2009	226.6097	229.7088	-208.8113	393.1104
Year2010	151.4804	278.2848	-273.6259	512.9287
Year2011	476.7109	179.6284	-152.7405	370.7262
Year2012				

Note: ***statistically significant at 0.01; **statistically significant at 0.05; *statistically significant at 0.1
Certain yearly dummies are omitted as there are no treated observations as HSR has started operations on all treated routes during these years.

Appendix C. Yearly fixed effect prior to HSR entry with double-log form (after matching)



Appendix D. Test for heteroscedasticity with modified Wald test

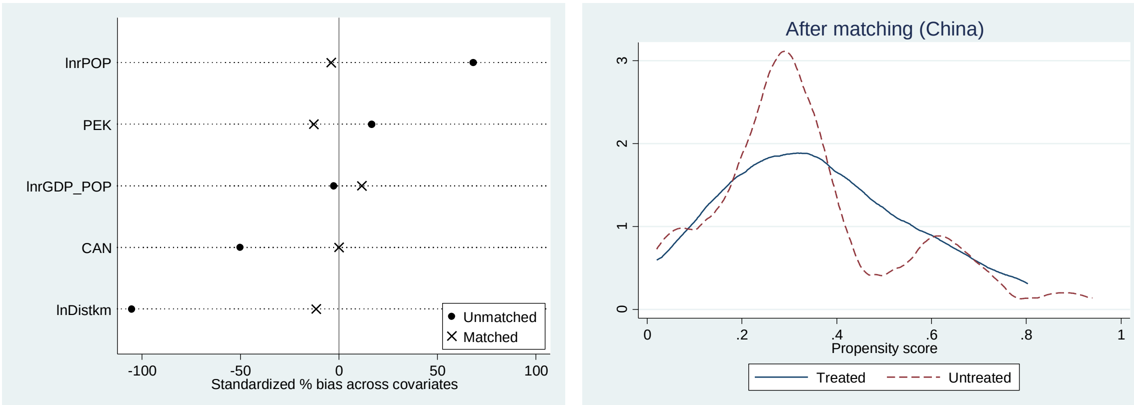
China		Double logs	Semi-logs	No logs
Before matching	chi2 (338) Prob>chi2	2.4e+05 0.0000	6.6e+05 0.0000	1.1e+06 0.0000
After matching (nearest neighbor)	chi2 (83) Prob>chi2	82947.25 0.0000	93381.28 0.0000	32603.14 0.0000

Japan		Double logs	Semi-logs	No logs
Before matching	chi2 (113) Prob>chi2	5.1e+05 0.0000	2.7e+05 0.0000	2.5e+05 0.0000
After matching (nearest neighbor)	chi2 (20) Prob>chi2	58515.60 0.0000	32640.71 0.0000	1902.00 0.0000

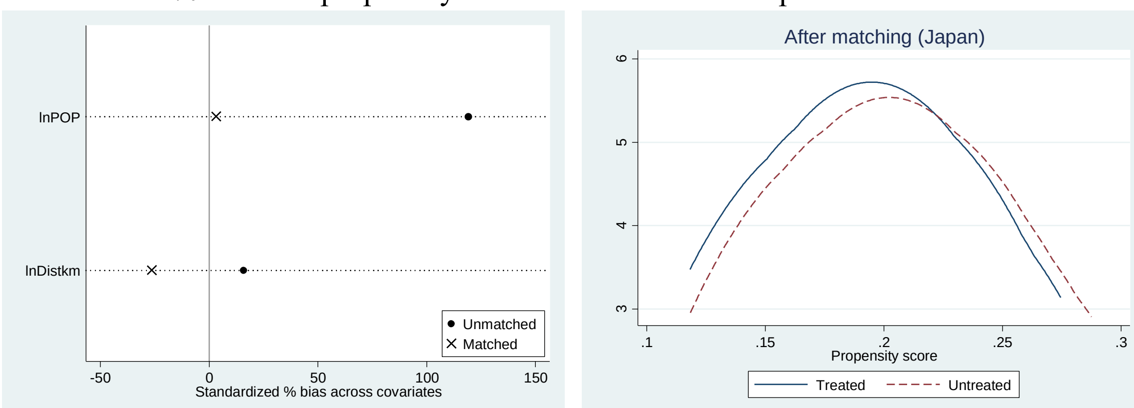
Korea (Year $\geq$ 2003)		Double logs	Semi-logs	No logs
Before matching	chi2 (16) Prob>chi2	1.3e+05 0.0000	1.2e+05 0.0000	3417.48 0.0000

Appendix E. Nearest neighbor without replacement matching outcomes (air distance)

Standardized percentage bias and propensity score distributions for China



Standardized % bias and propensity score distributions for Japan



Appendix F. Summary statistics on estimated HSR operating speed (km/h) for sampled treated routes

	Obs	Mean	Std. Dev.	Min	Max
China G-train	28	212.55	37.13	154.21	266.79
China G-train (distance ≤ 500km)	12	177.94	26.59	154.21	233.14
China G-train(500km < distance ≤ 800km)	9	216.52	33.42	164.50	266.79
China G-train (distance > 800km)	15	231.43	22.87	173.26	255.24
China D-train	21	143.77	17.37	105.30	175.32
China D-train (distance ≤ 500km)	17	134.60	20.64	97.56	175.32
China D-train(500km < distance ≤ 800km)	12	140.71	15.16	117.04	162.60
China D-train (distance > 800km)	14	142.19	10.06	126.11	165.63
Japan	12	223.03	25.65	170.82	246.90
Japan (distance ≤ 500km)	3	221.54	43.92	170.82	246.90
Japan (500km < distance ≤ 800km)	5	217.61	27.22	189.91	244.29
Japan (distance > 800km)	4	230.92	2.26	229.02	233.46
Korea	4	134.53	13.61	114.68	145.64

Appendix G. Distance distributions of treated air routes (China, Japan and Korea)