Estimation of Demand Function for Train Travel

Ke Ma, A0212524U
Mingzhe Xu, A0232022A
Nanhai Zhong, A0231953E
Xiao Liang, A0232007X

1 Abstract

The aim for this project is to estimate the demand function of train travel by using data on train ticket sales at a particular train station. The price, from experience, might be endogenous. The two-stage least squares method (2SLS) is used to handle this in a linear regression framework. Instrumental variables are identified and tested. 2SLS is applied to generate unbiased slope coefficients at a cost of variance increase.

2 Data Preprocessing

2.1 Data Review

There are 209,697 original data, and no missing values. *Train_Number_All, Train_Number_All, isNormCabin, isReturn, isOneway, Customer_Cat* are categorical variables; *dept_datetime* and *purch_datetime* are time series; other data are continuous variables.

2.2 Unify Date Format

For *dept_datetime* and *purch_datetime*, their format is not uniform: DD/MM/YY and YY/MM/DD. Therefore, these two columns are converted into the datetime data type of YY-MM-DD.

2.3 Add A New Variable: *Gap*

The date of purchase of the ticket will affect the price of each seat. In order to study the relationship between the ticket purchase date and the train departure date, this paper calculates the difference between the two, names it Gap and sets it as an int-type variable.

2.4 Add A New Categorical Variable: *isWeekend*

The ticket price on holidays is generally higher than the ticket price on weekdays. Therefore, a new categorical variable *isWeekend* is generated: if the train departure day is a working day, the return value is 0; if the train departure day is a weekend, the return value is 1.

2.5 Add A New Variable: *Price*

Mean_net_ticket_price represents the average price of each train ticket, but one train ticket corresponds to multiple seats. Therefore, the price of each seat should be calculated. Calculate *mean_net_ticket_price / num_seats_total* and name the result *price*.

2.6 Add A New Variable: *isdate_1819*

Noticed that prices of tickets dramatically jumped up on every purchased date on 18th, 19th monthly. Furtherly, the special purchased date is the same date in any month that the train on tickets departs. Therefore, a new binary variable *isdate_1819* is generated: if the purchase day is the same as the departure day in any month, the return value is 1, otherwise 0.

2.7 Processing Variable: *Quantity (Square Root)*

Culmulative_sales represents the number of train tickets sold, which is the Quantity in the economic model. Therefore, modify the column name to *Quantity*. After multiple times modeling, the square root of Q would advance the result accuracy, because the square root transformation for Q will reduce the variance of data.

Table. 2.1 Final Variable Table

Variable Name	Type	Explanation
Price	num	The price of each seat
Quantity	num	Quantity of ticket sales at end of day
Quantity(sqrt)	num	Square root of Quantity
Gap	num	Difference between train departure date and date the train ticket was purchased
isWeekend	categ	1 if the train departure day is a weekend, 0 if the train departure day is a weekday
isdate_1819	categ	1 if the train departure date is 18^{th} or 19^{th} of a month, else returns 0
Customer_Cat	categ	Customer segmentation done by the provider
Train_Number_All	categ	Train number/name (masked)
isNormCabin	categ	1 if the ticket is normal cabin, 0 if the ticket is for special cabin
isReturn	categ	1 if ticket corresponds to the return part of the trip
isOneway	categ	1 if the ticket corresponds to a one-way ticket (no return trip)

3 Data cleaning

3.1 Eliminating Records Conflicted with Truth

a) After observing the data, it can be found that the departure date of some trains is earlier than the date of ticket purchased. This is obviously abnormal, and this data should be eliminated. After eliminating, there are 155,211 rows contained.
b) Via Figure E1.3, the new variable gap is set within 360 days because it is not unusual to purchase the ticket 1 year earlier, so the data behind this point will be eliminated.

3.2 Eliminating Outliers

According to Figure E1.2, there are some outliers that need to be eliminated first. Then this diagram does not fit with the demand function accurately which is a smoothly concave up polynomial function, So in other words, there are some strange things that should be found out by data screening.

From Figure E1.4-E1.5, there are a few outliers that need to be eliminated, then these two plots also claim that *isNormCabin* and *Customer_cat* have some relationship with the price. Moverover, keep doing data cleaning for other binary variables by similar method. After data cleaning, there are 87854 rows in the data set.

4 Model Building

4.1 Correlation analysis between *Customer_Cat* and other exogenous variables

To use *Customer_Cat* verify low correlation between *Customer_Cat* and *isReturn, isOneway, Gap, Train_Number_All, isNormCabin, isWeekend*, calculate the correlation coefficient matrix between them, and make a heat map Figure E1.6 attached in the appendix.

According to the heat map, *Customer_Cat* has little correlation with them, so *Customer_Cat* can be used as a variable without Multicollinearity.

4.2 Demand Function and Supply Function Modeling

First, with the basic knowledge of microeconomics, the demand quantity and the supply quantity both have a linear relationship with price.

Secondly, according to the actual scenario, the variables collected above needed to be separated into different functions. *Train_Number_All* (indicating the origin and destination of the train), *isNormCabin* and *isWeekend* are included in both functions. Because for customers, these three variables will influence their willingness probably based on their purpose of traveling, their requirement for the comfort, their spare time and so on. Also, for the train company, due to the limitation of trains and different kinds of cabins, these variables will also be taken into account when providing services. *Customer_Cat*, *Gap* and *isdate_1819* are included in the demand function for the reason that these variables are only related to customers' behavior. *IsReturn* and *isOneway* are included in the supply function. When customers are buying tickets, as long as the origin and destination meet their needs, it doesn't matter whether this trip is defined as 'Return' or whether this train will go back. However, for the train company, they need to decide which station to send the train to complete each task, and at which station the train stops after completing all the tasks for each day.

Thirdly, considering the unknown influencing factors, error terms are added to each function. Then,

Demand Function:

$$\begin{aligned} Quantity \ &= \beta_{10} + \beta_{11} Price + \beta_{12} Train\_Number\_All \ + \ \beta_{13} isNormCabin \ + \ \beta_{14} isWeekend \\ &+ \ \alpha_1 Customer\_Cat + \alpha_2 \ Gap \ + \alpha_3 isdate\_1819 \ + \upsilon_{11} \end{aligned}$$

Supply Function:

$$\begin{aligned} Quantity \ &= \beta_{20} + \beta_{21} Price \ + \ \beta_{22} Train\_Number\_All \ + \ \beta_{23} isNormCabin \ + \ \beta_{24} isWeekend \\ &+ \ \gamma_1 isReturn + \gamma_2 isOneway + \ \upsilon_{21} \end{aligned}$$

4.3 Endogeneity derived by simultaneity

However, the quantity and price provided is only when the equilibrium is achieved, which means the two functions are simultaneous equations. Therefore, it is possible that the price in demand function is endogenous. Then, we need to find several instrumental variables to separate the price into the part correlated to the error term and the other part independent of the error term.

When market equilibrium is achieved, the demand quantity will equal the supply quantity. Then, *The Reduced Form of Demand Function:*

$$Price = \pi_0 + \pi_1 Train\_Number\_All + \pi_2 isNormCabin + \pi_3 isWeekend + \pi_4 Customer\_Cat + \pi_5 \, Gap + \pi_6 isdate\_1819 + \pi_7 isReturn + \pi_8 isOneway + \epsilon$$

From the equations above, *isReturn* and *isOneway* are only included in the supply function. For now, *isReturn* and *isOneway* are assumed to be exogenous to demand function, which will be tested after the estimation. *isReturn* and *isOneway*, therefore, become instrumental variables candidates.

4.4 Fitting model with 2SLS

After selecting instrumental variables, the 2SLS method will be applied to estimate demand function. The first step is to fit the reduced form model and use this model to predict the price. The second step is to substitute the price in the structural form equation with the prediction of price and fit the updated model.

$$\widehat{Price} = \widehat{\pi_0} + \widehat{\pi_1} Train\_Number\_All + \widehat{\pi_2} isNormCabin + \widehat{\pi_3} isWeekend + \widehat{\pi_4} Customer\_Cat + \widehat{\pi_5} \, Gap + \widehat{\pi_6} isdate\_1819 + \widehat{\pi_7} isReturn + \widehat{\pi_8} isOneway$$

$$Quantity = \beta_{10} + \beta_{11}\widehat{Price} + \beta_{12} Train\_Number\_All + \beta_{13} isNormCabin + \beta_{14} isWeekend + \alpha_1 Customer\_Cat + \alpha_2 \, Gap + \alpha_3 isdate\_1819 + u$$

4.5 2SLS Results and Corresponding Test

The accuracy (Adj. R-squared) of the reduced form model is 0.47 (Table E2.3)). From the fitting result of the 1st step of 2SLS (Table E2.3)), all the p-value is lower than 0.05, which means all the variables are significant in the reduced form. The coefficients of *Customer_Cat[T.B]*, *isNormCabin*, *Gap* and *isReturn* are negative, while the coefficients of *isdate_1819[T.True]*, *isWeekend* and *isOneway* are positive. Therefore, if the customer who is in category A buys a special-cabin weekend train ticket which is a one-way train from the origin to the destination, at the 18th or 19th of each month near the departure time, the price of this seat will be higher. The final model is shown as below.

Reduced form:

$$\begin{aligned}\widehat{Price} = {} & 318.2618 - 7.0841 Train\_Number\_All[T.B] + 24.3502 Train\_Number\_All[T.C] \\ & + 17.3149 Train\_Number\_All[T.D] - 28.2922 Train\_Number\_All[T.E] \\ & - 18.1252 Train\_Number\_All[T.F] - 4.121 Train\_Number\_All[T.G] \\ & - 18.0785 Train\_Number\_All[T.H] + 11.1702 Train\_Number\_All[T.I] \\ & - 5.4605 Train\_Number\_All[T.J] + 28.4299 Train\_Number\_All[T.K] \\ & + 17.0864 Train\_Number\_All[T.L] + 4.1103 Train\_Number\_All[T.M] \\ & + 64.715 Train\_Number\_All[T.N] - 132.499 isNormCabin + 7.7435 isWeekend \\ & + 83.2361 Customer\_Cat[T.B] - 0.1572 Gap + 76.8503 isdate\_1819[T.True] \\ & - 14.6745 isReturn + 48.26 isOneway\end{aligned}$$

The accuracy (Adj. R-squared) of the 2nd step model is 0.261 (Table E2.4). From the fitting result of the 2nd step of 2SLS (Table E2.4), all the variables are significant as well. The coefficient of predicted price is negative, which means with the price decrease the quantity will increase. The coefficients of *Customer_Cat[T.B]*, *isNormCabin* and *Gap*are negative, while the coefficients of *isdate_1819[T.True]* and *isWeekend* are positive. The final model is shown below.

2nd step demand function:

$$
\begin{aligned}
Quantity = {} & 5.8423 - 0.0129\widehat{Price} + 1.1566Train\_Number\_All[T.B] \\
& + 0.5756Train\_Number\_All[T.C] + 0.2125Train\_Number\_All[T.D] \\
& + 1.0725Train\_Number\_All[T.E] + 0.4077Train\_Number\_All[T.F] \\
& + 0.7534Train\_Number\_All[T.G] + 0.5798Train\_Number\_All[T.H] \\
& + 0.4443Train\_Number\_All[T.I] + 0.0494Train\_Number\_All[T.J] \\
& - 0.1396Train\_Number\_All[T.K] + 0.0537Train\_Number\_All[T.L] \\
& - 0.1135Train\_Number\_All[T.M] + 0.0802Train\_Number\_All[T.N] \\
& - 0.4721isNormCabin + 0.0301isWeekend - 0.4029Customer\_Cat[T.B] - 0.0041Gap \\
& + 0.9178isdate\_1819[T.True] + u
\end{aligned}
$$

4.5.1 Weak Instruments Test:

As stated above, it is assumed that *isReturn* and *isOneway* are IVs. Test here whether *isReturn* and *isOneway* are jointly significant in the endogenous variable *Price:*

H0: $\pi_1 = \pi_2 = \ldots = \pi_8 = 0$

H1: At least one coefficient not equal to 0

First stage OLS *F*-statistic equals to 3900 (Table E2.3), Moreover, slope coefficients are all not 0, reject *H0*. Thus, instruments are strongly correlated with *Price,* to have reliable 2SLS estimators.

4.5.2 Hausman Test: Testing for endogeneity

To test endogeneity, estimate the reduced form for *Price* by regressing on all exogenous variables. Obtain the residuals and add it to the structural equation whose coefficient is 0.0077 with P value equals to 0.000. Price is significantly correlated to residuals. Thus, conclude that *Price* is indeed endogenous..

4.5.3 Sargan Test: Test for overidentifying restrictions

Since we identify two IVs, *isReturn* and *isOneway*, Sargan test was done to verify instruments are not correlated with 2SLS error term. Set q = 1, test $nR^2 \sim \chi^2_1$. Outcome $\chi^2_1 = 2.060$, p = 0.151. Do not reject *H0*, i.e., all IVs are exogenous.

5 Discussion

5.1 Comparing results of OLS and 2SLS

Model	Structural Form (OLS)	Reduced Form (1st step of 2SLS)	2st step of 2SLS
Adj. R-squared	0.328	0.47	0.261

5.1.1 In terms of adjusted R^2, OLS model performs better than 2SLS model in explaining the demand function with the given dataset. However, because of endogeneity the OLS model is biased in the first place. 2SLS helps to fix this problem, using exogenous predicted *Price* to estimate the demand.

5.1.2 What should be noticed is that the standard errors summarized in (Table E2.4) from 2SLS are incorrect, usually smaller than the correct ones. Amendation can be done to fix them.

5.2 Slope of demand curve and IVs

a) The negative coefficient of *Price* in OLS (-0.0054, Table E2.2) is way bigger than one in 2SLS (-0.0129, Tabel E2.4), which is reasonable. According to the Q-P simultaneity model discussed in class, OLS estimation is biased because of endogeneity. Fake demand curve is flatter than the real one.

b) Through both two pass Sargan Test, *IsOneway* is better suggested by lower correlation with 2SLS residual. To improve, maybe the model can use *IsReturn* only as an instrument.

5.3 Improvement can be done furtherly

i. Time series treatment of purchase date and departure date: it is inferred that price various in different years as inflation, systematic adjustment of train price, technology advance, etc.

ii. Seasonal fluctuation of the ticket price can be taken into consideration.

Appendix

E1. Data Exploration Figures

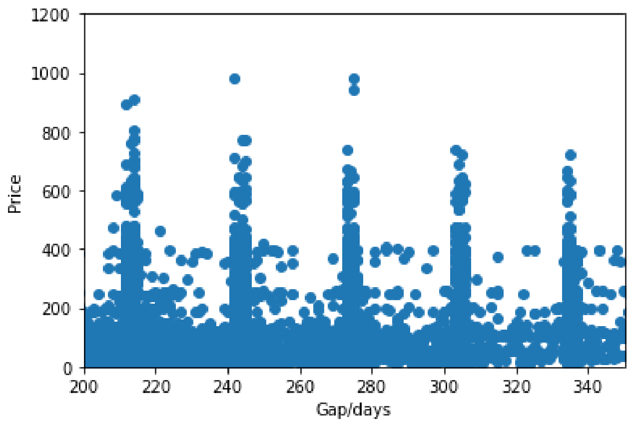

Figure E1.1 *Price* abnormally increase periodically with *Gap*

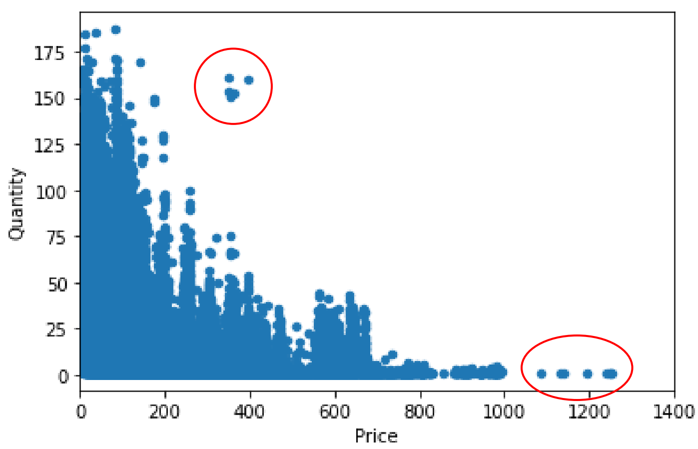

Figure E1.2 Outliers in Q-P plot

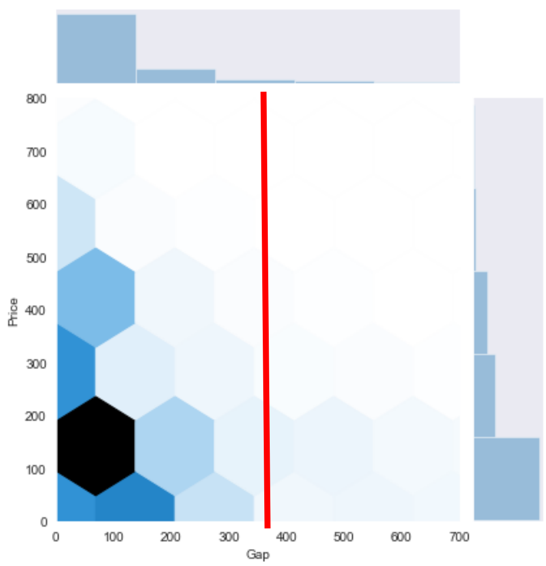

Figure E1.3 *Gap* with *Price*

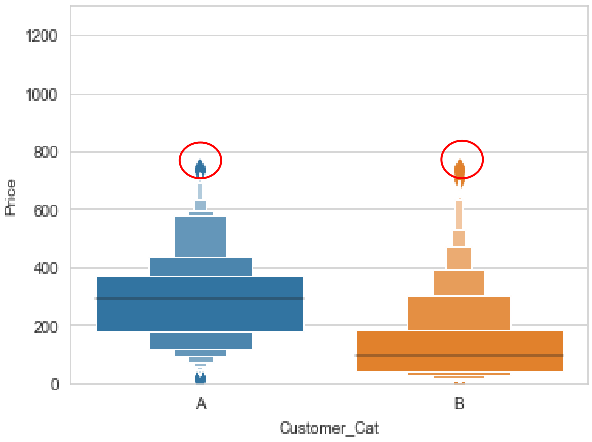

Figure E1.4 *Customer_Cat* with *Price*

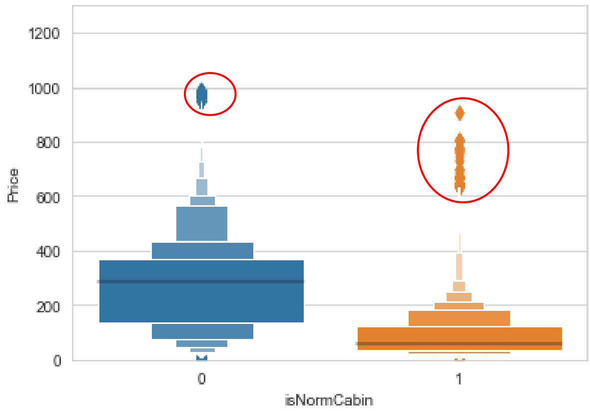

Figure E1.5 *isNormCabin* with *Price*

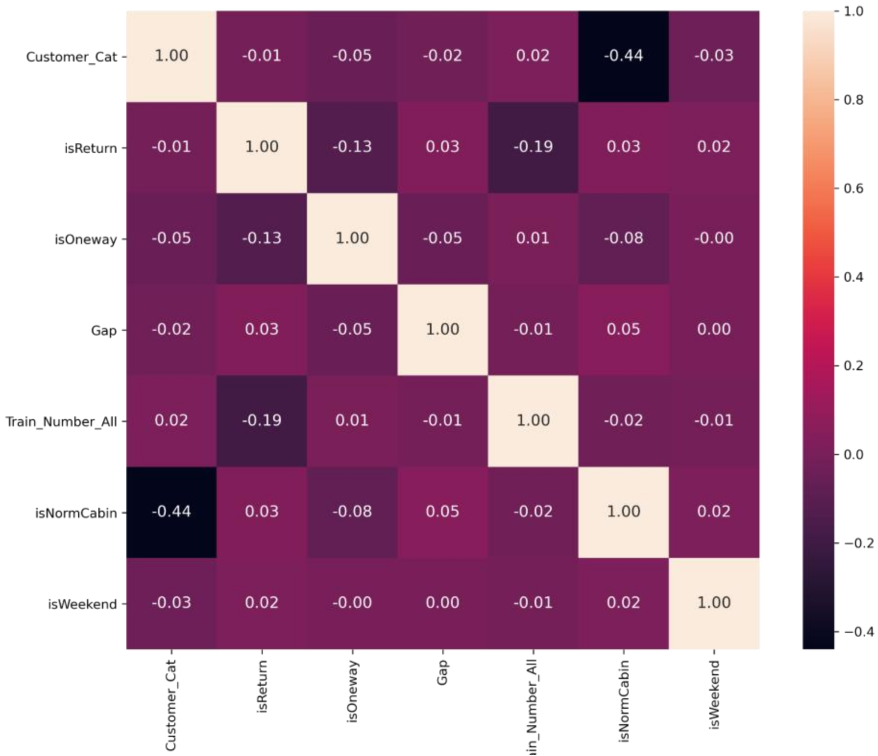

Figure E1.6 Data Correlation

E2. Test Result

Table E2.1 Data Transformation

	R^2 for Reduced Form	R^2 for the 2nd step of 2SLS	Hausman test (F-statistic)	Sargan test (p-value)
P & Q	0.470	0.212	1504	0.898
logP & Q	0.444	0.212	1992	0.357
P & logQ	0.470	0.275	2672	0.001
logP & logQ	0.444	0.275	3219	0.000
sqrtP & Q	0.487	0.212	1745	0.73
P & sqrtQ	0.470	0.261	2186	0.151
sqrtP & sqrtQ	0.487	0.261	2555	0.029
logP & sqrQ	0.444	0.261	2828	0.002
sqrP & logQ	0.487	0.274	3086	0.000

Table E2.2 OLS for Structural Form

Model	Adj. R-squared
OLS	0.328

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	3.4069	0.0313	108.8062	0	3.3456	3.4683
Train_Number_All[T.B]	1.2915	0.029	44.4814	0	1.2346	1.3484
Train_Number_All[T.C]	0.4765	0.032	14.8902	0	0.4138	0.5392
Train_Number_All[T.D]	0.1631	0.0331	4.9253	0	0.0982	0.228
Train_Number_All[T.E]	1.2863	0.029	44.3936	0	1.2295	1.343
Train_Number_All[T.F]	0.5479	0.0307	17.8397	0	0.4877	0.6081
Train_Number_All[T.G]	0.8655	0.0298	29.0779	0	0.8072	0.9238
Train_Number_All[T.H]	0.7183	0.0307	23.4266	0	0.6582	0.7784
Train_Number_All[T.I]	0.4458	0.0313	14.2266	0	0.3844	0.5073
Train_Number_All[T.J]	0.0915	0.0329	2.7805	0.0054	0.027	0.156
Train_Number_All[T.K]	-0.2805	0.0396	-7.0821	0	-0.3582	-0.2029
Train_Number_All[T.L]	-0.068	0.0397	-1.7133	0.0867	-0.1459	0.0098
Train_Number_All[T.M]	-0.1455	0.0353	-4.1178	0	-0.2147	-0.0762
Train_Number_All[T.N]	-0.3531	0.0533	-6.6255	0	-0.4576	-0.2487
Customer_Cat[T.B]	0.1965	0.0155	12.6663	0	0.1661	0.2269
isdate_1819[T.True]	0.1683	0.045	3.7371	0.0002	0.08	0.2566
Price	-0.0054	0.0001	-102.032	0	-0.0055	-0.0053
isNormCabin	0.5542	0.0151	36.7191	0	0.5246	0.5837
isWeekend	-0.0273	0.0129	-2.1272	0.0334	-0.0525	-0.0021
Gap	-0.0029	0.0001	-36.7057	0	-0.003	-0.0027

Table E2.3 Reduced form

Model	Adj. R-squared	F-statistic
OLS	0.47	3900

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	318.2618	1.67	190.534	0	314.988	321.536
Train_Number_All[T.B]	-7.0841	2.037	-3.478	0.001	-11.076	-3.092
Train_Number_All[T.C]	24.3502	2.208	11.028	0	20.022	28.678
Train_Number_All[T.D]	17.3149	2.272	7.622	0	12.862	21.768
Train_Number_All[T.E]	-28.2922	1.831	-15.455	0	-31.88	-24.704
Train_Number_All[T.F]	-18.1252	1.942	-9.332	0	-21.932	-14.318
Train_Number_All[T.G]	-4.121	2.079	-1.982	0.047	-8.195	-0.047
Train_Number_All[T.H]	-18.0785	1.939	-9.324	0	-21.879	-14.278
Train_Number_All[T.I]	11.1702	2.17	5.148	0	6.918	15.423
Train_Number_All[T.J]	-5.4605	2.082	-2.623	0.009	-9.541	-1.38
Train_Number_All[T.K]	28.4299	2.643	10.757	0	23.25	33.61
Train_Number_All[T.L]	17.0864	2.512	6.802	0	12.163	22.01
Train_Number_All[T.M]	4.1103	2.235	1.839	0.066	-0.271	8.492
Train_Number_All[T.N]	64.715	3.45	18.757	0	57.952	71.477
Customer_Cat[T.B]	-83.2361	0.952	-87.393	0	-85.103	-81.369
isdate_1819[T.True]	76.8503	2.87	26.778	0	71.225	82.475
isNormCabin	-132.499	0.844	-157.021	0	-134.152	-130.845
isWeekend	7.7435	0.813	9.529	0	6.151	9.336
Gap	-0.1572	0.005	-32.053	0	-0.167	-0.148
isReturn	-14.6745	1.165	-12.597	0	-16.958	-12.391
isOneway	48.26	1.109	43.528	0	46.087	50.433

Table E2.4 2nd Step of 2SLS

Model	Adj. R-squared
OLS	0.261

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	5.8423	0.112	51.965	0	5.622	6.063
Train_Number_All[T.B]	1.1566	0.031	37.272	0	1.096	1.217
Train_Number_All[T.C]	0.5756	0.034	17.004	0	0.509	0.642
Train_Number_All[T.D]	0.2125	0.035	6.108	0	0.144	0.281
Train_Number_All[T.E]	1.0725	0.032	33.703	0	1.01	1.135
Train_Number_All[T.F]	0.4077	0.033	12.43	0	0.343	0.472
Train_Number_All[T.G]	0.7534	0.032	23.835	0	0.691	0.815
Train_Number_All[T.H]	0.5798	0.033	17.713	0	0.516	0.644
Train_Number_All[T.I]	0.4443	0.033	13.517	0	0.38	0.509
Train_Number_All[T.J]	0.0494	0.035	1.43	0.153	-0.018	0.117
Train_Number_All[T.K]	-0.1396	0.042	-3.324	0.001	-0.222	-0.057
Train_Number_All[T.L]	0.0537	0.042	1.278	0.201	-0.029	0.136
Train_Number_All[T.M]	-0.1135	0.037	-3.061	0.002	-0.186	-0.041
Train_Number_All[T.N]	0.0802	0.059	1.358	0.175	-0.036	0.196
Customer_Cat[T.B]	-0.4029	0.031	-12.969	0	-0.464	-0.342
isdate_1819[T.True]	0.9178	0.058	15.913	0	0.805	1.031
pre_Price	-0.0129	0	-38.28	0	-0.014	-0.012
isNormCabin	-0.4721	0.048	-9.836	0	-0.566	-0.378
isWeekend	0.0301	0.014	2.197	0.028	0.003	0.057
Gap	-0.0041	9.93E-05	-41.667	0	-0.004	-0.004

Table E2.5 Hausman test

Model	Adj. R-squared	F-statistic
OLS	0.332	2186

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	5.8423	0.1068	54.6815	0	5.6329	6.0517
Train_Number_All[T.B]	1.1566	0.0295	39.22	0	1.0988	1.2144
Train_Number_All[T.C]	0.5756	0.0322	17.8923	0	0.5125	0.6386
Train_Number_All[T.D]	0.2125	0.0331	6.4273	0	0.1477	0.2774
Train_Number_All[T.E]	1.0725	0.0302	35.4645	0	1.0132	1.1318
Train_Number_All[T.F]	0.4077	0.0312	13.0795	0	0.3466	0.4688
Train_Number_All[T.G]	0.7534	0.03	25.0809	0	0.6945	0.8123
Train_Number_All[T.H]	0.5798	0.0311	18.6384	0	0.5189	0.6408
Train_Number_All[T.I]	0.4443	0.0312	14.2233	0	0.3831	0.5055
Train_Number_All[T.J]	0.0494	0.0328	1.5045	0.1324	-0.015	0.1138
Train_Number_All[T.K]	-0.1396	0.0399	-3.4975	0.0005	-0.2179	-0.0614
Train_Number_All[T.L]	0.0537	0.0399	1.3452	0.1786	-0.0245	0.1319
Train_Number_All[T.M]	-0.1135	0.0352	-3.2213	0.0013	-0.1826	-0.0445
Train_Number_All[T.N]	0.0802	0.0562	1.4286	0.1531	-0.0298	0.1903
Customer_Cat[T.B]	-0.4029	0.0295	-13.6466	0	-0.4608	-0.3451
isdate_1819[T.True]	0.9178	0.0548	16.7452	0	0.8104	1.0252
Price	-0.0129	0.0003	-40.2807	0	-0.0135	-0.0123
isNormCabin	-0.4721	0.0456	-10.3501	0	-0.5615	-0.3827
isWeekend	0.0301	0.013	2.3114	0.0208	0.0046	0.0557
Gap	-0.0041	0.0001	-43.8452	0	-0.0043	-0.004
res1stage	0.0077	0.0003	23.8335	0	0.0071	0.0084

Table E2.6 Sargan test

Model	Adj. R-squared	test_stat	p_value
OLS	0	2.059659	0.151244

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	-0.0029	0.0277	-0.1032	0.9178	-0.0571	0.0514
Train_Number_All[T.B]	-0.0203	0.0338	-0.6015	0.5475	-0.0865	0.0459
Train_Number_All[T.C]	-0.0203	0.0366	-0.5541	0.5795	-0.092	0.0515
Train_Number_All[T.D]	-0.0203	0.0377	-0.5384	0.5903	-0.0941	0.0535
Train_Number_All[T.E]	0.0004	0.0304	0.0122	0.9903	-0.0591	0.0599
Train_Number_All[T.F]	0.0008	0.0322	0.025	0.9801	-0.0623	0.0639
Train_Number_All[T.G]	-0.0203	0.0345	-0.5897	0.5554	-0.0879	0.0472
Train_Number_All[T.H]	0.0006	0.0321	0.0189	0.9849	-0.0624	0.0636
Train_Number_All[T.I]	-0.0202	0.036	-0.5626	0.5737	-0.0907	0.0503
Train_Number_All[T.J]	0.0001	0.0345	0.0021	0.9983	-0.0676	0.0677
Train_Number_All[T.K]	-0.0195	0.0438	-0.4441	0.657	-0.1053	0.0664
Train_Number_All[T.L]	-0.0009	0.0416	-0.022	0.9825	-0.0825	0.0807
Train_Number_All[T.M]	-0.0009	0.0371	-0.0235	0.9813	-0.0735	0.0718
Train_Number_All[T.N]	-0.0176	0.0572	-0.3085	0.7577	-0.1298	0.0945
Customer_Cat[T.B]	-0.0022	0.0158	-0.1394	0.8891	-0.0332	0.0287
isdate_1819[T.True]	0.003	0.0476	0.0635	0.9494	-0.0902	0.0963
isNormCabin	0.0004	0.014	0.0291	0.9768	-0.027	0.0278
isWeekend	0.0001	0.0135	0.0039	0.9969	-0.0264	0.0265
Gap	0	0.0001	-0.0061	0.9951	-0.0002	0.0002
isReturn	0.0268	0.0193	1.3872	0.1654	-0.0111	0.0646
isOneway	0.0123	0.0184	0.6697	0.503	-0.0237	0.0483

Model	Structural Form (OLS)	Reduced Form (1st step of 2SLS)	2st step of 2SLS
Adj. R-squared	0.328	0.47	0.261