Estimation of the elasticity of flight ticket prices to fuel prices

Statistical Consulting 2023/2024

by:

Jade Arpaliangeas, Alexis Bouteloup, Tadandjoa Kolani, Shen Shixuan

Supervised by: Stéphane Malkawi (Airbus) and Michel Simioni (TSE)

Abstract

This study investigates the elasticity of airline ticket prices concerning fuel prices. Our analysis focuses on two databases: one provides a global perspective on the effect of fuel price variations, while the other concentrates on the USA, allowing us to differentiate between short and long-term elasticity and periods of fuel price increases and decreases. We utilize panel data models such as Fixed Effects OLS, Panel Within model, as well as dynamic models such as ARDL and an attempted implementation of PanelVar. The models were separately created for the two databases; consequently, the modeling and variable selection processes will differ. We conclude with a positive elasticity ranging between 0.06% and 0.42% depending on the models and cases considered for the global database, and between 0.05% and 0.10% for the US database.

Table of Contents

Introduction

In a context of dwindling fossil energy resources and growing concerns regarding greenhouse gas emissions, it is crucial to examine the consequences of these issues on the transportation sector. Specifically, we have focused on the impact of fuel prices in the aviation industry. With increasing discussions surrounding kerosene taxes in the context of ecological transition, we seek to understand the extent to which an increase or introduction of fuel taxes on aircrafts would affect the prices of tickets paid by consumers. Thus, our aim is to measure the impact of fuel prices on airline ticket prices, via the fuel price-ticket price elasticity.

To achieve this, we have gathered two databases concerning ticket prices of major airlines in the USA and worldwide over recent decades. This enables us to better understand the factors influencing flight prices, differentiating the impact of kerosene prices from other variables of interest cited in the literature, such as business model, level of competitiveness, GDP per capita, and demand. We also focus on the temporal dimension of certain variables' impact. For instance, hedging practices by airlines regarding fuel necessitate the consideration of delayed market values of fuel to analyze ticket price variations. Additionally, we will study possible differences in elasticity during periods of fuel price decreases and increases, based on US airline data.

Our study begins with a review of existing literature on fuel price-ticket price elasticity, then proceeds with two main parts, each exploiting the information from one of the two databases, based on the relevant analyses for each. The first database provides an overview of the subject and estimates global parameters, while the second enables us to address questions of elasticity dependence on market trends (upward or downward), and to calculate short and long-term elasticities. Finally, we summarize our results to compare them both between the two databases and with existing literature.

Literature Review

Research on the impact of fuel price fluctuations on airline revenues or ticket fares remains underdeveloped. Most studies focus on demand, particularly on estimates of price or income elasticities of demand. However, the approaches taken to study this effect vary, ranging from forecasting (Bebonchu Atems et al. (2019)) to comparative analysis between airline business segments (Barbara Gaudenzi and Alessandro Bucciol (2016)), or to a market-level analysis (IATA, 2007). The studies also differ according to the region studied, scale, period, and type of data used. Furthermore, the measurement of the dependent variable varies from one study to another, and most studies have not estimated elasticities, focusing instead on level regressions. To our knowledge, only the study by Alexandre H. Wolter, Thorsten Ehlers et al. (2021) has investigated the impact of fuel price variations on airline revenues, but the estimated parameters are not elasticities. Thus, a summary of the characteristics of previous studies addressing this topic is presented in Table 1 to provide an overview of the existing literature.

Table 1: Summary of the Literature Review on Airfare Elasticity or Unit Revenue Elasticity

Auteurs/Article	**Données**	**Élasticités estimés (EE) / paramètres estimés (PE)**	**Variable dépendante**	**Variables indépendantes**	**Méthode économétrique**	**Résultats de l'étude**
Alexandre H. Wolter, Thorsten Ehlers et al (2021)	Données USA mensuelles et trimestrielles sur la période 1990 à 2019	PE (0, 27 à 0,51)	Le revenu moyen par unité de carburant (RUOF) brûlée.	Prix moyen décalé du carburéacteur, Taux de chômage,lnASK , Baggage_fee, SLF , Efficacité (litre de carburant pour 100 sièges-kilomètres), LCC_share, quarter_dummies et fixes time-period effect, binaire de rupture structurelle entre 2014 et 2015	OLS robustes (hétéroscédasticité et autocorrélation (HAC) erreurs-types Newy-West robustes)	Les résultats suggèrent que l'ampleur de la répercussion d'une augmentation exogène des prix du carburant sur les revenus dépendra de la situation concurrentielle spécifique à chaque compagnie aérienne..

Bebonchu Atems et al (2018)	Données trimestrielles de 2000 à 2016 provenant de plusieurs sources (Industrie de l'aviation, FRED, statistiques sur les transports, US-EIA)	Pas d'élasticité s estimés	tarifs aériens	prix du carburant, indice de prix à la consommation, embarquements, revenus (RPM), sièges disponibles en miles, vols effectués, coefficient de remplissage, données sur les employés	AR (benchmark) Bivariate VAR SVAR	Les prix du carburant ont une capacité de prévision limitée pour les tarifs aériens et la demande de voyages, avec des améliorations minimes par rapport au modèle de référence.
Dimas Putra Pamungkas Suhadak (2017)	Données secondaires annuelles sur 2006-2015 provenant des sites officiels de PT. Garuda Indonesia, Tbk. (Indonésie), Air India, Ltd. (Inde), Air China, Ltd. (Chine), BM et US-EIA).	PE (PT. Garuda Indonesia, Tbk. (- 0,219) Air India,Ltd. (0,775) Air China,Ltd. (- 0,492))	Marge bénéficiaire nette (NPM)	Jet Fuel Price, Exchange Rate , GDP, Inflation	OLS par Airlines	Effet différentiel du Jet fuel Price sur les profits selon les Airlines

Werner D. Kristjanpoller, Diego Concha (2016)	Données sur les clôtures d'actions journalières de 56 arilines sur la période 2008-2013 et les données US-EIA.	PE (−0,2708 à 0,2415)	Rendement de la compagnie aérienne	Variation du prix du carburant (WTI ou Jet Fuel Price), variables muettes du marché et du carburant créées, le rendement orthogonal du marché, le rendement décalé du carburant, indicateur de sensibilité du stock en présence de changements dans le prix du carburant	Modèles dérivés du Capital Asset Pricing Model (CAPM) contrôlant l'hétéroscédacité	La variation des prix du carburant a un impact sur les prix des actions des compagnies aériennes au quotidien, avec une réaction principalement positive. L'asymétrie dans la réaction aux variations des prix du carburant est observée dans quelques compagnies aériennes, avec une réaction plus forte à la baisse des prix qu'à la hausse, tandis que les variations de prix retardées ont un effet mineur sur les prix des actions.
Barbara Gaudenzi and Alessandro Bucciol (2016)	Bases de données financières telles que Freelunch et du MIT Airline Data Project, ainsi des prix du carburant de l'EIA.	PE (−0,270 à – 0,193)	Rendements journaliers des actions de compagnies aériennes	Jet fuel price, rendement du marché, coût total, rendement des actions, coût total, log (RPM) Log (ASM)	OLS à effets fixes individuelles	Les compagnies aériennes à bas coûts semblent mieux gérer l'impact des variations du prix du carburant sur leurs rendements par rapport aux compagnies régulières.

Woraphon Wattanatorn, Termkiat Kanchanapoom (2012)	Données trimestrielles de 2006 à 2010 provenant de SETSMART et de l'EIA.	PE (0,0177 à 0,0702)	Profitabilité du Secteur/Industrie : Mesurée à l'aide du retour sur actif (ROA) pour chaque industrie et secteur analysé	Jet fuel price, Taux d'intérêt politique, Taux de change, log (actif total)	Régression sur données de panel : GLS, *modèles à effets aléatoires et à effets fixes*	Les prix du pétrole brut influencent positivement la performance comptable des secteurs de l'énergie, de l'alimentation et dans une moindre mesure, du secteur pétrochimique. Pas d'effet pour le secteur de transport.
Estimating Air Travel Demand Elasticities (IATA, 2007)	Données trimestrielles sur la période de 1994 à 2005 provenant de : U.S. DB1A data (U.S. domestic), PaxIS (worldwide traffic), International Passenger Survey (UK outbound traffic).	Pas d'élasticités estimés	Trafic de passagers (en volume)	Tarif moyen (classe éco ou loisirs), niveaux de revenu (PIB), les niveaux de population, la distance parcourue, la saisonnalité	OLS, 2SLS et ARDL (décalage 1 juste pour la variable y, pas de décalage pour les variables de contrôles et Jet fuel price)	Les augmentations générales des tarifs aériens semblent être inélastiques au prix du carburant pour avions.

Research on the impact of fuel prices on airline profitability remains underdeveloped, with varied approaches and heterogeneous data. Despite previous studies, specific elasticities remain largely unestimated. Our study addresses this gap by providing new insights into the impact of fuel price fluctuations on airline revenues or ticket prices, emphasizing the importance of thorough analysis to understand the underlying mechanisms and guide strategic decisions in the airline industry. Furthermore, we estimate short-term and long-term elasticities and adopt a more methodological approach to analyze the robustness of the results to econometric specification.

MUST

Data description

The initial data comprises 1820 observations about 230 of the most important airlines for the period 2010-2019 and 58 variables, from which we select a subset of 830 observations according to the selection procedure described below.

For reliability purposes, and in agreement with the Airbus expert Mr.Malkawi we decided to directly remove the following airlines from our analysis (representing 41 observations) : rossiya airlines, yemenia, viva air colombia, iraqi airways, star flyer, yeti airlines.

The data does not contain duplicates. Only 193 observations do not present missing values for any of the 58 variables.
NB: We did not exactly base our observations/airlines cleaning on the selected variables (even if mostly) in the table above, as this choice of variables is consecutive to our primary data modelisation, that intervened after the cleaning.

Data Cleaning

- Missing years for airlines

Throughout the period 2010 - 2019, some airlines were created or disappeared. Thus, we can observe missing years in the data for some airlines. These missing values are not problematic. However, 8 airlines (11 if we consider the ones previously removed) have missing years inside during their lifetime, resulting in 'holes' in the data. Those airlines are: Transavia, Virgin Australia Regional Airlines, Air Arabia Jordan, Blue Air, Tigerair Australia, Cebgo, Cathay Dragon, Vlm Airlines. This could bring issues during our upcoming panel analysis. We first thought of imputing the whole missing observations but faced several difficulties, mostly to the nature of our data (panel data). Only a few documented procedures were found for this purpose but were not easily implementable. We eventually decided to drop the up-cited airlines from our database (54 observations).

- Missing values for observed data

The data presents a lot of missing values for the observed years, concerning variables stemming especially from financial reports from the airlines. Therefore, the interest variable *Yield* displays 800 missing values, which represents 46% of the remaining 1725 observations. Thus, we seriously considered using imputation methods to fill in the data.

VARIABLES	DESCRIPTION	STATUS	NAME USED
YEAR	Year of the observation	Used for the fixed effect model	YEAR
AIRLINE_ID	Unique Identifier of the Airline	Used for the fixed effect model	AIRLINE_ID
REGION	Continent of the airline	Used as control variable	Dummy for each Region. EUROPE is the reference
LH_RATIO	Long Haul Ratio. Ratio of Long haul to all flights	Used as control variable in percentage form	LH_RATIO_100
CTRL_TYPE	Ownership : Government, airline, ??	Used as control variable	Dummy for Airline(ref)/ Government/ Private investor
MKT_CONCENTRATION	Competition level O&D(origin and destination)	Used as control variable in percentage form form	MKT_CONCENTRATION_100
LOW_COST_FIN	Low cost score (converted to 0/1)	Used as control variable	IS_LOW_COST : 1 if low cost
FUEL_COSTS	Total fuel costs of the airline over the year	Has been overshadowed by the focus FUEL_COST_ASK	
CASK	Total costs per ASK	Replaced by a transformation	
CASK_SLA	Total costs per ASK (Stage Length Adjusted)		
ASK_m_final	Number of available seats * Flight distance(km)	Used as control variable in logarithm form	log_ASK
RPKs_m	Number of passengers * Flight distance(km)	Has been overshadowed by the focus ASK_m_final	
YIELD	Total Revenue per RPK ('straight' line distance)	Has been overshadowed by the focus UTKT_PRICE	
YIELD_SLA	Total Revenue per RPK (Stage Length Adjusted)		
UTKT_PRICE	Total Revenue per RPK (this distance takes into account flight connections)	Used as dependent variable in logartithm form, preferred to YIELD	log_UTKT_PRICE
FUEL_COSTS_SLA	Total fuel costs of the airline (SLA)		
FUEL_COSTS_ASK	Total fuel costs of the airline per ASK	Used as main independent variable in the first model in logartithm form	log_FUEL_COSTS_ASK
FUEL_COSTS_ASK_SLA	Total fuel costs of the airline per ASK (SLA)		
LOAD_FACTOR	RPK to ASK ratio	Used as control variable in percentage form	LOAD_FACTOR_100
ASK_REGION	Total ASK of the Region		
RPK_REGION	Total RPK of the Region		
MKT_CONCENTRATION	Competition level O&D based	Used as control variable in percentage form	MKT_CONCENTRATION_100
HHI_REGION	Competition level global RPK based - Region	Has been overshadowed by the focus MKT_CONCENTRATION	
FUEL_JET_GULF	Jet Fuel Price (Market value) in $ per unit (barrel)	Used as main independent variable in the second model in logartithm form	log_FUEL_JET_GULF
FUEL_CRUDE_AVG	Crude oil price	Has been overshadowed by the focus on FUEL_JET_GULF	
CPI_REGION	Consumer price index in the Region	Has been overshadowed by the focus on GDP	
GDPpCAPITA_CURRENT_REGION	GDP per Capita in the Region	Used as control variable in logarithm form	log_GDPpCAPITA_CURRENT_REGION
COST_management	All cost that do not depend on fuel per ASK. Is egal to CASK - FUEL_COST_ASK	Used as control variable in logarithm form	log_COST_management

Table 1 : variables for must

```
Missing values per variable of interest:
YIELD                  800
AIRLINE_ID               0
YEAR                     0
COUNTRY                  0
CURRENT_AIRLINE_NAME     0
FUEL_COSTS_SLA         588
FUEL_COSTS             588
FUEL_CRUDE_AVG           0
CASK_SLA               588
CASK                   588
REGION                   0
HHI_REGION             588
MKT_SHARE_RPK_REGION   800
RPKs_m                 800
CTRL_TYPE                0
ASK_m_final            588
UTKT_PRICE              31
LOAD_FACTOR            800
LH_RATIO                 0
```

However, most of the data is missing in blocks (sorted by ascending year), meaning that a missing value is hardly ever surrounded by non-missing values for adjacent years (this solution would facilitate interpolation). On the contrary, a block of existing values for a variable is usually followed by a block of missing values. The data is clearly not missing at random, and this prevents us from implementing a reliable imputation procedure, as we are not really to take the risk of damaging our database with unrealistic values. For example: we cannot predict the variable for year 2015 - 2019 for an airline if the last observed period is 2010 - 2014. We chose to remove the observations for which some of the selected variables are missing. Fortunately, as the data does not alternate between missing and nonmissing, we do not face the same gap issue as in the previous point. We end up with 922 rows. No missing year inside existing data was left by this suppression.

Only 5 airlines (54 observations) present some missing values for at least 1 variable that are chronologically surrounded by nonmissing values. We dropped those airlines too as we did not believe it was worth it to set up an imputation procedure for only a few rows: viva aerobus, virgin atlantic airways, interjet, egyptair, vanilla air.

- Airlines with only 1 year observed

The following airlines have to be removed from the database as it would be too risky to impute data while we only have 1 reference year because of the panel methods we will use. We would almost invent the data. We remove 6 observations. (916 left)

AIRLINE_ID	YEAR	COUNTRY	CURRENT_AIRLINE_NAME
12861	2010	United States	Continental Airlines
74	2017	United Kingdom	Thomas Cook Airlines
14253	2018	Germany	TUI AG
271	2019	Croatia	CROATIA AIRLINES
6429	2017	Russia	POBEDA
129	2019	Russia	URAL AIRLINES

- Outliers

Econometric modelisations are very often sensitive to outliers, it is thus important to identify outliers in our database in order to ensure the coherence and robustness of our estimates. We tested several procedures for the detection of anomalies in the data, from InterQuartile Range (IQR) measures to Machine Learning algorithms such as IsolationForest or Local Outlier Factor. However, these methods are unfortunately not interpretable and tend to identify different observations as outliers. In addition, those procedures were not readily applicable to panel data. Indeed, we would rather like to spot the outliers among airlines, and not solely among observations (1 year for 1 airline). Indeed, suppressing only 1 observation in the middle of existence period of an airline would be problematic for the modelisation part as stated in the previous section. We eventually decided to use more conventional methods - mostly relying on IQR - to detect outliers. Our final database is composed of 830 observations about 101 airlines.

Preliminary analysis and variable choice

Based on the literature and our reasoning, we have selected a set of relevant variables for our models. The correlation matrix associated with these variables is as follows:

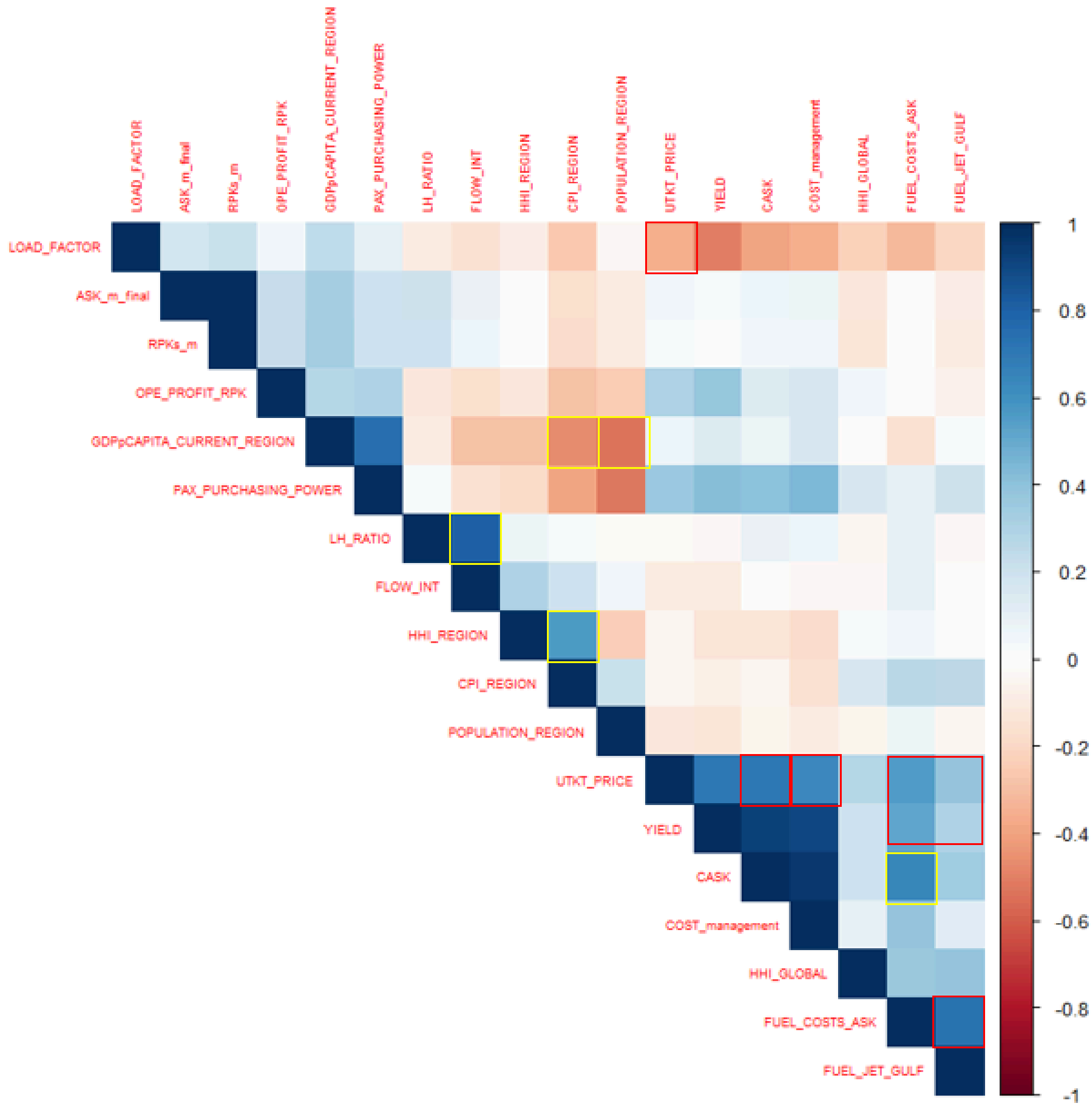


Table 2: Correlation matrix between variables

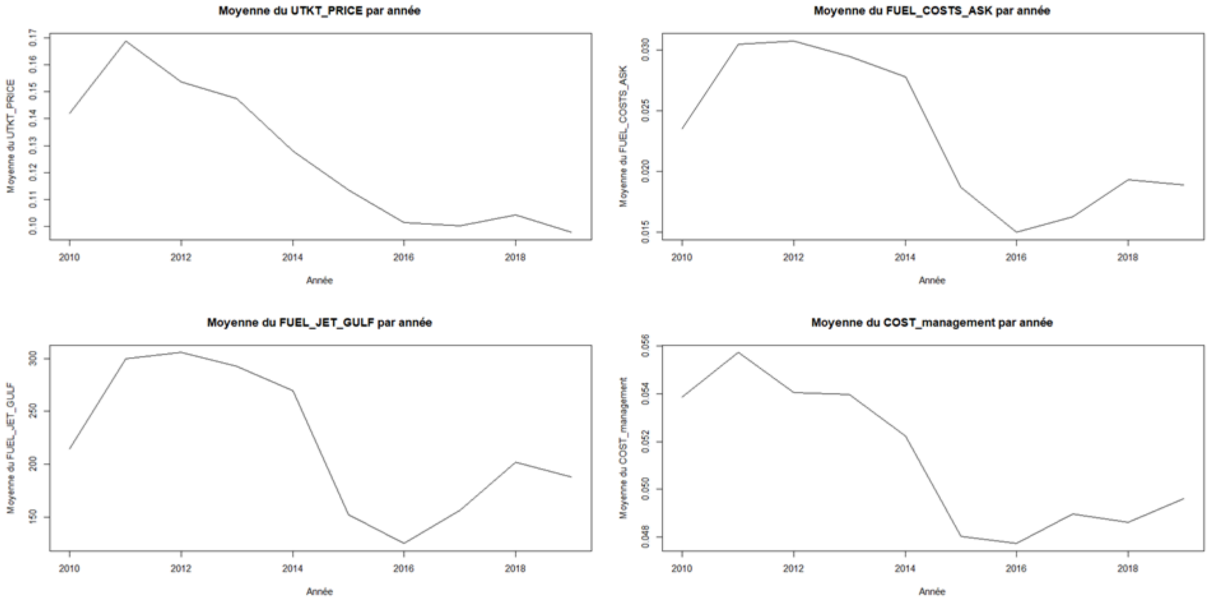
As assumed, the variables FUEL_COST_ASK and FUEL_JET_GOLF are highly correlated, so we can substitute either one in our analyses. They are also both strongly correlated with UTKT_PRICE and YIELD. Both variables UTKT_PRICE and YIELD could be considered of interest, but we will only retain UTKT_PRICE as it is more pertinent due to its construction. Additionally, the correlations with the two fuel-related variables are stronger with UTKT_PRICE.

Regarding our variable of interest, UTKT_PRICE, we already note that costs—both CASK and COST_management—are strongly positively correlated with UTKT_PRICE, as well as LOAD_FACTOR, which is strongly negatively correlated.

We also observe in yellow, variables that are highly correlated with each other, which we will need to consider when choosing control variables. For example, CPI (inflation) and population are highly correlated with GDP. Therefore, we may need to retain only one of these three variables. Additionally, CASK is strongly correlated with the variable FUEL_COST_ASK, which we will need to be mindful of.

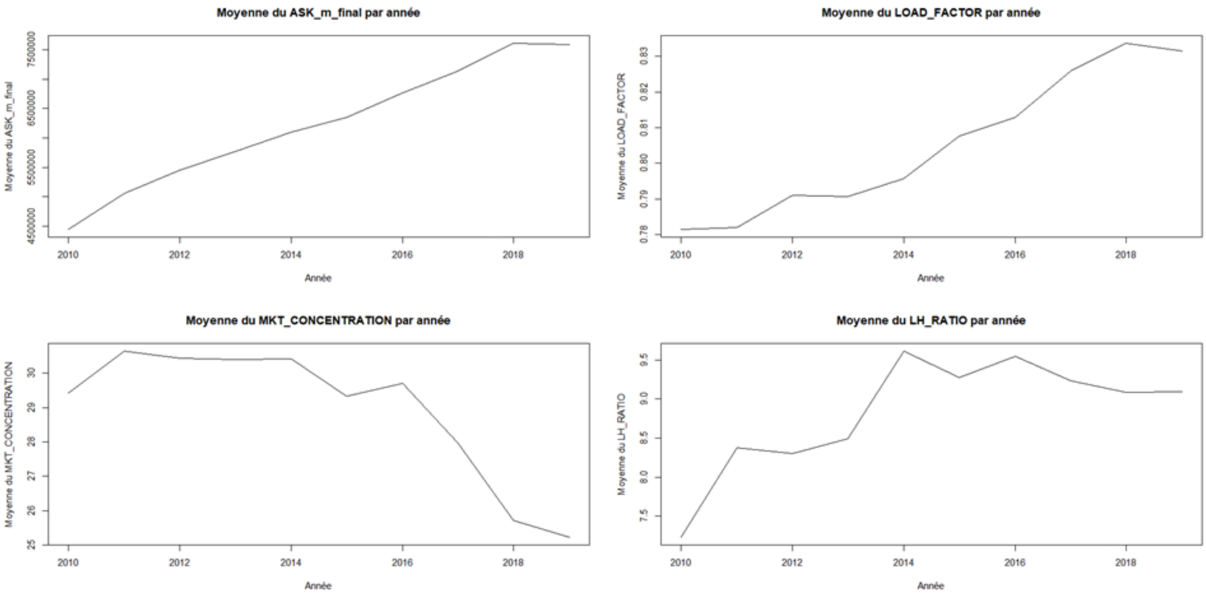
Now, let's take a closer look at some variables, particularly those mentioned earlier, and their evolution over time.

Graph 1



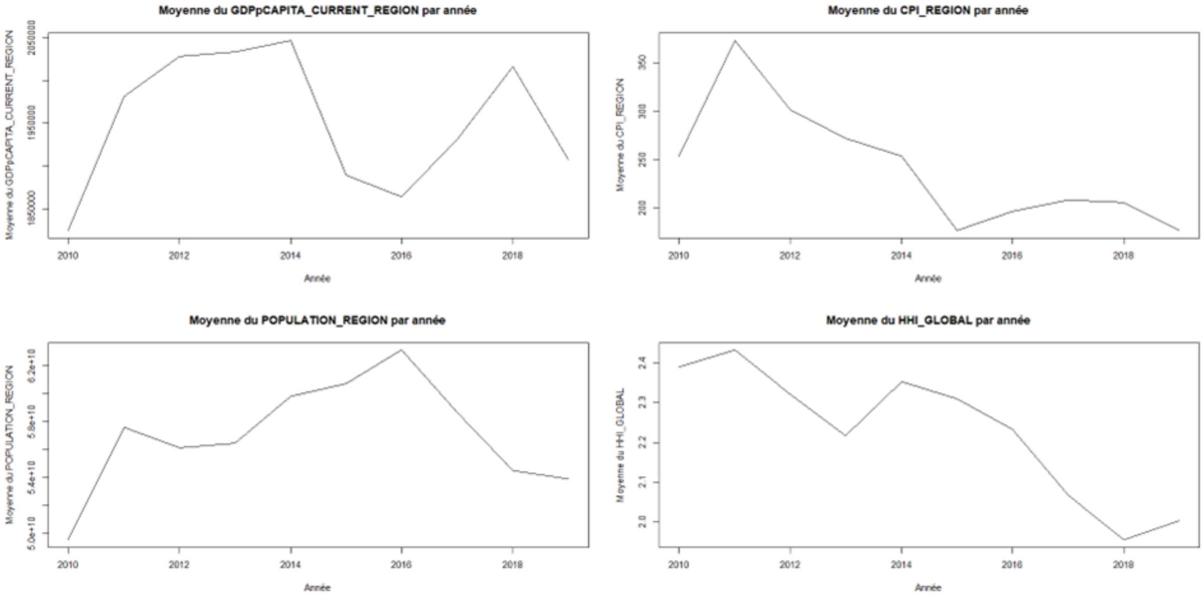
In the graphs above, we can see that these variables exhibit strongly similar trends over time, with a significant decrease in costs and revenues between 2011 and 2016.

Graph 2



In the graphs above,, we observe that over these years, airlines have greatly increased their flight capacity, as seen with ASK, and have also made significant efforts to increase aircraft load factor. Such variables will be important as controls. Additionally, we notice an increase in competition among airlines with MKT_CONCENTRATION, which likely leads to a decrease in prices and therefore UTKT_PRICE due to this growing competition. Airlines have also increased the proportion of long-haul flights, a behavior to control as airlines conducting more long-haul flights likely have different strategies in terms of cost management, especially fuel-related costs.

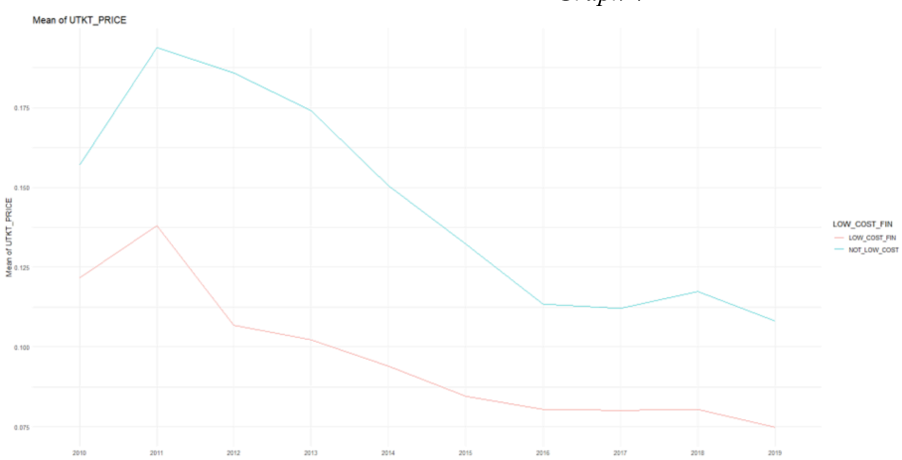
Graph 3



If we have to choose between GDP, CPI, and population to best represent the macroeconomic dynamics that could impact airlines, our choice could lean towards GDP because its behavior is the closest to FUEL_JET_GULF and UTKT_PRICE. Finally, the variable HHI exhibits a behavior similar to MKT_CONCENTRATION but is more fluctuating as we can observe in the graphics above. Therefore, we prefer to use MKT_CONCENTRATION as a representation of competitiveness.

Concerning the variable LOW_COST :

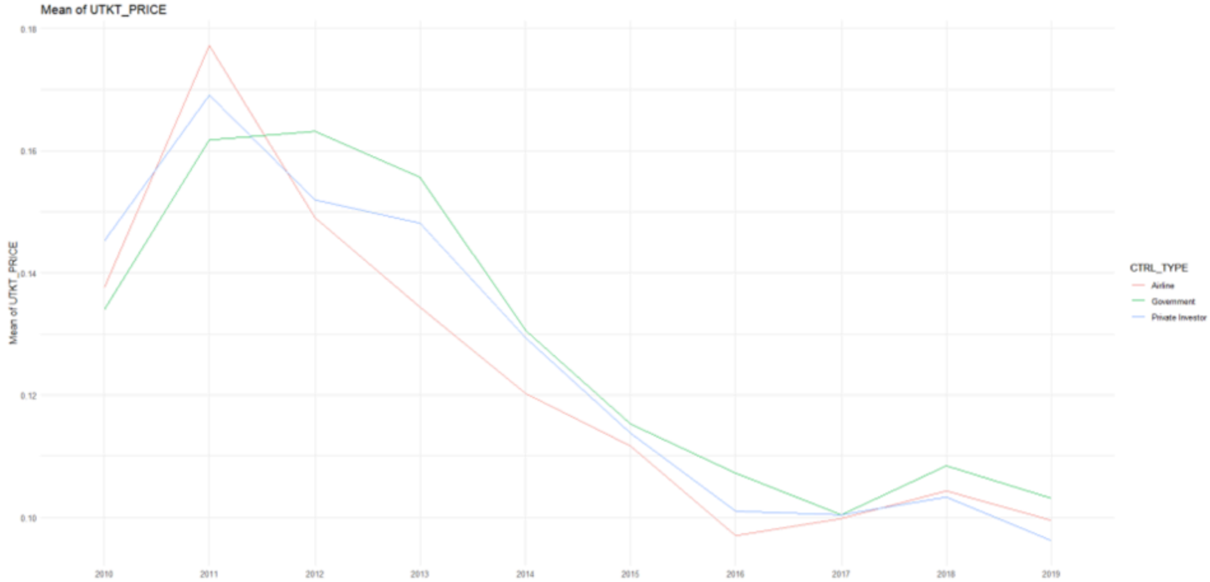
Graph 4



There is a notable disparity between low-cost and conventional airlines in terms of UTKT_PRICE. This observation is consistent with findings in the literature, suggesting the importance of controlling for this factor.

Concerning the variable CTRL_TYPE leading to three different type of Airline :

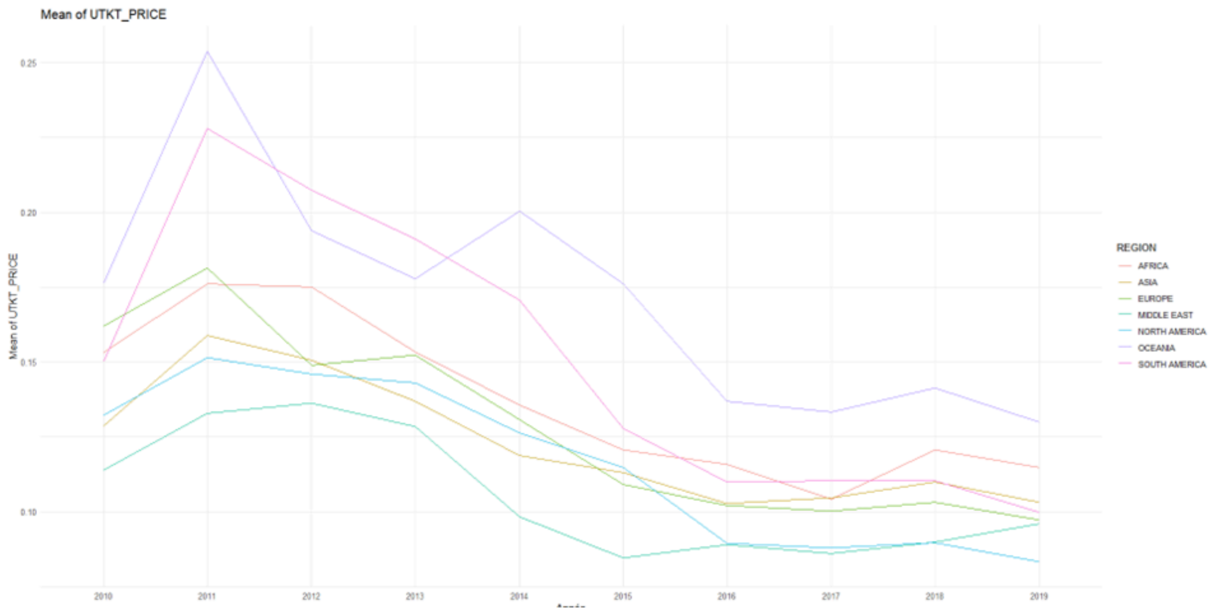
Graph 5



While not as clear-cut as for low-cost airlines, this variable could still be relevant. Indeed strategies could differ from one type to another. Even though including it in our models may result in null coefficients for the dummy variables of this variable, when combined with other control variables, it could prove to be interesting.

Concerning the variable REGION :

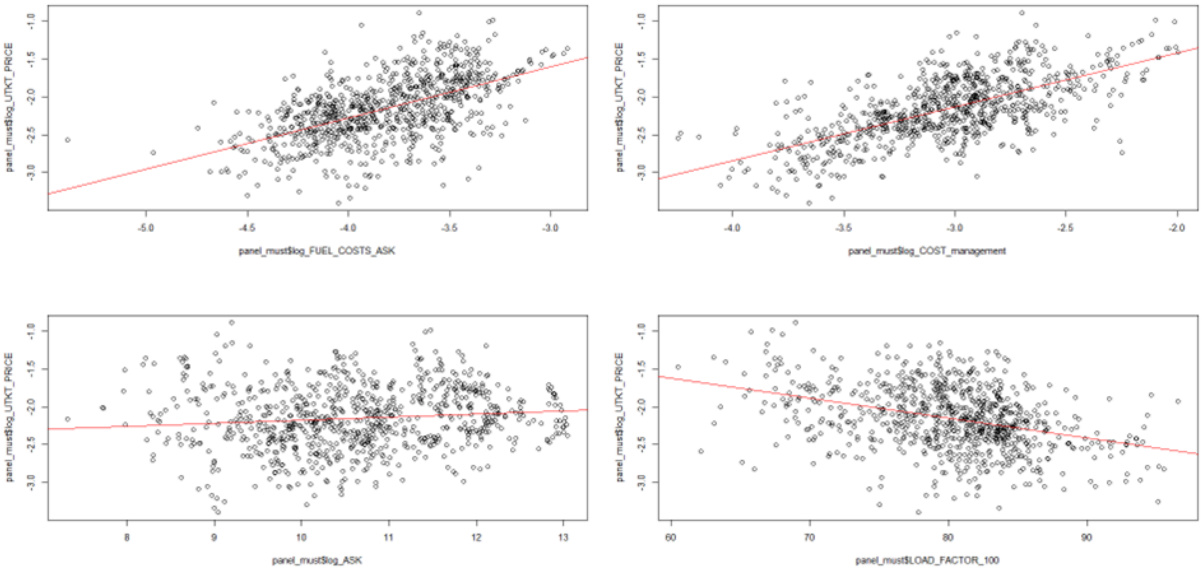
Graph 6



This variable seems highly relevant based on the graph. We expect effects based on the region in which the airline primarily conducts its business. For example, airlines in the Middle East might have more competitive fuel prices, leading to a lower UTKT_PRICE. It represents the price paid by travelers, so with lower fuel costs, an airline can lower its prices. We will control for this variable, REGION.

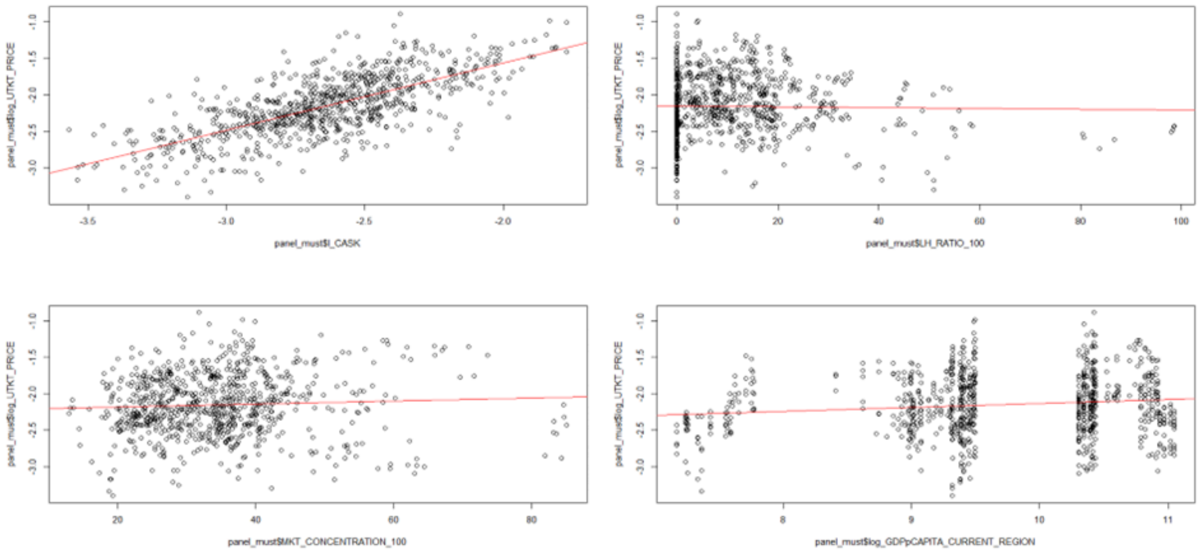
With a bivariate analysis between the logarithm of UTKT_PRICE and the logarithm of some interest variables:

Graph 7



If we examine simple regressions, FUEL_COST_ASK, COST_management, and LOAD_FACTOR appear to have an effect on UTKT_PRICE, which is not the case for ASK.

Graph 8



The impact of CASK appears comparable to that of COST_management. Upon reviewing these variables, we observe that LH_RATIO, MKT_CONCENTRATION, and GDP do not show a consistent relationship with UTKT_PRICE.

To sum up about the variables:

- We had the choice between YIELD and UTKT_PRICE, and we chose UTKT_PRICE.
- Some variables are in SLA (Length Adjusted Yield) such as CASK_SLA, but their simplified versions were retained if there was a choice.
- Regarding the variable representing macroeconomic dynamics, we opted for GDP only. This is due to strong multicollinearity between CPI and GDP.
- MKT_CONCENTRATION was preferred over HHI.

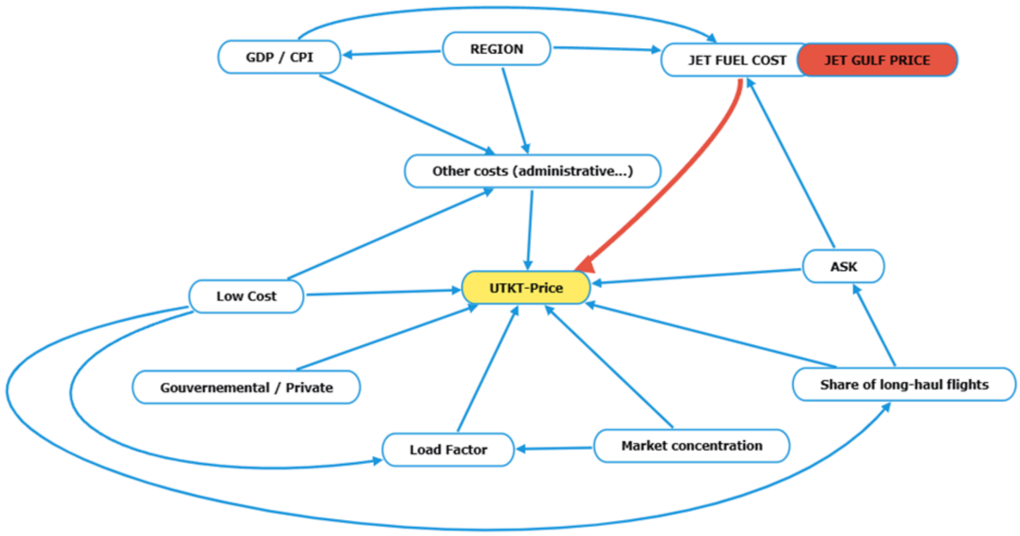
- In order to analyze the elasticity of fuel on the variable of interest (UTKT_PRICE), FUEL_COST_ASK, FUEL_JET_GULF, COST_management, and ASK were selected in logarithmic form. The other quantitative variables representing ratios were selected in their simple percentage form.
- Lastly, we need to choose a variable concerning costs between CASK and COST_management. In CASK, we include fuel cost since CASK represents all costs per ASK. If we subtract FUEL_COST_ASK to CASK to create a new variable, this new variable is almost perfectly correlated to CASK with 0.95. The correlation between UTKT_PRICE and COST_management is as good as it is with CASK (0.65 against 0.71) but we have in addition a smaller correlation between FUEL_COST_ASK and COST_management (0.39 against 0.65), which could be better to avoid multicollinearity in the models.

Table 3 : correlation

Tableau des corrélations	UTKT_PRICE	CASK	COST_management	FUEL_COST_ASK
UTKT_PRICE	1			
CASK	0.71	1		
COST_management	0.65	0.95	1	
FUEL_COST_ASK		0.65	0.39	1

The variables selected and their main links may be as follows :

Schema 1 : link between variables



Econometric analysis and models

In economics, elasticity quantifies the responsiveness or sensitivity of one economic variable to changes in another economic variable. Elasticities provide valuable insights to decision-makers about the impact of various economic actions. In our case, we are interested in the following equation for elasticity: elasticity = (% change in ticket price) / (% change in fuel price). .

We have developed several models to examine the elasticities of fuel prices on airline revenues. Our objective is to understand how fluctuations in fuel prices influence airline revenues and to predict these fluctuations for the future. To obtain elasticity, we will use the log-log transformation for both the explanatory and explained variables.

Linear models and within

We begin our analysis with panel fixed effects models and the within model. We have the following models (detailed explanations on how this best model was found can be found in the appendix):

OLS :

$$\log(UTKT\_PRICE_{it}) = \alpha_0 + \alpha_1 \log(FUEL\_COST\_ASK_{it}) +$$
$$\alpha_2 \log(COST\_MANAGEMENT_{it}) + \alpha_3 \log(ASK_{it}) + \alpha_4 LOAD\_FACTOR\_100_{it}$$
$$+\alpha_5 \log(MKT\_CONCENTRATION\_100_{it}) + \alpha_6 Lh\_Ratio\_100_{it} +$$
$$\alpha_7 \log(GDP\_CURRENT\_REGION_{it})$$
$$+\beta REGION_i + \gamma IS\_LOW\_COST_i + \delta_1 Government_i + \delta_2 Private\_investor_i +$$
$$\sum_j \mu_j \cdot AIRLINE\_ID_j + \epsilon_{it}$$

Where :

- i represents the individual (the airline company).
- t represents the time period.
- μit represents the individual fixed effect
- ϵit is the error term.

In order to get the temporal fixed effect remove $\sum_j u_j \, AIRLINEid_j$ and replace it by $\sum_j u_j \, t_j$.

Within

The within estimator is essentially ordinary least squares (OLS) applied to the transformed model as follow :

$$Y_{it} - \text{mean}(Y_i) = \beta_1(X_{it} - \text{mean}(X_i)) + \alpha(c_i - \text{mean}(c_i)) + (\epsilon_{it} - \text{mean}(\epsilon_i))$$

Where :

- Y is Log(UTKT_PRICE$_{i}$□).
- X are respectively the quantitative variables .
- ci are the qualitative variables and if c_i = mean(c_i) meaning that the attribute is constant over time then ci won't appear in the results.
- ϵit is the error term.

We expect the following results:

- A positive and significant coefficient for FUEL_COST_ASK and COST_management since they represent costs. Similarly, for MKT_CONCENTRATION, as a maximum value of 100 represents a monopolistic situation, according to economic theory, when competition decreases, prices tend to increase. Additionally, for GDP, if the wealth of the region increases, we can easily imagine a rise in prices. As for ASK, it is ambiguous; on one hand, more ASK potentially means longer flights, which entail higher costs, but on the other hand, one can consider economies of scale for airlines by increasing ASK.

- A negative and significant coefficient for LOAD_FACTOR and IS_LOW_COST. If the airline has a higher load factor, it reduces costs per passenger. And if the airline is low-cost, it is its strategy and its business model to propose lower prices leading naturally to smaller value in UTKT_PRICE. Concerning LH_RATIO, we might expect a negative coefficient because an airline with a high value in this variable is specialized in long flight and might have more economies of scale.
- For the variables REGION, GOVERNMENT, and PRIVATE_INVESTOR, we expect mixed results, especially for the latter two, as seen in a previous graph.

Results

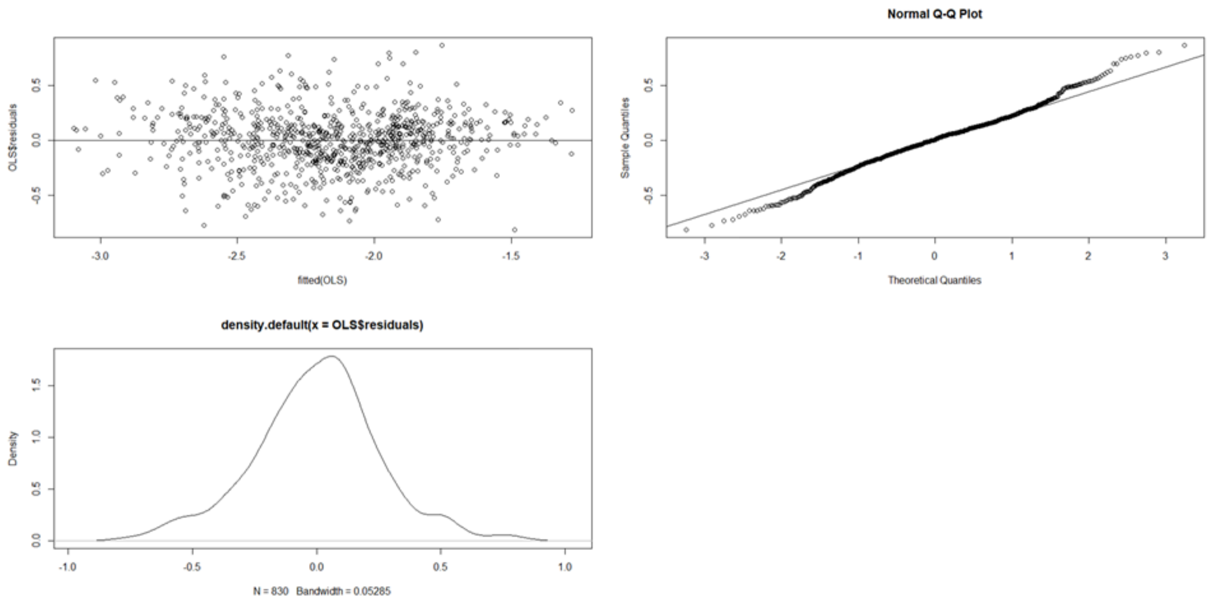
	log_UTKT_PRICE			
	OLS		panel linear	
	Inidividual	Temporal	Inidividual	Temporal
	(1)	(2)	(3)	(4)
log_FUEL_COSTS_ASK	0.424***	0.246***	0.347***	0.246***
	(0.031)	(0.053)	(0.032)	(0.053)
log_COST_management	0.364***	0.389***	0.244***	0.389***
	(0.036)	(0.035)	(0.072)	(0.035)
log_ASK	0.024*	0.035***	-0.142***	0.035***
	(0.010)	(0.010)	(0.030)	(0.010)
LOAD_FACTOR_100	-0.010***	-0.006**	-0.012***	-0.006**
	(0.002)	(0.002)	(0.003)	(0.002)
LH_RATIO_100	-0.005***	-0.005***	-0.014***	-0.005***
	(0.001)	(0.001)	(0.004)	(0.001)
MKT_CONCENTRATION_100	0.002*	0.002	0.005*	0.002
	(0.001)	(0.001)	(0.002)	(0.001)
log_GDPpCAPITA_CURRENT_REGION	0.066***	0.058***	0.192*	0.058***
	(0.016)	(0.016)	(0.081)	(0.016)
IS_LOW_COST	-0.163***	-0.211***	-0.102**	-0.211***
	(0.028)	(0.027)	(0.036)	(0.027)
BELONGS_to_GOVERNMENT	0.059	0.051		0.051
	(0.036)	(0.034)		(0.034)
BELONGS_to_PRIVATE_I	0.052	0.057*	-0.064	0.057*
	(0.027)	(0.026)	(0.075)	(0.026)
Constant	0.437	-0.409		
	(0.237)	(0.275)		
Variable omise	REGION	REGION		REGION
Variable omise	ID	YEAR	ID	YEAR
N	830	830	830	830
R2	0.618	0.656	0.450	0.584
Adjusted R2	0.610	0.645	0.367	0.571
Residual Std. Error	0.261 (df = 812)	0.249 (df = 804)		

Notes: ***Significant at the 0.1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

Table of results 1

The differences between OLS and within models with individual or time fixed effects are due to the specification of the within model, in which time-invariant variables are not taken into account.

Graph 9



Let's conduct some verification regarding the model specification and requirements. On the OLS with fixed effect model, the errors seem to be normally distributed. Moreover, from the results of the VIF function it seems that there is no multicollinearity (see appendix).

We perform the Hausman test to determine whether the specification of the model with fixed effects is valid or not on the within model with individual fixed effect :

- ❖ Null hypothesis (H0): There is no difference in the coefficients estimated between the random effects model and the fixed effects model, implying that the random effects model is consistent.
- ❖ Alternative hypothesis (H1): The coefficients estimated in the random effects model are inconsistent compared to the fixed effects model, suggesting that the random effects model is not appropriate and the fixed effects model is preferred.

We find a p-value lower than 0.05, we reject the null hypothesis and conclude that the within model is preferred over the random effect model.

We also perform the Breusch-Pagan test to check if the errors are homoscedastic on both the OLS and within model with fixed effect :

- ❖ Null hypothesis (H0): $Var(\varepsilon_i) = \sigma^2$ for all i, where Var(εi) represents the variance of the error term for observation i and σ^2 is a constant.
- ❖ Alternative hypothesis (H1): $Var(\varepsilon_i) \neq \sigma^2$ for at least one i, indicating that the variance of the error term differs across observations.

The p_value associated is lower than 0.01, we reject the null hypothesis and conclude that there is heteroskedasticity leading our coefficient to be not efficient.

Alternatively, this model shows satisfactory results with the selected control variables. We obtain significant coefficients for the fuel cost ranging from 0.24 to 0.42, which means that all else being equal, for a 1% increase in fuel cost, the expected increase in UTKT_PRICE ranges from 0.24% to 0.42%. Furthermore, low-cost airlines have an expected UTKT_PRICE lower by 0.16% to 0.21%, all else being equal. These results are consistent with the literature.

Only fuel_cost_ask is specific to each airline because it includes their hedging strategies and fleet composition. To gain a better understanding of the effect of fuel prices, we will replace fuel_cost_ask with fuel_jet_gulf, which represents the international price of aviation fuel. We hope to observe some stability in the coefficients despite this change in the variable of interest. Time fixed effects models are no longer relevant here, as the value of fuel_jet_gulf is unique for each year and does not vary among entities within the same year. However, we will retain the individual fixed effects.

Models are written as previously, but now with FUEL_JET_GULF instead.

Results

	log_UTKT_PRICE	
	OLS	panel linear
	Inidividual	Inidividual
	(1)	(2)
log_FUEL_JET_GULF	0.388***	0.305***
	(0.031)	(0.027)
log_COST_management	0.479***	0.304***
	(0.034)	(0.071)
log_ASK	0.038***	-0.162***
	(0.010)	(0.029)
LOAD_FACTOR_100	-0.009***	-0.012***
	(0.002)	(0.003)
LH_RATIO_100	-0.005***	-0.015***
	(0.001)	(0.004)
MKT_CONCENTRATION_100	0.002	0.005*
	(0.001)	(0.002)
log_GDPpCAPITA_CURRENT_REGION	0.035*	0.083
	(0.016)	(0.082)
IS_LOW_COST	-0.205***	-0.115**
	(0.029)	(0.036)
BELONGS_to_GOVERNMENT	0.034	
	(0.037)	
BELONGS_to_PRIVATE_I	0.065*	-0.101
	(0.028)	(0.074)
Constant	-2.843***	
	(0.310)	
Variable omise	REGION	
Variable omise	ID	ID
N	830	830
R2	0.604	0.455
Adjusted R2	0.595	0.372
Residual Std. Error	0.265 (df = 812)	
Notes:	***Significant at the 0.1 percent level.	
	**Significant at the 1 percent level.	
	*Significant at the 5 percent level.	

Table of results 2

Our results appear to remain stable. The variables MKT_CONCENTRATION and GDP seem to have lost significance, but the fuel variable of interest remained significant with a coefficient close to that found previously. The interpretation now is as follows: a 1% increase in fuel price in the year t leads to an increase in UTKT_PRICE in the year t of 0.305% to 0.388%, all else being equal.

To anticipate how airlines respond to fuel price changes, we examine the relationship between fuel_jet_gulf in the previous period (t-1) and UTKT_PRICE in the current period (t). Lagging fuel_jet_gulf is reasonable since airlines plan their hedging and pricing strategies in advance. They are influenced by fuel prices before making pricing decisions, attempting to forecast market dynamics.

This analytical approach accounts for potential delayed effects of fuel prices on airline revenues. It acknowledges that airlines may adjust their pricing strategies based on past fuel price fluctuations. By incorporating lagged variables, we aim to gain deeper insights into how prior fuel price variations impact current airline revenue outcomes. It seems logical and reasonable to also lag all control variables. And since we keep the same variables with the same reasoning, we are expecting results in agreement with previous ones.

Models are written as previously, but now with lagged variables instead.

Results

	log_UTKT_PRICE	
	OLS	panel linear
	Inidividual	Inidividual
	(1)	(2)
lag(log_FUEL_JET_GULF)	0.226***	0.260***
	(0.036)	(0.029)
lag(log_COST_management)	0.425***	0.128*
	(0.038)	(0.050)
lag(log_ASK)	0.036**	-0.030
	(0.012)	(0.018)
lag(LOAD_FACTOR_100)	-0.006**	-0.001
	(0.002)	(0.003)
lag(LH_RATIO_100)	-0.004***	-0.0004
	(0.001)	(0.001)
lag(MKT_CONCENTRATION_100)	0.001	0.001
	(0.001)	(0.002)
lag(log_GDPpCAPITA_CURRENT_REGION)	-0.009	-0.039
	(0.016)	(0.020)
IS_LOW_COST	-0.245***	-0.026
	(0.030)	(0.044)
BELONGS_to_GOVERNMENT	0.034	
	(0.042)	
BELONGS_to_PRIVATE_I	0.100**	0.212*
	(0.032)	(0.090)
Constant	-1.905***	
	(0.329)	
Variable omise	REGION	
Variable omise	ID	ID
N	829	829
R2	0.468	0.160
Adjusted R2	0.457	0.033
Residual Std. Error	0.308 (df = 811)	
Notes:	***Significant at the 0.1 percent level. **Significant at the 1 percent level. *Significant at the 5 percent level.	

Table of results 3

The results align closely with previous findings. We observed a decrease in significance for the lag of ASK and GDP, while MKT_CONCENTRATION remains nonsignificant, consistent with previous results. However, other variables remain stable. This model also yields satisfactory results, with a significant and positive elasticity for lagged fuel prices, affirming the importance of this factor in explaining variations in airline revenues. Specifically, a 1% increase in fuel prices is associated with an expected increase of 0.22% to 0.26% in UTKT_PRICE the following year. This model illustrates that airlines pass on fuel price increases to passengers, indicating a positive elasticity of flight ticket prices to Jet Fuel variations in our data.

As did previously, we perform the Hausman test to determine whether the specification of the model with fixed effects is valid or not. We find a p-value lower than 0.05, we reject the null hypothesis and conclude that the within model is preferred over the random effect model.
We also perform the Breusch-Pagan test to check if the errors are homoscedastic. The p_value associated is lower than 0.01, we reject the null hypothesis and conclude that there is heteroskedasticity leading our coefficient to be not efficient.

This final model is said to be dynamic because at least one of the regressors is the lagged dependent variable, as a consequence other methods might be preferable such as GMM estimation with the r function 'pgmm' or 'pvargmm' for simultaneous equations. We will explore these methods in a next part. But before going to the exploration of other methods, we have a look at 2 models with interaction.

For these 2 models, we will discuss the OLS form only.

1) Interaction term with the low_cost variable

The "low cost" attribute is one of the most important variables in the model. It is utilized in every study concerning airlines. Therefore, we will add an interaction term between FUEL_JET_GULF and IS_LOW_COST to explore a potentially different elasticity for low-cost carriers. This decision is supported by recent research conducted by Alexandre H. Wolter, Thorsten Ehlers, et al. (2021) on '*Commodity price pass-through in the US airline industry and the hidden perks of consolidation*'.

Results

	log_UTKT_PRICE	
	OLS	panel linear
	Inidividual	Inidividual
	(1)	(2)
lag(log_FUEL_JET_GULF)	0.318***	0.378***
	(0.045)	(0.035)
lag(log_COST_management)	0.415***	0.080
	(0.037)	(0.050)
lag(log_ASK)	0.034**	-0.042*
	(0.011)	(0.018)
lag(LOAD_FACTOR_100)	-0.007**	-0.001
	(0.002)	(0.003)
lag(LH_RATIO_100)	-0.004***	-0.0003
	(0.001)	(0.001)
lag(MKT_CONCENTRATION_100)	0.001	0.001
	(0.001)	(0.002)
lag(log_GDPpCAPITA_CURRENT_REGION)	-0.005	-0.029
	(0.016)	(0.020)
IS_LOW_COST	1.106**	1.758***
	(0.388)	(0.313)
BELONGS_to_GOVERNMENT	0.039	
	(0.041)	
BELONGS_to_PRIVATE_I	0.103**	0.204*
	(0.032)	(0.088)
lag(log_FUEL_JET_GULF):IS_LOW_COST	-0.253***	-0.332***
	(0.072)	(0.058)
Constant	-2.413***	
	(0.358)	
Variable omise	REGION	
Variable omise	ID	ID
N	829	829
R2	0.476	0.197
Adjusted R2	0.464	0.074
Residual Std. Error	0.306 (df = 810)	
Notes:	***Significant at the 0.1 percent level.	
	**Significant at the 1 percent level.	
	*Significant at the 5 percent level.	

Table of results 4

In the context of the interaction looking at the OLS model, the elasticity of UTKT_PRICE to fuel price is 0.318 for non-low-cost airlines and 0.318 - 0.253 = 0.065 for low-cost airlines. Each coefficient is significant, but we will now test if the combined effect (FUEL_JET_GULF | IS_LOW_COST=1) is different from 0.

For this purpose:

- Null hypothesis (H0) : α of FUEL_JET_GULF + α_2 of (FUEL_JET_GULF * IS_LOW_COST) = 0
- Alternative hypothesis (H1) : α of FUEL_JET_GULF + α_2 of (FUEL_JET_GULF * IS_LOW_COST) ≠ 0 ⇒ 0.065 ≠ 0

Test statistic : F = $((RSS_{sans\ interaction} - RSS_{avec\ interaction})/m\ )\ /\ RSS_{avec\ interaction}/\ (n - k - m)$

Where m is the number of coefficients in the interaction (1 in this case), n is the number of observations, and k is the number of coefficients estimated in the model with the interaction.

F follows a Fisher-Snedecor distribution with m and n-k-m degrees of freedom.

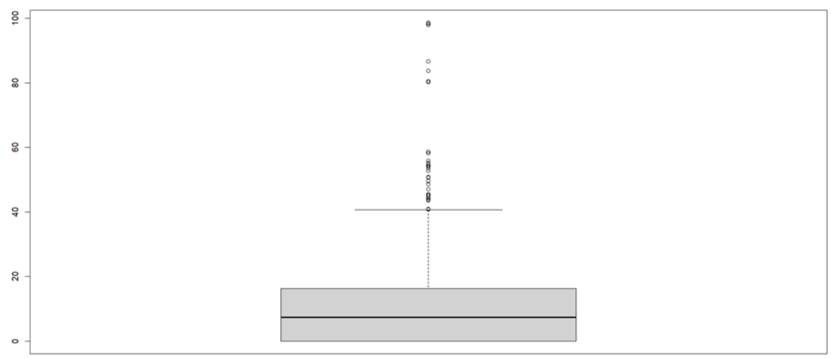
And in our case we get a p_value < 0.01, we can then reject the null hypothesis. These findings suggest that low-cost airlines may exhibit a unique response to fuel cost changes compared to non-low-cost carriers, highlighting the importance of considering pricing dynamics within the context of airline business models.

2) Interaction term with proportion of Long-Haul flights

Introducing an interaction term between fuel prices (represented by FUEL_JET_GULF) and the proportion of long-haul flights (LH_RATIO) in our model holds significant relevance in the context of airline economics. The LH_RATIO variable reflects the percentage of long-haul flights operated by the airline, which is indicative of its operational strategy and network configuration. Fuel costs constitute a substantial portion of the operating expenses for airlines, and the mix of short-haul and long-haul flights can influence the overall fuel consumption efficiency. By incorporating an interaction between fuel prices and the proportion of long-haul flights, we aim to capture the differential impact of fuel price fluctuations on airlines based on their flight distance composition. This interaction can provide deeper insights into how airlines adjust their pricing strategies and route planning in response to changes in fuel costs, especially considering that the mix of flight distances affects fuel consumption rates and operational costs. Thus, modeling this interaction enables a more nuanced understanding of the complex relationship between fuel prices, route composition, and operational decision-making within the airline industry.

In order to create this interaction, we generated three dummy variables based on the continuous variable LH_RATIO. We categorized flights as "small" if LH_RATIO equal 0%, "medium" if LH_RATIO was less than 10%, and "big" otherwise. The choice of segmentation is subject to debate and will significantly influence the outcomes.

Boxplot of LH_RATIO



Results

	log_UTKT_PRICE	
	OLS	panel linear
	Inidividual	Inidividual
	(1)	(2)
lag(log_FUEL_JET_GULF)	0.038	-0.004
	(0.067)	(0.053)
lag(log_COST_management)	0.414***	0.096
	(0.038)	(0.049)
lag(log_ASK)	0.033**	-0.042*
	(0.012)	(0.018)
lag(LOAD_FACTOR_100)	-0.006**	-0.001
	(0.002)	(0.003)
lag(LH_RATIO_100)	-0.003**	-0.0004
	(0.001)	(0.001)
lag(MKT_CONCENTRATION_100)	0.001	0.001
	(0.001)	(0.002)
lag(log_GDPpCAPITA_CURRENT_REGION)	-0.003	-0.028
	(0.016)	(0.020)
IS_LOW_COST	-0.263***	-0.034
	(0.034)	(0.044)
BELONGS_to_GOVERNMENT	0.037	
	(0.041)	
BELONGS_to_PRIVATE_I	0.097**	0.177*
	(0.032)	(0.089)
big_lh	-1.609***	-2.135***
	(0.456)	(0.366)
medium_lh	-1.157*	-1.895***
	(0.496)	(0.396)
lag(log_FUEL_JET_GULF):big_lh	0.294***	0.364***
	(0.085)	(0.067)
lag(log_FUEL_JET_GULF):medium_lh	0.217*	0.338***
	(0.092)	(0.073)
Constant	-0.911*	
	(0.443)	
Variable omise	REGION	
Variable omise	ID	ID
N	829	829
R2	0.477	0.205
Adjusted R2	0.463	0.079
Residual Std. Error	0.306 (df = 807)	

Notes: ***Significant at the 0.1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

Table of results 5

The main results are display in the following table :

	small % of long haul (ref)	medium % of LH	big % of LH
Elasticity of UTKT_PRICE with respect to fuel cost	0.038	0.217	0.294
It is significant ?	NO	Yes at 5% level	Yes at 0.1% level

In the elasticity of medium and big dummies we do not add the value of the coefficient of the reference because it is not significant so It would be like adding zero.

Based on the results obtained from the model incorporating the interaction term with the percentage of long-haul flights (LH_RATIO), we observe varying elasticities of UTKT_PRICE with respect to fuel cost across different categories of LH_RATIO. Specifically, for airlines categorized under "small % of long haul," the elasticity is 0, indicating no sensitivity of ticket prices to changes in fuel cost. In contrast, airlines with a "medium % of LH" exhibit a higher elasticity of 0.217, suggesting a more significant impact of fuel cost fluctuations on ticket prices. Similarly, airlines characterized by a "big % of LH" demonstrate an even higher elasticity of 0.294.

Considering the significance levels, we find that the elasticity is not statistically significant for airlines with a "small % of long haul," while it is significant at the 5% level for airlines with a "medium % of LH" and at the 0.1% level for airlines with a "big % of LH."
Overall, these findings underscore the importance of considering the composition of flight distances (long-haul vs. short-haul) when analyzing the relationship between fuel costs and airline revenues. Airlines with a higher proportion of long-haul flights tend to experience more pronounced effects of fuel cost changes on revenues, highlighting the relevance of route composition in shaping pricing strategies within the airline industry.

‘PGMM’

In our study on price elasticity in the airline industry, we utilize estimation methods such as the Generalized Method of Moments (GMM) system and the Panel Generalized Method of Moments (PGMM) function to analyze the relationships between explanatory variables and the target variable. The GMM system is a widely adopted approach used to estimate models where explanatory variables may be endogenous and correlated with regression errors. It addresses this issue by incorporating additional instrumental variables, which serve as proxies for the endogenous variables, thereby facilitating efficient estimation of model parameters.

The PGMM function, on the other hand, extends the GMM system to accommodate panel data structures commonly encountered in longitudinal studies. By applying the GMM method within a panel data framework, the PGMM function effectively controls for individual-specific effects present in the data. This ensures that any endogenous explanatory variables correlated with regression errors are adequately controlled for, leading to robust estimates of model coefficients.

The significance of instrumental variables lies in their role in addressing endogeneity concerns in econometric models. When certain explanatory variables are endogenous, meaning they are influenced by factors not captured in the model and are correlated with the error term, standard regression techniques may produce biased and inconsistent parameter estimates. Instrumental variables provide a means to circumvent this issue by leveraging external variables that are correlated with the endogenous variables but unrelated to the error term. By including these instruments in the estimation process, we can isolate the exogenous variation in the endogenous variables, thus obtaining unbiased and efficient estimates of the model parameters.

In summary, the use of instrumental variables, as implemented in the GMM system and PGMM function, allows us to overcome endogeneity concerns and obtain reliable estimates of the relationships between airline revenues and cost-influencing factors. This enables us to conduct rigorous analyses of the determinants of airline profitability while properly accounting for the complexities of panel data structures and potential endogeneity in the explanatory variables.

a) If we define every lagged of log_UTKT_PRICE as GMM instruments and only them, with all the other variables serving as normal instruments, we obtain the following results:

   | lag(l_UTKT_PRICE, 1:99) only

```
Coefficients:
                                        Estimate  Std. Error z-value  Pr(>|z|)
lag(log_FUEL_JET_GULF, 1)             0.09159378  0.04986995  1.8367  0.066261
lag(log_COST_management, 1)           0.60277755  0.08870620  6.7952 1.082e-11
lag(log_ASK, 1)                       0.02579933  0.03528636  0.7311  0.464693
lag(LOAD_FACTOR_100, 1)              -0.01113633  0.00432037 -2.5776  0.009948
lag(LH_RATIO_100, 1)                 -0.00382475  0.00201209 -1.9009  0.057317
lag(MKT_CONCENTRATION_100, 1)         0.00080387  0.00258191  0.3113  0.755537
lag(log_GDPpCAPITA_CURRENT_REGION, 1) -0.02499666  0.04584351 -0.5453  0.585574
low_cost                             -0.16537892  0.08792212 -1.8810  0.059976
Government                           -0.01932913  0.09982839 -0.1936  0.846471
Private_Investor                      0.04976730  0.06416825  0.7756  0.438000
ASIA                                  0.08233023  0.07887772  1.0438  0.296592
NORTH_AMERICA                         0.10228982  0.08866517  1.1537  0.248638
AFRICA                                0.09178139  0.16837391  0.5451  0.585682
OCEANIA                               0.35064579  0.13769132  2.5466  0.010878
MIDDLE_EAST                          -0.08550977  0.15619730 -0.5474  0.584072
SOUTH_AMERICA                         0.19569451  0.09910276  1.9747  0.048306

lag(log_FUEL_JET_GULF, 1)             .
lag(log_COST_management, 1)           ***
lag(log_ASK, 1)
lag(LOAD_FACTOR_100, 1)               **
lag(LH_RATIO_100, 1)                  .
lag(MKT_CONCENTRATION_100, 1)
lag(log_GDPpCAPITA_CURRENT_REGION, 1)
low_cost                              .
Government
Private_Investor
ASIA
NORTH_AMERICA
AFRICA
OCEANIA                               *
MIDDLE_EAST
SOUTH_AMERICA                         *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(26) = 70.73829 (p-value = 5.1769e-06)
Autocorrelation test (1): normal = 0.6016731 (p-value = 0.54739)
Autocorrelation test (2): normal = -1.406831 (p-value = 0.15948)
Wald test for coefficients: chisq(16) = 10127.23 (p-value = < 2.22e-16)
```

Table of results 6

Since we reject the null hypothesis from the Sargan test, it suggests that the instruments are overidentifying, which implies that the model may not be relevant or suitable for the data.

b) We utilize all the lagged values of each quantitative variable as GMM instruments. The approach is outlined as follows:
| lag(l_UTKT_PRICE, 1:99)
+ lag(log_COST_management, 2:99)
+ lag(log_ASK, 2:99)
+ lag(MKT_CONCENTRATION_100, 2:99)
+ lag(LH_RATIO_100, 2:99)
+ lag(LOAD_FACTOR_100, 2:99)
+ lag(log_GDPpCAPITA_CURRENT_REGION, 2:99)

```
Coefficients:
                                        Estimate  Std. Error z-value  Pr(>|z|)
lag(log_FUEL_JET_GULF, 1)             0.15457209  0.05752588  2.6870   0.00721
lag(log_COST_management, 1)           0.89038306  0.14600938  6.0981 1.073e-09
lag(log_ASK, 1)                      -0.15596118  0.08463958 -1.8427   0.06538
lag(LOAD_FACTOR_100, 1)               0.00586833  0.00847021  0.6928   0.48842
lag(LH_RATIO_100, 1)                 -0.00923568  0.00796037 -1.1602   0.24596
lag(MKT_CONCENTRATION_100, 1)         0.01513008  0.01181118  1.2810   0.20019
lag(log_GDPpCAPITA_CURRENT_REGION, 1) 0.03746766  0.14327486  0.2615   0.79370
low_cost                             -0.29978594  0.10995037 -2.7266   0.00640
Government                            0.13540305  0.19029849  0.7115   0.47676
Private_Investor                      0.00090685  0.11027128  0.0082   0.99344
ASIA                                  0.34520957  0.16877919  2.0453   0.04082
NORTH_AMERICA                         0.28368408  0.13671412  2.0750   0.03798
AFRICA                                0.11586253  0.36170214  0.3203   0.74872
OCEANIA                               0.25032814  0.32237963  0.7765   0.43745
MIDDLE_EAST                          -0.06562786  0.30708007 -0.2137   0.83077
SOUTH_AMERICA                         0.27170113  0.19431387  1.3983   0.16204

lag(log_FUEL_JET_GULF, 1)            **
lag(log_COST_management, 1)          ***
lag(log_ASK, 1)                      .
lag(LOAD_FACTOR_100, 1)
lag(LH_RATIO_100, 1)
lag(MKT_CONCENTRATION_100, 1)
lag(log_GDPpCAPITA_CURRENT_REGION, 1)
low_cost                             **
Government
Private_Investor
ASIA                                 *
NORTH_AMERICA                        *
AFRICA
OCEANIA
MIDDLE_EAST
SOUTH_AMERICA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(68) = 80.02626 (p-value = 0.15093)
Autocorrelation test (1): normal = -0.9855656 (p-value = 0.32435)
Autocorrelation test (2): normal = -2.137941 (p-value = 0.032522)
Wald test for coefficients: chisq(16) = 7035.226 (p-value = < 2.22e-16)
```

Table of results 7

Following the presentation of the model test results, it is essential to delve into their implications. The Sargan test yields a chi-square statistic of 80.03 with a corresponding p-value of 0.15. This indicates that the instruments used in the model are not over-identifying, suggesting that the model's specification is satisfactory. Moving on to the autocorrelation tests, the first test yields a normal statistic of -0.99 with a p-value of 0.32, while the second test yields a normal statistic of -2.14 with a p-value of 0.03. These results suggest that there may be some autocorrelation present in the model, particularly indicated by the significant p-value in the second test. Additionally, the Wald test for coefficients returns a chi-square statistic of 7035.23 with an extremely low p-value of less than 0.01, indicating strong evidence against the null hypothesis. Overall, these test results provide valuable insights into the adequacy of the model specification and the presence of autocorrelation, warranting further investigation and potential adjustments to the model.

In that case the elasticity of UTKT_PRICE to FUEL_JET_GULF is about 0.15, significant at 1% level.

What we aim to highlight here is the alternative offered by the GMM system. However, we encountered some theoretical and technical difficulties while using the R software. Finding the right model specification proved challenging, as the model is highly sensitive to the arguments provided, as demonstrated in this small example. Despite its potential, the GMM system requires careful consideration and expertise in model specification to yield accurate and reliable results.

Panel Var model

As explained in the previous section, the price of flight tickets not only depends on the current state of the explanatory variables, but also on their lagged values. For instance, airlines tend to apply hedging strategies in order to smooth the impact of the variations of the market fuel price. Therefore, a big variation in the fuel price may only be reflected on the ticket prices 1 year later. Hence, taking into account the lags of the FUEL_JET_GULF variable is relevant and important. However, and in addition, we have seen in the graph (*Schema 1 : link between variables*) that there is a large interdependence between covariates, introducing some endogeneity. More particularly, UTKT_PRICE is theoretically linked to the demand variable RPKs. As a result, the value of this variable at time T-1 should impact the value of UTKT_PRICE at time T. This endogeneity issue is well tackled by the Panel-VAR model which is able to capture the dynamic dependencies between endogenous variables. We implemented this class of models using the PanelVar package developed on R by Micheal Sigmund (see bibliography)
We tested a lot of various variable selections for our panelVar Model, among which models including the Regions' Dummies and LOAD_FACTOR. The regions' dummies were insignificant and did not change the significance nor the extent of the other variables. A deeper analysis could be made on this topic, but our results tend to indicate that the UTKT_PRICE is not highly influenced by the origin region of the airline. Concerning LOAD_FACTOR, it caused significance issues on all other variables when setting both RPK and UTKT_PRICE as endogenous variables. Consequently, we chose to remove it.

We are going to compare 2 models:

M0: No interaction, this is our baseline Pvar Model
M1: M0 + Interaction between IS_LOW_COST and JET_FUEL_GULF at time T-1, to ascertain whether the airline reaction to a Fuel price rise depends on its business model or not.
We also tried a model with both the lag and the actual value of FUEL_JET_GULF but it led to incoherent results and tends to hinder the interpretation of the coefficients.

Models specifications : Log(*UTKT_PRICE_it*)= α0 + α1 Lag1(log(*UTKT_PRICE_it*))+α2 Lag1(log(*RPK_it*) + α3 log(*ASK_it*)+ α4 *log(COUT_AUTRE_it)* + α5 log(*MKT_CONCENTRATION_100_it*)+ α6 *Lh_Ratio_100_it* + α7 log(*GDPper capita_CURRENT_REGION_it*)+ α8 log(*FUEL_JET_GULF_it)* + α9 *IS_LOW_COST_i* + α10 *Government_i* + α11 *Private_investor_i* + ϵ_it
M0:

$$\log(UTKT\_PRICE_{it}) = \alpha_0 + \alpha_1 \mathrm{Lag1}(\log(UTKT\_PRICE_{it})) + \alpha_2 \mathrm{Lag1}(\log(RPK_{it})) + \alpha_3 \log(ASK_{it}) + \alpha_4 \log(COUT\_AUTRE_{it}) + \alpha_5 \log(MKT\_CONCENTRATION\_100_{it}) + \alpha_6 Lh\_Ratio\_100_{it} + \alpha_7 \log(GDPper\_capita\_CURRENT\_REGION_{it}) + \alpha_8 \log(FUEL\_JET\_GULF_{it}) + \alpha_9 \mathrm{IS\_LOW\_COST}_i + \alpha_{10} \mathrm{Government}_i + \alpha_{11} \mathrm{Private\_investor}_i + \epsilon_{it}$$

$$\log(RPK_{it}) = \beta_0 + \beta_1 \mathrm{Lag1}(\log(UTKT\_PRICE_{it})) + \beta_2 \mathrm{Lag1}(\log(RPK_{it})) + \beta_3 \log(ASK_{it}) + \beta_4 \log(COUT\_AUTRE_{it}) + \beta_5 \log(MKT\_CONCENTRATION\_100_{it}) + \beta_6 Lh\_Ratio\_100_{it} + \beta_7 \log(GDPper\_capita\_CURRENT\_REGION_{it}) + \beta_8 \log(FUEL\_JET\_GULF_{it}) + \beta_9 \mathrm{IS\_LOW\_COST}_i + \beta_{10} \mathrm{Government}_i + \beta_{11} \mathrm{Private\_investor}_i + \epsilon_{it}$$

M1:

$$\log(UTKT\_PRICE_{it}) = \text{Formula M0} + \alpha_9 \mathrm{Lag1}(\log(UTKT\_PRICE_{it})) \times \mathrm{IS\_LOW\_COST}_i$$

	M1		M0	
	l_UTKT_PRICE	l_RPK	l_UTKT_PRICE	l_RPK
lag1_l_UTKT_PRICE	0.5569 ***	-0.0303 *	0.5980 ***	-0.0260
	(0.0584)	(0.0140)	(0.0520)	(0.0150)
lag1_l_RPK	-1.0236 ***	0.0850	-0.7097 **	0.0523
	(0.2575)	(0.0535)	(0.2518)	(0.0554)
low_cost	1.5259 **	-0.0658	-0.1275	-0.0254
	(0.5672)	(0.1153)	(0.0988)	(0.0261)
Government	0.1623	-0.1187	0.0662	-0.1065
	(0.2381)	(0.1029)	(0.2061)	(0.0815)
Private_Investor	0.1930	-0.0477	0.1829	-0.0669
	(0.1680)	(0.0863)	(0.1255)	(0.0937)
l_FUEL_JET_GULF	0.2323 ***	-0.0004	0.1417 ***	0.0015
	(0.0431)	(0.0103)	(0.0325)	(0.0079)
l_cout_autre	0.2504 ***	-0.0574 *	0.2040 **	-0.0806 *
	(0.0672)	(0.0284)	(0.0780)	(0.0315)
l_ASK_m_final	1.0531 ***	0.9373 ***	0.7344 **	0.9655 ***
	(0.2654)	(0.0586)	(0.2555)	(0.0580)
l_MKT_CONCENTRATION	-0.0003	0.0003	-0.0006	-0.0002
	(0.0019)	(0.0009)	(0.0017)	(0.0009)
l_LH_RATIO	-0.0005	-0.0008	-0.0007	-0.0013
	(0.0014)	(0.0007)	(0.0015)	(0.0009)
l_GDPpCAPITA_CURRENT_REGION	0.0210	0.0063	0.0020	0.0173
	(0.0399)	(0.0156)	(0.0369)	(0.0165)
l_FUEL_JET_GULF_low_cost	-0.3044 **	0.0125		
	(0.1034)	(0.0213)		

*** p < 0.001; ** p < 0.01; * p < 0.05

- Model M0: As we observe the results of the first model M0, we notice that FUEL_JET_GULF is significant and positive and the elasticity is around 0.14%. However, the lag of UTKT_PRICE captures a lot of the variation of the UTKT_PRICE at time T. We observe that both ASK and RPK coefficients are significant and of same magnitude, but of inverse sign. As we can observe in the second column of M0, the demand (RPK) is almost exclusively predicted by the ASK. It is then coherent that the impacts of these 2 variables almost cancel each other in the UTKT_PRICE column.

- Model M1: However, it is worthy to note that when we add the interaction between FUEL_JET_GULF and LOW_COST, we observe that both the interaction and the FUEL_JET_GULF alone are significant, and that their sum drops below zero (-0.07%). THis would suggest that the elasticity of price tickets to fuel price is negative for low cost airlines. This appears quite counter intuitive, and in contradiction with our previous results. In addition, in this second model, the positive and significant coefficient for LOW_COST is disturbing. This would imply that the low cost airlines tend to have a higher UTKT_PRICE than non low costs, all other things being equal.

Overall, this M1 model with interaction does not seem relevant to describe the relationship between our variables. The M0 model appears more rational but the dynamic nature of the PanelVar setting makes the interpretation of coefficients challenging. This Model family seems inadequate to solve our problematic.

Study on the US Database

The US base dataset initially consists of quarterly data from 66 US airlines, spanning from the third quarter of 1995 to the second quarter of 2023. It includes financial indicators, operational efficiency, market performance, and macroeconomic indicators. The US base dataset provides richness in terms of temporal dimension.

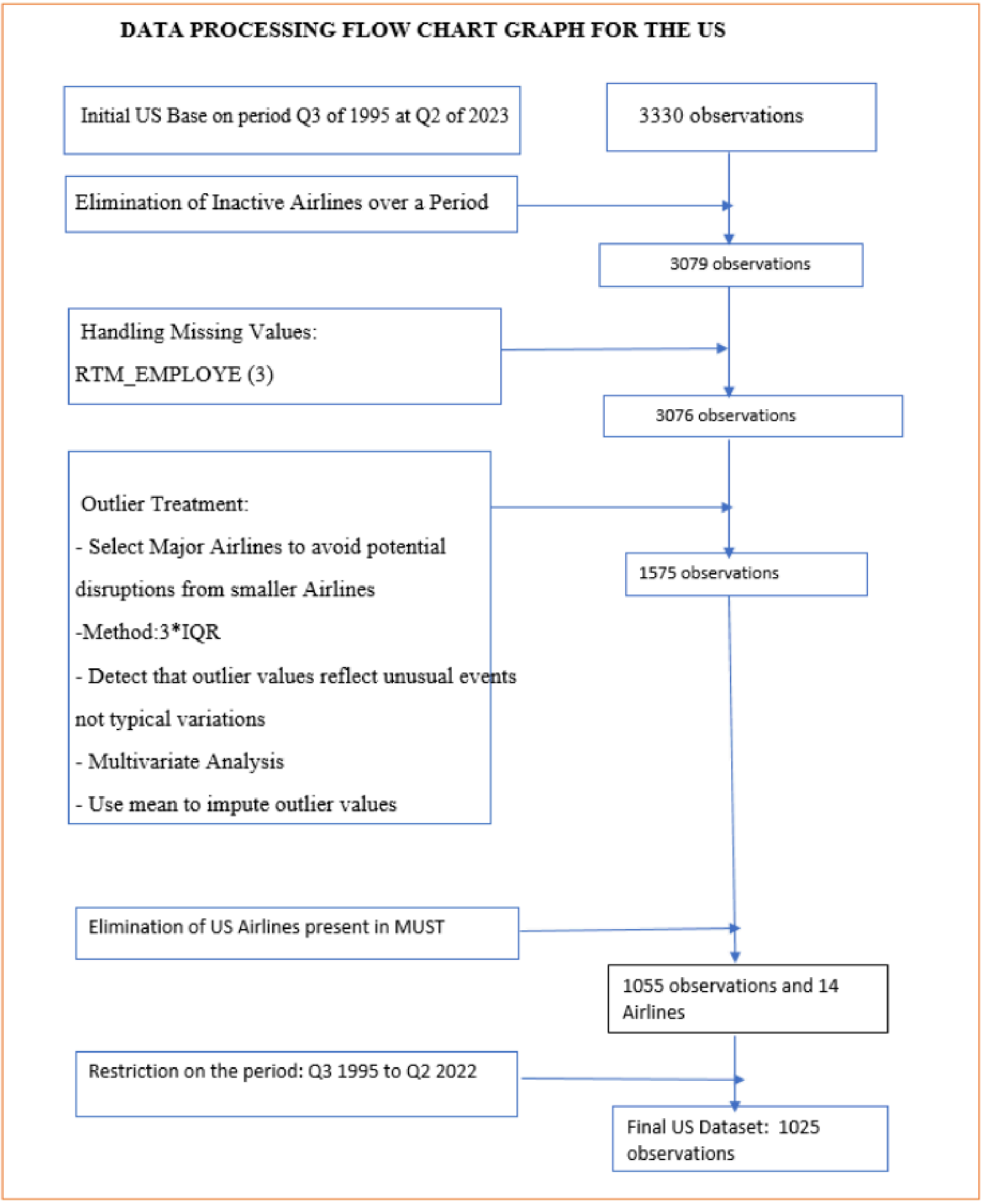
Data Presentation

Data for the US base are sourced from three different data sources. Fuel price data are obtained from the Energy Information Administration (EIA), while operational revenue data and other relevant variables are extracted from the United States Department of Transportation (US DOT) data. Macroeconomic data such as Gross Domestic Product per capita, Consumer Price Index, and population growth are provided by the International Monetary Fund (IMF). The entirety of these variables, their descriptions, and respective sources are presented in the following table:

VARIABLES	**DESCRIPTION DE LA VARIABLE**	**SOURCE**
Variable dépendante		
PAX_REV_RPM	Pax Revenue per RPM (includes AR)	US DOT
Variables indépendantes		
FUEL_JET_GULF	Jet Fuel price (UScents / US gallon)	EIA
OPE_EXP_RTM	Operating Expenses per RTM (includes AR)	US DOT

HHI_RPM	Competition level global RPM based	US DOT
MKT_SHARE_RPM	Market share in RPM	US DOT
RPM	Market traffic in RPM growth	US DOT
LOAD_FACTOR	Airline Load Factor based on RPM	US DOT
GDPpCAPITA_CURRENT_US	Region GDP p. capita current price in $	IMF
CPI_US	CPI Region	IMF
POPULATION_US	Population growth - Region	IMF
RTM_EMPLOYEE	Revenue Tonne Mile (RTM) per employee	US DOT

Data Processing/Cleaning



Upon examining the distribution of missing values, we noticed a high rate (7%) for key variables such as PAX_REV. This suggests that the absence of passenger revenue may reflect temporary inactivity of certain airlines. Therefore, our analysis is limited to active airlines, excluding cases with missing PAX_REV. Additionally, we removed three cases where the

RTM_EMPLOYEE variable was missing while OPE_EXP_RTM was present, indicating a contradictory situation. After cleaning, our database comprises 3076 observations and 41 variables.

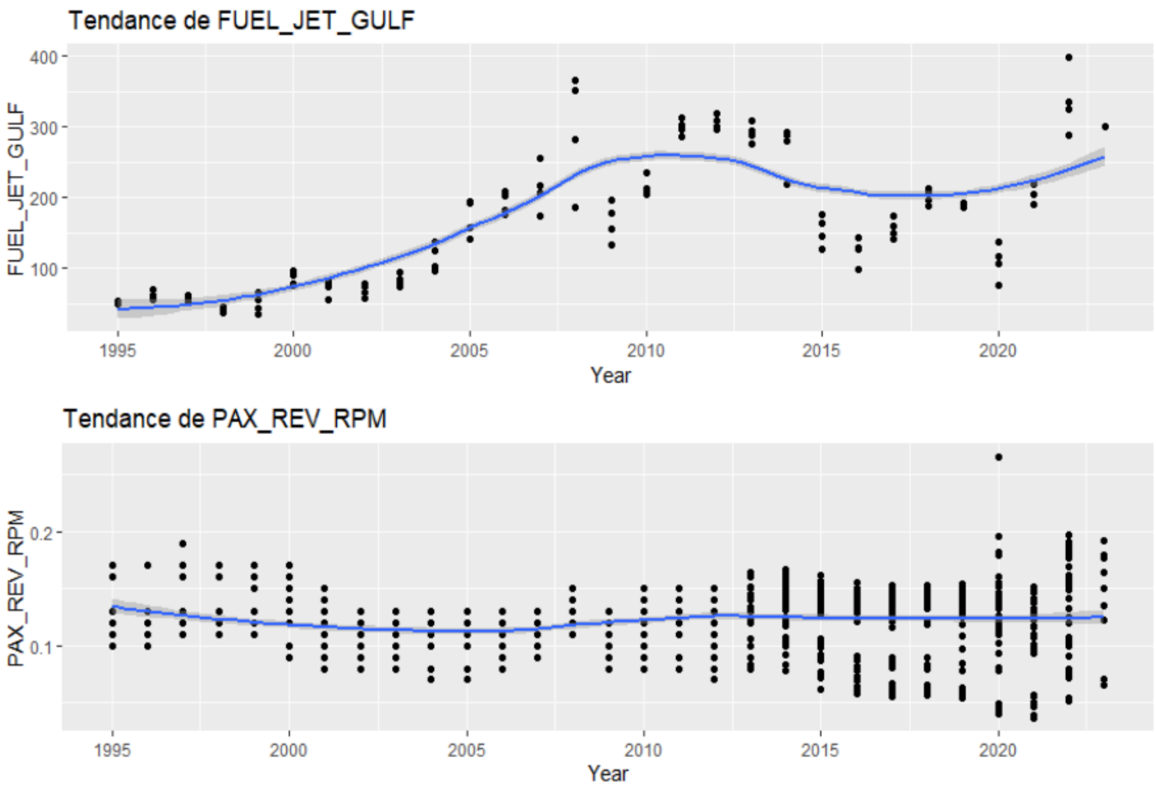
In analyzing outliers, we favor major airlines to avoid biases stemming from events such as bankruptcies or mergers. We employ the annual 3*IQR method to filter extremes while discarding neighboring data to minimize distortions. Subsequently, we compare these anomalies to company trends through temporal graphs, ensuring that detected deviations reflect specific irregularities rather than normal fluctuations. Moreover, we conduct multivariate analysis of annual data to verify consistency among correlated variables. This comprehensive approach confirms the statistical reliability of our results and their alignment with the specific context of the airline. To facilitate result comparison and maintain consistency, we retained only US airlines present simultaneously in MUST (14 in total). Ultimately, we obtained 1055 observations.

Methodology

1. Preliminary Analysis

a. Trend of FUEL JET PRICE and PAX REV RPM

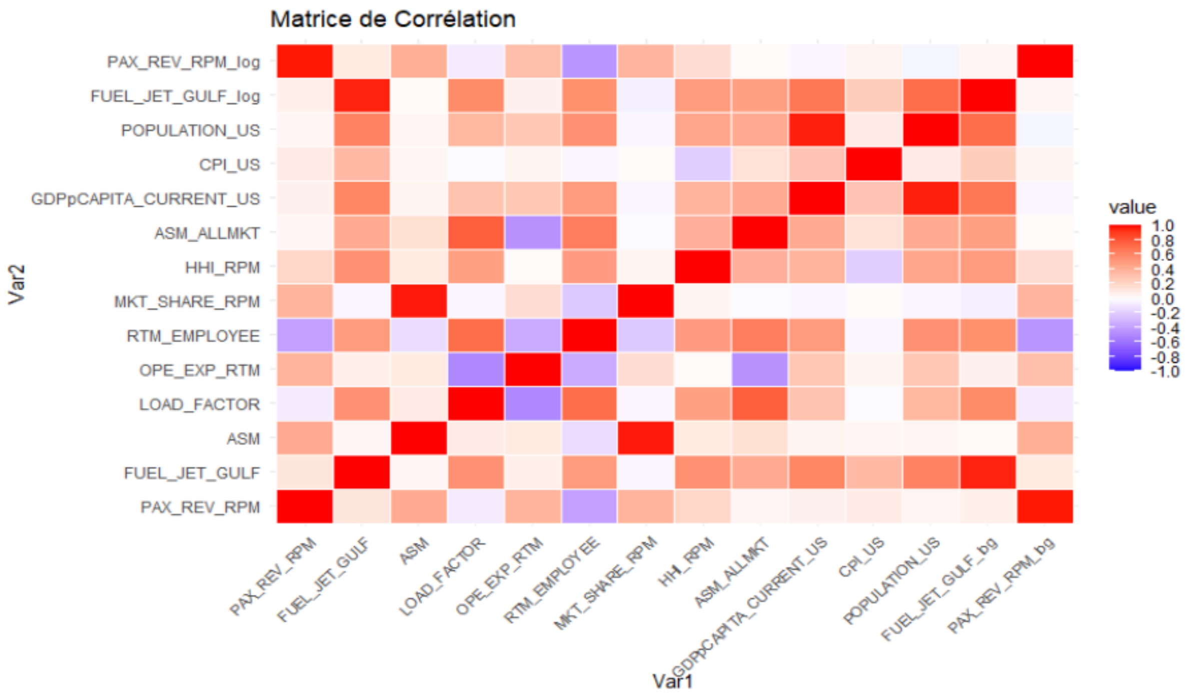
Graph 2.1



The graphical analysis of jet fuel prices over the study period reveals an overall trend of growth with phases of both increase and decrease, whereas passenger revenues (PAX rev RPM) exhibit a relatively constant average trend.

b. Analysis of Correlation between Variables

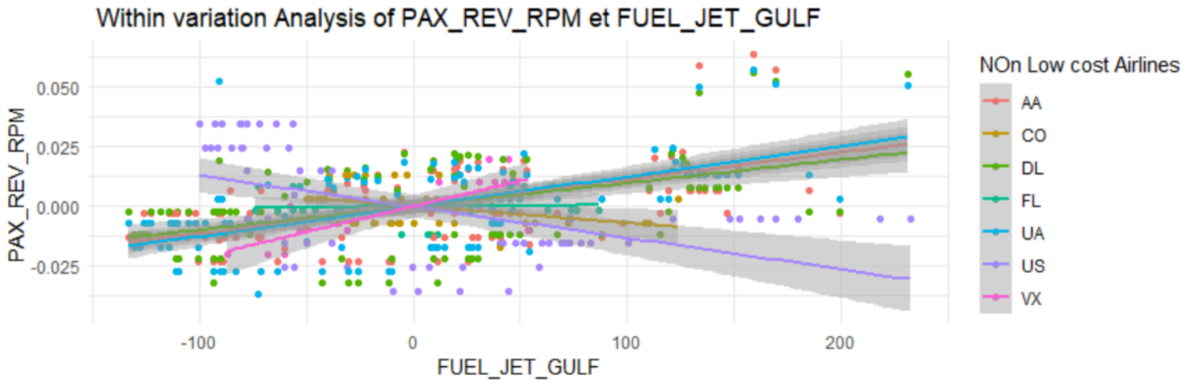
Table 2.1

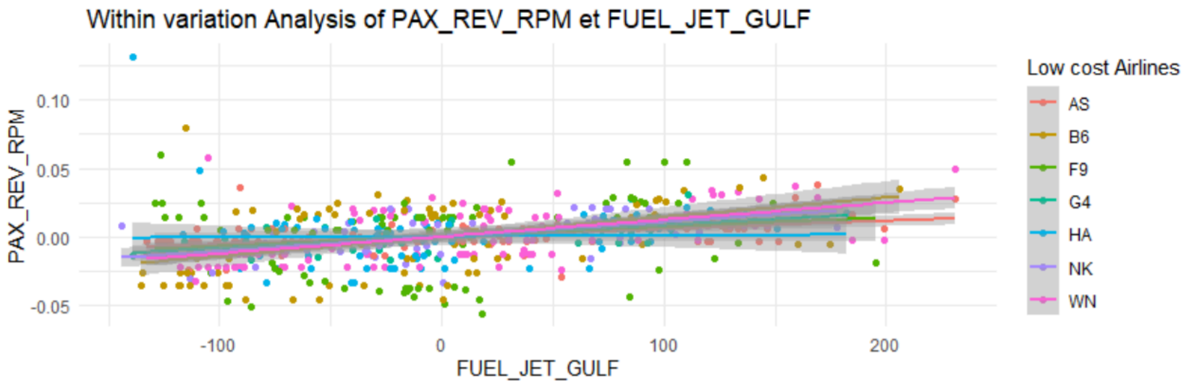


The correlation matrix analysis reveals that the majority of correlations with the variable PAX_REV_RPM are not strong. The variables most correlated are RTM_EMPLOYEE, RPM, and OPE_EXP_RTM. However, we have also included some less correlated variables that are economically relevant to our model, such as market competition (HHI_RPM, MKT_SHARE_RPM), operational efficiency (LOAD_FACTOR), and macroeconomic variables (GDPpCAPITA_CURRENT_US, CPI_US).

c. Within Bivariate Variation Analysis

Graph 2.2





The bivariate graphical analysis within the variation between PAX revenue and jet fuel reveals interesting insights. Firstly, there is a positive and weak slope for low-cost carriers, indicating that increases in jet fuel prices are associated with slight increases in passenger revenue for this category of airlines. On the other hand, for non-low-cost carriers, the slope is steeper, suggesting a more pronounced impact of jet fuel price changes on passenger revenue. However, it's noteworthy that negative slopes are observed in some cases among non-low-cost carriers, indicating that higher jet fuel prices might lead to reduced passenger revenue for certain airlines in this category. These findings highlight the differential responses of airlines to fluctuations in jet fuel prices based on their operational models, with low-cost carriers exhibiting a more resilient revenue pattern compared to non-low-cost carriers.

2. Econometric Analysis

In this section, we describe the methodology employed for our study using the US dataset. Variables were selected based on their economic relevance and were integrated into the model as control variables. We conducted several econometric analyses to estimate the elasticities of FUEL_JET_GULF on PAX_REV_RPM, including multiple regression models and OLS with individual fixed effects. Additionally, we tested interactions between JET_FUEL_GULF and IS_LOW_COST on one hand, and between JET_FUEL_GULF and trend variables (growth, decline, and stagnation) on the other hand, to determine if the elasticity coefficients of fuel and airfare prices vary depending on the company's cost structure and phases of fuel price fluctuations in the market. Finally, we complemented our analysis by estimating an ARDL model to examine the long-term dynamics between variables and evaluate the short and long-term effects of fuel price variations on airline revenues.

Below are the econometric specifications of the different estimated models:

OLS MODEL

$$\log(PAX\_REV\_RPM_{it}) = \beta_o + \beta_1 \log(FUEL\_JET\_GULF_t) + \beta_2 \log(RPM_{it}) + \beta_3 \log(LOAD\_FACTOR_{it}) + \beta_4 \log(HHI\_RPM_{it}) + \beta_5 \log(RTM\_EMPLOYEE_{it}) + \beta_6 MKT\_SHARE\_RPM_{it} + \beta_7 OPE\_EXP\_RTM_{it} + \beta_8 \log(GDPpCAPITA\_CURRENT\_US_t) + \beta_9 \log(POPULATION\_US) + \beta_{10} \log(CPI\_US) + \beta_{11} IsLowCost_i + \varepsilon_{it}$$

Where IS_LOW_COST is a binary variable, taking 1 if the Airline is low-cost, and $_{it}$ represents the error term of the model.

MODEL OLS à EFI

$$\log(PAX\_REV\_RPM_{it}) = \alpha_o + \alpha_1 \log(FUEL\_JET\_GULF_t) + \alpha_2 \log(RPM_{it}) + \alpha_3 \log(LOAD\_FACTOR_{it}) + \alpha_4 \log(HHI\_RPM_{it}) + \alpha_5 \log(RTM\_EMPLOYEE_{it}) + \alpha_6 MKT\_SHARE\_RPM_{it} + \alpha_7 OPE\_EXP\_RTM_{it} + \alpha_8 \log(GDPpCAPITA\_CURRENT\_US_t) + \alpha_9 \log(POPULATION\_US) + \alpha_{10} \log(CPI\_US) + \alpha_{11} IsLowCost_i + \sum_{i=1}^{n} \gamma_i D_i + u_{it}$$

 Where Di is a dummy variable associated with the airline, and it represents the error term of the model.

MODEL OLS à EFI et interaction avec tendance

$$\log(PAX\_REV\_RPM_{it}) = \theta_o + \theta_1 \log(FUEL\_JET\_GULF_t) + \theta_2 \log(RPM_{it}) + \theta_3 \log(LOAD\_FACTOR_{it}) + \theta_4 \log(HHI\_RPM_{it}) + \theta_5 \log(RTM\_EMPLOYEE_{it}) + \theta_6 MKT\_SHARE\_RPM_{it} + \theta_7 OPE\_EXP\_RTM_{it} + \theta_8 \log(GDPpCAPITA\_CURRENT\_US_t) + \theta_9 \log(POPULATION\_US) + \theta_{10} \log(CPI\_US) + \theta_{11} IsLowCost_i + \theta_{12} \log(FUEL\_JET\_GULF_t) * Croissance + \theta_{13} \log(FUEL\_JET\_GULF_t) * Decroissance + \sum_{i=1}^{n} \gamma_i D_i + v_{it}$$

Where "Croissance" and "Décroissance" represent the periods of increasing and decreasing fuel jet prices during the study period, respectively. The period of stagnation is taken as the reference.

MODEL ARDL

$$\log\_PAX\_REV\_RPM\_mean_t = \varphi + \alpha \log\_PAX\_REV\_RPM\_mean_{t-1} + \beta_1 \log\_FUEL\_JET\_GULF\_mean_t + \beta_2 \log\_RPM\_mean_t + \beta_3 \log\_RTM\_EMPLOYEE\_mean_t + \beta_4 Market\_Share\_Islowcost_t + \beta_5 HHI\_RPM_t + \lambda_1 \log\_FUEL\_JET\_PRICE_t * Croissance + \lambda_2 \log\_FUEL\_JET\_PRICE_t * Decroissance + \varepsilon_t$$

The logarithm (log) is used for both the dependent variable and the explanatory variable of interest (Fuel_jet_gulf) in all models, to facilitate the interpretation of the estimated coefficient as elasticity.

a. OLS Regression Results

Récapitulatif de Modèle de Régression multiple OLS, Base Trimestre

	Dependent variable:
	OLS Multiple Compl
log(FUEL_JET_GULF)	0.070***
	(0.017)
log(RPM)	0.057***
	(0.013)
log(LOAD_FACTOR)	0.878***
	(0.083)
HHI_RPM	4.959***
	(0.424)
log(RTM_EMPLOYEE)	-0.636***
	(0.034)
MKT_SHARE_RPM	-0.522***
	(0.181)
OPE_EXP_RTM	0.003
	(0.014)
log(GDPpCAPITA_CURRENT_US)	-0.570**
	(0.272)
CPI_US	0.015**
	(0.008)
log(POPULATION_US)	2.489**
	(1.074)
IsLowCost_binary	-0.088***
	(0.019)
Constant	-39.770**
	(18.067)
Observations	1,025
R^2	0.521
Adjusted R^2	0.515
Residual Std. Error	0.165 (df = 1013)
F Statistic	100.036*** (df = 11; 1013)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table of results 2.1

The initial results from the OLS regression indicate that the variable Fuel_jet_gulf exhibits the expected sign, with a statistically significant and positive association. This observation suggests a consistent relationship with theoretical expectations regarding its impact on airline revenues.

However, given that our study deals with panel data, it's important to acknowledge that OLS regression results may be biased. While the sign of the coefficients aligns with our theoretical expectations, it's likely that these results are affected by bias and inefficiency issues associated with the Ordinary Least Squares (OLS) method in the context of panel data.

Therefore, it's imperative to explore alternative methods better suited for panel data analysis. Techniques such as Fixed Effects models (OLS with Fixed Effects) can be considered to correct for these biases and obtain more precise estimates of the coefficients. This approach will enhance the robustness and validity of our results, ensuring more reliable conclusions in our study.

b. Results of Individual Fixed Effects OLS Regression

In this section, we present comparative results between Ordinary Least Squares (OLS) regression and Individual Fixed Effects OLS regression.

Récapitulatif de Modèle de Régression multiple OLS, Base Trimestre		
	Dependent variable:	
	OLS Multiple Complet	OLS à Effets Fixes
	(1)	(2)
log(FUEL_JET_GULF)	0.070^{***}	0.047^{***}
	(0.017)	(0.013)
log(RPM)	0.057^{***}	0.070^{***}
	(0.013)	(0.017)
log(LOAD_FACTOR)	0.878^{***}	0.776^{***}
	(0.083)	(0.067)
HHI_RPM	4.959^{***}	5.127^{***}
	(0.424)	(0.334)
log(RTM_EMPLOYEE)	-0.636^{***}	-0.612^{***}
	(0.034)	(0.034)
MKT_SHARE_RPM	-0.522^{***}	-0.853^{***}
	(0.181)	(0.254)
OPE_EXP_RTM	0.003	-0.021^{*}
	(0.014)	(0.013)
log(GDPpCAPITA_CURRENT_US)	-0.570^{**}	-0.868^{***}
	(0.272)	(0.208)
CPI_US	0.015^{**}	0.022^{***}
	(0.008)	(0.006)
log(POPULATION_US)	2.489^{**}	4.012^{***}
	(1.074)	(0.819)
IsLowCost_binary	-0.088^{***}	-0.426^{***}
	(0.019)	(0.054)
Observations	1,025	1,025
R^2	0.521	0.733
Adjusted R^2	0.515	0.727
Residual Std. Error	0.165 (df = 1013)	0.124 (df = 1001)
F Statistic	100.036^{***} (df = 11; 1013)	119.764^{***} (df = 23; 1001)
Note:		*p<0.1; **p<0.05; ***p<0.01
		reg1 : Modèle OLS multiples
		reg2 : Modèle OLS à effets individuels

Table of results 2.2

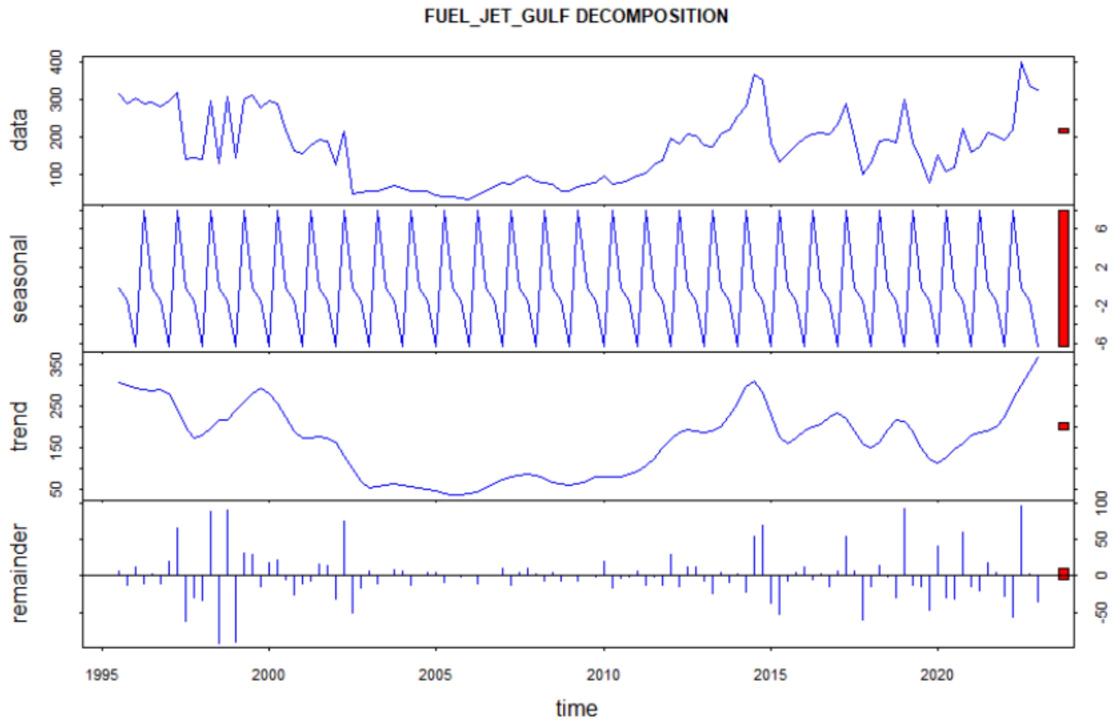
The results indicate that in the multiple linear regression OLS model, the coefficient of log(FUEL_JET_GULF) is 0.070, while it decreases to 0.047 in the Individual Fixed Effects OLS model. This decrease suggests that after controlling for individual fixed effects, the impact of fuel price (log(FUEL_JET_GULF)) on the dependent variable is less significant than initially expected. In other words, each 1% increase in fuel price in the market now only leads to a 0.047% increase in airline PAX_REV_RPM.

Furthermore, accounting for fixed effects reveals a negative impact of OPE_EXP_RTM on the dependent variable. This result underscores the importance of considering unobservable and time-invariant characteristics of each entity to obtain more precise estimates of coefficients in our model.

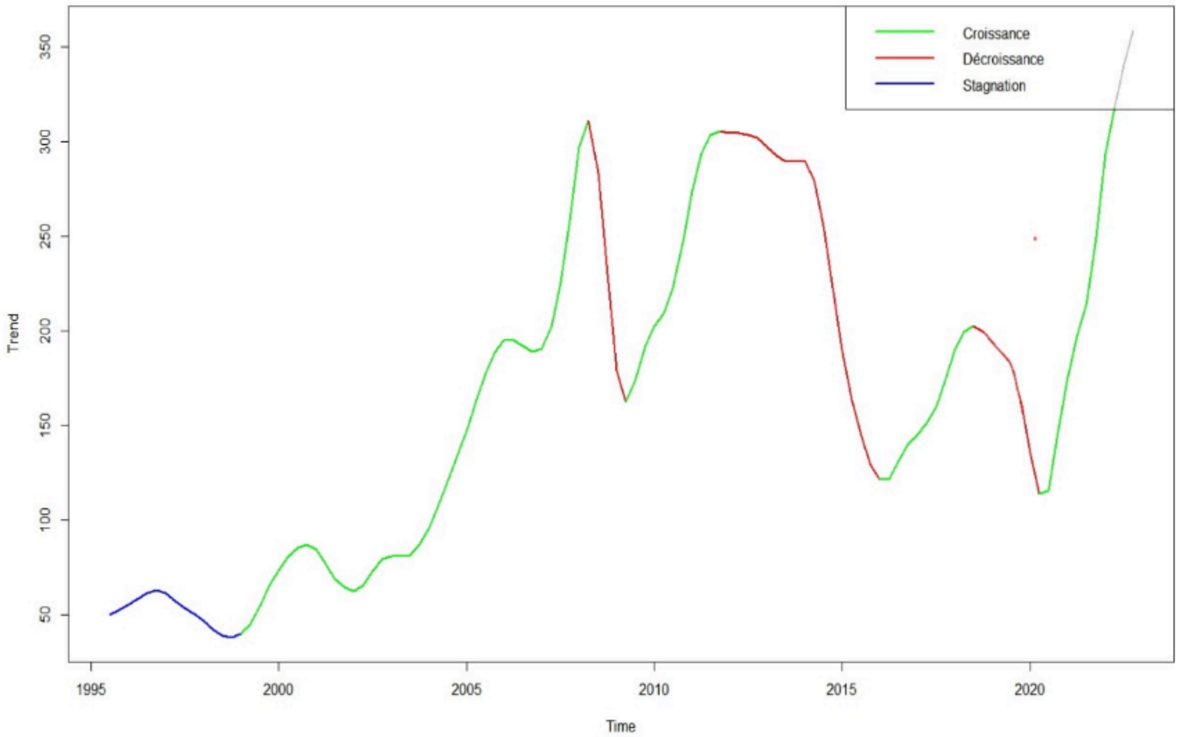
c. Results of Individual Fixed Effects OLS Regression with Interaction

The time series of FUEL_JET_GULF is decomposed into trend, seasonality, and random component, with emphasis on the trend. Using the moving average method, the main periods of variation in fuel prices are identified and analyzed, including growth, decline, and stagnation. By integrating these trends into the previous Individual Fixed Effects OLS model, trend indicators of fuel prices are incorporated. This allows for a more precise analysis of the impact of fuel prices on airline revenues, taking into account long-term fluctuations.

Graph 2.3



Graph 2.4



Below is the table presenting the results of the different specifications of the Individual Fixed Effects OLS model, considering both with and without interaction between trend variables, the Islowcost variable, and fuel price (Fuel_jet_gulf).

Récapitulatif des Modèles de Régression EFI			
	Dependent variable:		
	log(PAX_REV_RPM)		
	OLS	OLS	OLS
	(1)	(2)	(3)
log(FUEL_JET_GULF)	0.047***	0.025*	0.055***
	(0.013)	(0.015)	(0.018)
log(RPM)	0.070***	0.053***	0.047***
	(0.017)	(0.018)	(0.017)
log(LOAD_FACTOR)	0.776***	0.775***	0.801***
	(0.067)	(0.067)	(0.066)
HHI_RPM	5.127***	5.053***	3.985***
	(0.334)	(0.333)	(0.371)
log(RTM_EMPLOYEE)	-0.612***	-0.602***	-0.614***
	(0.034)	(0.034)	(0.033)
MKT_SHARE_RPM	-0.853***	-0.786***	-0.798***
	(0.254)	(0.254)	(0.249)
OPE_EXP_RTM	-0.021*	-0.024*	-0.032***
	(0.013)	(0.012)	(0.012)
log(GDPpCAPITA_CURRENT_US)	-0.868***	-0.814***	-0.912***
	(0.208)	(0.208)	(0.206)
log(POPULATION_US)	4.012***	3.906***	4.498***
	(0.819)	(0.816)	(0.807)
CPI_US	0.022***	0.021***	0.026***
	(0.006)	(0.006)	(0.006)
IsLowCost_binary	-0.426***	-0.243***	-0.216***
	(0.054)	(0.082)	(0.080)
log(FUEL_JET_GULF):IsLowCost_binary		0.046***	0.039**
		(0.015)	(0.015)
log(FUEL_JET_GULF):Croissance			-0.019***
			(0.005)
log(FUEL_JET_GULF):Decroissance			-0.008
			(0.005)
Observations	1,025	1,025	1,025
R^2	0.733	0.736	0.747
Adjusted R^2	0.727	0.729	0.740
Residual Std. Error	0.124 (df = 1001)	0.124 (df = 1000)	0.121 (df = 998)
F Statistic	119.764*** (df = 23; 1001)	116.054*** (df = 24; 1000)	113.307*** (df = 26; 998)
Note:			*p<0.1; **p<0.05; ***p<0.01
			reg1 : Modèle EFI
			reg2 : Modèle EFI et interaction avec is
			reg3 : Modèle EFI et interaction avec la tendance et is

Table of results 2.3

The results show a significant and negative association of low-cost airlines (Inslowcost) with revenue per passenger per mile (PAX_REV_RPM), suggesting that low-cost airlines tend to generate lower revenue per passenger per mile compared to full-service airlines.

Furthermore, the significant and positive sign of the interaction between log(Fuel_jet_gulf) and Inslowcost suggests that low-cost airlines are particularly sensitive to fuel price fluctuations. They benefit from a flexible pricing system, allowing them to quickly pass on changes in fuel costs to the fares offered to consumers.

The coefficient for log(Fuel_jet_gulf) during the growth period is -0.019***, which means that during periods of increasing fuel prices, the increase in unit price per passenger mile is lower by 0.019% for each 1% increase in fuel prices, compared to a period of stagnation. This suggests that during periods of price increase, the increase in unit price is lower compared to a period of stagnation, possibly because airlines absorb some of the fuel cost increase through other means (such as improving operational efficiency, reducing other costs, or through market strategies).

On the other hand, the coefficient for log(Fuel_jet_gulf) during the decline period is negative but not significant, indicating that during periods of decline, the change in unit price is not notable compared to a period of stagnation, meaning that cost reductions are not fully passed on to consumers through price reductions.

The results of the control variables reveal significant trends regarding their impact on revenue per passenger-mile (PAX_REV_RPM). The load factor (LOAD_FACTOR) exhibits a statistically significant positive association with revenues, emphasizing the importance of optimizing this parameter to increase the financial performance of airlines. Similarly, higher market concentration (measured by HHI_RPM) is associated with increased revenues, while increased employee productivity (log(RTM_EMPLOYEE)) tends to reduce revenue per passenger-mile, possibly due to cost reduction strategies. Market share (MKT_SHARE_RPM) shows a negative association with revenues, likely due to increased competition. Additionally, GDP per capita appears to have a significant positive impact, while inflation (CPI_US) has a modest but positive impact on revenues. Finally, low-cost airlines display lower revenue per passenger-mile compared to traditional airlines, confirming the business model focused on competitive pricing of the former.

d. Results of Trend-Stratified Analysis

Récapitulatif des Modèles EFI dans trois bases			
	Dependent variable:		
	log(PAX_REV_RPM)		
	Stagnation	Croissance	Decroissance
	(1)	(2)	(3)
log(FUEL_JET_GULF)	0.109^{***}	0.045^{**}	0.091^{**}
	(0.041)	(0.022)	(0.038)
log(RPM)	0.162	0.056^{***}	-0.194^{***}
	(0.139)	(0.021)	(0.046)
log(LOAD_FACTOR)	0.409^{**}	0.925^{***}	0.773^{***}
	(0.181)	(0.091)	(0.125)
HHI_RPM	-0.074	4.105^{***}	12.049^{***}
	(1.356)	(0.418)	(1.556)
log(RTM_EMPLOYEE)	-0.357^{***}	-0.596^{***}	-0.441^{***}
	(0.130)	(0.047)	(0.054)
MKT_SHARE_RPM	-2.115^{**}	-0.698^{**}	-0.937^{*}
	(0.988)	(0.325)	(0.521)
OPE_EXP_RTM	0.393^{***}	0.044	-0.069^{***}
	(0.082)	(0.029)	(0.015)
log(GDPpCAPITA_CURRENT_US)	4.457	-1.545^{***}	1.274^{***}
	(4.039)	(0.262)	(0.440)
log(POPULATION_US)	-17.814	6.266^{***}	-0.946
	(16.058)	(1.010)	(1.786)
CPI_US	-0.037	0.047^{***}	-0.002
	(0.025)	(0.008)	(0.010)
IsLowCost_binary	0.419^{*}	0.077	0.286
	(0.235)	(0.119)	(0.182)
log(FUEL_JET_GULF):IsLowCost_binary	-0.147^{***}	-0.011	-0.076^{**}
	(0.052)	(0.023)	(0.036)
Observations	105	638	282
R^2	0.908	0.725	0.881
Adjusted R^2	0.890	0.715	0.871
Residual Std. Error	0.044 (df = 87)	0.125 (df = 613)	0.094 (df = 259)
F Statistic	50.531^{***} (df = 17; 87)	67.458^{***} (df = 24; 613)	87.269^{***} (df = 22; 259)
Note:			*p<0.1; **p<0.05; ***p<0.01
			reg1 : Modèle EFI dans la base de Stagnation
			reg2 : Modèle EFI dans la base de Croissance
			reg3 : Modèle EFI dans la base de decroissance

Table of results 2.4

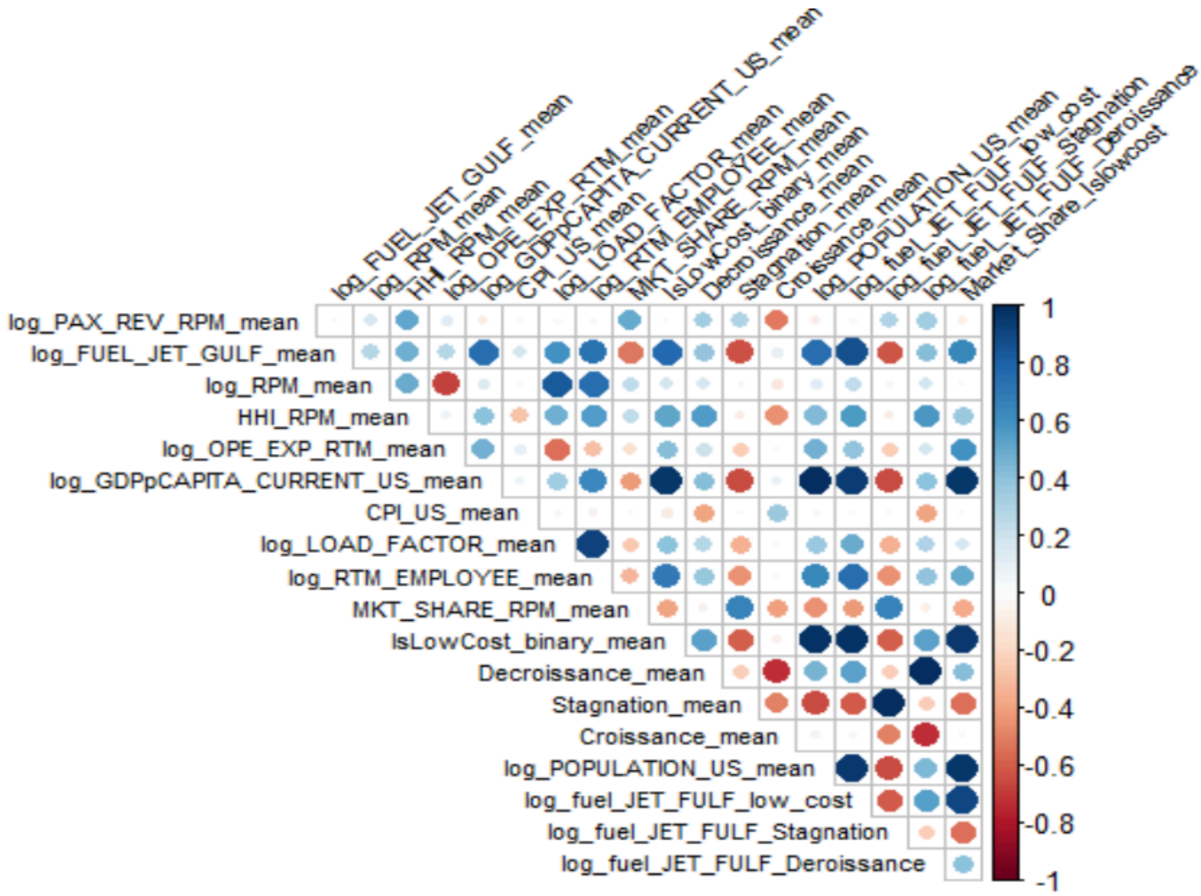
The obtained results confirm previous observations, highlighting significantly different impacts on PAX_REV_RPM depending on periods of fuel price fluctuations. To assess these variations, we utilized the Chow test to compare coefficients associated with different periods. Overall, we observe that the effect of periods of stable fuel prices on revenue per passenger mile (PAX_REV_RPM) is greater than that of periods of declining fuel prices, which in turn is greater than that of periods of increasing fuel prices. This finding underscores the importance of considering fuel price fluctuations in the analysis of airline financial performance, providing valuable insights for decision-making and risk management. However, it is worth noting the possibility of low statistical power of these results due to the limited number of observations, which could affect the reliability of the conclusions drawn.

e. Results of the ARDL Model

Given the richness in the time dimension of our database, we chose to estimate an AutoRegressive Distributed Lag (ARDL) model to analyze the long-term relationship between the explanatory variables and the dependent variable. This approach, which considers both short and long-term dynamic relationships, provides a comprehensive perspective on the impact of variables on our variable of interest.

Before estimating the model, we conducted various tests to ensure the robustness of our analysis. We specifically used the Augmented Dickey-Fuller unit root test to check for the stationarity of variables. Additionally, a graphical analysis of the series was performed to visualize the existence of a co-integration relationship, allowing us to detect long-term relationships between variables. Diagnostic tests to verify the absence of autocorrelation and heteroscedasticity were also conducted. These tests are essential to ensure the reliability of the results obtained.

Furthermore, we calculated individual means of the different variables and conducted a correlation analysis to better understand the relationships between them and detect multicollinearity issues. This additional step strengthens the validity of the results and helps obtain precise estimations of the ARDL model.



The correlation analysis highlights several cases of potential multicollinearity between variables, with a correlation threshold set at 0.8. Firstly, a strong correlation is observed between CPI_US_mean and log_GDPpCAPITA_CURRENT_US_mean, suggesting a close relationship between the Consumer Price Index and the logarithm of GDP per capita.

Similarly, a significant correlation is identified between Log_RPM_mean and log_LOAD_FACTOR_mean, indicating a relationship between the logarithm of market traffic and the logarithm of the load factor of airlines. Furthermore, an association is noticed between MKT_SHARE_RPM_mean, HHI_RPM, and CPI_US_mean, highlighting an interconnection between the market share in market traffic, the concentration index, and the consumer price index.

A strong correlation is also observed between log_GDPpCAPITA_CURRENT_US_mean and Market_Share_Islowcost. Lastly, a significant correlation is noticed between Log_RTM_EMPLOYEE_mean and log_LOAD_FACTOR_mean, as well as between log_LOAD_FACTOR_mean and OPE_EXP_RTM_mean, highlighting potential relationships between the logarithm of employees per revenue ton-mile, the logarithm of the load factor, and operational expenses per revenue ton-mile. Among these correlations, Log_RPM_mean and log_LOAD_FACTOR_mean stand out as the variables most likely to generate multicollinearity, being strongly correlated with each other. To address this issue, two distinct models were constructed, one including Log_RPM_mean and the other including log_LOAD_FACTOR_mean. We retained the model that minimizes the AIC while providing robust and significant results for our analysis.

The choice of lags or no lags for the PAX_REV_RPM variable and the JET_FUEL_GULF variable was based on the Akaike Information Criterion (AIC). In accordance with the literature, no lags were retained for the control variables.

The results of the different chosen specifications are presented in the following table:

	ardl_RPM		ardl_RPM_INT		ardl_RPM_INT_GDP	
(Intercept)	-1.687*	(0.753)	-2.072**	(0.645)	-2.459*	(1.079)
L(log_PAX_REV_RPM_mean, 1)	0.502***	(0.084)	0.440***	(0.087)	0.452***	(0.091)
log_FUEL_JET_GULF_mean	0.035*	(0.015)	0.060**	(0.018)	0.058**	(0.019)
log_GDPpCAPITA_CURRENT_US_mean	0.000	(0.035)			0.064	(0.142)
log_RPM_mean	0.078*	(0.037)	0.095*	(0.042)	0.082	(0.051)
HHI_RPM_mean	1.705***	(0.469)	1.267*	(0.500)	1.296*	(0.506)
log_RTM_EMPLOYEE_mean	-0.149**	(0.046)	-0.164***	(0.048)	-0.158**	(0.050)
Market_Share_Islowcost			0.042	(0.079)	-0.092	(0.309)
log_fuel_JET_FULF_croissance			-0.012*	(0.005)	-0.013*	(0.006)
log_fuel_JET_FULF_Deroissance			-0.008	(0.005)	-0.009	(0.006)
Deviance	0.254		0.238		0.238	
AIC	-327.164		-329.864		-328.086	
N	107		107		107	

Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05

Table of results 2.5

The results obtained from the ARDL models confirm the conclusions previously drawn with the individual fixed effects OLS model. Firstly, the lagged variable of the log of revenue per passenger-mile (L(log_PAX_REV_RPM_mean, 1)) displays a significant positive coefficient, suggesting a positive feedback in the system, indicating that previous variations in revenue per passenger-mile have a significant effect on its current variations. Similarly, an increase in the average fuel price (log_FUEL_JET_GULF_mean) is associated with a significant increase in the log of average revenue per passenger-mile, highlighting the significant impact of fuel costs on airline profitability. However, the log of GDP per capita (log_GDPpCAPITA_CURRENT_US_mean) does not show a significant influence on the log of revenue per passenger-mile, probably due to its strong correlation with Market_Share_Islowcost, as indicated by the AIC analysis between the models in columns 2 and 3.

Furthermore, the significant positive coefficients of log_RPM_mean and HHI_RPM_mean underline the positive effects of the market traffic log and the market concentration index on the log of revenue per passenger-mile, highlighting the importance of market demand and competitive structure for the financial performance of airlines. On the other hand, increased employee productivity (log_RTM_EMPLOYEE_mean) is associated with a significant decrease in the log of revenue per passenger-mile, which may reflect cost-cutting strategies that can lead to lower fares.

Regarding the ARDL model with interaction between trend variables and fuel price (column 2), the results remain consistent with the previous model (column 1) for the shared variables, with the addition of an interaction with Market_Share_Islowcost. Although this interaction slightly influences the coefficients of other variables, their statistical significance remains stable.

Finally, the results of the model with the addition of log_GDPpCAPITA_CURRENT_US_mean suggest similar trends to ardl_RPM_INT. However, the log of GDP per capita is not significant in this model, which may be attributed to the strong correlation with Market_Share_Islowcost.

Overall, these ARDL models provide valuable insights into the factors influencing revenue per passenger-mile, shedding light on the importance of fuel price, market traffic, market concentration, and employee productivity in the airline industry. The estimated elasticities from our selected ARDL model (column 2), which was more parsimonious and also minimized the AIC, are as follows:

	Estimated Elasticities	
ardl_RPM_INT	Short term	Long term
	0 ,06	0,1071

The estimated elasticities indicate the sensitivity of revenue per passenger mile (PAX_REV_RPM) to variations in fuel prices (FUEL_JET_PRICE). In the short term, for a 1%

increase in fuel prices, revenue per passenger mile increases by 0.06%. This suggests a relatively weak response of revenue per passenger mile to immediate changes in fuel prices. However, in the long term, a 1% increase in fuel prices leads to a 0.1071% increase in revenue per passenger mile. This higher elasticity in the long term indicates a more significant response of revenue per passenger mile to sustained variations in fuel prices, likely due to gradual adjustments in airlines business and pricing strategies. In summary, these elasticities suggest a complex dynamic between fuel prices and revenue per passenger mile, with more pronounced effects in the long term compared to the short term.

Summary of the findings

In this study, we aimed to estimate the elasticity of airline ticket prices concerning fuel prices. To achieve this, we utilized two panel databases that were cleaned and explored to extract the most reliable and relevant observations for our econometric analysis.

The first database, comprising data from the largest airlines globally over the decade 2010-2019, allowed us to obtain overall results and test differences in elasticity among regions worldwide. Across all models tested, dummy variables representing the seven continents and the business model consistently demonstrated non significance. In the two main categories of models we estimated (Fixed Effects OLS and Panel Within Model), the elasticity of ticket prices/fuel costs ranged between 0.24% and 0.42%. In other words, a 1% increase in airline fuel costs was followed by a 0.24% to 0.42% increase in ticket prices paid by consumers. The elasticity with respect to market fuel prices ranged between 0.30 and 0.38%. Additionally, the elasticity calculated based on the lag of market fuel prices ranged from 0.22 to 0.26%. Finally, in line with the literature, we introduced interactions between our variables. The first interaction involved fuel prices (market) and the variable 'low_cost'. This interaction was found to be negative, suggesting that low-cost airlines pass on fuel price fluctuations to ticket prices less than non-low-cost carriers. Thus, the elasticity obtained was 0.32% for non-low-cost airlines and only 0.065% for low-cost carriers. The second interaction segmented airlines based on their propensity to offer long-haul flights. We concluded that the longer the flight distance, the more sensitive ticket prices are to fuel price fluctuations. An attempted implementation of a PanelVar model yielded inconclusive results.

Our second database concerns the 14 largest airlines in the United States, covering the period 1995-2023. These panel data are available quarterly for each airline. The nearly 30-year span of observations with high frequency allowed us to estimate both short and long-term elasticities and highlight the variability of ticket price/fuel price elasticity depending on whether fuel prices are rising or falling. To model these elasticities, we evaluated Fixed Effects OLS and ARDL models. In the Fixed Effects OLS model, the overall elasticity calculated was 0.047%. By adding a first interaction of fuel price with the low_cost variable, we observed that it was significant and positive, suggesting that ticket price elasticity is higher among low-cost airlines, contradicting the results obtained from the MUST database. Another interesting result concerns the variations in elasticity depending on the dynamics of fuel price changes (growth, stagnation, or decline). Through Fixed Effects OLS models with interactions, stratified analysis, and ARDL, we observed that all results pointed to a lower elasticity during periods of fuel price growth than during periods of stagnation or decline. The robustness of this result underscores the fact that airlines pass on fuel price increases less when prices are rising. Finally, the ARDL model highlighted a contrast between the long-term elasticity, which is around 0.06%, and the short-term elasticity, which reaches 0.10%.

Other variables also appear to have a positive impact on ticket prices, such as other costs (excluding fuel costs). However, the significance and sign of many variables depend on the specification of each model, similar to ASK, RPK, market concentration/competition, and GDP per Capita. These variables were primarily used as controls; further analysis would be necessary to ascertain the exact effect of these variables on ticket prices.

MUST	
Modèle	**Elasticité estimée**
OLS effets fixes inviduels	0.39
Within	0.31
OLS lag FUEL_JET_GULF	0.23
Within lag FUEL_JET_GULF	0.26
OLS lag + Interaction is_lowcost	0.07 lowcost / 0.32 non lowcost
OLS lag + Interaction LH_ratio	0.21 medium LH_ratio / 0.29 Big LH_ratio
US	
Modèle	**Elasticité estimée**
OLS à effets fixes individuels	0.047
OLS interaction is_lowcost	0.071 lowcost / 0.025 non lowcost
OLS interaction tendances marché	0.055 / 0.036 en période de croissance
ARDL	0.06 court terme/ 0.10 long terme

Comparison with literature

Our study, in line with the findings of Alexandre H. Wolter, Thorsten Ehlers et al. (2021), highlights the positive impact of fuel price variations on airline ticket prices. We estimate an elasticity between 0.30% and 0.38% in year T, slightly lower when considering the fuel price lagged by one year. The literature, particularly the work of Barbara Gaudenzi and Alessandro Bucciol (2016), suggests that airlines pass on fuel price fluctuations to ticket prices less aggressively. Our results from the MUST database support this notion, as the elasticity for non-low-cost airlines is 0.32% compared to 0.065% for low-cost carriers. However, this conclusion appears to be contradicted by our analysis of US data.

Furthermore, our study introduces an interesting difference in elasticity between upward and downward phases of fuel prices. We conclude that the elasticity is lower during periods of price increases than during periods of decrease or stagnation. This observation echoes the findings of Werner D. Kristjanpoller, Diego Concha (2016) on airline stock price variations based on underlying trends in the fuel market.

Limitation/Future work

Our study also has certain limitations and spaces for further research. The discrepancy in the results of the two database could be caused by a variety of factors: the structure of the global airline market and the U.S. airline market may differ significantly, the U.S. market might be more concentrated, subject to less regulation, or have a higher degree of competition, whereas the global market might include a wide range of market structures, from highly competitive to nearly monopolistic markets, which could affect airlines' sensitivity to changes in fuel prices.

Moreover, different airlines may have different cost structures. For instance, some international airlines may rely more on long-haul flights, making fuel costs a higher proportion of their total costs, thereby making them more sensitive to changes in fuel prices. In contrast, airlines operating in the U.S. domestic market may have more short-haul flights, with fuel costs constituting a lower proportion of their total costs. Additionally, global and U.S. markets might be influenced by different macroeconomic environments and external events (such as oil price shocks, geopolitical events). These factors could indirectly affect airlines' cost pass-through mechanisms and pricing strategies, thereby impacting the estimation of elasticity coefficients.

Furthermore, according to forecasts by the International Energy Agency (IEA), energy prices are expected to rise significantly over the next 20 years. This change is structural, not a fluctuation that can be reversed in the short term, but a stable trend in the long term. Establishing a model capable of explaining structural changes (such as the long-term rising trend in energy prices) on economic variables (such as inflation, consumer behavior) could become a direction for improvement, such as using the beginning of these 20 years as a "stagnation period" as the baseline for observing and comparing future price changes, etc.

In light of our analysis, it has become apparent that further research is warranted to address several key issues and explore more nuanced aspects of our model. Firstly, we have identified a potential problem of heteroskedasticity, which could impact the reliability of our results. This issue underscores the need for more robust statistical techniques or data transformations to mitigate its effects. Additionally, while our current model incorporates linear interactions, there is scope for investigating more complex non-linear interactions that may better capture the relationships between variables. Exploring such interactions could offer deeper insights into the dynamics of the phenomena under study. Moreover, the methods we have employed thus far represent only a preliminary exploration of the research questions at hand. There is considerable potential for delving deeper into the topic through alternative methodologies or by refining our current approach. Therefore, future research endeavors should focus on addressing these considerations to enhance the comprehensiveness and validity of our findings.

Bibliography

Atems, B., Bachmeier, L., & Williams, C. (2019). Do jet fuel price movements help forecast airline fares and the demand for air travel? *Applied Economics Letters*, *26*(11), 877‑882. https://doi.org/10.1080/13504851.2018.1510466

Gaudenzi, B., & Bucciol, A. (2016). Jet fuel price variations and market value : A focus on low-cost and regular airline companies. *Journal of Business Economics and Management*, *17*(6), 977‑991. https://doi.org/10.3846/16111699.2016.1209784

Kristjanpoller, W. D., & Concha, D. (2016). Impact of fuel price fluctuations on airline stock returns. *Applied Energy*, *178*, 496‑504. https://doi.org/10.1016/j.apenergy.2016.06.089

Pamungkas, D. P., & Suhadak, S. (2017). *The Effect of Jet Fuel Price and Macroeconomics Variables on Profitability of Airline Industry in Asia (Study at Airline Companies in Indonesia, India, and China Period 2006-2015)* (Numéro 3) [Journal:eArticle, Brawijaya University]. https://www.neliti.com/publications/187755/

Sigmund, Michael and Ferstl, Robert, Panel Vector Autoregression in R with the Package Panelvar (February 18, 2019). Quarterly Review of Economics and Finance, 2019, Available at SSRN: https://ssrn.com/abstract=2896087 or http://dx.doi.org/10.2139/ssrn.2896087)

Wattanatorn, W., & Kanchanapoom, T. (2012). Oil Prices and Profitability Performance : Sector Analysis. *Procedia - Social and Behavioral Sciences*, *40*, 763‑767. https://doi.org/10.1016/j.sbspro.2012.03.263

Wolter, A. H., Ehlers, T., Luetjens, K., & Gollnick, V. (2021). Commodity price pass-through in the US airline industry and the hidden perks of consolidation. *Journal of Air Transport Management*, *95*, 102100. https://doi.org/10.1016/j.jairtraman.2021.102100

Estimating-air-travel-demand-elasticities—By-intervistas.pdf. (s. d.). Consulté 28 mars 2024, àl'adressehttps://www.iata.org/en/iata-repository/publications/economic-reports/estimating-air-travel-demand-elasticities---by-intervistas/

Appendix

We began with a comprehensive model containing more variables than necessary and employed the drop1() function in R. This function enabled us to systematically remove variables from the model and compare it with the full model, as shown in the following screenshot :

```
                              Df Sum of Sq    RSS     AIC  F value    Pr(>F)    
<none>                                     52.912 -2240.8                       
log_FUEL_COSTS_ASK             1   12.5230 65.435 -2066.5 191.2346 < 2.2e-16 ***
log_COST_management            1    7.7639 60.676 -2129.2 118.5604 < 2.2e-16 ***
log_ASK                        1    0.5759 53.488 -2233.8   8.7948 0.0031099 ** 
LOAD_FACTOR_100                1    1.1532 54.065 -2224.9  17.6105 3.011e-05 ***
LH_RATIO_100                   1    0.6828 53.595 -2232.2  10.4273 0.0012918 ** 
MKT_CONCENTRATION_100          1    0.3540 53.266 -2237.3   5.4064 0.0203099 *  
log_GDPpCAPITA_CURRENT_REGION  1    0.0770 52.989 -2241.6   1.1754 0.2786135    
HHI_ASK_REGION                 1    0.0279 52.940 -2242.4   0.4268 0.5137714    
log(POPULATION_REGION)         1    1.1952 54.107 -2224.3  18.2520 2.166e-05 ***
CPI_REGION                     1    0.3260 53.238 -2237.7   4.9782 0.0259424 *  
FLOW_INT                       1    0.0169 52.929 -2242.6   0.2580 0.6116378    
IS_LOW_COST                    1    2.0356 54.948 -2211.5  31.0850 3.368e-08 ***
BELONGS_to_GOVERNMENT          1    0.3177 53.230 -2237.8   4.8522 0.0278920 *  
BELONGS_to_PRIVATE_I           1    0.2262 53.138 -2239.3   3.4539 0.0634669 .  
ASIA                           1    0.5034 53.415 -2235.0   7.6870 0.0056900 ** 
NORTH_AMERICA                  1    0.3918 53.304 -2236.7   5.9832 0.0146550 *  
AFRICA                         1    0.8424 53.754 -2229.7  12.8642 0.0003551 ***
OCEANIA                        1    0.2766 53.188 -2238.5   4.2236 0.0401865 *  
MIDDLE_EAST                    1    0.8132 53.725 -2230.2  12.4173 0.0004493 ***
SOUTH_AMERICA                  1    0.0068 52.919 -2242.7   0.1037 0.7475612    
AIRLINE_ID                     1    0.0016 52.914 -2242.8   0.0245 0.8756014    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value indicates whether the removal of a particular variable leads to a significant reduction in the model's fit. A p_value higher than 0.05 signifies then that the variable is useless in the model. In this case, we removed HHI_ASK_REGION at the first stage.

```
> vif(OLS)
           log_FUEL_COSTS_ASK           log_COST_management                       log_ASK 
                     1.523529                      2.388497                      1.632570 
              LOAD_FACTOR_100                  LH_RATIO_100         MKT_CONCENTRATION_100 
                     1.745870                      5.323004                      1.408340 
log_GDPpCAPITA_CURRENT_REGION                HHI_ASK_REGION        log(POPULATION_REGION) 
                     7.656757                      7.346154                     11.623327 
                   CPI_REGION                      FLOW_INT                   IS_LOW_COST 
                     3.696257                      6.502565                      2.273141 
        BELONGS_to_GOVERNMENT          BELONGS_to_PRIVATE_I                          ASIA 
                     2.757808                      2.129971                      4.824591 
                NORTH_AMERICA                        AFRICA                       OCEANIA 
                     3.283262                      1.812075                      5.326653 
                  MIDDLE_EAST                 SOUTH_AMERICA                    AIRLINE_ID 
                     3.074814                      2.317407                      1.059900 
```

As evidenced by the VIF (Variance Inflation Factor) function, many variables exceed the indicative threshold of 5, potentially indicating collinearity among our variables. By controlling for collinearity, we ensure that the model is robust, stable, and that the coefficient estimates are accurate and interpretable.

We do the process again :

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			52.940	-2242.4			
log_FUEL_COSTS_ASK	1	12.5147	65.455	-2068.3	191.2432	< 2.2e-16	***
log_COST_management	1	7.7497	60.690	-2131.0	118.4265	< 2.2e-16	***
log_ASK	1	0.6607	53.601	-2234.1	10.0968	0.0015417	**
LOAD_FACTOR_100	1	1.1253	54.065	-2226.9	17.1962	3.727e-05	***
LH_RATIO_100	1	0.7346	53.674	-2232.9	11.2261	0.0008439	***
MKT_CONCENTRATION_100	1	0.3450	53.285	-2239.0	5.2721	0.0219241	*
log_GDPpCAPITA_CURRENT_REGION	1	0.0503	52.990	-2243.6	0.7683	0.3810173	
log(POPULATION_REGION)	1	1.7545	54.694	-2217.3	26.8108	2.834e-07	***
CPI_REGION	1	0.2989	53.239	-2239.7	4.5683	0.0328687	*
FLOW_INT	1	0.0146	52.954	-2244.2	0.2231	0.6368247	
IS_LOW_COST	1	2.0418	54.982	-2213.0	31.2020	3.177e-08	***
BELONGS_to_GOVERNMENT	1	0.4100	53.350	-2238.0	6.2647	0.0125129	*
BELONGS_to_PRIVATE_I	1	0.2448	53.185	-2240.6	3.7410	0.0534398	.
ASIA	1	0.4897	53.430	-2236.7	7.4828	0.0063656	**
NORTH_AMERICA	1	0.3650	53.305	-2238.7	5.5781	0.0184227	*
AFRICA	1	0.8146	53.754	-2231.7	12.4476	0.0004421	***
OCEANIA	1	0.2576	53.197	-2240.3	3.9368	0.0475789	*
MIDDLE_EAST	1	0.7941	53.734	-2232.0	12.1345	0.0005215	***
SOUTH_AMERICA	1	0.0484	52.988	-2243.6	0.7390	0.3902412	
AIRLINE_ID	1	0.0018	52.942	-2244.4	0.0272	0.8691183	

Similarly, here, we decide to remove the variable FLOW_INT.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			52.954	-2244.2			
log_FUEL_COSTS_ASK	1	12.5039	65.458	-2070.2	191.2613	< 2.2e-16	***
log_COST_management	1	7.8261	60.781	-2131.8	119.7093	< 2.2e-16	***
log_ASK	1	0.6875	53.642	-2235.4	10.5158	0.0012320	**
LOAD_FACTOR_100	1	1.1490	54.103	-2228.3	17.5757	3.065e-05	***
LH_RATIO_100	1	3.1920	56.146	-2197.6	48.8253	5.844e-12	***
MKT_CONCENTRATION_100	1	0.3313	53.286	-2241.0	5.0676	0.0246443	*
log_GDPpCAPITA_CURRENT_REGION	1	0.0618	53.016	-2245.2	0.9451	0.3312675	
log(POPULATION_REGION)	1	1.7710	54.725	-2218.8	27.0901	2.463e-07	***
CPI_REGION	1	0.2989	53.253	-2241.5	4.5722	0.0327925	*
IS_LOW_COST	1	2.0300	54.984	-2214.9	31.0512	3.422e-08	***
BELONGS_to_GOVERNMENT	1	0.4288	53.383	-2239.5	6.5589	0.0106162	*
BELONGS_to_PRIVATE_I	1	0.2618	53.216	-2242.1	4.0039	0.0457282	*
ASIA	1	0.4785	53.433	-2238.7	7.3192	0.0069655	**
NORTH_AMERICA	1	0.3908	53.345	-2240.1	5.9780	0.0146972	*
AFRICA	1	0.8529	53.807	-2232.9	13.0465	0.0003226	***
OCEANIA	1	0.2595	53.214	-2242.1	3.9691	0.0466781	*
MIDDLE_EAST	1	1.2891	54.244	-2226.2	19.7183	1.022e-05	***
SOUTH_AMERICA	1	0.0527	53.007	-2245.3	0.8058	0.3696302	
AIRLINE_ID	1	0.0026	52.957	-2246.1	0.0397	0.8421157	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Based on these new results, we should remove GDP. However, as you can see below, the coefficients of POPULATION and CPI from the regression without GDP are not of the expected sign. CPI represents price increases, so we would expect an increase in the CPI variable to lead to a significant increase in UTKT_PRICE, not a decrease. Regarding population, according to economic concepts, if, at constant supply, demand increases, then prices should increase. However, when the population increases, it potentially indicates an increase in the number of travelers and thus an increase in demand, yet the coefficient is also negative.

```
Results
=================================================================
                                    log_UTKT_PRICE
                                         OLS
                                     Inidividual
-----------------------------------------------------------------
log_FUEL_COSTS_ASK                     0.442***
                                       (0.031)

log_COST_management                    0.396***
                                       (0.035)

log_ASK                                0.034***
                                       (0.010)

LOAD_FACTOR_100                       -0.008***
                                       (0.002)

LH_RATIO_100                          -0.005***
                                       (0.001)

MKT_CONCENTRATION_100                  0.002*
                                       (0.001)

log(POPULATION_REGION)                -0.183***
                                       (0.035)

CPI_REGION                            -0.021***
                                       (0.006)

IS_LOW_COST                           -0.149***
                                       (0.027)

BELONGS_to_GOVERNMENT                  0.087*
                                       (0.035)

BELONGS_to_PRIVATE_I                   0.052
                                       (0.027)

Constant                               4.793***
                                       (0.692)

Variable omise                          REGION
Variable omise                            ID

N                                       830
R2                                     0.632
Adjusted R2                            0.624
Residual Std. Error               0.256 (df = 811)
=================================================================
Notes:                   ***Significant at the 0.1 percent level.
                          **Significant at the 1 percent level.
                           *Significant at the 5 percent level.
```

```
> vif(OLS)
   log_FUEL_COSTS_ASK    log_COST_management                log_ASK        LOAD_FACTOR_100
             1.489813               2.305165               1.495154               1.690872
         LH_RATIO_100  MKT_CONCENTRATION_100 log(POPULATION_REGION)             CPI_REGION
             1.504085               1.367030               9.008338               1.960277
          IS_LOW_COST  BELONGS_to_GOVERNMENT   BELONGS_to_PRIVATE_I                   ASIA
             2.204784               2.469282               2.077592               3.348889
        NORTH_AMERICA                 AFRICA                OCEANIA            MIDDLE_EAST
             2.898000               1.592561               5.221461               2.128920
        SOUTH_AMERICA             AIRLINE_ID
             1.452716               1.050885
```

Furthermore, we can see again from the VIF that certain variables exhibit strong collinearity, notably POPULATION.

Alternatively, if we replace CPI and POPULATION with GDP, then the results for the VIF are satisfactory, indicating better control of collinearity. Additionally, GDP has the expected sign.

We conclude that the best model, based on our analysis, is the one shown previously in the analysis section

=> Hausmann test on the first model

```
> phtest( within, random )

        Hausman Test

data:  log_UTKT_PRICE ~ log_FUEL_COSTS_ASK + log_COST_management + log_ASK +  ...
chisq = 66.256, df = 9, p-value = 8.216e-11
alternative hypothesis: one model is inconsistent
```

.=>VIF on the first model

```
> vif(OLS)
          log_FUEL_COSTS_ASK          log_COST_management                      log_ASK               LOAD_FACTOR_100
                    1.365566                     2.277370                     1.508947                      1.651850
                LH_RATIO_100        MKT_CONCENTRATION_100 log_GDPpCAPITA_CURRENT_REGION                  IS_LOW_COST
                    1.467965                     1.366978                     3.078264                      2.228247
       BELONGS_to_GOVERNMENT          BELONGS_to_PRIVATE_I                         ASIA                NORTH_AMERICA
                    2.433195                     2.080242                     2.764046                      1.778789
                      AFRICA                      OCEANIA                  MIDDLE_EAST                 SOUTH_AMERICA
                    1.627226                     1.135790                     1.484768                      1.372976
                  AIRLINE_ID
                    1.049351
```