Available online at www.iujtl.com

JTL

Journal of Transportation and Logistics
7 (1) 2022

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/JTL.2022.994262

RESEARCH ARTICLE

The Liberalization Effect on Air Transport Expenditure Choices of Turkish Households: The Perspective of Micro Data Analysis

Gizem Kaya Aydın[1], Hidayet Beyhan[2], Umut Aydın[3], Berker Pandır[4]

ABSTRACT

After the liberalization of the air transportation sector in 2003, we investigate the factors affecting the preference of air transportation as a mode of transportation and the changes in the air transportation preferences of the Turkish households over the years. For this purpose, we analyze the micro datasets of the household budget surveys using logistic regression and decision tree methods. We find that the most critical factors affecting the air transportation preference of the households for the 2003-2017 period are the income level and the occupation of the household head. The fuel expenditures and the existence of transportation subsidies reduce the air transportation preference of the households. In addition, since 2003, there has been a significant increase in the rate of households preferring air transportation in Turkey. So much so that air transportation is the only mode of transportation that has risen over time. A remarkable finding is that air expenditure of the low-income group is not observed in the 2003 survey, whereas it was a small amount in 2017. The literature on transportation preferences is generally based on primary survey data. This study contributes to the literature as it is applied to a relatively large sample of household micro datasets.

Keywords: Air Transportation Expenditure, Household Preferences, Liberalization Effect

Submitted: 14.09.2021 • **Revision Requested:** 14.03.2022 • **Last Revision Received:** 19.03.2022 • **Accepted:** 20.03.2022

[1] **Corresponding author:** Gizem Kaya Aydın (Ph.D.), Istanbul Technical University, Department of Management Engineering, Besiktas, Istanbul, Turkiye. E-mail: kayagizem@itu.edu.tr ORCID: 0000-0002-6870-7219

[2] Hidayet Beyhan (Ph.D. Candidate), Istanbul Technical University, Department of Management Engineering, Besiktas, Istanbul, Turkiye . E-mail: beyhan17@itu.edu.tr ORCID: 0000-0002-0219-7076

[3] Umut Aydın (Ph.D . Candidate), Bandırma Onyedi Eylül University, Department of Transportation Engineering, Intelligent Transportation System, Balıkesir, Turkiye. E-mail: uaydin@bandirma.edu.tr ORCID: 0000-0003-4802-8793

[4] Berker Pandır (Ph.D. Candidate), Istanbul Technical University Department of Management Engineering, Besiktas, Istanbul, Turkiye. E-mail: pandir@itu.edu.tr ORCID: 0000-0003-3869-2244

Citation: Kaya Aydin, G., Beyhan, H., Aydin, U., & Pandir, B. (2022). The liberalization effect on air transport expenditure choices of turkish households: the perspective of micro data analysis. *Journal of Transportation and Logistics, 7*(1), 1-18. https://doi.org/10.26650/JTL.2022.994262

1. Introduction

The airline industry has grown in importance with its effect on the economic situation of the countries and a world-perceived country profile. However, this sector suffered losses due to Covid-19 in last two years, overall traveller numbers in 2021 were 47% lower than in 2019. Nevertheless, according to IATA's expectations, this will rise to 83 percent in 2022, 94 percent in 2023, 103 percent in 2024, and 111 percent in 2025. Therefore, as long as there are no major shocks and crises, the growth potential of the airline industry is high (IATA, 2022).

Turkey has shown improvements in the aviation sector along with worldwide developments in the air transport sector. The directorate general of civil aviation of Turkey reveals that the total number of passengers using air transportation in Turkey has increased over four-fold from 2004 to 2017 (Turkstat, 2019). According to the consumption statistics of Turkstat (2021), the share of household consumption expenditures in transportation was 9.8 per cent in 2003, whereas it was 18.7 per cent in 2017. In particular, when income groups are analyzed to determine a differentiation in terms of their airline expenditure, the high-income group's budget share of transportation expenditure is the highest as expected, and it has the highest percentage of their budget (with a share of 29.4 per cent in 2017). These micro-level statistics are essential as they provide an insight into the decomposition of household transportation spending.

With the implementation of liberal policies for air transport in the 2000s in Turkey, there has been a big step forward for the entrance of private airlines in the domestic market that leads to a competitive domestic market. Along with the amendments to the regulations, some special incentives have been offered to airlines that consider entering the market, such as reducing airport usage fees, reducing some of the additional taxes, or removing some. These applications are aimed at reducing the operating costs of enterprises entering the market, offering lower passenger ticket prices, and making the airline operations market attractive. A single airline had been operated under state ownership until 2003; seven airlines entered the domestic market within seven years of liberalization and were opened to competition (Gerede, 2015). With this competition, some advantages have been also experienced in favor of Turkish households (such as differing prices in airline tickets and increasing route options), and airlines have started to come to the fore in the transportation preferences of households living in Turkey.

2. Theorical Chapter

Airline industry liberalization has led to a competitive market that increases low-cost ticket options and increases household air transport preferences, so do they gain. (Vieira, Câmara, Silva and Santos, 2019). There are underlying reasons that increase the preference for air transport, such as economic reasons, marketing strategies, and government policies. These reasons directly affect the choice of household mode of transport. Not only the economic and marketing policies, the economic and demographic characteristics of the households also have significant effects on choosing the air transport as a transport mode and spending money on which is used by airlines to determine their marketing strategies. The literature is therefore primarily concerned with studies focusing on determinants

of preferences in the mode of transportation (Buehler, 2011; Zenina and Borisov, 2011; Lee, Yoo and Song, 2016; Saygılı and Türkcan, 2018) and determinants of preferences in airlines (Fourie and Lubbe, 2006; Chiou and Chen, 2010; Ukpere, Stephens, Ikeogu, Ibe, and Akpan, 2012). Understanding the degree of determinants that leads customers to buy services is of great importance, and companies operating in air transport seek to maximize their market share based on these results.

The share of transportation expenditures in household consumption expenditures in Turkey is in an increasing trend. With this, between 2002 and 2019, the share of transportation expenditures in the total is approximately 14.8% on average. In the changes in the share of transportation expenditures in the total, factors such as income level, land use and urbanization, regional socio-economic concentrations, mobility, population and automobile ownership are effective. When the share of other expenditures in household expenditures is considered, it is seen that transportation expenditures rank third after "housing and rent" and "food and non-alcoholic beverages" expenditures (UNDP, 2021).

This study aims to concentrate on households' air transport preferences as a transportation mode and to examine the change in air transport expenditure by decomposing total transportation expenditure after the liberalization in 2003. We also aim to find the essential socio-economic influential factors that affect the choice of air transport in detail by considering the impact of other transport modes and using such a big micro data. For this reason, we contribute to the literature by defining significant household preference factors for air transportation and designing policies for the airline sector.

3. Literature Review

To the best of our knowledge, no studies are investigating the socioeconomic factors affecting airway expenditures based on the household dataset. However, many studies are aiming to find variables that affect the choice of airport or airline companies. There are some studies in the literature examining Turkey's data. Yaylalı, Dilek and Çelik (2015) find the practical factors in choosing airlines by using a multistate logit model over the March-April 2012 period with a face-to-face survey of 2,473 participants. As a result of their study, the likelihood of choosing "A" type airlines is found higher than others in the event of a rise in either ticket price or passenger income. In another work, Yaylalı and Dilek (2017) analyzed the factors that affect the choice of passenger for Turkey. On the other hand, Çelikkol, Uçkun, Tekin, and Çelikkol (2012) examined the preference of airports, and they conducted a survey of customers in the Sabiha Gökçen Airport case to find out the factors that determine the customers' choice of service purchase and investigate the underlying reasons for preference of customers.

There are also studies examining preferences for airlines and airports in other countries. Milioti, Karlaftis, and Akkogiounoglou (2015) examined the factors affecting passenger decisions on airline choice by using a multivariate probit model based on a survey of 853 respondents in Athens International airport Eleftherios Venizelos. In the case of Nigeria, Ukpere et al. (2012) examined the domestic air transport industry with the same research question by using Nlogit.

Besides these studies, there is literature on the selection of transportation modes. Zenina and Borisov (2011) investigated the performance of mode choice analysis with classification methods - decision trees, discriminant analysis, and multinomial logit. On the other hand, Chee and Fernandez (2013) investigated factors related to the choice of mode of either public or private transport in Penang (in Malaysia) with primary data obtained in 2012. They found that males use transportation more. The presences of a driving license and access to a private vehicle are significant in influencing private mode choice behavior. Zhou et al. (2019) examined the choice of transport mode in the regional sense by using multinomial and nested logit models in Western Australia. Based on regional household data for the UK, Ryley and Davison (2007) found out people who fly very often prefer low-cost airlines. Nevertheless, most of the respondents think that aviation is beneficial to the national economy but harmful to the environment.

The literature is based primarily on survey data, since it has been rarely studied. This analysis contributes to the literature by using a large sample size direct household data.

4. Methodology and Data

In this study, the descriptive statistics of household transportation expenditures are primarily given. Subsequently, the preferences of households on air transportation expenditure are examined with a binary logistic regression method by taking socioeconomic variables. Besides logistic regression, a decision tree method is used to determine the best bunch of variables for our model. Sensitivity analysis and cross-validation are applied to confirm the validities of the model used.

4.1 Logistic regression

The application of linear regression is unlikely to be used as a method of classification as the probability boundaries could be out of [0,1]. Logistic and probit regression models, however, escape this problem as they are based respectively on the sigmoid and the normal distribution curve. The dependent variable is defined as a discrete binary with codes 0 and 1 in the logistic regression model, while both continuous and nominal variables may be the independent variables.

There are two main advantages which are that logistic regression does not rely on assumptions of normality for the predictor variables or the errors, and that it allows the selection effect to vary nonlinearly (Janzan and Stern, 2004). Logistic regression is a simple concept to grasp and necessitates less training. It works well with simple datasets as well as linearly separable data sets, and makes no assumptions about the class distributions in the feature space. Lastly, it provides different elasticity of the independent variables at each different point (Gujarati, 2009).

The specific form of the logistic regression model is given in Equation 1.

$$\ln\left(\frac{p}{1-p}\right) = Z_i = B_0 + B_1 X_i \quad (1)$$

The coefficients do not give direct marginal effects as in the linear regression. Because the P / (1-P) shows the odds ratio, the coefficient of B_1 shows the increase in the logarithmic

odds ratio. By using the formula given in Equation 2, the probabilities are obtained from logistic regression.

$$P_i = \frac{e^{z_i}}{1+e^{z_i}} \quad (2)$$

Mode selection models have been studied extensively in the literature and are usually based on a single year primary data survey. However, collecting for each year is difficult, and data may not represent the entire country. By using Turkstat's household budget survey data (secondary data), it is possible to compare households' air transport expenditure choices by years. In this study, expenditure values have not been analyzed directly; even if the real expenditure could be examined, the elasticity would probably exceed one, and air expenditure would be found as luxury goods. Moreover, due to a large number of studies about it in the literature (see Alperovich and Machnes, 1994; Taplin 1997; Gillen, Morrison, and Stewart, 2003), we have chosen a logistic regression model that will provide information to airlines about under which socioeconomic conditions households prefer an airline. Accordingly, airlines will develop their marketing policies.

4.2 Decision tree

A decision tree is a hierarchical relationship group in a tree-like structure starting from a variable called a root node. This root node is divided into two parts in a plurality of branches representing individual classes of the root node or specific intervals along the scale of the node. In each division, a question is asked, which has a response in terms of the classes or range of the divided variable. Decision trees can also be created with multiple partitions. The questions asked in each division are defined in terms of some measure of uncertainty that reflects the extent to which the resulting cases should be uniform in the divisions. Each branch is further divided by using classes or ranges of other variables. The process goes on until the cutting rule takes place. (Nisbet, Elder and Miner, 2009).

Decision trees are based on several examples that are supposed to represent the population, and the results obtained are tested on the so-called population and the most appropriate nodes achieved by the control automation method (Aytekin, Sütçü, and Özfidan, 2018). Although there are many algorithms used in the decision tree method, Chi-Squared Automatic Interaction Detector (CHAID) is used to perform division in this study. The branch number varies from two to the number of classes to be formed.

Logistic regression is a parametric regression method based on assumptions and provides policy suggestions with the help of direction and magnitude effects of coefficients. Decision trees are non-parametric methods and are used more often for classification, but do not provide much policy recommendations as implied as logistic regression. The CHAID algorithm, as applied in our study, shows the ranking of the most critical variables in the classification of the data. It has been therefore used to determine the most important variables that affect the airway expenditure choices of households. In the logistic regression, a variable could be interpreted as either statistically significant or not; however, with the decision tree, variables can be ranked according to their importance levels of classifying data.

5. The Data

Household Budget Surveys are critical sources to test the validities of the socioeconomic policies implemented and provide information about the socio-economic structures,

Table 1. Descriptive statistics

Income Quintiles	**Frequency**	**Percentage**
First Quintile	21,210	15.6
Second Quintile	25,994	19.1
Third Quintile	28,141	20.6
Fourth Quintile	29,888	21.9
Fifth Quintile	31,087	22.8
Transportation Usage	**Frequency**	**Percentage**
Air (Dependent Variable)		
No	134,288	98.5
Yes	2,032	1.5
Road		
No	18,598	13.6
Yes	117,722	86.4
Rail		
No	132,016	96.8
Yes	4,304	3.2
Maritime		
No	134,775	98.9
Yes	1,545	1.1
Fuel		
No	80,548	59.1
Yes	55,772	40.9
Presence of Transportation Subsidy	**Frequency**	**Percentage**
No	108,353	79.5
Yes	27,967	20.5
Occupation	**Frequency**	**Percentage**
Lawmakers, Senior Executives, and Managers	11,352	11.7
Professionals	7,614	7.8
Technicians and Assisting Professional Profession	5,503	5.7
Office Services	4,762	4.9
Service and Sales	12,025	12.4
Qualified Agriculture, Forestry and Water	18,666	19.2
Artisans and Related Works	15,422	15.8
Plant and Machine Operators	11,197	11.5
Jobs That Do Not Require Qualifications	10,796	11.1
Year	**Frequency**	**Percentage**
2003	19,717	14.5
2004	6,857	5.0
2005	7,267	5.3
2006	7,455	5.5
2007	7,506	5.5
2008	7,190	5.3
2009	8,724	6.4
2010	8,878	6.5
2011	8,769	6.4
2012	8,822	6.5
2013	9,048	6.6
2014	9,070	6.7
2015	8,682	6.4
2016	9,139	6.7
2017	9,196	6.8

life levels, and consumption patterns of households. Turkstat applies these surveys to different households every month from January to December and collects the monthly consumption expenditures of these households. In this study, household budget surveys with a micro data set of transportation expenditures are used. The cross-sectional data is retrieved from Turkstat from 2003 to 2017, which covers the airline liberalization steps in Turkey. The dataset includes the monthly expenditure on air, maritime, rail, road transport, and fuel of each household examined. In addition to these variables, the economic and demographic characteristics of the households, which may affect the air transportation preferences, are also used in the model.

For both logistic regression and decision tree methodologies, the dependent variable is a nominal variable that is defined as "1" if the household spends on air transport in the relevant month, otherwise it is defined as "0". The independent variables are railway preference, road preference, fuel expenditure preference, and maritime transportation preference, which are defined as "1" if the household spends on these transports in the relevant month, otherwise it is "0". Having the transportation subsidy is also included as explanatory variables. Finally, the socioeconomic variables also included the income group to which the household belongs, and the occupation of the household head (for studies using similar variables see Chee, and Fernandez, 2013; Yaylalı et al. 2015; Soltanzadeh, and Masoumi, 2014). All these explanatory variables are used as dummy variables. The number of observations was 136,320.

Table 1 displays the frequencies and percentages of the variables used for the study. When the dataset is analyzed, the percentage of households spending on road transport is seen to be higher than other forms of transport while the amount of maritime expenditure is seen to be the least. The household data is available more for those who belong to the middle and lower level occupational groups. The 2003 data is the first of the dataset, and the other years are nearly equal.

As it is in table 1, more than 6,000 households are surveyed each year on average. For a total of 15 years, there are 136,320 observations in the model as a big micro data. There are about 2000 households who have airway expenses in the surveyed month. The data are secondary data collected by Turkstat by a stratified random sampling method by monthly household expenditure. It is noteworthy that different households are included in the dataset for each year. Since a household could spend on air, road, or rail transportation in the same month, it would not be appropriate to apply transportation choice mode models here. Therefore, logistic regression is employed to investigate the substitution case. A household may not have made an airline expenditure during the survey month. Since the dataset is sufficiently large, however, it can be offset and seen in other households ' expenses. As an example, the 2003 Household Budget Survey was applied to a sample of 2,160 households, that varied every month from January 1 to December 31, 2003, and a total of 19,717 households throughout the year. Household budget surveys for 2004-2017 were applied to about 6,000 or 7,000 sample households. Unfortunately, the dataset does not contain any flight frequency data. Due to the reasons mentioned earlier and the availability of data, logistic regression is preferred as the model.

According to these, the logistic regression model can be written as in Equation 3,

$$Air_{it} = B_0 + B_1 * Rail_{it} + B_2 Road_{it} + B_3 Sea_{it} + B_4 Fuel_{it} + B_5 Transportation\ Subsidy_{it} + \sum_{k=6}^{9} B_k\ Income\ Group_{it} + \sum_{k=10}^{17} B_k\ Occupation_{it} + \sum_{k=18}^{31} B_k\ Year_{it} + u_{it} \quad (3)$$

where *i* is the household dimension, *t* is the year dimension, and *k* represents the index of summation for coefficients. The logistic regression method is applied to the pooled data. Besides, the variable year is also added to the model as a dummy in terms of examining the changes in preferences over the years. We also use robust standard errors, which are more "robust" to the problem of heteroscedasticity.

These variables are also included in the decision tree algorithm, and the models are solved with the help of Stata 14 software.

6. Results

In Table 2, the share of transportation expenditures in total expenditure and shares of transportation modes in total transportation expenditure are given. When transportation expenditures are analyzed, over the 2003-2015 period, it is seen that the share of transportation expenditure in total expenditures increases from 13.28 per cent to 21.4 per cent. In this study, income groups are formed as income quintile groups in each year by dividing the number of observation into five groups equally represented by 20 per cent of households and the first quintile group represents 20 per cent of the households with the lowest income, and the fifth quintile group represents the 20 per cent of the households with the highest income. Based on income groups, it is observed that the budget shares of transportation expenditures increase in each income group. In particular, the budget share of the transportation expenditure of the highest 20 per cent income group reaches 30 per cent.

When evaluating the share of household expenditure on transportation modes in total transportation expenditure, it is seen that households generally prefer road transportation as a mode of transportation and have a high share of fuel expenditure in transportation expenditure. However, the shares of road and fuel expenditures in transportation expenditures decreased from 2003 to 2017, and road expenditures are mostly allocated to low and middle-income households with high transportation share. The shares of transport expenditures of households allocated to rail, air, and maritime routes are relatively low and decreased over time. However, the increase in the share of airline expenditures is statistically significant at a 5 per cent significance level. (Z-score =3.55). A striking finding is that households in the low and middle-income groups did not make any expenditure on air transport in 2003, whereas they spent on air transportation even with a small amount in each income group in 2017.

Table 3 shows the results of the logistic regression given in Equation 1. In this analysis, because of the missing data, the number of data decreased to 97,337. Firstly, the equation is found as significant according to the Wald test result, and Pseudo R^2 value is obtained

as 0.17. Since the dataset consists of pooling multi-year cross-sectional data, the R^2 value is low as expected.

Table 2. Expenditure shares of transportation modes by income quintiles (20 %)

2003						
	First Quintile	**Second Quintile**	**Third Quintile**	**Fourth Quintile**	**Fifth Quintile**	**Total**
Transportation	7.71	8.38	9.59	10.69	17.00	13.28
Railway Transportation	0.52	0.64	0.68	0.61	0.44	0.51
Road Transportation	70.44	68.48	63.17	55.14	26.20	38.96
Air Transportation	0.00	0.00	0.00	0.08	1.24	0.79
Maritime Transportation	0.00	0.04	0.05	0.08	0.23	0.16
Fuel	14.19	22.26	28.28	32.91	27.65	27.96
2017						
	First Quintile	**Second Quintile**	**Third Quintile**	**Forth Quintile**	**Fifth Quintile**	**Total**
Transportation	9.57	10.87	12.21	15.47	30.20	21.14
Railway Transportation	0.08	0.27	0.28	0.14	0.08	0.12
Road Transportation	58.08	58.47	50.68	34.46	9.46	20.88
Air Transportation	1.05	0.50	1.55	1.49	1.24	1.26
Maritime Transportation	0.13	0.07	0.18	0.17	0.05	0.09
Fuel	10.09	12.45	15.01	15.04	8.13	10.16

According to Table 3, it is seen that households prefer mostly air transportation as the income level increases. When it is especially compared to the lowest income level, the coefficients of other income levels are positive and have a significant effect. It is observed that preference for railway transportation expenditure has no significant effect on the preference of air transportation expenditures, whereas preferences on maritime and road transportation of households have a positive and significant effect on the preference for air transportation expenditure. There is a significant decrease in the expenditure preferences of air transportation, where households spend on fuel for their vehicles. Also, there is a significant decline in expenditure preferences on air transportation, if there is a subsidy for transportation to any individual in the household. In terms of occupation, the probability of expenditure on air transportation is expected to decrease as the level of occupation increases. The coefficients are found as significant when they are compared to the Lawmakers, Senior Executives, and Managers group. When the year variable is examined, it is seen that expenditure preferences on air transportation have increased since 2003. Notably, we tested the effect of years on household expenditure choice preferences in recent years, 2012-2017, by using t-tests and F tests. We found that there were no significant differences in the years between 2012 and 2015 on the expenditure preferences of households, while the effects of the years 2016 and 2017 were significantly different from them and found to be higher.

The ROC (Receiver Operating Characteristic) curve is a graphic that displays the two types of errors for all possible thresholds. The overall performance of a classifier, summarized over all possible thresholds, is given by the area under the (ROC) curve. An ideal ROC curve will hug the top left the corner, so the more significant the AUC, the better the classifier (James, Witten, Hastie, and Tibshirani, 2013). The ROC curve shown in Figure

1 shows the sensitivity and (1-specificity) values for different thresholds in the logistic regression analysis, which is given in Table 3. The area under the ROC curve is found sufficiently high as 0.85.

Table 3. Results of logistic regression

Dependent: Usage of Air Transportation	**Coefficient**	**Robust SE**	**Z**	**P**	**Odds Ratio**
Income Quintiles					
Second Quintile	1.34	0.32	4.25	0.00	3.82
Third Quintile	2.01	0.30	6.61	0.00	7.46
Forth Quintile	2.57	0.30	8.59	0.00	13.07
Fifth Quintile	3.51	0.30	11.8	0.00	33.45
Transportation Usage					
Road	-0.12	0.09	-1.37	0.17	0.89
Rail	0.46	0.12	3.75	0.00	1.58
Maritime	0.44	0.14	3.06	0.00	1.55
Fuel	-0.42	0.07	-5.96	0.00	0.66
Presence of Transportation Subsidy	-0.29	0.07	-4.36	0.00	0.75
Occupation					
Professionals	-0.21	0.08	-2.74	0.01	0.81
Technicians and Assisting Professional Profession	-0.52	0.10	-5.04	0.00	0.59
Office Services	-0.78	0.13	-6.14	0.00	0.46
Service and Sales	-0.82	0.09	-8.77	0.00	0.44
Qualified Agriculture, Forestry and Water	-1.03	0.11	-9.46	0.00	0.36
Artisans and Related Works	-1.26	0.11	-11.11	0.00	0.28
Plant and Machine Operators	-1.36	0.13	-10.66	0.00	0.26
Jobs That Do Not Require Qualifications	-1.36	0.15	-9.01	0.00	0.26
Year					
2004	0.85	0.35	2.47	0.01	2.34
2005	1.40	0.30	4.62	0.00	4.06
2006	1.73	0.28	6.07	0.00	5.64
2007	1.85	0.28	6.60	0.00	6.36
2008	2.02	0.27	7.37	0.00	7.54
2009	2.33	0.26	9.00	0.00	10.28
2010	2.34	0.26	9.08	0.00	10.38
2011	2.84	0.25	11.47	0.00	17.12
2012	3.00	0.25	12.14	0.00	20.09
2013	3.17	0.24	12.95	0.00	23.81
2014	3.00	0.25	12.19	0.00	20.09
2015	3.04	0.24	12.53	0.00	20.91
2016	3.30	0.24	13.77	0.00	27.11
2017	3.22	0.24	13.38	0.00	25.03
Constant	-8.27	0.39	-21.41	0.00	0.00
LL	-6494.57				
Number of obs	97337.00				
Wald chi2(31)	2128.67				
Prob > chi2	0.00				
Pseudo R2	0.17				

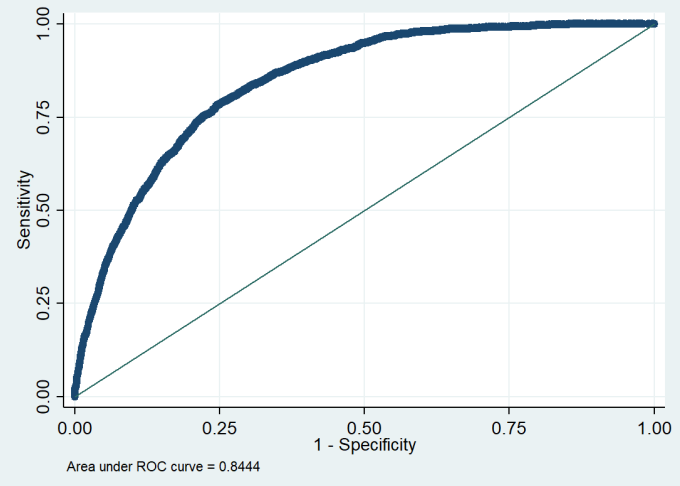

Figure 1. ROC Curve of logit estimates

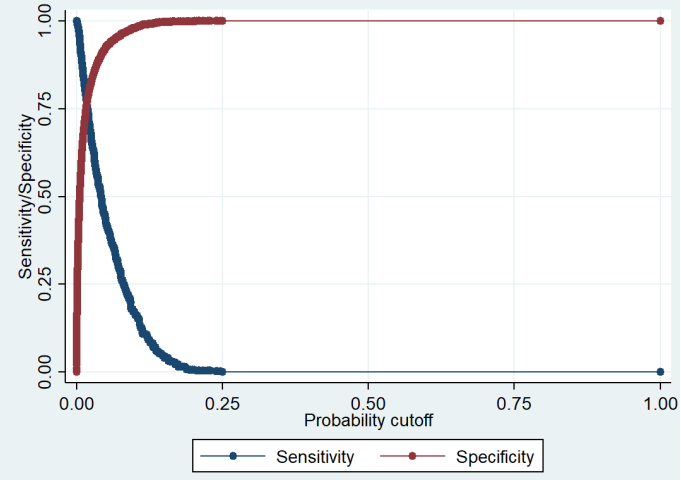

Figure 2. Sensitivity and specificity values of logit estimates

Figure 2 shows the sensitivity and specificity values for different thresholds. In our dataset, 1,513 of 97, 131 households have expenditures on air transport, so the appropriate threshold value is taken as approximately 0.01 (1,513/97,337), and the estimation success is determined according to this value.

Table 4. Classification table of logistic regression

	Original		
Classified	1	0	Total
1	1,319	34,207	35,526
0	194	61,617	61,811
Total	1,513	95,824	97,337
Classified + if predicted Pr(D) >= 0.01			
True D defined as Air != 0			
Sensitivity	Pr(+ D)	87.18 %	
Specificity	Pr(-~D)	64.30 %	
Correctly classified	64.66 %		

The classification table, in which the predictive success of logistic regression analysis is evaluated, is shown in Table 4. The important thing is to determine the correct classification

of the households that make airline spending, which is essential to achieve high sensitivity. As a result of the classification, the sensitivity of the analysis is obtained as high as 87 per cent. Although the data is large and contains many zeros, the success of the model is high. Therefore, it can be said that this model can be used for the determination of the households that spend on air transportation.

Table 5. Ten-fold cross-validation results

	MAE
Estimation 1	0.028
Estimation 2	0.029
Estimation 3	0.030
Estimation 4	0.031
Estimation 5	0.031
Estimation 6	0.030
Estimation 7	0.030
Estimation 8	0.029
Estimation 9	0.029
Estimation 10	0.028

The purpose of separating the dataset as a training and test set is to avoid possible overfitting, and to understand how the model performs on a dataset that it has not seen before. However, there may be some errors due to distribution during the training and testing phase of the model. In this study, the k-fold cross-validation technique is used to minimize these errors. Here, the training divides the dataset into random k segments. K-1 is used for training, 1 part is used for the test set, and k is repeated this time. The values obtained in each round are summed up, and the performance of the model is evaluated. That is why, after the sensitivity analysis, we use the k-fold cross-validation method to measure the cross-validation success of the model. The mean absolute values of the error terms are obtained for each fold, and they are given in Table 5. It can be seen that the values do not change according to different folds. It means that there is no overfitting problem in our model.

After obtaining the socioeconomic characteristics of the households who make expenditures on air transportation by logistic regression analysis, the decision tree method is made by using the CHAID algorithm to determine the essential variable that leads households to make an air transportation expenditure. The result of decision tree analysis is shown in Figure 3. Table 6 also shows the ranking of the most influential variables in the allocation. The most critical factor in the expenditure preference on air transportation is the income level in which they belong. It is also seen that the variable of the year has a significant effect on the preferences. Then, the occupation of the household's head has a significant impact on expenditure preference on air transportation. The household expenditure preferences on rail transportation and fuel consumption, the presence of transportation subsidy, determine their expenditure preference on air transportation relatively low. On the other hand, having maritime and road transportation expenditures does not affect classification.

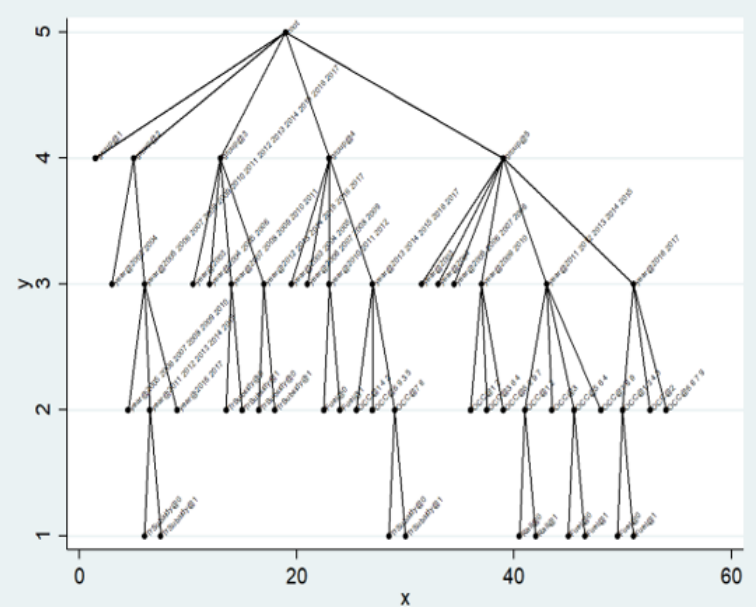

Figure 3. Decision tree results
Here, X-axis shows the number of nodes, while Y-axis shows the number of clustered stems.

Table 6. Splitting variable permutation importance

	Income Quintiles	Year
Raw	0.49	0.35
Rank	1	2
	Occupation	**Rail**
Raw	0.14	0.01
Rank	3	4
	Fuel	**Presence of Transportation Subsidy**
Raw	0.009	0.004
Rank	5	6
	Road	**Maritime**
Raw	0	0
Rank	8	8

Apart from these effects, a total of 36 clusters has occurred as a result of the decision tree analysis. According to Figure 3, for the lowest income group, it is found that there is no effect in determining the expenditure preference on air transportation of them, and they are not generally classified as having any airline expenditure. For the second 20 percent income group, the presence of transportation subsidy and the "year effect" determine their expenditure preferences on air transportation. Especially after 2005, they preferred spending on air transportation. While these two variables are also explanatory for the third 20 percent income group, the variable of "year" affects airline preference since 2003. For the fourth 20 percent income group, year and occupational variables are found as valid on the expenditure preferences on air transportation. Finally, for the top income group, the occupation of the household head, expenses on fuel, and road transportation affect the making of expenditure choices on air transportation.

7. Discussion

The domestic liberalization process in the air transportation sector took place in Turkey in 2003. In October 2003, Flyair became the first domestic airline, flying from Istanbul to Trabzon. Later, it was possible to provide services other than Turkish airlines to many destinations within the country by other airlines, which received domestic flights, mainly from Istanbul. With this development, some policies became a motto such as "an airport every 100 km" and "Every Turkish citizen will fly at least once" by the government. From 2003 to 2017, Turkey has experienced 5.8 per cent of economic growth (IMF, 2019), while the number of passengers who used airlines was 34.443 million in 2003, whereas it reached 193.577 million in 2017 (DHMI, 2019).

There are eight airline business activities for passenger flights in Turkey (SHGM, 2019). After the liberalization, the airline sector has been more competitive, some airline companies have started cross-flights with new regional aircraft. In addition to the reduction in the taxes on airline tickets, the airline companies have been allowed to determine the ticket prices with the new arrangements, and all the airlines are operating in the domestic lines. Moreover, Turkish airlines have provided lower prices on domestic flights (source the price of air tickets), and the airline companies increased their service diversification and revenues by taking advantage of the increase in traffic. But by the end of 2019, Borajet, which made regional flights and in early 2020, Atlas global announced their bankruptcy. The wrong business model of Atlas (low cost charter flights and full service; THY-like operation) brought them to bankruptcy. Other external factors accelerated this process. Besides, Pegasus, which has a low cost model, affordable prices, and basic services that can only be obtained at additional costs, announced the that it made a profit in 2019.

In this study, we examine the socioeconomic factors affecting the expenditure choice of households on air transportation in the case of Turkey and the changes in their preferences after the liberalization in the airline industry. Despite the low level of expenditure preference for air transportation, it has increased compared to other transportation modes since the liberalization. Notably, only the high-income group had air transportation expenditures in 2003, while the low-income group also had air transportation expenditures in 2017 as a result of the increase in the general income level households and lower ticket prices. The increase in competition has led to an increase in long-distance travel, obligates airlines to have technological innovations, and airlines give great importance to customer satisfaction, thereby, they increased their revenues (see Investor, 2019). As a part of the liberalization policy process, Istanbul Airport was opened to air traffic in 2018. Istanbul has the potential of being a hub for air transport due to its geographical location. It is expected that there will be an increase in both external and internal demand, and there will be 300 available destinations after all phases of the new airport are completed. However, the policy of reasonable ticket prices would attract people here in the early days of the new airport, and that would keep domestic demand alive.

Despite the advantages of situations, there are significant risks in the aviation sector. For instance, the deterioration of political relations with other countries or the deterioration in the security perception in Turkey could adversely affect the aviation industry. In

the domestic market, the weakening of the purchasing power of households due to macroeconomic conditions, the rise in the exchange rate, and the increasing oil prices may weaken the sector. On the other hand, domestic companies carry their risks, for example, the Pegasus airline company which is characterized by low cost and young fleet characteristics has high financial costs, and this poses a threat to the sector since it is on dangerous ground (see Investor, 2019). In order to hedge the financial risks such as the exchange rate, it is essential to benefit from derivatives markets more productively. Not only airlines carry risk, but also airports in Turkey, since most of them are not suitable for the expedition on wide-body aircraft, and that is an obstacle to the development of the sector.

In this study, household preferences as domestic demand in the air transportation sector are investigated, and the most critical factors are found to be the income level of household and the occupation level of the household head. Increasing the income and occupational level significantly increases the choice of air transportation. It is also observed that households are more likely to choose air transportation in proportion to the reasons mentioned earlier, such as low-price tickets on airlines, and perception on the security level of air transportation. The fact that the possibility of air transportation choice of households increases with an increase in railway and maritime transportation preferences shows that people care about other public transport facilities besides air transportation, and there is no substitution effect from the railway and maritime transportation to air transportation. In particular, the use of the ferry and subway facilities with access to airports in Istanbul is also one of the qualities that improve this situation. In Turkey, households owning a car are spending more on fuel rather than on airline travel, which indicates that they prefer their vehicles to air travel. At the same time, the presence of a transportation subsidy in households has a negative impact on households' preference for air transportation as there are subsidies for cars issued by the private sector. Therefore, the increase in income levels of households is a need to increase the likelihood of households preferring air transportation, and improvement of the railway and maritime transportation sector will also stimulate the demand for the air transportation sector in terms of access to airports. Since the most important factor in households' airline preference is income, the liberalization movement should be aimed at low-cost flights and optional services.

8. Conclusion

Following the liberalization of the airline industry, private airline firms began to compete in Turkey's domestic market. Thus, airlines have developed many improvements to assure consumer satisfaction as a result of greater competition, and as a result, service quality has improved. The domestic market in Turkey has become more diversified as people's living standards have risen, ticket policies have become more affordable, and new aircraft companies have sprung up. In this study, we focus on determinants that affect the household expenditure preferences on air transportation based on socio-economic characteristics of households and preference changes after the liberalization process in Turkey. This study contributes to the literature by using a large sample size direct household data.

We observed that only the high-income group used to make airline expenditures in 2003; however, it comprises all income groups at present. Throughout the period observed, the share of airline expenditures has increased significantly, although the share of transportation expenditures is still low. On the other hand, the other transportation modes expenditures have decreased, thereby that may lead to a further increase in airline expenditures in the years ahead.

The key factors affecting the preference of air transportation of households are found as the income group and the occupation group of the household head. The increase in income and occupational group leads households to prefer airlines. Since the most important factor in households' airline preference is income, the liberalization movement should be aimed at low-cost flights and optional services, which will make the sector more competitive. Thus, companies will continue their existence with high occupancy rates. Moreover, these effects will rebound the tourism and economic growth rates in the long term. As a striking result, an increase in railway and maritime transportation expenditure positively affect the choice of airlines without a substitution effect from these transportation modes to air transportation. However, the presence of a transportation subsidy in households and spending on fuel for their vehicles reduces the likelihood of air transportation preference.

An increase in household income level would put them in a "can afford" position for air transport, and that would lead a resistant aviation sector to the shocks of oil prices and currency. Additionally, reductions in the transportation tax will drive households to prefer air transportation leading to a growing domestic demand for the sector. Considering that, not only the studies at the micro-level but also studies at the macro level are necessary for further investigation. It should be considered that liberalization in aviation will promote growth and domestic tourism, as well as that increased air traffic will have negative consequences for CO_2 emissions. On the other hand, following a series of disasters, our world was eventually confronted with Covid-19, which first appeared in Wuhan, China, in December 2019. The World Health Organization designated the disease as a pandemic after it quickly spread globally. It has wreaked havoc on the global economy. The aviation industry felt the brunt of the catastrophe the most during a time when most of the aviation sector experienced significant financial losses and the number of passengers carried worldwide. Therefore, in future studies, it should be observed how the effect of Covid-19 changes household transportation preferences by using up-to-date data.

Informed Consent: Written consent was obtained from the participants.
Peer Review: Externally peer-reviewed.
Author Contributions: Author Contributions: Conception/Design of Study- G.K.A.; Data Acquisition- G.K.A.; Data Analysis/Interpretation- G.K.A., H.B., U.A., B.P.; Drafting Manuscript- H.B., G.K.A.; Critical Revision of Manuscript- U.A., B.P.; Final Approval and Accountability- G.K.A., H.B., U.A., B.P.
Conflict of Interest: Authors declared no conflict of interest.
Financial Disclosure: Authors declared no financial support.

References

Alperovich, G., & Machnes, Y. (1994). The role of wealth in the demand for international air travel. Journal of transport economics and policy, 163-173.

Aytekin, Ç., Sütçü, C. S., & Özfidan, U. (2018). Karar ağacı algoritmasi ile metin sınıflandırma: müşteri yorumlari örneği. Journal of International Social Research, 11(55).

Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. Journal of Transport Geography, 19(4), 644-657.

Çelikkol, E. S., Uçkun, C. G., Tekin, V. N., & Çelikkol, Ş. (2012). A research on factors that have effect on satisfaction and preferences of customers with regard to airline transport within domestic flights in Turkey. İşletme Araştırma Dergisi, 4(3), 70-81.

Chee, W. L., & Fernandez, J. L. (2013). Factors that influence the choice of mode of transport in Penang: A preliminary analysis. Procedia-Social and Behavioral Sciences, 91, 120-127.

Chiou, Y. C., & Chen, Y. H. (2010). Factors influencing the intentions of passengers regarding full service and low cost carriers: A note. Journal of Air Transport Management, 16(4), 226-228.

DHMI (2019, January 1). Passenger statistics. Retrieved from https.//www.dhmi.gov.tr/Lists/IstatistikList/Attachments/314/YOLCU.pdf.

Fourie, C., & Lubbe, B. (2006). Determinants of selection of full-service airlines and low-cost carriers—A note on business travelers in South Africa. Journal of Air Transport Management, 12(2), 98-102.

Gerede, E. (2015). *Havayolu Taşımacılığı Ve Ekonomik Düzenlemeler Teori Ve Türkiye Uygulaması* [Air freight and economic regulations theory and Turkey practice]. Art Ofset Matbaacılık, Ankara.

Gillen, D. W., Morrison, W. G., & Stewart, C. (2003). Air travel demand elasticities: Concepts, issues and measurement. Department of Finance, Government of Canada, 2, 365-410.

Gujarati, D. N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.

IATA (2022, March 19). Air passenger numbers to recover in 2024. Retrieved from https://www.iata.org/en/pressroom/2022-releases/2022-03-01-01/#:~:text=In%202022%2C%20passenger%20numbers%20will,%25)%2C%20ahead%20of%20other%20regions.

IMF (2020, January 23). Data mapper. Retrieved from https.//www.imf.org/external/datamapper/NGDP_RPCH@WEO/TUR.

Investor (2020, January 23). Financial statements. Retrieved from https.//investor.turkishairlines.com/documents/ThyInvestorRelations/2018_12_Months_Financial_Statements_-_TL.PDF.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 112, p. 18). New York: springer.

Janzen, F. J., & Stern, H. S. (1998). *Logistic regression for empirical studies of multivariate selection*. Evolution, 52(6), 1564-1571.

Lee, J. K., Yoo, K. E., & Song, K. H. (2016). A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the Seoul-Jeju route. Journal of Air Transport Management, 56, 131-137.

Milioti, C. P., Karlaftis, M. G., & Akkogiounoglou, E. (2015). Traveler perceptions and airline choice: A multivariate probit approach. Journal of Air Transport Management, 49, 46-52.

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Academic press.

Ryley, T., & Davison, L. (2008). UK air travel preferences: Evidence from an East Midlands household survey. Journal of Air Transport Management, 14(1), 43-46.

Saygili, H., & Turkcan, K. (2018). Transportation mode choice and international fragmentation of production: Evidence from a developing country. Working Papers 1813, Research and Monetary Policy Department, Central Bank of the Republic of Turkey.

SHGM (2019, January 2). Statistics of aviation sector in Turkey. Retrieved from http.//web.shgm.gov.tr/tr/kurumsal/4547-istatistikler.

Soltanzadeh, H., & Masoumi, H. E. (2014). The determinants of transportation mode choice in the middle eastern cities: The Kerman case, Iran. TeMA-Journal of Land Use, Mobility and Environment, 7(2), 199-222.

Taplin, J. H. (1997). A generalised decomposition of travel-related demand elasticities into choice and generation components. Journal of Transport Economics and Policy, 183-191.

Turkstat (2021, September 11). Tables of Household Budget Survey Micro Data Set. Retrieved from https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Tuketim-Harcamasi-2017-27840.

Turkstat (2019, January 12). Air transportation statistics. Retrieved from http.//www.tuik.gov.tr/PreTablo.do?alt_id=1051.

Ukpere, W. I., Stephens, M. S., Ikeogu, C. C., Ibe, C. C., & Akpan, E. O. (2012). Determinants of airline choice-making: The Nigerian perspective. African Journal of Business Management, 6(15), 5442-5455.

UNDP (2021, March 19). Logistics sector analysis report and guide tr31 region (Izmir), Ministry of Industry and Technology Development Agencies General Directorate, 2021, Ankara.section

Vieira, J., Câmara, G., Silva, F., & Santos, C. (2019). Airline choice and tourism growth in the Azores. Journal of Air Transport Management, 77, 1-6.

Yaylali, M., & Dilek, Ö. (2017). Havayolu ulaşımında tüketici tercihlerine etki eden faktörlerin belirlenmesi. Uluslararası Ekonomi İşletme ve Politika Dergisi, 1(2), 75-88.

Yaylalı, M., Dilek, Ö., & Çelik, A. (2015). Türkiye hava ulaşımında tüketici tercihinin çok durumlu logit modeli analizi. Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, 37(1), 261-288.

Zenina, N., & Borisov, A. (2011). Transportation mode choice analysis based on classification methods. Computer Science (1407-7493), 50.

Zhou, H., Xia, J., Norman, R., Hughes, B., Nikolova, G., Kelobonye, K., ... & Falkmer, T. (2019). Do air passengers behave differently to other regional travellers?: A travel mode choice model investigation. Journal of Air Transport Management, 79, 101682.