Web Appendix for “Planes, Trains, and Co-Opetition: Evidence from China”

I Additional Descriptive Statistics

Table I.1: Average Route Characteristics by Year

	2007	2008	2009	2010	2011	2012	2013	2014	2015
Airline present (1=Y, 0=N)	0.25	0.25	0.26	0.28	0.29	0.31	0.34	0.35	0.37
Number of airlines	0.44	0.47	0.50	0.53	0.56	0.60	0.64	0.67	0.70
Number of flights	1.06	1.11	1.26	1.39	1.52	1.63	1.78	1.91	2.02
Number of airline connections	65.01	69.07	75.04	80.43	87.50	93.29	101.17	106.33	113.86
HSR present (1=Y, 0=N)	0.07	0.07	0.08	0.09	0.10	0.11	0.12	0.19	0.21
Fast train present (1=Y, 0=N)	0.07	0.07	0.08	0.09	0.10	0.10	0.11	0.19	0.21
Bullet train present (1=Y, 0=N)	0.00	0.00	0.00	0.00	0.02	0.02	0.07	0.07	0.10
Number of HSR connections	0.26	0.31	0.40	0.65	0.86	0.89	1.11	1.23	1.54
City-pair average population (millions)	5.13	5.19	5.22	5.30	5.36	5.40	5.43	5.49	5.55
City-pair average GDP (billion U.S. dol.)	30.00	35.52	40.14	47.50	56.52	63.19	69.59	75.72	81.84

Note: There are 2,278 city-pairs in the data each year. For the definition of the variables please refer to Table 1's note.

II Robustness Checks for Difference-in-Differences Analysis

Here we check the robustness of the results from Table 3 by carrying out three additional analyses. Specifically, in columns 2 through 4 of Table II.1, we report the estimation results at the airline-route-month level (as opposed to at the airline-route year, as in the main specification) and restrict the set of treated routes such that we only include data for treated routes for one month before the entry of the HSR and one month after the introduction of the HSR. In columns 5 through 7, we show estimation results when the dependent variable is the log of the number of flights (as opposed to the number of airlines, as in the main analysis). Lastly, in columns 8 through 10, we check the robustness of our results with respect to our data sample definition and re-run our analysis using all flights that operate longer than three months (as opposed to including only those that operate longer than one year, as in the main specification).

For each analysis, we report the results for the same specifications as in Table 3 for ease of comparison. The definitions for the variables used in the tables can be found in Table 3's note. Robust and clustered (at the route level) standard errors are reported in parentheses. (***), (**) and (*) denote statistical significance at the 1%, 5% and 10% level, respectively.

Table II.1: Determinants of Airline Entry

	Month-level			Log of number of flights			Flights operating longer than 3 months		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
HSR (Yes/No)	−0.081**	−0.087**	−0.070*	−0.022	−0.133***	−0.114**	−0.204***	−0.226***	−0.181**
	(0.034)	(0.044)	(0.042)	(0.027)	(0.039)	(0.048)	(0.076)	(0.072)	(0.089)
No. of HSR connections	0.026***	0.046***	0.047***	0.010***	0.033***	0.046***	0.039***	0.023***	0.039***
	(0.006)	(0.013)	(0.013)	(0.004)	(0.006)	(0.008)	(0.008)	(0.007)	(0.009)
HSR × No. of HSR connections	−0.023*	−0.034**	−0.027**	−0.019***	−0.033***	−0.040***	−0.077***	−0.027***	−0.028***
	(0.013)	(0.016)	(0.013)	(0.006)	(0.009)	(0.009)	(0.013)	(0.009)	(0.010)
HSR × Medium Distance	0.215***	0.285***	0.314***	0.077**	0.295***	0.316***	0.522***	0.516***	0.553***
	(0.044)	(0.056)	(0.049)	(0.038)	(0.042)	(0.047)	(0.089)	(0.082)	(0.094)
HSR × Long Distance	0.284***	0.278***	0.243***	0.222***	0.413***	0.449***	0.687***	0.585***	0.603***
	(0.056)	(0.075)	(0.069)	(0.048)	(0.051)	(0.057)	(0.106)	(0.098)	(0.120)
No. of airline connections	0.010***	0.012***	0.012***	0.002***	0.012***	0.012***	0.009***	0.011***	0.011***
	(0.001)	(0.002)	(0.002)	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Average GDP	0.014***	0.003	0.003	0.009***	0.103	0.264	0.008	−0.001	−0.006
	(0.005)	(0.011)	(0.011)	(0.003)	(0.114)	(0.167)	(0.006)	(0.011)	(0.013)
Year fixed effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Route fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	No	No	No	No	No	No	No	No	No
Year-group fixed effects	No	No	Yes	No	No	Yes	No	No	Yes
Observations	237,345	61,653	61,653	20,502	7,884	7,884	20,502	7,884	7,884
R^2	0.887	0.131	0.143	0.601	0.945	0.963	0.893	0.892	0.929

III Robustness Checks for Structural Model Results

Here we check the robustness of the results from Table 4 by carrying out two additional analyses. Specifically, in columns 1 and 2 of Table IV.1, we check the robustness of our results with respect to our data sample definition and re-run our analysis using all flights that operate longer than three months (as opposed to including only those that operate longer than one year, as in the main specification). In columns 3 and 4 we report the estimation results for the case in which we relax the symmetry assumption (as discussed in Appendix B.3) allowing each airline to have both different baseline flow payoffs and different entry costs (note that, in this last case, compared to the main specification, we do not include the variable "exempt" because otherwise the state space becomes unmanageable). The definitions for the variables used in the tables can be found in Table 4's note. Standard errors are calculated via bootstrapping. (***), (**) and (*) denote statistical significance at the 1%, 5% and 10% level, respectively.

Table III.1: Structural Model Estimation Results

		(1) Coef.	(2) s.e.	(3) Coef.	(4) s.e.
Flow Payoffs' Parameters					
	No. of own routes connected	0.58***	0.04	0.11**	0.04
Strategic Effects	No. of competitors	−0.82***	0.18	−0.88***	0.33
	No. of competitors' routes connected	0.09***	0.02	0.11**	0.04
	Bullet Train (Y/N)	−0.74***	0.20	−0.90***	0.31
	Bullet Train × Medium distance	0.51**	0.21	0.56*	0.30
	Bullet Train × Long distance	0.73***	0.20	0.95***	0.34
	No. of Bullet Train line connections	0.17***	0.05	0.19**	0.08
Impact of HSR	Bullet Train × No. of Bullet train line connections	−0.57***	0.11	−0.37**	0.15
	Fast Train (Y/N)	−1.31***	0.24	−1.36***	0.29
	Fast Train × Medium distance	1.27***	0.22	1.26***	0.26
	Fast Train × Long distance	2.06***	0.27	2.01***	0.43
	No. of Fast Train line connections	0.26***	0.05	0.23***	0.06
	Fast Train × No. of Fast Train line connections	−0.54***	0.08	−0.34***	0.11
	Average GDP	0.08	0.06	0.16**	0.07
	Medium distance	−0.56***	0.21	−0.55**	0.26
Market Characteristics	Long distance	−1.18***	0.27	−1.14***	0.34
	Unobserved Type	1.84***	0.27	1.81***	0.31
	Constant	−1.10***	0.28	−1.27***	0.48
	CZ			−0.05	0.04
	MU			0.002	0.06
	HU			−0.08	0.06
Entry Costs' Parameters					
	Constant	−2.48***	0.18	−3.10***	0.71
	CZ			−0.33**	0.16
	MU			0.14	0.11
	HU			0.53***	0.19
	Unobserved Type	0.75**	0.35	0.68***	0.22
	Regulated	−1.44***	0.25	−1.67***	0.45
	Exempt	0.20	0.28		
	No. of own routes connected	0.07*	0.04	0.41	0.41

IV Structural Model's Goodness of Fit

To assess the goodness of fit of the structural model we compare the predicted with the observed airline network configurations using two different metrics. The first metric tries to capture the ability of the model to predict airline route presence (i.e., whether each route is served by at least one airline), while the second metric assesses the quality of the model with respect to airline route density (i.e., how many airlines operate in each route). More specifically, for the first metric we calculate the correlation between the vector (of length equal to the number of routes) of the expected number of airlines (capped at one if the expected number is greater than one) predicted by the model and the observed vector of indicator variables which take the value one if there is at least one airline present in a given route. We repeat this for each year-end and report the correlations between the predicted and the observed route-level values. In what concerns the second metric, we compare the vectors (with length equal to the number of routes) of the observed and the expected number of airlines serving each of the different routes. The expected numbers used in both metrics are calculated based on predictions about route presence decisions obtained from $1,000$ simulations of each airline's policy functions using the estimated parameters from the structural estimation. Table III.1 present the results. The model does a reasonably good job at predicting both the route presence and the number of airlines present with average correlations (across all years in the sample) of 83 and 92 percent for the first and second metrics, respectively.

Table IV.1: Goodness of Fit

	Correlations between Model Predictions and Sample Observations	
Year	(1)	(2)
2007	0.923	0.953
2008	0.865	0.937
2009	0.825	0.919
2010	0.811	0.918
2011	0.821	0.916
2012	0.812	0.906
2013	0.806	0.907
2014	0.804	0.906
2015	0.815	0.905

Note: Columns (1) and (2) report correlations between the predicted and observed values of two metrics, Route Presence and Route Density, respectively.

V Algorithm for Computing Counterfactual Equilibria

Here we describe the algorithm employed to compute the counterfactual equilibria in Section 7. To compute the equilibrium outcomes in the counterfactual scenarios, we proceed in a manner similar to Aguirregabiria and Ho (2012) and use a forward simulation method to update the beliefs regarding the transition of the sufficient statistics which capture information on the connecting routes. The specific steps for solving for the equilibrium are described as follows:[1]

1. Start with some initial belief of the transition probabilities for the sufficient statistics, denoted by σ'_0.

2. Given σ'_0, obtain the policy functions for each route (unobserved) type.

3. Use the recovered policy functions to simulate the evolution of the entire route network over the entire time period.

4. With the simulation results from the previous step, update airlines' beliefs of the transition probabilities of the sufficient statistics, denoted by σ'_1.

5. If $|\sigma'_1 - \sigma'_0 > \epsilon|$, replace σ'_0 with σ'_1 and obtain the policy functions for each route's (unobserved) type and repeat from step 3. Here, ϵ is set to $1e^{-6}$.

Once the equilibrium in the counterfactual scenario is found, we use the airlines' policy functions to simulate $1,000$ times the evolution of airlines' route networks over the period that is covered by the data (i.e., 2007–2015) and then average across the equilibrium outcome variables of interest (such as entry probabilities and flow profits). We then compare these averages with those simulated for the baseline scenario, in which the HSR is present as observed and the parameters are the ones estimated in the structural model. Note that, in each simulation, we keep the initial airline network configurations the same as observed. Further, to ensure a fair comparison, we also solve for the equilibrium in the baseline scenario using the steps listed above.

[1] There is no guarantee that the equilibrium is unique. However, as discussed in ABBE, the continuous time framework helps to eliminate simultaneity as a likely source of multiplicity in the equilibrium. In addition, we solve for the equilibrium using different starting points for the value function as well as for the airlines' expectations in what concerns the evolution of the sufficient statistics, and find that the results always converge to the same equilibrium.

VI Removal of the HSR: Profit Decomposition

Table VI.1: Routes Most Affected by the HSR

Panel A: Routes most negatively affected by the HSR

	Baseline	No HSR	Difference
Average No. of Airlines	0.41	3.15	-2.75
Profit	-0.39	0.53	-0.92

Profit Decomposition

	Gains/Losses	Fraction of Total Gains (Losses)	Gains/Losses	Fraction of Total Gains (Losses)	Gains/Losses	Fraction of Total Gains (Losses)
Own Network	0.76	62%	0.79	100%	-0.04	(2%)
Competitors' Network	0.18	15%	-0.06	(24%)	0.25	46%
Negative Spillovers from HSR	-1.41	(88%)	0.00	0%	-1.41	(98%)
Positive Spillovers from HSR	0.29	23%	0.00	0%	0.29	54%
Market Characteristics	-0.20	(12%)	-0.20	(76%)	0	0%
Total Gains	1.23		0.79		0.53	
Total Losses	-1.61		-0.26		-1.45	

Panel B: Routes most positively affected by the HSR

	Baseline	No HSR	Difference
Average No. of Airlines	2.38	1.54	0.83
Profit	0.46	0.24	0.22

Profit Decomposition

	Gains/Losses	Fraction of Total Gains (Losses)	Gains/Losses	Fraction of Total Gains (Losses)	Gains/Losses	Fraction of Total Gains (Losses)
Own Network	0.73	54%	0.77	91%	-0.04	(10%)
Competitors' Network	-0.01	(1%)	0.08	9%	-0.09	(23%)
Negative Spillovers from HSR	-0.26	(30%)	0	0%	-0.26	(67%)
Positive Spillovers from HSR	0.61	46%	0	0%	0.61	100%
Market Characteristics	-0.61	(69%)	-0.61	(100%)	0.00	0%
Total Gains	1.34		0.85		0.61	
Total Losses	-0.88		-0.61		-0.39	

Note: This table reports simulated airline presence and profit decomposition for the two groups of top 10 routes which have the largest difference in the predicted number of airlines between the two scenarios "Baseline" and "No HSR". Panel A reports results for the routes which are the most negatively affected by the HSR in terms of airline presence, and Panel B for the routes which are the most positively affected by the HSR. "Profit" refers to the average profits calculated across routes (10 routes) and airlines (4 airlines). The profit decomposition items are defined based on the structural model flow payoff variables listed in Table 4. "Own network" refers to the effect associated with the number of own routes connected. "Competitors' Network" includes the effects from the number of competitors and competitors' connections. The classification of the effects associated with the impact of the HSR into "Negative" and "Positive" Spillover effects is done based on the sign of the estimated coefficients from the structural model. "Market characteristics" includes the effects associated with the constant term of profits, GDP, and the unobserved market type. The numbers reported are obtained by averaging across $1,000$ simulations for each scenario at the end of 2015.

VII Corridor Prioritization: Additional Tables

Table VII.1: Major HSR Corridors

Corridor	Direction	First Segment Completion Date	Last Segment Completion Date	Number of Stations	Length (km)
Beijing-Harbin	North-South	12/1/2012	9/12/2013	28	1,700
Beijing-Guangzhou-Shenzhen-Hongkong	North-South	12/26/2009	12/30/2015	44	2,260
Beijing-Shanghai	North-South	6/30/2011	7/1/2013	45	1318
Hangzhou-Fuzhou-Shenzhen	North-South	2/6/2010	12/28/2013	23	1,400
Shanghai-Kunming	East-West	10/26/2010	6/18/2015	56	2,080
Shanghai-Wuhan-Chengdu	East-West	4/18/2008	12/28/2013	79	1,600
Tsingtao-Taiyuan	East-West	12/21/2008	4/1/2009	20	770
Xuzhou-Lanzhou	East-West	9/28/2009	12/28/2013	64	1,600

Note: This table presents the main features of the HSR corridors in China.

Table VII.2: Corridor Prioritization: Effect on Airline Presence and Profits

	Airline Presence (Non-unique Routes) (Baseline=1,785)		Airline Presence (Unique Routes) (Baseline=894)		Flow Profits (Baseline=639)	
Corridor	△	△(%)	△	△(%)	△	△(%)
Beijing-Harbin	22	1.23	6.86	0.77	9	1.41
Beijing-Guangzhou-Shenzhen-Hongkong	6	0.34	1.62	0.18	4	0.63
Beijing-Shanghai	82	4.59	25.33	2.83	35	5.48
Hangzhou-Fuzhou-Shenzhen	-19	-1.06	-7.61	-0.85	-6	-0.94
Shanghai-Kunming	45	2.52	14.18	1.59	26	4.07
Shanghai-Wuhan-Chengdu	57	3.19	17.31	1.94	24	3.76
Tsingtao-Taiyuan	10	0.56	2.81	0.31	4	0.63
Xuzhou-Lanzhou	13	0.73	4.07	0.45	5	0.78

Note: This table reports changes in airline presence and flow profits resulting from corridor prioritization relative to the baseline.

Table VII.3: Corridor Prioritization: Effect on Airline Profits by Region

Corridor	Northeast	North	Northwest	East	Central	South	Southwest
Beijing-Harbin	1.747	1.566	0.808	3.803	1.777	2.608	-5.134
Beijing-Guangzhou-Shenzhen-Hongkong	0.321	0.576	-0.193	-0.390	0.547	1.277	-6.161
Beijing-Shanghai	5.924	8.176	4.372	19.787	6.842	12.385	0.500
Hangzhou-Fuzhou-Shenzhen	2.162	0.222	0.254	0.552	-0.172	0.692	-5.185
Shanghai-Kunming	-1.287	-2.233	-1.345	-5.661	-2.112	-3.323	-8.228
Shanghai-Wuhan-Chengdu	3.798	4.798	2.854	13.098	4.700	6.845	4.377
Tsingtao-Taiyuan	4.090	5.467	2.782	12.308	4.355	8.655	-1.895
Xuzhou-Lanzhou	0.342	-0.009	0.040	-0.071	0.479	1.188	-6.162

Note: This table reports changes in airline flow profits resulting from corridor prioritization, by region, relative to the baseline.