



HAL
open science

Estimation of consumer demand on the air transport market

Alexandra Belova

► **To cite this version:**

Alexandra Belova. Estimation of consumer demand on the air transport market. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2018. English. NNT : 2018PA01E040 . tel-02015575

HAL Id: tel-02015575

<https://theses.hal.science/tel-02015575v1>

Submitted on 12 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**UNIVERSITE PARIS 1 PANTHEON SORBONNE
UFR D'ECONOMIE**

THESE

Pour l'obtention du titre de Docteur en économie
Présentée et soutenue publiquement le 19 décembre 2018
par Alexandra BELOVA

Estimation of consumer demand on the air transport market

Sous la direction de M. Philippe GAGNEPAIN

Professeur Université Paris 1 Panthéon-Sorbonne

Membres du Jury:

Mme Catherine Muller, Professeur à Toulouse Business School

Mme Claudine Desrieux, Professeur à l'Université Paris II Panthéon-Assas

M. Angelo Secchi, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Summary

Nowadays one of peculiarities of the liberalized airline market is a huge divergence of ticket prices for the same flights. Mostly it reflects the companies' being unable to easily change the volumes of production or/and store them, so they pay special attention to the behavior of other companies and their volumes of production. The development and use of the yield management models (seat allocation models) have centered on airlines offering a variety of different types of fares for travel on the same flight. The goal of this dissertation is to construct a number of economic models to explain the price dispersion on the airline market from the different points of view. In Part 3, I create a direct price model which explains how different product and consumer characteristics influence the price level. It is shown how different attributes like the moment of ticket reservation, ticket class, weekday of the departure and number of coupons define the price and how it corresponds to the consumer characteristics (gender, income, age, etc.). Part 4 is devoted to the differences of the price level from the competition point of view. In a strategic game where firms compete against each other the set of rationalizable strategies for each player entails all the best responses to the others' decisions. This chapter proposes an empirical test of the existence of the unique Nash equilibrium in a Cournot oligopoly. In Part 5 I treat an airline passenger market as a market with the product differentiation and apply a multinomial logit model to calculate price elasticities. The logit model (with a special focus on the consumers heterogeneity) estimates how the different product characteristics influence the market shares.

Résumé

Une des particularités du marché des compagnies aériennes libéralisé est la grande divergence des prix des billets pour les mêmes vols. Cela reflète principalement l'incapacité des entreprises à modifier facilement les volumes de production et / ou à les stocker, de sorte qu'elles accordent une attention particulière au comportement des autres entreprises et à leurs volumes de production. Le développement et l'utilisation des modèles de "yield management" (modèles d'attribution des sièges) ont été centrés sur les compagnies aériennes proposant différents types de tarifs pour un même vol. L'objectif de cette thèse est de construire un certain nombre de modèles économiques pour expliquer la dispersion des prix sur le marché du transport aérien à partir de différents points de vue. Dans le chapitre 3, je crée un modèle de prix direct qui explique comment différentes caractéristiques du produit et du consommateur influencent le niveau de prix. Il est montré comment différents attributs tels que le moment de la réservation du billet, la classe du billet, le jour de départ et le nombre de coupons définissent le prix et comment il correspond aux caractéristiques du consommateur (sexe, revenu, âge, etc.). Le chapitre 4 est consacré aux différences de niveau de prix du point de vue de la concurrence. Dans un jeu stratégique où les entreprises se font concurrence, l'ensemble de stratégies rationalisables pour chaque joueur implique toutes les meilleures réponses aux décisions des autres. Ce chapitre propose un test empirique de l'existence de l'équilibre de Nash unique dans un oligopole de Cournot. Dans le chapitre 5, je traite le marché des passagers aériens comme un marché différenciant les produits et applique un modèle logit multinomial pour calculer les élasticités-prix. Le modèle logit (mettant particulièrement l'accent sur l'hétérogénéité des consommateurs) estime de quelle manière les différentes caractéristiques du produit influencent les parts de marché.

Keywords

Airlines - hedonic price function - willingness to pay - demand

Competition - price policies - price elasticities - costs

Cournot - Nash equilibrium - rationalizable strategies - market stability

Mot-clés

Airlines - fonction de prix hédonique - volonté de payer - demande

Concurrence - politiques des prix - élasticités des prix - coûts

Cournot - équilibre de Nash - stratégies rationalisables - stabilité du marché

“L’université Paris I Panthéon-Sorbonne n’entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.”

Acknowledgements

A thesis may look like an individual effort and in many ways it is. However, its outcome does depend (sometimes greatly) on the contributions of many other people than its author. In my case, there were many, and I hope I do not forget any. If I do, please forgive my failing memory.

Foremost, I am very grateful to my supervisor Philippe Gagnepain. Philippe has supported me from the very beginning to this day. He has been a wonderful supervisor, always very supportive, encouraging and patient. For all this, I am thousand times thankful.

I want also to thank greatly Stéphane Gauthier with whom I worked on the second chapter of this thesis. He has been always so inspiring, motivated, so bright. I have learnt a great number of things from him.

I would also like to thank all the people who are or were in charge of the Doctoral School in university Paris 1. My thoughts go especially to Loic Sorel, who has always been ready to help on this uneasy way.

Of course, I am thankful to my institutions in which I spent this time: the university Paris 1 and the Paris School of Economics. I am also grateful to all the professors of Master program, who gave me the knowledge and competences to fulfill this task and to many of my professors during my school and university years in Moscow. Without the education they gave me the PhD degree would be unachievable for me.

Many researchers from different universities around the world made some comments which were valuable for this thesis. I am thankful to all of them for making this thesis more accurate and scientific.

Many thanks go to my colleagues from the doctoral school. I changed a few offices,

but in all of them it felt like home. Special thanks to Leontine and Marko for guiding me through all the bureaucratic barriers I faced.

I am forever grateful to my comrades, my best friends, my dear girls: Anastasia and Anna. It is absolutely impossible to describe in a few phrases all the things you have done for me, it is just so nice to know that you have somebody to have you back.

Another person I am grateful is Anastasia Popova for being my personal the most effective “project manager”.

Probably he does not know it, but I am very grateful to my Maxim, my husband, for waiting, supporting and helping me.

I would like to thank my wonderful family for support and inspiration: my parents Irina and Valerii, my brother Nikita. A separate gratitude goes to my sister, Ulyana, who has always believed in me unconditionally, and to whom I would like to dedicate this thesis.

And the last thanks go to the amazing city of Paris for welcoming me all these years.

Contents

I	Introduction Générale	10
1	Stratégies de prix sur le marché aérien international	10
1.1	Introduction	10
1.2	Le modèle	11
1.3	Les données	12
1.4	Les résultats d'estimation	13
1.5	Les extensions proposés	15
2	Pouvoir de marché et volatilité dans le secteur du transport aérien	16
2.1	Le modèle empirique et les résultats d'estimation	17
2.1.1	L'équation de coûts	19
2.1.2	L'équation de demande	19
2.2	L'indice de stabilité	20
2.3	Les autres constatations	20
3	Estimation de la demande de services de transport aérien	22
3.1	Introduction	22
3.2	La stratégie d'estimation	23
3.3	Les données	24
3.4	Les résultats d'estimation	25
3.4.1	Le modèle de demande pour l'itinéraire Los Angeles - Taipei	25
3.4.2	Les élasticités-prix	27
3.5	Les extensions proposés	28
II	General Introduction	29
III	Price Strategies on the International Airline Market	32
1	Introduction	32
2	Yield management	36
3	Model	38
3.1	Hedonic price function	38
3.2	First step. Price function estimation	39
3.3	Second step. Consumers willingness to pay	41
4	Data	43
4.1	Data handling and descriptive statistics	45
4.2	Reservation moment and the personal characteristics of the consumers	49
5	Estimation results	50

6	Summary, Conclusion and Proposed extensions	58
IV	Market power and volatility in the airline industry	61
1	Introduction	62
2	Theoretical benchmark	65
2.1	General framework	65
2.2	Linear-quadratic specification	72
2.3	Stability from sub-market observations	75
3	Empirical illustration to the airline industry	79
3.1	Data	80
3.2	Costs	83
3.3	Demand	88
4	Stability index	90
4.1	Assessing the Nash equilibrium	97
5	Conclusion	100
V	Demand estimation for airline passenger services	102
1	Introduction	102
2	Econometric model	104
2.1	Consumers Utility	105
2.2	Choice probabilities	107
2.3	Estimation Strategy	108
2.4	Price elasticities of demand	109
3	Data	110
3.1	Survey of International Air Travelers	110
3.2	Data handling and descriptive statistics	112
3.3	Choice set	114
4	Estimation results	117
4.1	Demand models for the route Los Angeles - Taipei	117
4.2	Price elasticities	119
4.3	Demand models for all the routes ¹	121
5	Conclusions	125

¹The list of routes is in the Table 27, in Annex

List of Tables

1	Descriptive statistics (1)	48
2	Descriptive statistics (2)	48
3	Descriptive statistics per route	50
4	Hedonic price function	54
5	Reservation moment, WTP	55
6	Business class, WTP	56
7	Descriptive statistics per carrier.	81
8	Direct and indirect routes	83
9	Descriptive statistics. Route level data.	83
10	Estimated cost function C_{aft}	87
11	Estimated demand function	89
12	Market fundamentals sufficient statistics for stability	91
13	Determinants of the stability index	94
14	Airlines network from the main airlines	96
15	Nash departure and stability index	99
16	Descriptive statistics	116
17	Descriptive statistics (2)	116
18	Multinomial logit model, market Los Angeles, US - Tapei, Taiwan	118
19	Mixed logit model, market Los Angeles, US - Tapei, Taiwan	120
20	Semi-price elasticity of the options	121
21	Price elasticities of the options	122
22	Semi - price elasticities between direct-indirect tickets	122
23	Price elasticities between direct-indirect tickets	122
24	Significance of variables in the multinomial logit models on all markets	123
25	Significance of variables in the estimated mixed logit models on all markets	124
26	Average influence of the parameters in the models	125
27	List of routes with descriptive statistics	132

Part I

Introduction Générale

1 Stratégies de prix sur le marché aérien international

1.1 Introduction

Cette thèse s'intéresse à la forte dispersion des prix des billets d'avion qui est apparue avec une nouvelle politique de tarification appelée "yield management". Elle propose une analyse empirique du lien entre les prix des billets sur le marché international, les caractéristiques du vol, les caractéristiques personnelles des consommateurs et leur disposition à payer. Il s'agit d'appréhender le marché des services aériens sous un angle totalement nouveau. Je propose une analyse de prix en contrôlant les principaux outils "yield management" et utilise pour cela une approche hédonique spécialement développée pour les marchés à produits différenciés.² Je propose également un modèle afin d'étudier le lien entre les attributs du service, les prix et les caractéristiques du consommateur. Pour atteindre cet objectif, je considère un système d'équations : la première est une fonction de prix hédonique où le prix dépend des caractéristiques du produit. Les autres sont des équations qui capturent la disposition à payer des consommateurs pour un produit en fonction de différentes variables telles que le revenu, le sexe, ou l'âge.

Cette thèse s'appuie sur l'exploitation d'une nouvelle base de données, le Survey of International Air Travelers (2005). Cette base de données nous permet de contrôler les paramètres qui peuvent avoir une influence cruciale sur les prix, tels que le moment de la réservation et les caractéristiques personnelles des consommateurs. Le moment de la réservation joue un rôle prépondérant dans notre analyse car il s'agit du principal

²Pour plus de détails, voir Rosen (1974), Nesheim (2006)

instrument des stratégies de prix des compagnies aériennes.

1.2 Le modèle

J'utilise l'approche hédonique spécialement développée pour les marchés avec des produits différenciés. Proposée initialement par Rosen (1974), la fonction de prix hédonique décrit la relation d'équilibre entre les caractéristiques économiquement pertinentes d'un produit (ou d'un service) et son prix. Il est utilisé pour mesurer les évaluations des consommateurs et des producteurs de produits différenciés, et pour calculer des indices de prix ajustés en fonction de la qualité des services proposés. Rosen décrit un modèle de différenciation des produits fondé sur l'hypothèse hédonique selon laquelle les biens sont évalués en fonction de leurs attributs ou caractéristiques. Selon l'approche hédonique, toute classe de produits différenciés peut être complètement décrite par un vecteur de caractéristiques mesurables. Les prix et les caractéristiques de produit observés définissent un ensemble de prix implicites (ou «hédoniques») qui guident les consommateurs et les producteurs dans la prise de décision dans l'espace des caractéristiques. En fonction de leurs caractéristiques, différentes entreprises choisissent de produire des biens avec des attributs particuliers. Par exemple, dans le cas de l'industrie du transport aérien, les compagnies à bas coûts ont tendance à fournir uniquement des produits avec un nombre minimal de services, tandis que les compagnies traditionnelles trouvent plus rentable de vendre les billets avec une gamme complète de services. En même temps, différents types de consommateurs choisissent des produits différents selon leurs préférences. Les consommateurs maximisent leur utilité et les entreprises maximisent leurs profits en choisissant le meilleur ensemble d'attributs et de

prix. Les prix d'équilibre et les formes fonctionnelles sont fondamentalement déterminés par la répartition des acheteurs et de leurs préférences, la répartition des vendeurs et leurs coûts, ainsi que par la structure de la concurrence sur un marché particulier.

1.3 Les données

Cette thèse utilise une nouvelle base de données qui n'a pas encore été exploitée pour l'analyse de la concurrence dans l'industrie du transport aérien. Il s'agit de la base de données Survey of International Air Travelers (2005), établie par l'Office américain du commerce et des industries du commerce international. Elle est construite à partir d'une enquête menée dans plusieurs aéroports des États-Unis. Cette enquête exploite une longue liste de questions sur le vol, les prix, les activités pendant le voyage, les caractéristiques personnelles des personnes interrogées et leur avis sur la qualité des services. L'enquête fournit les caractéristiques du consommateur et tous les détails sur les billets vendus. Cela me donne la possibilité de contrôler la classe du billet et la date de départ. On a également demandé aux voyageurs combien de jours avant le départ le billet avait été acheté. Cela permet d'établir une distinction entre les consommateurs qui achètent leurs billets plus tôt et plus près de la date de départ. Les bases de données similaires à l'Enquête sur les passagers aériens internationaux sont généralement utilisées pour les études de la demande de déplacements, comme dans Reece (2001), Divisekera (2003), Racello et tous (2005), Laesser et Crouch (2006), ou Cortes-Jimenez Blake (2010).

1.4 Les résultats d'estimation

La première équation du modèle estimé est une fonction de prix hédonique. Les résultats sont stables dans toutes les spécifications, ce qui suggère une bonne qualité du modèle. Le R^2 de la régression des prix est d'environ 50%, ce qui suggère que les caractéristiques du billet choisies peuvent expliquer la moitié de la dispersion des prix. Les signes des coefficients sont ceux attendus. L'augmentation de 1% de la période entre l'achat du billet et la date de départ conduit à une baisse de prix d'environ 0,04%. Les billets en classe affaires sont 80% plus chers que les billets similaires en classe économique. Une augmentation de 1% de la distance sur un itinéraire signifie une augmentation du prix de 0,6% en moyenne. En moyenne, les billets achetés avec les avantages du programme de fidélisation bénéficient d'une réduction de 50%. Le PIB du pays de destination et la population des points d'origine et de destination n'ont pas d'influence significatives sur les prix.

Pour certains facteurs explicatifs, la différence entre les billets en classe affaires et la classe économique est fortement significative. Une augmentation de 1% de la distance parcourue augmente le prix de 0,8% pour la classe affaires, contre 0,6% pour la classe économique. De même, le rabais du programme de fidélisation est 15% plus élevé en classe affaires. Ce fait a plusieurs explications logiques : tout d'abord, de nombreux clients de la classe affaires voyagent pour des raisons professionnelles; ils effectuent donc généralement des voyages plus souvent, en particulier les voyages internationaux de longue distance. Ils conviennent donc mieux au programme de fidélisation. Deuxièmement, l'une des approches populaires de la gestion du rendement consiste à faire passer les grands voyageurs de la classe économique à la classe affaires. Une analyse plus détaillée de l'influence de la classe affaires ne modifie pas les résultats de manière

spectaculaire et n'a pratiquement aucun impact sur la qualité de la régression.

Enfin, les coefficients pour les valeurs fixes affectent les valeurs estimées attendues. Les billets sont les plus chers pendant les mois d'été et en décembre: environ 20% plus chers pendant ces quatre mois. Pour les autres mois, la différence entre eux n'est pas statistiquement significative. Les billets les moins chers sont pour mardi et mercredi, les plus chers sont pour samedi. Il est à noter que j'ai essayé d'introduire plus de variables de contrôle dans la régression mais les coefficients de ces variables ne sont pas significativement différents de zéro.³

La deuxième régression du système est consacrée à l'analyse du lien entre la volonté des consommateurs de payer pour leur flexibilité et certaines caractéristiques personnelles. La variable dépendante est la volonté des consommateurs de payer pour un jour supplémentaire de flexibilité, ce qui signifie combien ils sont prêts à payer un billet pour avoir le droit de l'acheter un jour plus près de la date de départ. En moyenne, cette variable est égale à 4,5 dollars par jour. Le sexe des consommateurs joue un rôle important. En moyenne, les femmes sont prêtes à payer plus pour leur flexibilité: 0,6 à 0,8 dollar US en plus par jour supplémentaire. Les personnes riches apprécient également leur flexibilité: chaque groupe de revenu suivant a tendance à payer 0,3 dollar de plus pour chaque jour supplémentaire. L'âge n'a pas d'influence significative. Je teste également les influences quadratiques de l'âge mais je ne trouve aucun impact significatif. Un voyage supplémentaire par an augmente la volonté de payer pour la flexibilité de 0,08 dollar. Les effets de l'occupation et de l'objectif ont eu une influence significative sur le consentement à payer.

La dernière équation de notre modèle est la volonté de payer pour la classe affaires. Pour un billet moyen de 1000 dollars, les clients de la classe affaires paient 800 dollars

³Nombre de coupons, durée du trajet, nombre de villes visitées.

de plus. La volonté des femmes de payer pour la classe affaires est inférieure à celle des hommes, mais la différence n'est pas énorme: environ 18 dollars par billet. De manière surprenante, le revenu des consommateurs a un impact significatif négatif sur la volonté de payer pour la classe affaires. L'âge a également un impact négatif : les personnes âgées apprécient moins la classe affaires puisque la volonté de payer diminue de 1,5 dollar par client. L'influence du nombre de voyages est négative mais quadratique. Enfin, le modèle développé couvre deux principaux instruments de "yield management" qui sont la date d'achat et la classe du ticket. Malheureusement, je n'ai pas la possibilité de contrôler le dernier instrument principal de la gestion du rendement qui est la volonté de payer pour les bonus du programme de fidélisation.

1.5 Les extensions proposés

L'analyse effectuée peut être étendue de plusieurs manières différentes. Tout d'abord, il y a encore des problèmes de variables non observés. Plus de données peuvent me permettre de résoudre certains problèmes. Deuxièmement, la volonté à payer peut être mesuré séparément pour différents types de billets, ce qui peut améliorer la qualité de l'estimation. Troisièmement, dans le cadre de cette thèse, je compte élaborer une recommandation explicite pour les compagnies aériennes sur la base des résultats obtenus. Disposant d'information plus concrètes sur les préférences des consommateurs, les compagnies aériennes peuvent développer un système de rabais plus flexible pour différents groupes de clients.

2 Pouvoir de marché et volatilité dans le secteur du transport aérien

Dans un jeu stratégique où les entreprises se font concurrence, l'ensemble des stratégies rationalisables pour chaque joueur englobe toutes les meilleures réponses aux décisions de l'autre. La littérature théorique a suggéré que le caractère unique des résultats rationalisables coïncide avec l'équilibre de Nash du jeu. Ce second chapitre de la thèse propose un test empirique de l'existence d'un l'équilibre de Nash dans un oligopole de Cournot. Nous nous concentrons sur le secteur aérien américain et développons un modèle théorique de concurrence sur chaque route. On suppose que les compagnies aériennes ne sont pas toujours en mesure de prévoir parfaitement le comportement des concurrents, ce qui peut entraîner une multiplicité de résultats rationalisables. Sur la base des ingrédients de l'offre et de la demande de notre modèle, nous construisons un critère de stabilité garantissant l'unicité de l'équilibre. Nous concluons que plus de 90% des marchés locaux observés dans l'industrie du transport aérien aux États-Unis ont atteint le seul équilibre possible de Nash. En plus, nous identifions également les principaux déterminants qui empêchent les entreprises d'atteindre cet équilibre. Nous montrons en particulier que les marchés locaux comprenant un nombre plus élevé de concurrents sont ceux où la quantité produite est plus volatile.

Les modèles économiques axés sur la concurrence des entreprises supposent traditionnellement que les entreprises sont capables de prédire correctement le comportement de leurs concurrents. Une telle situation se présente par exemple dans un équilibre Cournot – Nash, où les entreprises se font concurrence en quantité. Le critère de rationalisabilité élargit au contraire l'ensemble des décisions possibles à prendre par les acteurs et englobe tout niveau de production autour de l'équilibre: il permet ainsi

l'existence de résultats rationalisables potentiellement distincts de l'équilibre de Nash. Ce critère suggère par exemple qu'une entreprise peut s'attendre à ce que les autres produisent au-dessus de l'équilibre, ce qui incite les premières à produire au-dessous de l'équilibre. Par conséquent, la notion d'équilibre est une hypothèse comportementale beaucoup plus forte que le critère de rationalisabilité qui découle de la connaissance commune de la rationalité des acteurs.

La question de la rationalisabilité a fait l'objet d'une attention considérable de la part de la littérature théorique (Bernheim, 1984, Moulin, 1984, Basu, 1992, Guesnerie, 1992). L'aspect empirique de l'analyse s'est toutefois avéré rare jusqu'à présent. Ce second chapitre vise à combler le fossé : sur la base d'un modèle Cournot structurel décrivant les interactions stratégiques des entreprises, nous construisons un test qui permet de déterminer si les concurrents ont atteint ou non un équilibre de Nash. La méthodologie présentée ici généralise les résultats de Desgranges et Gauthier (2016) au cas d'unités de production hétérogènes. Nous identifions également de manière empirique les principaux déterminants qui rendent plus probable une convergence vers l'équilibre.

2.1 Le modèle empirique et les résultats d'estimation

Notre analyse empirique porte sur le secteur du transport aérien américain sur la période 2003-2016. Nous avons supposé que les entreprises étaient en concurrence sur les marchés locaux, définis comme des paires de villes. Les compagnies aériennes maximisent leurs bénéfices à court terme et définissent la quantité optimale de passagers à transporter en fonction d'un nombre donné de créneaux horaires dans les aéroports et

d'une capacité d'avion censée être fixe. Sans surprise, il existe un équilibre de Nash si le revenu marginal de chaque entreprise diminue dans la production totale des autres entreprises. Ensuite, nous envisageons la possibilité que les compagnies aériennes ne prédisent pas correctement le nombre de passagers transportés par leurs concurrents. Pour évaluer cette affirmation, nous faisons appel au concept de rationalisabilité, qui stipule que l'ensemble des stratégies rationalisables pour chaque concurrent implique toutes les meilleures réponses aux décisions de l'autre. Dans ce contexte particulier, un équilibre de Nash est plus susceptible de se produire s'il s'agit localement du résultat rationalisable unique. Notre test spécifie la condition qui garantit la convergence vers l'équilibre de Nash. Il repose sur un indice de stabilité qui exploite les principaux ingrédients d'une fonction de coût décrivant la technologie des compagnies aériennes et d'une équation de la demande liant la quantité totale de passagers transportés sur un marché spécifique et un prix moyen. Notre prédiction est qu'un indice de stabilité supérieur à 1 indique que les productions théoriques à l'équilibre de Nash diffèrent considérablement des productions réelles observées, ce qui suggère un échec de la coordination entre les concurrents. En revanche, un indice de stabilité inférieur à un indique que les entreprises ont réussi à atteindre un équilibre.

Nous combinons les informations démographiques et climatiques avec trois bases de données accessibles au public publiées par le Bureau of Transportation Statistics du ministère des Transports des États-Unis : the Air Carrier Financial Reports, the Air Carrier Statistics et the Airline Origin and Destination Survey (DB1B).

2.1.1 L'équation de coûts

Les résultats sont ceux attendus : les prix des facteurs de production ont un impact significatif positif sur les coûts. La saisonnalité joue un rôle important, avec un coût plus élevé au cours du premier et du dernier trimestre de l'année, et les mauvaises conditions météorologiques entraînent des dépenses supplémentaires. Les effets fixes des compagnies aériennes sont également très significatifs, ce qui montre que les caractéristiques non observées des compagnies aériennes importent. Ces caractéristiques non observées couvrent des éléments tels que l'efficacité productive des entreprises, la productivité des intrants, l'effort de gestion ou les stratégies de marketing. Nous recouvrons l'impact positif attendu une fois que les effets fixes sur les aéronefs sont introduits, ce qui reflète le fait que les compagnies aériennes affectent des types d'avion spécifiques à des segments plus courts et que, en maintenant le niveau de production et le type d'avion, le transporteur fait face à des coûts plus élevés sur des segments plus grands. Ce résultat suggère la présence d'économies de densité, c'est-à-dire que l'augmentation de la production tout en maintenant la taille du réseau de transport fixe entraîne une augmentation moins que proportionnelle des coûts d'exploitation (Caves et al., 1984). L'augmentation des flux de passagers dans le réseau existant permet aux compagnies aériennes d'exploiter des avions plus gros sur des segments à forte densité et de réduire leurs coûts.

2.1.2 L'équation de demande

Les coefficients estimés sont généralement significatifs et les signes sont ceux attendus. Premièrement, l'impact sur la demande de la taille de la population dans les villes

d'origine et de destination est positif et significatif. Deuxièmement, le paramètre de prix propre est globalement négatif et significatif, mais l'élasticité du prix de la demande dépend également des caractéristiques des itinéraires. En particulier, une population plus importante à l'origine ou à destination implique une demande plus élastique : comme les grandes villes sont plus systématiquement connectées à des moyens de transport alternatifs, les compagnies aériennes sont confrontées à une concurrence plus vive sur les marchés qui impliquent un hub important à l'origine ou à la destination. De plus, il est à noter que la demande est moins élastique sur les grands segments, parce que la concurrence intermodale étant probablement moins intense dans ce cas.

2.2 L'indice de stabilité

Globalement, nos résultats empiriques suggèrent que plus de 90% des marchés considérés dans notre base de données ont atteint un équilibre de Nash au cours de notre période d'observation. Comme prévu, la convergence vers l'équilibre est moins probable sur les grands marchés, qui impliquent un plus grand nombre de concurrents, un plus grand volume de passagers transportés ou une population plus importante de la ville d'origine ou de destination. Il est suggéré que, sur les marchés où l'équilibre n'est pas atteint, les entreprises sont généralement confrontées à des prix de facteurs de production plus élevés (notamment les salaires et le prix du carburant) par rapport aux marchés dans lesquels la convergence est obtenue.

2.3 Les autres constatations

En plus, notre cadre structurel révèle des informations intéressantes. Premièrement, il a été démontré qu'une certaine asymétrie dans les capacités des compagnies aériennes

pourrait compliquer leur capacité à atteindre un équilibre de Nash. D'une part, augmenter ex nihilo la capacité d'une des entreprises concurrentes sur un marché donné offre à l'entreprise plus de flexibilité pour réorganiser son processus de production; il est donc plus rentable, mais au prix d'une perte de stabilité du marché. Si, au contraire, l'augmentation de la capacité de la plus grande (la plus petite) est rendue possible aux dépens de la plus petite (la plus grande), la stabilité du marché augmente (diminue). Deuxièmement, notre indice de stabilité peut également être utilisé comme un outil empirique pour tester la définition du marché, ce qui constitue une étape importante dans l'évaluation de la concurrence (Davis et Garcés, 2009). Définir le marché pertinent de la politique de concurrence consiste à identifier l'ensemble des produits des compagnies aériennes qui imposent des contraintes sur leurs prix ou d'autres dimensions de la concurrence, telles que la qualité. Les techniques standard consistent à estimer le rapport de détournement entre tous les services de substitution potentiels afin de récupérer des informations sur la réaction des consommateurs à une modification des conditions des services proposés. Nous préconisons ici une approche différente qui ne nécessite pas l'évaluation des ratios de détournement, mais dépend plutôt de la distance entre le nombre théorique de Nash et les quantités observées sur des marchés associés à un indice de stabilité inférieur à un. Nous soutenons dans ce cas que, sur ces marchés, la quantité théorique de Nash et les quantités observées peuvent s'éloigner l'une de l'autre si les services concurrentiels potentiels sont omis de l'analyse, ce qui suggère à son tour que le marché examiné n'est pas le marché pertinent.

3 Estimation de la demande de services de transport aérien

3.1 Introduction

Ce troisième chapitre utilise l'enquête Survey of International Air Travelers (US) pour estimer les élasticités de la demande de billets d'avion par rapport aux prix et en tirer des implications pour la définition du marché. L'objectif est de construire un modèle de demande de billets d'avion pour mesurer l'influence du changement de prix sur la part de marché de ce produit. L'accent est mis principalement sur l'élasticité entre les itinéraires directs et indirects, étant donné que les consommateurs apprécient différemment les caractéristiques du produit. Je considère ici un modèle de type logit multinomial qui inclut les caractéristiques observables des consommateurs et un modèle de type logit mixte dans lequel les ménages choisissent entre des groupes de billets présentant des caractéristiques différentes telles que le nombre de coupons, le moment de la réservation et le prix du billet.

L'industrie du transport aérien a fait l'objet d'une attention particulière dans la littérature depuis sa déréglementation dans les années 70. Néanmoins, la contribution scientifique sur l'étude des politiques de prix et sur les comportements des consommateurs est plutôt faible. Cela est principalement dû au manque de données détaillées et à la difficulté de modéliser la stratégie de "yield management". La plupart des articles présentent soit une analyse qualitative (Smith et tous (1992), Weatherford et Bodily (1992), Weatherford et Belobaba (2002)), soit un cadre théorique (Botimer et Belobaba (1999), Coughlan (1999), Dana (1999), Desiraju et Shugan (1999), Feng et Xiao (2000), Chen et tous (2003)).

Les estimations de l'élasticité de la demande par rapport aux prix à l'aide de modèles à choix discrets sont couramment utilisées à diverses fins comme dans le cas de la

définition du marché de produits ou la prévision des effets probables des fusions sur les prix. Par exemple, Ivaldi et Verboven (2005), Grzybowski et Pereira (2007), Durand et Pesaresi (2007) utilisent les modèles de choix discrets à coefficient aléatoire pour estimer l'effet de fusion, Grzybowski et al. (2014) estiment les élasticités des prix et contribuent à la définition du marché pour internet haut débit. Lee et al. (2006) et Grzybowski et Pereira (2011) estiment les élasticités de prix et les coûts de commutation pour le secteur des télécommunications.

3.2 La stratégie d'estimation

L'analyse des modèles à choix discret est un cadre naturel pour analyser la décision du consommateur qui fait face à des produits différenciés. Un service d'un point A à un point B est différencié par le nombre de ses caractéristiques qui peuvent être : la date du départ, l'identité de la compagnie aérienne, le nombre de coupons, la classe du billet, ou la date d'achat. Chaque consommateur choisit parmi un ensemble d'alternatives discrètes en fonction de ses propres caractéristiques et des attributs du produit. La première étape du modèle de choix discret consiste à définir un ensemble de choix exhaustif et mutuellement exclusif. La deuxième étape consiste à décrire le processus de décision des consommateurs. Les modèles de substitution estimés entre les variantes dépendent de la définition de l'ensemble de choix et du processus de prise de décision. La méthode de choix discret nécessite des informations sur les prix de tous les types disponibles pour le consommateur au moment du choix. Comme les différences de prix influencent de manière significative les décisions du client, une matrice des élasticités croisées et propres entre différents produits peut être estimée pour chaque client. Les élasticités des prix individuels sont ensuite agrégées à l'ensemble

des clients. Dans ce chapitre, j'essaie de spécifier un processus de décision réaliste qui crée des modèles de substitution flexibles entre les types de produits. L'approche la plus simple consiste à estimer le logit multinomial, ce qui permet d'estimer des modèles de substitution d'agrégats flexibles lorsque les caractéristiques des clients font partie des déterminants de choix. Cependant, ce modèle a la propriété d'indépendance des solutions de rechange non pertinentes (IIA - independence of irrelevant alternatives), ce qui peut donner lieu à des schémas de substitution proportionnels non plausibles d'une solution à l'autre, s'il existe une grande hétérogénéité des consommateurs non observés. Sur la base de ces considérations, j'estime d'abord un modèle logit multinomial standard avec toutes les caractéristiques des clients disponibles pour contrôler l'hétérogénéité observée des consommateurs. Ensuite, j'estime un modèle logit mixte qui permet de contrôler l'hétérogénéité non observée entre les clients.

3.3 Les données

J'exploite un ensemble détaillé de microdonnées fournissant des informations sur les caractéristiques des consommateurs. Il s'agit de la même base que celle utilisée dans le premier chapitre, à savoir le Survey of International Air Travelers réalisée par l'Office américain du commerce et des industries du voyage et du tourisme (2005, 2007). Pour les besoins de ce troisième chapitre, les variables particulièrement importantes sont l'identité de la compagnie aérienne, les points d'origine et de destination, le nombre de coupons (vol direct ou indirect), la date de départ, la date d'achat et le prix du billet. Comme le prix du billet varie énormément d'un consommateur à l'autre, seules quelques bases de données peuvent fournir un prix réel payé pour un type de voyage particulier. L'enquête fournit également les caractéristiques personnelles des consommateurs, qui

sont utilisées comme variables explicatives dans le modèle et sont l'âge du répondant, le revenu (11 groupes de revenu au total), ou le sexe.

Dans le modèle du logit multinomial, l'ensemble des choix comprend tous les types de billets disponibles sur un itinéraire donné. Il existe de légères différences dans les types de produits en fonction d'un itinéraire spécifique. Le choix varie donc entre les itinéraires. Comme option de choix pour le modèle, je considère un billet direct ou indirect par une entreprise donnée sur un itinéraire donné. Je commence l'analyse des estimations du modèle pour le plus grand itinéraire de l'ensemble de données qui est l'itinéraire Los-Angeles (États-Unis) - Taipei (Taiwan). Ensuite, j'estime également le modèle sur d'autres route. Sur chaque route, les estimations sont conservées indépendamment. Les itinéraires doivent donc également répondre à un certain nombre de conditions. Ils devraient avoir plus de 80 observations par route et au moins trois options de choix. Il devrait également y avoir au moins une compagnie aérienne qui propose des vols directs et des vols indirects sur cette route. Avec toutes ces restrictions, j'obtiens 18 itinéraires. L'introduction des caractéristiques de route dans le modèle n'a pas de sens, une régression séparée étant estimée pour chaque route.

3.4 Les résultats d'estimation

3.4.1 Le modèle de demande pour l'itinéraire Los Angeles - Tapei

Le logit multinomial est très sensible au nombre de variables. Comme je n'ai pas beaucoup d'observations par itinéraire, j'inclus uniquement l'interaction des caractéristiques du consommateur avec les caractéristiques de l'option, mais pas la variable d'option elle-même. Cette approche permet également une interprétation plus claire des résultats de l'estimation. Les variables continues telles que le prix et la date de la réservation

(nombre de jours avant le départ où la réservation est faite) sont utilisées comme logarithmes. Les résultats montrent que tous les coefficients significatifs ont les signes attendus, néanmoins les caractéristiques du consommateur sont non-significatives. Sur cette liaison, seule l'interaction des prix et des revenus a une influence significative sur la part de marché. L'impact prix est strictement négatif : plus le prix est élevé, moins il est probable que le ticket soit acheté et cette influence diminue avec la croissance du revenu des consommateurs. La date de la réservation a également une influence importante : les consommateurs préfèrent rester flexibles et acheter des billets plus près de la date de départ. Cela explique pourquoi les prix des billets augmentent généralement à l'approche de la date de départ. La variable binaire des vols directs est également très significative : les vols directs sont préférables par les consommateurs.

Le deuxième ensemble d'équations est estimé avec le modèle logit mixte, où je contrôle l'hétérogénéité non observée des consommateurs. J'utilise 100 Halton draws pour estimer les écarts-types des coefficients estimés. Les modèles prévoient des écarts-types importants pour tous les coefficients estimés, mais néanmoins significatifs pour les variables de prix et la date de réservation. Les coefficients de la variable itinéraire directe ne montrent aucun lien avec l'hétérogénéité des consommateurs. Les résultats des modèles logit mixtes sont similaires à ceux du logit multinomial standard : le prix et le nombre de jours avant le départ ont une influence fortement négative et le coefficient de la variable du vol direct est fortement positive. Parmi toutes les interactions des attributs de produit avec les caractéristiques des consommateurs, comme dans le logit multinomial standard, seule l'interaction du revenu avec le prix est significative : l'élasticité des prix diminue avec la croissance du revenu des consommateurs. Bien que les résultats semblent être similaires au logit multinomial standard, les valeurs absolues des coefficients

pour le prix et les vols directs sont plus grandes que dans le logit multinomial standard. L'effet du nombre de jours avant le départ est presque identique. Cela signifie qu'il existe une certaine hétérogénéité non observable chez les consommateurs, qu'il convient de prendre en compte pour obtenir une estimation plus précise de l'élasticité des prix.

3.4.2 Les élasticités-prix

Les élasticités-prix estimées sont calculées pour toutes les options du trajet Los Angeles - Tapei. Avec l'augmentation de 10% du prix du billet direct avec China Airlines, la part des billets directs offerts par cette compagnie aérienne diminue d'environ 0,1%. Dans le cas des autres options, l'influence est beaucoup moins grande. Alors que les semi-élasticités sont beaucoup plus petites pour les options indirectes, le pourcentage de variation de la part de marché est beaucoup plus impressionnant. Dans les modèles multinomiaux standard, les élasticités-prix sont très proches d'une option à l'autre. Une variation de 1% des prix entraîne une variation de 1 à 2% de la part de marché de ce type de produit. Dans le cas des modèles logit mixtes, les élasticités-prix sont beaucoup plus élevées, en particulier pour les routes indirectes. Une augmentation du prix de 1% entraîne une augmentation de la part de marché de 20-30%, bien que la part de marché de la voie indirecte soit très faible et que cette modification n'entraîne pas de changement considérable sur le marché.

Je calcule également une élasticité pour une option de groupe : changement de part directe avec la variation de prix de toutes les options directes. L'élasticité-prix des modes logit mixtes est plusieurs fois supérieure à l'élasticité-prix estimée avec le modèle sans contrôle de l'hétérogénéité des consommateurs. Les résultats de l'estimation

montrent la nécessité de contrôler l'hétérogénéité non observable des consommateurs, sinon les résultats du modèle peuvent considérablement sous-estimer l'élasticité des prix sur le marché.

3.5 Les extensions proposés

Pour les recherches ultérieures sur ce sujet, il est nécessaire d'obtenir une base de données plus étendue, qui, d'une part, fournit une estimation plus précise du marché et, d'autre part, fournit davantage d'observations pour améliorer la qualité du modèle. Aussi, afin de contrôler le problème possible d'endogénéité, il est nécessaire de rechercher des variables instrumentales pour le prix du service.

Part II

General Introduction

The airline industry has been recently deregulated. Before the liberalization in the 1980s in Europe and in the 1970s in the U.S. the whole air service market was strictly regulated by a range of the government agreements and the airlines had constant state support. The deregulation gave companies a possibility to choose their own prices, routes and production volumes and also opened the market for new companies. Consequently, an increased competition forced price adjustments⁴, better efficiency⁵, creation of alliances and code-sharing agreements⁶. In the 1990s a range of agreements made possible for the European and U.S. companies to compete on the transatlantic market.

Nowadays one of peculiarities of the liberalized market is a huge divergence of ticket prices for the same flights. Mostly it reflects the companies' being unable to easily change the volumes of production or/and store them, so they pay special attention to the behavior of other companies and their volumes of production. Also, with an increased competition the airline companies introduced a "yield management"⁷. Yield management is the process of understanding, anticipating and influencing the consumer behavior in order to maximize profits (or yield) from a fixed, perishable resource (such as airplane seats). In the airline industry it implies a strategic seat management as in selling them to "the right customer at the right time for the right price"⁸. The

⁴The overall decrease in fares is documented in John R. Meyer and Clinton V. Oster, Jr., eds., *Airline Deregulation: The Early Experience* (Boston: Auburn House, 1981); Elizabeth E. Bailey, David R. Graham, and Daniel P. Kaplan, *Deregulating the Airlines* (MIT Press, 1985); and Steven Morrison and Clifford Winston, *The Economic Effects of Airline Deregulation* (Brookings, 1986)

⁵Gagnepain and Marin (2006) argue that facing increased competition the airline companies have to increase their efficiency.

⁶There is a number of articles devoted to the alliances and code sharing agreements: Oum and all (1996), Park and Zhang (2000), Armantier and Richard (2003), Bilotkach (2005, 2007), Gagnepain and Marin (2010)

⁷The expression "yield management" was invented by Robert Crandall, former Chairman and CEO of American Airlines, who has called it "the single most important technical development in transportation management since we entered deregulation".

⁸American Airlines annual report, 1987, definition of the yield management

development and use of the yield management models (seat allocation models) have centered on airlines offering a variety of different types of fares for travel on the same flight.

The goal of this dissertation is to construct a number of economic models to explain the price dispersion on the airline market from the different points of view.

In Chapter 1, I create a direct price model which explains how different product and consumer characteristics influence the price level. It is shown how different attributes like the moment of ticket reservation (number of days between the flight and ticket purchase), ticket class, weekday of the departure and number of coupons define the price and how it corresponds to the consumer characteristics (gender, income, age, etc.). In my model I employ a hedonic price framework and the Survey of International Airline Travelers (made in the U.S.) which allows to control a great number of characteristics.

Chapter 2 is devoted to the differences of the price level from the competition point of view. It is a joint work with Philippe Gagnepain and Stephane Gautier. In a strategic game where firms compete against each other the set of rationalizable strategies for each player entails all the best responses to the others' decisions. The theoretical literature has suggested that the uniqueness of the rationalizable outcomes coincides with the Nash equilibrium. This chapter proposes an empirical test of the existence of the unique Nash equilibrium in a Cournot oligopoly. We focus on the U.S. airline industry and develop a theoretical model of competition on each route. By assumption the airlines cannot always perfectly predict the behavior of the competitors, which results in a multiplicity of rationalizable outcomes. Based on the supply and demand ingredients of our model, we construct a stability criterion which guarantees the uniqueness. Several extensive databases from the Bureau of Transportation Statistics are employed for the

empirical part of the paper.

In Chapter 3 I treat an airline passenger market as a market with the product differentiation and apply a multinomial logit model to calculate price elasticities. The logit model (with a special focus on the consumers heterogeneity) estimates how the different product characteristics influence the market shares. At first I control for the consumers heterogeneity with consumers characteristics: age, income and gender. But as they show no significant influence on the market share (except the price-income interaction), the random coefficient model is estimated to control for unobserved consumers heterogeneity. The results show that the failure to control for unobserved consumers heterogeneity can lead to significant underestimation of the price elasticities. For the purpose of the empirical analysis I apply the Survey of the International Air Travelers which provides detailed product and consumer characteristics.

Part III

Price Strategies on the International Airline Market

Abstract

The focus of this paper is on the high dispersion of the airline ticket prices which appeared with a new pricing policy, called yield management. The paper empirically investigates the connection between ticket prices on the international market, characteristics of the flight, consumer personal characteristics and their willingness to pay. The main contribution of the paper is connected with the employment of a new database: Survey of International Air Travelers (2005). This data base allows me to control for the parameters which can have crucial influence on the prices, like reservation moment and personal characteristics of consumers. The moment of reservation plays a leading role in our analysis as it is the main instrument of the firm's price strategies.

1 Introduction

Few decades ago the world market of the passenger airline service was deregulated.⁹ A lot of articles concerning the competition analysis of the deregulated market have appeared since liberalization. The majority of them employ the average price on the market for their research. However, several years after the liberalization the ticket prices became highly diverse on any given route, so using one price level is not always

⁹Originally the whole air service market was strictly regulated by the range of government agreements. In the 1970s the U.S. and later in the 1980s Europe deregulated the air service market enabling companies to choose their own prices, routes and volumes of production and allowing new companies to enter the market. Transatlantic market was deregulated in 1990s.

justified. In the current paper I try a new approach and construct a pricing model in order to explain such a price diversity.

Fierce competition caused by liberalization made the companies introduce a new marketing approach based on consumers preferences, named yield management.¹⁰ The idea of yield management is to differentiate the airline service products in order to have the possibility to charge different types of consumers with different prices. The development and use of yield management models have resulted in companies offering a variety of different types of fares for the tickets on the same route. Its main instruments: ticket class, service restriction and reservation moment - allow airlines to divide consumers into groups according to their elasticities.¹¹ Therefore price diversity has a direct connection with the consumer preferences. That is why in my model I introduce consumer characteristics. The main goal of this paper is to estimate a relevant pricing model taking into consideration principal instruments of the yield management and clients behavior. I analyze the connection between prices, characteristics of the flight and consumer personal attributes.

For any kind of price model it is necessary to control the huge price dispersion and product differentiation introduced by the yield management policies. Most of the first articles devoted to airline competition analysis with a unique price level: an average per a market.¹² Later in a number of articles the authors pay attention to a huge dispersion in prices on the same route. Borenstein (1989), Dresner and Tretheway (1992), Evans and Kessides (1993) try to control for several level of prices in their research.¹³ A

¹⁰The expression "yield management" was invented by Robert Crandall, former Chairman and CEO of American Airlines, who has called it "the single most important technical development in transportation management since I entered deregulation".

¹¹"A reservation moment" - how many days prior to departure a ticket was bought

¹²Barley, Pazar (1981), Strassmann (1990), Dresner, Tretheway (1992), Brueckner, Dyer, Spiller (1992), Brueckner, Spiller (1994), Marin (1995).

¹³However, the main goal of Borenstein (1989) was to analysis influence of market share and airport share on prices, among the conclusions I see that low end fares appears to be more responsive to the competition. Similar conclusions

paper of Borenstein and Rose (1994) is the first one with the main focus on price dispersion. They prove that the pattern of the observed price dispersion cannot be explained by cost differences alone. They argue that the data support the models of price discrimination and also agree that they do not have sufficient data to test all the theories and make their analysis exhaustive.¹⁴ There are also a few following papers analyzing the connection between competition and price dispersion on the air service market.¹⁵

Unlike Borenstein and Rose (1994) I explain high dispersion by the consequences of the yield management approach. There are two possible reasons, why this topic has not received enough attention in the literature. Firstly, it is not easy to find a reliable data with detailed information about flights and consumers like reservation moment or personal characteristics. Secondly, in order to employ a model of differentiated market I consider the reservation moment being a product characteristic. The product stays the same regardless how many days prior to the departure it was bought, but companies discriminate consumers by the moment of purchase. I assume that consumers are aware of companies policies, so in equilibrium maximizing their utility they choose the best moment of purchase as another product characteristic.

The principal contribution of the current article is connected with a new database, which has not been employed before for the competition analysis of the airline market. It is the Survey of International Air Travelers (2005) database, made by American International Trade Administration Office of Travel and Tourism Industries. This sur-

are made in articles of Dresner and Tretheway (1992) and Evans and Kessides (1993): deregulation has reduced the discounted prices but have not influenced the highest fares on the market. All these articles fail to explain how this price discrimination is possible, as they do not have access to data which allows distinguishing the consumers types.

¹⁴Destination and Origin Survey (DB1B, by Department of Transportation) is the database employed by Borenstein and Rose (1994)

¹⁵ Stavins (2001), Clemons and all (2002), Giaume and Guillou (2004), Kristopher and Shapiro (2009)

vey was conducted in the airports of the U.S. and it involved a long list of questions about the flight, prices, activities during the trip, personal characteristics of the polled and their opinion about quality of the services. The main advantage of the Survey of International Air Travelers (2005) is that it provides the consumers' characteristics and all the details about the sold tickets. It gives me possibility to control for ticket class and date of departure. The travelers were also inquired how many days prior to the departure the ticket had been bought. That allows distinguishing between the consumers who buy tickets earlier and closer to the departure date. The databases, similar to the Survey of International Air Travelers, are usually employed for the travel demand studies, like Reece (2001), Divisekera (2003), Racello and all (2005), Laesser and Crouch (2006), Cortes-Jimenez and Blake (2010).¹⁶

The current article analyzes the market of the airline services from a totally new perspective. I conduct a price analysis controlling for the main tools of the yield management. For this purpose I employ a hedonic approach which is specially developed for the markets with product differentiation.¹⁷ I construct a model in order to study the connection between product attributes, prices and consumers characteristics. To achieve this goal I estimate a system of equations. The first one is a hedonic price function: price as a function of the product characteristics. The others are equations of the consumers willingness to pay (WTP) for one or another product characteristic depending on the consumers individual attributes like income, sex or age.

The rest of the paper is organized as follows. The next section presents a few details about the yield management. Section 3 is devoted to the model. Section 4 provides

¹⁶Reece (2001) analyzes how the demographic, lifestyle and distance variables explain household travel to Los Vegas and Atlantic Cape May. Divisekera (2003) study the substitution effects between different destination of the US tourists. Racello and all (2005) develop an dynamic model of demand employing the information variable. Laesser and Crouch (2006) estimate also a hedonic price function for overall expenditures of the travelers. Cortes Jimenez and Blake (2010) use a country level aggregate data to construct a dynamic model of touristic demand.

¹⁷For details see Rosen (1974), Nesheim (2006)

a detailed data description and descriptive statistics of the variables. The empirical results are described in Section 5. Section 6 concludes.

2 Yield management

In different industries companies use a great number of approaches in order to understand the consumers behavior and sell their products with the highest profit margin. One of them is “yield management”, a set of special pricing strategies developed by airline companies after the deregulation of the market. This strategy let airline companies survive without government support. For the last few decades the airline companies all over the world have started to use yield management. Yield management instruments are also employed by a number of other industries, e.g. hotels, transportation, telecommunication, insurance, rental, advertising companies, etc.

Yield management is the process of understanding, anticipating and influencing the consumer behavior in order to maximize profits (or yield) from a fixed, perishable resource (such as an airline seat on a plane or a hotel room or an advertising space). So the product can be bought in advance for some particular moment, and if it is not sold, it cannot be stored. By another definition yield management involves strategic control of the available resources, meaning to sell the product to “the right customer at the right time for the right price”.¹⁸

The general idea of the yield management is to divide the consumers into the groups according to their elasticities and charge every group with a different price. There is a number of ways to identify the groups of consumers: business and economy classes,

¹⁸American Airlines annual report, 1987, definition of the yield management

service restrictions, loyalty program, etc.¹⁹ A consumer choosing the services rendered to her provides a company with the information about her preferences that helps to define which group this consumer belongs to. For examples, business class travelers are less price sensitive than the economy class passengers. But even within the single service compartment case, for examples, the economy cabin on an airplane, the companies sells the products for different fares. In this case they employ intertemporal price discrimination, which means the consumers are offered different prices if they purchase the products further or closer to the departure date.²⁰

The companies change the prices during the whole period of sales as the different type of consumers prefer to buy or book the products at different moments.²¹ For example, in case of airline or hotel industry early arrivals are generally individuals or families who are planning vacations or other trips far in advance. These customers are usually price sensitive and have lower costs of committing to an early purchase. On the other hand for business or other urgent trips the consumers have to book their tickets or room closer to the date of departure because they did not plan their travel earlier. In general the consumers who buy product at the last moment are characterized by a lower price elasticity of demand. In a number of yield management models, described in the literature, the demand for each fare product is treated even as completely separable and independent.²² Although these assumptions are clearly unrealistic, the products cannot be considered being absolutely homogeneous either. Despite the fact that in the

¹⁹see Smith and all (1992), Weather and Bodily (1992), Botimer and Belobaba (1999), Coughlan (1999), Desiraju and Shugan (1999), Feng, Xiao (2000), Chen and all (2003)

²⁰In the most of the articles, focused on yield management, the differentiation of the prices in time is mentioned as the one of the principal tools of the yield management: Smith and all (1992), Weather and Bodily (1992), Dana (1999), Desiraju and Shugan (1999), Feng and Xiao (2000), Chen and all (2003)

²¹Borenstein and Rose (1994) called it self selective price discrimination: "self selective price discrimination relies on product heterogeneity, since it is carried out by offering consumers a set of alternatives and allowing their choices to reveal information about their characteristics".

²²This kind of approach can be found in the following articles: Botimer and Belobaba (1999), Coughlan (1999), Dana (1999), Feng and Xiao (2000), Chen and all (2003).

literature it is called “intertemporal price discrimination” yield management is not an example of price discrimination.²³ In case of price discrimination companies also use differences in consumer elasticities, but they are not obliged to sell the fixed number of tickets before departure.²⁴

The total number of economic papers devoted to the yield management strategy is very small. Moreover the majority of them presents either qualitative analysis (Smith and all (1992), Weatherford and Bodily (1992), Weatherford and Belobaba (2002)) or develop theoretical framework (Botimer and Belobaba (1999), Coughlan (1999), Dana (1999), Desiraju and Shugan (1999), Feng and Xiao (2000), Chen and all (2003)).²⁵

3 Model

3.1 Hedonic price function

For the purpose of this paper I use hedonic approach which is specifically developed for markets with a differentiated products. Firstly introduced by Rosen (1974), a hedonic price function describes the equilibrium relationship between the economically relevant characteristics of a product (or service) and its price. It is used to measure consumer and producer valuations of differentiated product and to calculate quality adjusted price indexes for goods. Rosen describes a model of product differentiation based on the hedonic hypothesis that goods are valued for their utility-bearing attributes or

²³ Borenstein and Rose (1994).

²⁴Dana (1999) devotes an article to this topic showing the difference between price discrimination and yield management.

²⁵In some of these articles there are several simulations but there is no real data estimations.

characteristics. According to hedonic approach any class of the differentiated products can be completely described by a vector of measurable characteristics. Observed product prices and product characteristics define a set of implicit (or “hedonic”) prices which guides both consumers and producers to make their decision in the characteristics space. According to their characteristics different companies choose to produce goods with some particular set of attributes. For examples, in case of the airline market the low-cost companies tend to provide only products with a minimum number of services, while the huge well-established companies find more profitable to sell the tickets with a full range of services. At the same time different types of consumers choose different products according to their preferences. Consumers maximize their utility and companies maximize their profit choosing the best set of product attributes and prices. Competition prevails because a single agent adds zero weight to the market. The equilibrium prices and the functional forms fundamentally are determined by the distribution of buyers and their preferences, the distribution of sellers and their costs, and structure of competition in the market.

3.2 First step. Price function estimation

This chapter is devoted to the model specification chosen for the empirical estimation.

The following notation will be used in the model.:

- $j = 1 \dots J$ - product index,
- $i = 1 \dots I$ - consumers index,
- $t = 1 \dots T$ - market (route) index,
- $l = 1 \dots L$ - airline company index,
- $p_{j,t,l}$ - product price,
- x_j - vector of the product characteristics
- $k = 1 \dots K$ - product characteristic index (K is the size of the vector x).

As it is recommended by Rosen, as the first step the standard hedonic price function should be estimated by a standard hedonic method regardless the firm and consumers characteristics:

$$p_{j,t,l} = f(x_j, \delta_t, \alpha_l) \quad (1)$$

where α_l and δ_t are the company and route effects. These fixed effects are also considered being product characteristics, as they are observable to the consumers and they have an influence on their choice. The specific fixed effects are incorporated to account for unobservable characteristics of routes and companies.

The main problem of hedonic price that there is no theoretical specification for the price function (1). A number of researches argue that hedonic price function should be estimated nonparametrically or semiparametrically.²⁶ But in this article I do not apply the nonparametric estimation. Firstly, in case of nonparametric function I do not have possibility to interpret coefficients. Secondly, there are not enough observations for any kind of the nonparametric estimation.

For the estimation of the hedonic price function I chose log-linear functional form:²⁷

$$\log(p_{j,t,l}) = \sum_k \log(x_{j,k})\gamma_k + \delta_t + \alpha_l + \epsilon_{j,t,l} \quad (2)$$

These is a general form of the estimated functions. Following Pakes (2003), I am going to estimate several alternative specification of price function in order to find the best specification. Some product attributes can be additive combinable.²⁸ For example, price for each additional kilometer can be higher for business class than for the economy

²⁶Bajari, Benkard (2001), Ekeland, Heckman, Neisheim (2004), Neisheim (2006)

²⁷see Coulson, Bond (1990), Requena Silvente, Walker (2006),

²⁸Chan (2006)

class.

3.3 Second step. Consumers willingness to pay

According to the Rosen the second step is to incorporate the consumers and firm individual characteristics. I assume that all the firm face the same costs and the same production function, so there is no need to estimate an additional function for companies.

As far as consumers are concerned, the utility maximization problem provides me with an additional equation. In case of continuous choice version the consumers willingness to pay is:

$$\frac{\partial p(x)}{\partial x} = - \left(\frac{\partial u(x, z, p(x))}{\partial x} / \frac{\partial u(x, z, p(x))}{\partial p} \right) \quad (3)$$

where z_i is a vector of the individual characteristics for each consumer.

With the quasi-linear utility function $u(x, z, p(x)) = u(x, z) - p(x)$, the equation (3) is simplified to:²⁹

$$\frac{\partial p(x)}{\partial x} = \frac{\partial u(x, z)}{\partial x} \quad (4)$$

In case of discrete choice version, when the product attributes are not continuous, and also for quasi-linear utility function consumers willingness to pay to x_k over x_n is

$$p_k - p_n = u(x_k, z) - u(x_n, z) \quad (5)$$

²⁹Quasi linear utility function is used in a large number of the articles on the similar topics: Berry (1994), Berry and all (1995), Berry and all (2004), Nevo (2000), Ekeland and all (2004), Neigheim (2006)

So knowing the functional form of the utility I am able to estimate the equation (4). For simplicity I assume linear form of the utility function: ³⁰

$$u_{i,j,t} = \sum_k x_{j,k} \tilde{\beta}_{i,k} - p_{j,t} \quad (6)$$

The coefficient $\beta_{i,k}$ is consumer specific, which means each consumer has her own individual value for one or another product attribute. And this individual attribute valuation depends on clients personal characteristics:³¹

$$\tilde{\beta}_{i,k} = \bar{\beta}_k + \sum_r z_{i,r} \beta_{k,r}^0 + \nu_{i,k} \quad (7)$$

The dependent variable can be found from the equation (4):

$$\tilde{\beta}_{i,k} = \frac{\partial p(x'_j)}{\partial x_k}, \quad (8)$$

where both sides represent consumers willingness to pay for product characteristic k.

In case of discrete choice model willingness to pay also can be estimated for different types of discrete variables.

Connected through utility maximisation problem (equation 8) the equation 1 and equation 7 are estimated as a simultaneous system. So, finally I will be able to analyze the influence of the different product attribute on the prices and analyze the influence of the consumers characteristics on their willingness to pay.

³⁰Linear form of the utility function is used by Berry (1994), Berry and all (1995), Berry and all (2004), Nevo (2000)

³¹see Berry and all (2004)

4 Data

For the purpose of this analysis I use the Survey of International Air Travelers conducted by the American International Trade Administration Office of Travel and Tourism Industries. This survey took place in several airports of the U.S. The passengers were asked series of questions about their trip. First of all, the respondents gave the details about the flight: airline company, destination point, departure date, etc. All the flights are international: from the U.S. to abroad. If a passenger has any connecting flights, the survey provides information about the starting points and principal destinations of their travel. Secondly, there is also a part of the survey dedicated to the personal characteristics of the consumer: country of origin, age, income, etc. Moreover there is also a number of the additional questions: the purpose of the trip, with whom the respondent is traveling, hotel references, etc. Finally, the passengers were asked about their expenses that gives me the most interesting parameter: ticket price. They also gave their opinion about quality of all the types of the provided services.

This database is unique in a way that it provides ticket prices and dates of the departure. The respondents were also asked how many days prior to the departure the ticket had been purchased, which gives me a possibility to control for the reservation time. The combination of these three variables cannot be found in the databases used in the previous research.³² The previous articles devoted to the airline competition research employ other databases for the estimations. The commonly used is Destination and Origin Survey (DB1B, by Department of Transportation), a 10% random sample of the all tickets sold in the U.S.³³ This database is highly reliable but it does not include

³²for details see section 1 Introduction

³³see Borenstein (1989), Borenstein and Rose (1994), Dresner and Tretheway (1992), Evans and Kessides (1993), Brueckner and Spiller (1994), etc.

any details about the passengers or product characteristics. It also concerns the other databases used for competition analysis of the airline market before: Digest Statistics (ICAO), World Air Transport Statistics (IATA), etc.³⁴

Although the Survey of International Air Travelers is a large database, it does not contain all the information necessary for our analysis. For examples, the flight distance is an important attribute. I use geographical distance between two points, not the number of flown kilometers, in order to have the same distance for every couple origin-destination points. From the consumers' point of view they have to cover the distance from one point to another regardless of how many kilometers the aircraft flies. The Distance variable is taken from another American database - All Carrier Statistics, T-100 International Market - a highly reliable database of Bureau of Transportation Statistics (all the data are received directly from the airlines). The T-100 International Market database has a variable called "Distance", which provides the geographical distance between two airports.

The social-economic characteristics of the cities are thought to influence the ticket price, therefore several more variables are added: GDP per capita of the destination country and the population of the origin and destination cities.³⁵ There is no need to include GDP of the origin country as it is always the U.S. GDP data are taken from the World Bank database, population from the site www.citypopulation.de.

As it was mentioned above the survey I employ has not been used for competition analysis before. It can be explained by a number of disadvantages of the database. First of all, it is survey type data so there is a great number of missing information. Secondly, supposedly, there are mistakes in variables. For example, not all the respondents

³⁴see Gagnepain and Marin (2006, 2010), Marin (1995), Oum and all (1996)

³⁵see Berry (1990), Brueckner and Spiller (1994), Marin (1995), Peteraf (1995), Zhang and Park (1998), Ciliberto and Tamer (2009), Forbes and Lederman (2009), others

remember the exact day of the ticket purchase, so they give an approximate number of days. The same problem can arise with the most important variable for our analysis, the ticket price. The respondent can make a mistake providing the information about the ticket price, so I don't possess the true value of the ticket price (P_t^{obs}); I have a price which can contain a possible measurement mistake ($P_t^{obs} = P_t^{real} + \varepsilon_t$). The third disadvantage of the Survey is a sample problem. I cannot be sure that this database is a perfect approximation of the real market. For example, it is not possible to obtain any plausible information about the volumes of production. Nevertheless, the similar databases are used for travel demand analysis.³⁶ I employ this database regardless all the disadvantages, as only survey type of data can provide me with all the ticket details and consumers characteristics.

4.1 Data handling and descriptive statistics

A huge amount of work has been done in order to prepare the database for the necessary estimations. I start the analysis with the definition of a route. The principal characteristic of a route is the origin and destination points. The database provides codes of the origin and destination airports of an international direct flight originated from the U.S. But there are a lot of passengers traveling by indirect flights and I have to take them into consideration as well. During the survey respondents answered the questions about the starting and destination points of their journey. The database provides the origin airports which are situated on the territory of the United States. As far as destination point is concerned, I know only the city (town) or region of destination, not the airports. Therefore it was decided to use the cities or towns as the

³⁶see I. Introduction

origin-destination points. The lack of the information does not allow me to use the airport in our analysis. Moreover, this approach allows having more observations for every route that could make our results more consistent.³⁷

After the unification of the origin and destination information, I receive a database of a number of direct and indirect flights covering the whole journey of a passenger. The number of coupons vary from one (for direct flight) to three. In database I have only variable that indicates what airline served the direct international flight, so I have to assume that the same airline which is indicated for the direct flight served all the connecting flights. It is a reasonable assumption, as it is usually much cheaper to buy the whole trip from the same company than to buy the tickets separately from point to point.³⁸ Some of the flights were bought as a part of the package tour.³⁹

Then I have to delete a number of observations for several reason. First of all, as in any survey there is a lot of omissions when the interviewees failed to answer some questions, so I eliminate the observations without information for crucial variables, like price or reservation moment. Two more types of observations are also deleted: one way tickets and first class tickets. For both of them the airline companies employ special pricing strategy and these tickets present a very small percentage of the data. Calculation of the descriptive statistics shows a huge dispersion of the price per kilometer, so in order to minimize the measurement mistake I deleted the extreme values: 5% quantile for the highest values and 5% quantile for the smallest price values. After the

³⁷The both approaches are used in the literature. Some authors use the airports not cities as the origin destinations points (Borenstein (1989) Morrison, Winston (1990), Kim, Singal (1993), Borenstein, Rose (1994)). Others prefer cities pairs for the analysis of competition on the airline market (Berry (1990), Forbes and Lederman (2009) Barley, Pazar (1981), Dresner, Tretheway (1992), Evans, Kessides (1993) Brueckner, Spiller (1994), Sinclair (1995), Marin (1995), Oum, Park, Zhang (1996)). There is also a number of article with the control for both: cities and airports (Brueckner, Dyer, Spiller (1992), Ciliberto and Tamer (2009)).

³⁸Sometimes the tickets for an indirect flight are more expensive than direct flights by the same company and for the same route, however they cost more for the company. It also a yield management strategy allowing to distinguish consumers more sensitive for price.

³⁹In the models below I test if the fact that ticket was bought as a part of the package tour influences the price. I have not found any significant influence

data cleaning, less than a half of all the observations has left.

Tables 1-2 present descriptive statistics of the main variables. There are different types of routes included in the database. All of them originate from the U.S., but they have various destinations points starting from cities in Canada and Mexico to Japan and South Korea. That gives me a huge dispersion in distance from 1,000 kilometers to 16,000 kilometers. There is also a large variety in Origin and Destination population: from 10 thousand people in Shannon, Ireland to huge metropolises like Shanghai and Istanbul with more than 10 billion people population.

As another factor for the model I used GDP per capita in country of destination and it is also has a large dispersion. Prices are various, which can be explained by the diversity of routes and their characteristics. Moreover, two ticket classes - business and economy - have also a serious effect on price dispersion. The price per kilometer is also calculated but it is not used in regression.

The maximum period of the reservation is 250 days in the restricted base. In the original base this figure is much bigger (730 days, but less than 1% of the tickets reserved more than 300 days before departure). The reservations made more than eight months before departure are usually exceptional.

75% of the tickets correspond to the economy class and 25 % of them are business class tickets. One half of the consumers travel by direct flights, the other with one or two stops. In reality the part of direct flights is less (In the unrestricted database this ration is: 43% of direct flights). All the major routes have at least several direct flights and the restricted database capture the biggest routes so direct routes part is slightly higher.

Table 1: Descriptive statistics (1)

		Mean	Std. Dev	Min	Max	Obs.
Dist	Distance, km	8 596	3676	728	17 684	11831
PopO	Population of origin city/town, thousands people	3 447	3 242	2383	8 175	11831
PopD	Population of destination city/town, thousands people	4 224	3 813	9673	14 200	11831
GDP	GDP of the destination country per capita, US dollars	16 832	14 023	294	62 143	11831
Price	Price per person, US dollars	1 306	1279	46	11 000	11831
Price/Km	Price per kilometer	0.16	0.14	0.013	0.88	11831
Res	The time before departure the reservation was made	45.5	50.0	0	365	11831

Table 2: Descriptive statistics (2)

Ticket class		Freq.	Percent
1	Business	2 448	20.69
0	Economy	9 383	79.31

FF	Freq.	Percent
No	10 941	92.5
Yes	890	7.5

4.2 Reservation moment and the personal characteristics of the consumers

It is possible to offer different prices for the same flight only if the different types of consumers tend to reserve their tickets at different moments. In this subsection I try to show the connection between the personal characteristic of the consumers and reservation time.

During the survey the respondents answered a question how many days prior to the departure the ticket had been bought. Using this information I create a discrete variable from 1 to 9 corresponding to reservation time. Separation for nine groups are made according to the recommendations attached to the data base. The second and third columns of the Table 3 demonstrate how these groups are created.

The same table (Table 3) shows the connection between personal characteristics of consumers and reservation moment. The statistics presented in the Table 3 are calculated using the unrestricted database. The *Income* variable is a discrete variable from 1 to 11, group 1 represents the lowest income, group 11 – highest income group. The *Sex* variable is equal to 1 for men and 0 for women, the *Bus* variable is equal 1 for business class tickets and 0 for economy class tickets.

Table 3 shows that women and older people prefer to make their reservation as early as possible. In general men buy tickets more often than women. The majority of the business class tickets are bought from 3 days to two weeks before departure. Surprisingly there is a considerable share of business class among tickets bought more than half of a year before departure. It may be connected with some important business events or conferences. As far as income is concerned, it does not have a monotonic influence on the booking time. It can be explained by the fact that richer people have probably more business trip, reserving their tickets closer to departure but at the same time they

Table 3: Descriptive statistics per route

Rest	ResTime (days)	Obs.	%	Variable	Sex	Income	Age	Bus
1	1-3 days	783	7.18	Mean	0.31	5.95	42.43	0.20
				Std.dev		3.06	11.49	
2	4-7 days	1100	10.14	Mean	0.27	6.49	42.99	0.27
				Std.dev		3.06	11.70	
3	8-14 days	1438	13.31	Mean	0.29	6.48	43.66	0.29
				Std.dev		3.14	12.07	
4	15-30 days	3181	29.93	Mean	0.36	6.18	43.62	0.22
				Std.dev		3.10	13.24	
5	31-60 days	2051	19.22	Mean	0.43	6.02	44.32	0.15
				Std.dev		3.05	13.93	
6	61-90 days	1105	9.39	Mean	0.47	5.99	44.26	0.13
				Std.dev		3.01	14.79	
7	91-120 days	476	4.63	Mean	0.44	6.16	45.05	0.16
				Std.dev		3.03	15.41	
8	121-180 days	494	4.58	Mean	0.49	6.38	48.31	0.16
				Std.dev		3.11	14.69	
9	180 or more	173	1.62	Mean	0.63	6.91	48.54	0.21
				Std.dev		3.11	14.56	

travel more for leisure purpose and buy their tickets more in advance. The largest part of the respondents buys their tickets from two week to a month in advance. More than 75% buy their tickets in the last two months and only 3% of the consumers obtain their tickets more than half a year before departure.

5 Estimation results

This section is devoted to the estimation results. For the purpose of our analysis, a system of the nonlinear equations is estimated simultaneously.

$$\ln Pr_t = \beta_0 + \beta_1 * \ln ResT_t + \beta_2 * Bus_t + \beta' X_t + \alpha_t + \delta_t + \epsilon_t \quad (9)$$

$$WTP_t = \gamma_0 + \gamma'Y_t + \lambda'Z_t + \vartheta_t \quad (10)$$

where $\ln Pr_t$ - logarithm of the price for the ticket, $\ln ResT_t$ - logarithm of number of days prior to the departure the ticket bought, Bus_t - dummy variable for business class, X_t - vector of control variables (route and ticket characteristics), α_t - company fixed effects, δ_t - fixed effects for some routes (biggest),⁴⁰ Y_t - vector of consumer personal characteristics, Z_t - vector of the ticket characteristics (controls), ϵ_t, ϑ_t - error term. Vector of control variables X_t : geographical distance between origin and destination points (Dist), population of the origin and destination points (PopO, PopD), dummy variable if ticket has been bought with the help of the Frequent Flyer program (=1 if Yes), GDP of the destination country, month and weekday fixed effects, holiday dummies. Vector of consumer personal characteristics, Y_t : age of the consumer (Age), income group (Income), number of international trips made for the last 12 months (Trip12m), fixed effects of the main purpose of the trip, fixed effects of the consumer's principal occupation. Vector of ticket characteristics (controls), Z_t - $\ln Dist, \ln PopO, \ln PopD, \ln ResT$

I also try another specification of the first function in the system:

$$\ln Pr_t = \beta_0 + \beta_1 * \ln ResT_t + \beta_2 * Bus_t + \beta'X_t + \gamma'X_t * Bus_t + \alpha_t + \delta_t + \epsilon_t \quad (11)$$

Estimating this function I check the theory that influence of the same factors is different for business and economy class. For example every additional kilometer can have larger influence on price for business class than for the economy.

⁴⁰The number of route is too large to include all of them in the estimation

As it was mentioned before there was no theoretical specification for a hedonic price function. So before proceed with the estimation of the system (equation 9; equation10) I run several tests in order to chose the best specification of the model. For these tests the equations are estimated separately. The previous chapter shows us that data is very heterogeneous, so it is necessary to control for the route, company, season influence. Therefore first of all I need to treat these effects. The standard approach to panel data estimation (our data can be treated as a panel of routes or companies) implies testing model without effects against model with fixed and random effects. Made for route and companies effects F tests and Hausmann tests mostly support the model with fixed effects.⁴¹ Moreover, the employment of the standard fixed effects is commonly used in all types of analysis. Therefore I introduce route, company and some seasonal fixed effects in the nonlinear system. Secondly, there are still some doubts about the linear specification. So I try to test the linear and loglinear specification against specifications with higher power members using the RESET test. It shows that introduction of the quadratic or higher power members does not improve the specification.

After specification tests I proceed with the system estimation. The results of estimation of a system of nonlinear equations are presented in Tables 4-5. For this purpose I employ a standard method: two-step FGNLS estimator. There are four models presented in both tables with a slightly different specifications. Model 1 - Model 3 have different fixed effects. In Model 4 is devoted to more detailed analysis of business class influence, namely estimation of the equation 11. It is necessary to keep in mind that Tables 4-5 present two parts of the same model. The results are stable in all the chosen specifications which indicates good quality of the model. The R^2 in the price regression is about 50%, so the chosen ticket characteristics are able to explain half of the price

⁴¹There are a few doubtful results of Hausmann test for some spesifications.

dispersion. The signs of the coefficients are expected. The increase of the reservation time by 1% leads to price decrease for about 0.04%. This figure means that when ticket is reserved three months in advance one days does not play a role, but if ticket reserved three days before departure every day can slightly change the price. However the received coefficient is smaller than I expected. Business class tickets are 80% more expensive than the similar tickets of economy class. Increase in route distance by 1% means increase in price by 0.6% in average. In average ticket bought with the advantages of the frequent flyer program have a 50% discount. GDP of the destination country and population of the origin and destination points has no significant influence on price. I expected to obtain the positive coefficient for population and GDP, as larger and economically developed cities produce much more demand for air service. But the obtained results have also a logical explanation: these kind of cities also provide a huge supply. Moreover for the airline companies working on a huge route can be more efficient than a small route. There are also touristic destinations like small islands, which can be very expensive regardless population and GDP.

The results of the Model 4 show us that indeed for some factors the difference between business and economy class is strongly significant. Every addition percent to distance increase price by 0.8% for business class comparing to 0.6% increase for economy class. Similarly, the discount of the frequent flyer program is higher for business class by 15 percentage points. This fact has several logical explanations. First of all, a lot of business class consumers travel for business purpose, so they usually make trips more often, especially long distance international trips, therefore they are more suitable for frequent flyer program. Secondly, one of the popular approaches of the yield management is to upgrade the frequent flyers from economy to business class. On average

Table 4: Hedonic price function

Price	M1	M2	M3	M4
lnResT	-0,03*** (0.005)	-0,03*** (0.005)	-0,04*** (0.005)	-0,05*** (0.005)
Bus	0.78*** (0.01)	0.80*** (0.01)	0.79*** (0.01)	-0.90*** (0.31)
Bus*lnDist				0.19*** (0.03)
Bus*FF				-0.15*** (0.04)
FF	-0.48 *** (0.02)	-0.49 *** (0.02)	-0.48*** (0.02)	-0.40*** (0.03)
lnDist	0.55*** (0.01)	0.59*** (0.02)	0.60*** (0.02)	0.58*** (0.02)
lnPopO	-0.004 (0.004)	-0.003 (0.004)	-0.003 (0.004)	-0.002 (0.004)
lnPopD	0.008 (0.005)	0.001 (0.006)	-0.002 (0.006)	-0.001 (0.006)
lnGDP	0.012 (0.004)	0.009 (0.006)	0.006 (0.006)	0.006 (0.006)
Xmas	0.15** (0.06)	0.16** (0.06)	0.10** (0.06)	0.10** (0.06)
Easter	0.02 (0.06)	0.02 (0.06)	0.02 (0.06)	0.004 (0.06)
Weekdays fixed effects			+	+
Months fixed effects			+	+
Big routes effects		+	+	+
Airline fixed effects	+	+	+	+
Constant	1.64*** (0.12)	1.43*** (0.18)	1.39*** (0.18)	1.58*** (0.18)
Obs.number	9147	9147	9147	9147
R2	0.48	0.51	0.53	0.53

The standard deviations are in the parentheses. * significant at 10% level.
 ** significant at 5% level. *** significant at 1% level.

Table 5: Reservation moment, WTP

Price	M1	M2	M3	M4
Sex	0.51*** (0.20)	0.66*** (0.24)	0.80*** (0.29)	0.82*** (0.29)
Income	0.19*** (0.04)	0.22*** (0.05)	0.28*** (0.06)	0.29*** (0.06)
Age	0.009 (0.007)	0.014 (0.010)	0.017 (0.012)	0.018 (0.012)
Trip12m	0.04* (0.03)	0.07* (0.04)	0.08* (0.04)	0.08* (0.04)
Controls	+	+	+	+
Occupation effects		+	+	+
Purpmain effects		+	+	+
Constant	-1.46 (1.95)	-4.34* (2.39)	5.20* (2.90)	-5.17* (2.93)
Obs.number	9147	9147	9147	9147
R2 (within)	0.32	0.32	0.32	0.32

The standard deviations are in the parentheses. * significant at 10% level.

** significant at 5% level. *** significant at 1% level.

Table 6: Business class, WTP

Price	M1	M2
Sex	-17.9*** (6.93)	-18.4*** (6.92)
Income	-3.35*** (1.15)	-3.05** (1.15)
Age	-1.36*** (0.29)	-1.33*** (0.29)
Trip12m	-3.78 (1.12)	-11.67*** (2.49)
Trip12m ²		0.47*** (0.13)
Not 1st trip	10.7 (15.0)	4.86 (15.1)
Controls	+	+
Occupation effects	+	+
Purpmain effects	+	+
Constant	-5900*** (102.6)	-5900*** (102.5)
Obs.number	9147	9147
R2 (within)	0.67	0.68
F statistics	373.1	359.6

The standard deviations are in the parentheses. * significant at 10% level.

** significant at 5% level. *** significant at 1% level.

in Model 4 the business class is more expensive for about 80% (on average means for distance of 8 500 and full price payed). More detailed analysis of the business class influence does not change the results dramatically and has almost no impact on the quality of the regression.

Finally, the coefficients for fixed effects are not presented in the table but the estimated values are also expected. The tickets are the most expensive during the summer months and in December: about 20% more expensive during these four months. For other months the difference between them is not statistically significant. The cheapest tickets are for Tuesday and Wednesday, the most expensive ones are for Saturday.

It is worth to mention that I tried to introduced more of the control variables in the regression (4) but the coefficients of these variable are not significantly different from zero.⁴²

The second regression of the system is devoted to the analysis of connection between consumers willingness to pay for their flexibility and certain personal characteristics. Dependent variable is consumers willingness to pay for one additional day of their flexibility, which means how much more they are ready to pay for a ticket to have the right to buy it one days closer to the departure date. In average this variable is equal to 4.5 dollars per days. The results presented in the Table 5 show that the quality of the regression is quite high, R^2 is about 30%. The gender of the consumers plays a significant role. In average the women are ready to pay more for their flexibility: 0.6 - 0.8 dollar plus for every additional day. These results are surprising because statistically women buy tickets earlier than men. Rich people also appreciate their flexibility: every next income group tends to pay 0.3 dollar more for every additional day. Age has no significant influence. I also tested quadratic influences of age but I failed to find any

⁴²Number of coupons, length of trip, number of visited cities

significant impact. One additional trip per year increase the willingness to pay for flexibility by 0.08 dollar. Occupation and purpose effects have the expected significant influence on willingness to pay.

The last equation of our model is willingness to pay for business class. For average ticket of 1000 dollars business class customers pay 800 dollars more. Women's willingness to pay for business class is less than men's, but the difference is not huge: about 18 dollars per ticket. Surprisingly consumer income has negative significant effect on willingness to pay for business class. Age has also a negative impact: elderly people appreciate business class less: the willingness to pay decrease by 1.5 dollar with every year of the consumer. Influence of number of trip is negative but quadratic. Occupation and purpose effects are jointly significant and they have the expected signs.

The developed model covers two main instruments of yield management: moment of purchase and ticket class. Unfortunately I do not have a possibility to control for the last principal instrument of yield management: willingness to pay for frequent flyer program bonuses.

6 Summary, Conclusion and Proposed extensions

This article presents an extensive analysis of the pricing policies on the airline market. Our research is devoted to the analysis of ticket price heterogeneity. The airline market is shown as a market with a differentiated product. I manage to estimate a price function taking into consideration both sides of the market: firms strategies and consumers personal preferences. A special attention is paid to the moment of reservation as it is the main instrument of the airline pricing policies. This analysis becomes possible only

with a new database: Survey of International Air Travelers which allows me to obtain all the necessary variables for the estimation.

The first part of a system of theoretical equations is a hedonic price function. It explains the connection between product characteristics and price. The obtained results are expected. The earlier reserved tickets cost less. The business class is more expensive in average for 80%. The tickets are also more expensive during the summer months and just before Christmas. More detailed analysis of the business class influence does not change the results dramatically and has almost no impact on the quality of the regression. However it has introduced a couple of new significant coefficients.

The second part of the estimated system is devoted to the consumers willingness to pay for certain ticket attributes. The estimations show that there is a significant connection between personal characteristics of the consumers and their buying patterns. Firstly, consumers willingness to pay for their flexibility is estimated. The women buy their tickets more in advance but at the same time they are ready to pay more for their flexibility. The richer people also prefer to stay more flexible. If consumers travel a lot they also tend to pay more for a possibility to buy ticket later. I found no significant influence of the age on the preference of the reservation moment.

The willingness to pay for business class tickets is also investigated. Women appreciate business class less than man. The more passengers travel less they are ready to pay for the business class, however the influence is quadratic. Surprisingly the model finds negative significant connection between income and consumers willingness to pay for business class.

The conducted analysis can be extended in several different ways. First of all, there are still some specification, database, unobserved variables problems. More data

can allow us to solve some of the problems. Secondly, the willingness to pay can be measured separately for different tickets types, which can improve the quality of the estimation. Thirdly, within the framework of this paper I plan to develop an explicit recommendation to the airline companies based on the obtained results. Having more concrete information about consumer preferences airline companies can develop more flexible system of discounts for different groups of customers.

Part IV

Market power and volatility in the airline industry

Joint work with Philippe Gagnepain and Stéphane Gauthier ⁴³

Abstract In a strategic game where firms compete against each other, the set of rationalizable strategies for each player entails all the best responses to the other's decisions. The theoretical literature has suggested that the uniqueness of the rationalizable outcomes coincides with the Nash equilibrium of the game. This paper proposes an empirical test of the existence of the uniqueness of the Nash equilibrium in a Cournot oligopoly. We focus on the U.S. airline industry and develop a theoretical model of competition on each route. It is assumed that airlines are not always able to predict perfectly the behavior of the competitors which can result in multiplicity of rationalizable outcomes. Based on the supply and demand ingredients of our model, we construct a stability criterion which guarantees uniqueness. We conclude that more than 90% of the local markets observed in the U.S. airline industry have reached the unique possible Nash equilibrium. As a by-product, we also identify the main determinants which prevent firms from reaching an equilibrium. We show in particular that local markets which include a higher number of competitors are the ones where the quantity produced is more volatile.

⁴³Paris School of Economics Université Paris 1.

1 Introduction

Economic models focusing on firms' competition have traditionally assumed that firms are able to predict correctly the behavior of their competitors. Such a situation arises for instance in a Cournot Nash equilibrium, where firms compete in quantities. The rationalizability criterion, instead, enlarges the set of possible decisions to be taken by players and encompasses any production level around the equilibrium: it thus allows for the existence of rationalizable outcomes that are potentially distinct from the Nash equilibrium. This criterion suggests for instance that a firm may expect the others to produce above the equilibrium, which induces the former to produce below the equilibrium. Hence, the notion of equilibrium is a much stronger behavioral assumption than the rationalizability criterion that follows from common knowledge of players' rationality.

The issue of rationalizability has received an extensive attention from the theoretical literature (Bernheim, 1984, Moulin, 1984, Basu, 1992, Guesnerie, 1992). The empirical side of the analysis has however proved scarce so far. This paper aims at filling the gap: based on a structural Cournot model that describes firms' strategical interactions, we construct a test that sheds light on whether or not the competitors have reached a Nash Equilibrium. The methodology presented here generalizes the results in Desgranges and Gauthier (2016) to the case of heterogeneous production facilities. We also identify empirically the main determinants that make a convergence towards the equilibrium more likely.

Our empirical analysis focuses on the U.S. airline industry over the period 2003-2016. We assumed that firms compete in local markets, which are defined as pairs of cities. Airlines maximize their short-run profit and define the optimal quantity of

passengers to be carried conditional on a given number of airport slots and an airplane capacity which are supposed to be fixed. Unsurprisingly, there exists a Nash equilibrium if the marginal revenue of each firm is decreasing in the aggregate production of the other firms. In a second step, we consider the possibility that airlines may not predict correctly the number of passengers transported by their competitors. To assess for this assertion we appeal to the concept of rationalizability, which states that the set of rationalizable strategies for each competitor entails all the best responses to the other's decisions. In this particular context, a Nash equilibrium is more likely to arise if it is locally the unique rationalizable outcome. Our test specifies the condition which guarantees convergence towards the Nash equilibrium. It is based on a stability index that exploits the main ingredients of both a cost function that describes the airlines' technology and a demand equation that relates the total quantity of passengers carried in a specific market and an average price. Our prediction is that a stability index greater than one signals that the theoretical Nash equilibrium productions significantly differ from the actual observed productions, suggesting thus a coordination failure between the competitors. On the other hand, a stability index lower than one indicates that firms have managed to reach an equilibrium.

As a by-product, our structural framework also sheds light on interesting insights. First, it is shown that some asymmetry in the airlines' capacities may complicate their ability to reach a Nash equilibrium. On the one hand, increasing *ex nihilo* the capacity of one of the competing firms on a specific market provides the firm with more flexibility to reorganize its production process; it is thus more cost efficient, but this comes at the cost of a loss in market stability. If, on the other hand, the increase of the largest (smallest resp.) firm's capacity is made possible at the expense of the smallest (largest resp.) firm,

the market stability increases (decreases resp.). Second, our stability index can also be used as an empirical tool to test for market definition, which is an important step in a competition assessment (Davis and Garcés, 2009). Defining the relevant competition policy market consists in identifying the set of airline products that impose constraints on each other's pricing or other dimension of competition such as quality. Standard techniques entail estimating the diversion ratio between all potential substitute services in order to retrieve information on how consumers would react to a change in the terms of the services on offer. We advocate a different approach here which does not require evaluating diversion ratios, but depends instead on the distance between the theoretical Nash and the observed quantities in markets that are associated with a stability index that is lower than one. We argue in this case that, in these markets, theoretical Nash and observed quantities may depart from each other if potential competitive services are omitted from the analysis, and this in turns suggests that the market under scrutiny is not the relevant one.

Overall, our empirical results suggest that more than 90% of the markets considered in our database have reached a Nash equilibrium over our period of observation. As expected, convergence toward the equilibrium is less likely in larger markets which entail a higher number of competitors, a larger volume of passengers carried, or where the population of the origin or destination city is more important. More interestingly, it is shown that, in markets where an equilibrium is not reached, firms generally face higher input prices (in particular wages and fuel price) compared to markets where convergence is obtained.

The rest of the paper is organized as follows. Section 2 provides a theoretical setup. Section 3 discusses the details of the estimation, data cleaning and the estimation results

of the cost and demand functions. The discussion of the estimated Stability Index is provided in Section 4. Section 5 concludes.

2 Theoretical benchmark

This first section discusses the theoretical setting that serves as the foundation of our empirical test. We present first the general framework which is based on a Cournot interaction between several airlines in a specific market. As firms may fail in predicting accurately their competitors behavior, the market may be unable to reach a Nash equilibrium. We construct a stability index which allows us to shed light on whether or not convergence toward the equilibrium is effective. We then discuss interesting insights related to the connection between the stability of the equilibrium and airlines heterogeneity; in particular, we suggest that a greater firms' capacity may come at the cost of market instability. Finally, we show that our testing procedure is also valid in situations where the sample used by the researcher does not allow observing all the products or airline services supplied in the so-called relevant market. We explain then how the conditions of the test should be altered accordingly.

2.1 General framework

In the airline industry a market is usually defined as a route consisting of all the flights between two endpoint airports or cities. The airlines active in a route are firms competing for carrying freight and passengers transportation. They all face the same demand function for transportation services but they may differ according to their technological and organizational characteristics summarized by their cost structure. For instance some airlines use a few large capacity aircraft to spread the cost of book-

ing airport slots, boarding passengers and operating flights on few departures, whereas others instead rely on lower capacity aircraft but more departures. Such choices result from airlines long-run capacity investment policies and intermittent contractual negotiations between airlines and airports. They can be considered as parts of the medium and long-run strategies giving rise to the actual shape of the connection network of the airlines: over a shorter horizon, the fact that Delta Airlines in the route Chicago-Atlanta, say, allocates a A320 Airbus to a booked slot at O'Hara airport from 8h00 to 8h30am every Monday can be viewed as given.

Over a shorter horizon, one can regard airlines as relying on yield management strategies to control prices and quantities of passengers given the available aircraft allocated to the route. We let $c_{af}(q)$ be firm f cost of transporting q passengers using a type a aircraft. Given the number of passengers, costs vary with the aircraft type a as well as additional firm f characteristics: fuel expenditures and total labor wages are higher for larger capacity aircraft, and we observe in the data huge variations of input prices across airlines that may reflect, e.g., heterogeneity in the airlines bargaining power against downstream input suppliers. We assume that the cost function is increasing and convex with the number of transported passengers. A polar but plausible function that fits such a shape involves a low marginal cost until the number of passengers meets the aircraft capacity. Around full capacity an additional passenger implies a high possibly infinite marginal cost.

Let \mathcal{A}_f be the set of aircraft types used by airlines f in the route. Let also n_{af} be the number of flights operated by this airlines using type a aircrafts. The total cost of

firm f for transporting q_f passengers is

$$C_f(q_f) = \min_{(q_{af})} \left\{ \sum_{a \in \mathcal{A}_f} n_{af} c_{af}(q_{af}) \mid \sum_{a \in \mathcal{A}_f} n_{af} q_{af} \geq q_f \right\}, \quad (12)$$

an increasing convex function of the number q_f . Assuming a Cournot-Nash behavior is a quite reasonable assumption in the airline industry (see for instance Brander and Zhang, 1990, Brueckner, 2002, or Basso, 2008); in this case, firm f takes as given the number Q_{-f} of passengers transported by the other airlines and produces

$$q_f \in \arg \max_q P(Q_{-f} + q)q - C_f(q),$$

where $P(Q)$ is the inverse demand function, relying on the total number Q of passengers transported in the route. A necessary first-order condition for q_f to maximize firm's f profit is

$$P'(Q)q_f + P(Q) - C'_f(q_f) = 0. \quad (13)$$

The second-order conditions are always satisfied if the marginal revenue $P(Q) + P'(Q)q_f$ is decreasing in the aggregate production Q_{-f} of the other firms,

$$P''(Q)q_f + P'(Q) < 0 \quad (14)$$

for all q_f and Q . Under this assumption, (13) implicitly defines firm f best-response function

$$q_f = R_f(Q_{-f}).$$

In the main strand of the literature the outcome of competition among airlines is assumed to be a Cournot-Nash equilibrium, i.e., a F -profile $\mathbf{q}^* = (q_f^*)$ such that

$q_f^* = R_f(Q_{-f}^*)$ for all f , with $Q^* = Q_{-f}^* + q_f^*$. It was shown by Novshek (1984) that a Nash equilibrium exists if (14) is satisfied for every firm. In such an equilibrium, every airline f is thought of as being able to predict correctly the number Q_{-f}^* of passengers transported by its competitors. Our paper provides an empirical assessment of this assumption by appealing to the rationalizability criterion.

Following Guesnerie (1992) we argue that firms should succeed to form correct predictions if the Nash equilibrium is the unique rationalizable outcome. Rationalizable outcomes can be characterized by referring to an iterative process of elimination of dominated strategies (see, e.g., Osborne and Rubinstein, 1994). Firms eventually convince themselves that the Nash equilibrium should occur if this process converges toward a single outcome, which must then coincide with the Nash equilibrium. In the present case, this iterative process starts from the assumption of common knowledge among airlines that production satisfies

$$q_f \in [q_f^{\text{inf}}(0), q_f^{\text{sup}}(0)] \quad (15)$$

for every f , with $q_f^{\text{inf}}(0) \leq q_f^* \leq q_f^{\text{sup}}(0)$. Assuming the occurrence of a Nash equilibrium amounts to a priori restrict attention to the case where $q_f^{\text{inf}}(0) = q_f^* = q_f^{\text{sup}}(0)$. In (15) we relax this assumption by allowing the bounds $q_f^{\text{inf}}(0)$ and $q_f^{\text{sup}}(0)$ to differ from the production q_f^* in the Nash equilibrium. If firms are ready to consider only a neighborhood of the equilibrium, then one may set $(q_f^{\text{inf}}(0), q_f^{\text{sup}}(0))$ close to, but different from (q_f^*, q_f^*) . In the absence of any a priori restrictions on firms' initial beliefs, $q_f^{\text{inf}}(0)$ instead could be set to 0 while $q_f^{\text{sup}}(0)$ would relate to the total number of available seats.

By individual rationality airlines only select production levels that are best-response to aggregate productions of others consistent with (15). The best-response function R_f

displays strategic substitutability: if f believes that the others are going to increase their production, thus implying a lower price, then f best-response is to reduce its own production. Indeed, differentiating (13) yields

$$dq_f = R'_f(\mathbf{q})dQ_{-f} \tag{16}$$

where

$$R'_f(\mathbf{q}) = -\frac{P''(Q)q_f + P'(Q)}{P''(Q)q_f + 2P'(Q) - C''_f(q_f)} \tag{17}$$

is indeed negative at any production profile $\mathbf{q} = (q_f)$ satisfying (13) and (14). Therefore, using (15), airline f chooses a production level in a new interval $[q_f^{\text{inf}}(1), q_f^{\text{sup}}(1)]$, with⁴⁴

$$q_f^{\text{inf}}(1) = R_f\left(\sum_{k \neq f} q_k^{\text{sup}}(0)\right), \quad q_f^{\text{sup}}(1) = R_f\left(\sum_{k \neq f} q_k^{\text{inf}}(0)\right).$$

In the immediate vicinity of the Nash equilibrium \mathbf{q}^* , this system behaves as

$$q_f^{\text{inf}}(1) - q_f^* = R'_f(\mathbf{q}^*) \sum_{k \neq f} [q_k^{\text{sup}}(0) - q_k^*],$$

$$q_f^{\text{sup}}(1) - q_f^* = R'_f(\mathbf{q}^*) \sum_{k \neq f} [q_k^{\text{inf}}(0) - q_k^*].$$

The same reasoning applies to every airlines. Under common knowledge of rationality and the best-response functions, one can iterate the argument. The process of elimina-

⁴⁴If $q_f^{\text{inf}}(1) \leq q_f^{\text{inf}}(0)$ then we set $q_f^{\text{inf}}(1) = q_f^{\text{inf}}(0)$. Similarly, if $q_f^{\text{sup}}(1) \geq q_f^{\text{sup}}(0)$ then we set $q_f^{\text{sup}}(1) = q_f^{\text{sup}}(0)$. The process thus remains at (15) if both $q_f^{\text{inf}}(1) \leq q_f^{\text{inf}}(0)$ and $q_f^{\text{sup}}(1) \geq q_f^{\text{sup}}(0)$. The same procedure applies to every step $\tau \geq 1$ of elimination.

tion of dominated strategies is then locally governed by

$$\begin{pmatrix} q_1^{\text{inf}}(\tau + 1) - q_1^* \\ \vdots \\ q_F^{\text{inf}}(\tau + 1) - q_F^* \end{pmatrix} = \mathbf{B} \begin{pmatrix} q_1^{\text{sup}}(\tau) - q_1^* \\ \vdots \\ q_F^{\text{sup}}(\tau) - q_F^* \end{pmatrix},$$

and

$$\begin{pmatrix} q_1^{\text{sup}}(\tau + 1) - q_1^* \\ \vdots \\ q_F^{\text{sup}}(\tau + 1) - q_F^* \end{pmatrix} = \mathbf{B} \begin{pmatrix} q_1^{\text{inf}}(\tau) - q_1^* \\ \vdots \\ q_F^{\text{inf}}(\tau) - q_F^* \end{pmatrix}$$

for all $\tau \geq 0$, with \mathbf{B} a $F \times F$ matrix whose every entry in the f -th row equals $R'_f(\mathbf{q}^*)$ except the diagonal entry (in the f -th column) which is 0.

The process is locally contracting toward the Nash equilibrium \mathbf{q}^* if and only if the spectral radius of \mathbf{B} is less than 1. In this case the Nash equilibrium is locally the only rationalizable outcome and firms using the above iterative process eventually convince themselves that their competitors should play Nash: we say that the equilibrium is locally stable. Otherwise it is locally unstable: then there are multiple rationalizable outcomes and the iterative process can no longer justify that firms eventually pin down their Nash productions. The following proposition gives a condition for local contraction of the process.

Proposition 1. *The Nash equilibrium is locally the only rationalizable outcome if and only if*

$$S(\mathbf{q}^*) = \sum_f \frac{R'_f(\mathbf{q}^*)}{R'_f(\mathbf{q}^*) - 1} < 1 \tag{18}$$

where the slope $R'_f(\mathbf{q}^*)$ is given by (17), with

$$C''_f(q_f^*) \sum_{a \in \mathcal{A}_f} \frac{n_{af}}{c''_{af}(q_{af}^*)} = 1.$$

Proof. By Lemma 3 in Desgranges and Gauthier (2016) the matrix \mathbf{B} has a spectral radius less than 1 if and only if (18) is satisfied. The explicit expression of $C''_f(q_f^*)$ obtains from applying the generalized envelope theorem to the cost minimization program (12). This yields $C'_f(q_f) = \lambda$, where λ is the non-negative Lagrange multiplier associated with the production constraint in (12). Therefore,

$$C''_f(q_f) = \frac{d\lambda}{dq_f}.$$

The first-order conditions for problem (12) are

$$\lambda = c'_{af}(q_{af}) \text{ for all } a \in \mathcal{A}_f \quad \text{and} \quad \sum_{a \in \mathcal{A}_f} n_{af} q_{af} = q_f$$

for every f . Differentiating these first-order conditions yields

$$d\lambda = c''_{af}(q_{af}) dq_{af} \text{ for all } a \in \mathcal{A}_f \quad \text{and} \quad \sum_{a \in \mathcal{A}_f} n_{af} dq_{af} = dq_f$$

for every f . Reintroducing the expression of dq_{af} into the last equality we finally obtain

$$d\lambda \sum_{a \in \mathcal{A}_f} \frac{n_{af}}{c''_{af}(q_{af})} = dq_f.$$

This completes the proof. ■

Proposition 1 generalizes the results in Desgranges and Gauthier (2016) to the case

of heterogeneous production facilities (aircraft). Local stability of the Nash equilibrium obtains if firms are not too sensitive to the production of others, $|R'_f(\mathbf{q}^*)|$ is close enough to 0. This accords with the early insights in Guesnerie (1992) developed in the competitive case. The intuition is that firms find it difficult to predict correctly the behavior of the others if individual production decisions are very reactive to the expected production of others: in this case, small expected departures from Nash production possibly trigger large changes in individual productions that cannot be easily accounted for.

Proposition 1 forms the basis of our empirical illustration by providing us with a simple criterion for the plausibility of the occurrence of the Nash equilibrium. Indeed the prediction to be tested is that the spread between the theoretical Nash equilibrium productions \mathbf{q}^* and the actual observed productions should be magnified if $S(\mathbf{q}^*)$ is greater than 1, the discrepancy reflecting a coordination failure.

2.2 Linear-quadratic specification

The specification used in the empirical illustration builds on the particular case where demand is linear and cost is quadratic,

$$P(Q) = \delta_0 - \delta Q, \quad \delta_0 > 0, \delta > 0, \quad (19)$$

and

$$c_{af}(q) = \frac{q^2}{2\sigma_{af}}, \quad \sigma_{af} > 0,$$

for every aircraft type a and airline f . In this variant, the cost function solution to the

program (12) is

$$C_f(q) = \frac{q^2}{2\sigma_f}, \quad \sigma_f = \sum_{a \in \mathcal{A}_f} n_{af}\sigma_{af}. \quad (20)$$

The index σ_f can be interpreted as follows: note that the marginal cost $C'_f(q)$ associated with (20) and its derivative $C''_f(q)$ are decreasing with σ_f for any given production q . Since $C_f(0) = 0$ a higher value of σ_f implies a production efficiency gain (a lower production cost), which is made possible thanks to an increase in the individual σ_{af} , or because the capacity n_{af} goes up. Cost efficiency and capacity are positively correlated given that a greater capacity provides the airline with more flexibility to organize its production process efficiently. Thus, this gain comes with a flexibility improvement captured by a best-response more sensitive to the production of others,

$$\frac{\partial R'_f(\mathbf{q})}{\partial \sigma_f} = \frac{\partial}{\partial \sigma_f} \left(-\frac{\delta \sigma_f}{2\delta \sigma_f + 1} \right) = -\frac{\delta}{(2\delta + 1)^2} < 0.$$

The fact that efficiency gains relate to greater flexibility in the linear-quadratic setup yields an interesting trade-off between surplus maximization in the Nash equilibrium and stability of this equilibrium (the condition for the equilibrium to be the unique rationalization outcome) illustrated by Proposition 2.

Proposition 2. *Consider the transfer of an additional type a aircraft to some airlines f in the linear-quadratic setup. This transfer locally destabilizes the Nash equilibrium, i.e., it increases $S(\mathbf{q}^*)$, and it increases the aggregate production Q^* in equilibrium.*

Proof. In the linear-quadratic specification,

$$S(\mathbf{q}^*) = \sum_f \frac{\delta\sigma_f}{1 + 3\delta\sigma_f} \quad \text{and} \quad Q^* = \frac{\sum_f \frac{\delta_0 \sigma_f}{1 + \delta\sigma_f}}{1 + \sum_f \frac{\delta\sigma_f}{1 + \delta\sigma_f}}$$

are increasing functions of σ_f (which itself increases with σ_{af}). An additional marginal aircraft transfer corresponds to $d\sigma_f > 0$. The result follows. ■

An additional aircraft implies an increase in the σ_f index. The property in Proposition 2 thus accords with the interpretation of this index as implying both an efficiency gain (which yields a higher production in equilibrium) and more flexibility. Flexibility is detrimental to stability since it makes more difficult the formation of an accurate belief about others behavior. Ranking routes according to total aircraft capacity one should therefore observe in the data that large route with high traffic display a higher spread between the theoretical Nash equilibrium and the actual production levels.

The next result allows us to control for route size by considering a transfer of aircrafts between two airlines in the same route.

Proposition 3. *Consider the transfer of a type a aircraft from airlines f to airlines f' , $f' \neq f$, in the linear-quadratic setup. This transfer locally stabilizes the Nash equilibrium, i.e., it reduces $S(\mathbf{q}^*)$, if and only if $\sigma_f < \sigma_{f'}$. It increases the aggregate production Q^* in equilibrium if and only if $\sigma_f > \sigma_{f'}$.*

Proof. It is similar to the proof of proposition 2 but now exploits second-order derivatives of $S(\mathbf{q}^*)$ and Q^* . In the linear-quadratic specification, $S(\mathbf{q}^*)$ and Q^* are increasing concave functions of σ_f (which is itself increasing linear in σ_{af}). A marginal

aircraft transfer $dn_{af} = -dn_{af'} = -1$ yields $d\sigma_f = -\sigma_{af} < 0$ and $d\sigma_{f'} = -d\sigma_f > 0$. The result follows. ■

The trade-off between efficiency and stability illustrated in Proposition 2 is still valid. However, the insights from Proposition 2 would not allow us to discuss the impact of the transfer considered in Proposition 3 since the contributing airlines f yields an efficiency loss and a stability gain whereas airlines f' that can use one additional aircraft yields an efficiency gain and a stability loss. Proposition 3 obtains by comparing the magnitudes of these marginal changes and so relies on second-order derivatives of $S(\mathbf{q}^*)$ and Q^* : both actually are increasing concave in σ_{af} . This leads to the new testable prediction that some asymmetry in the airlines capacities in a given route should be associated with a theoretical Nash equilibrium production closer to the actual one.

2.3 Stability from sub-market observations

The stability condition discussed in Proposition 1 requires that the econometrician interested in testing convergence toward a Nash equilibrium works on a sample that includes all the potential airline services available in a specific market. In other words, all the airline services of the relevant market should be observed in the data. In the airline industry, there are many circumstances in which a market cannot be considered in isolation. One may think of multimarket contacts through common endpoints or transportation services as a composite good that in fact comprises differentiated items, e.g., economy versus business class services, or non-stop direct versus indirect flights. Neglecting part of the relevant market implies an underestimated overall capacity. Proposition 2 thus suggests a downward bias of the magnitude of the efficiency gains of competition in equilibrium. Simultaneously the low capacity taken into account

may lead us to predict stability of the equilibrium whereas the relevant larger capacity makes the equilibrium unstable in the process of elimination of dominated strategies.

In order to highlight the main biases associated with missing competition, we examine a variant of the previous setup with two substitutable items thought of as direct and indirect flights in a given route. The demand for item m ($m = 1, 2$) is $P^m(Q^m, Q^{-m})$, where Q^m and Q^{-m} stand for the aggregate demands for the two items. On the demand side we assume that the set of airlines producing the two items are disjoint. Traditional carriers often offer both direct and indirect flights within the same market (reference to be added). In addition, we simplify the exposition by working under two strong symmetry assumptions:

Assumption A1. Symmetry of demand: $P^{-m}(Q^{-m}, Q^m) = P^m(Q^{-m}, Q^m)$ for all (Q^{-m}, Q^m) and all m .

Assumption A2. Symmetry of supply: every firm f producing item 1 has a mirror firm $F + f$ producing item 2, i.e., $\sigma_f = \sigma_{F+f}$ for every $f = 1, \dots, F$, with F the total number of airlines producing any given item.

Firm f in market m produces q_f that maximizes its profit given the aggregate production Q_{-f}^m of the other firms in market m and the aggregate production Q^{-m} in the other market. Its best-response can be written $q_f = R_f^m(Q_{-f}^m, Q^{-m})$, which is decreasing with its two arguments if the two items are substitutes. In view of Assumptions A1 and A2, we focus attention on a symmetric Nash equilibrium where $q_f^* = q_{F+f}^*$ for every $f \leq F$. In equilibrium the aggregate output is Q^* in each sub-market. The iterated process of elimination of dominated strategies in the neighborhood of such an

equilibrium is now driven by the partial derivatives

$$R'_{f1}(\mathbf{q}^*) = \frac{\partial R_f^m}{\partial Q_{-f}^m}(Q_{-f}^*, Q^*), \quad \text{and} \quad R'_{f2}(\mathbf{q}^*) = \frac{\partial R_f^m}{\partial Q_{-m}^m}(Q_{-f}^*, Q^*)$$

of the best-response function of firms f and $F + f$ with respect to its first and second argument, respectively. Symmetry implies that they do not depend on the sub-market m under scrutiny. In this richer setup the stability condition given in Proposition 1 becomes:

Proposition 4. *A symmetric Nash equilibrium is locally the unique rationalizable outcome if and only if*

$$\sum_{f \leq F} \frac{R'_{f1}(\mathbf{q}^*) + R'_{f2}(\mathbf{q}^*)}{R'_{f1}(\mathbf{q}^*) - 1} < \frac{1}{2}.$$

Proof. The new matrix \mathbf{B} that governs the iterated process of elimination of dominated strategies now is a $2F \times 2F$ matrix whose f -th row, $f \leq F$, is

$$\left(R'_{f1} \ \dots \ R'_{f1} \ 0 \ R'_{f1} \ \dots \ R'_{f1} \ R'_{f2} \ \dots \ R'_{f2} \right),$$

where all derivatives are evaluated at a symmetric Nash equilibrium. Entry 0 is at f -th column. Similarly, its $(F + f)$ -th row is

$$\left(R'_{f2} \ \dots \ R'_{f2} \ R'_{f1} \ \dots \ R'_{f1} \ 0 \ R'_{f1} \ \dots \ R'_{f1} \right),$$

with 0 at $(F + f)$ -th column. Matrix \mathbf{B} is contracting if and only if the spectral radius of the positive matrix $-\mathbf{B}$ is lower than 1. Let e some eigenvalue e of $-\mathbf{B}$ and \mathbf{v} the

associated $2F$ -eigenvector (v_1, \dots, v_{2F}) . From $e\mathbf{v} = -\mathbf{B}\mathbf{v}$, we have: for all $f \leq F$,

$$\begin{aligned} ev_f - R'_{f1}v_f &= - \sum_{k \leq F} R'_{f1}v_k - \sum_{F < k \leq 2F} R'_{f2}v_k \\ \Leftrightarrow v_f &= - \frac{R'_{f1}}{e - R'_{f1}} \sum_{k \leq F} v_k - \frac{R'_{f2}}{e - R'_{f1}} \sum_{F < k \leq 2F} v_k \end{aligned}$$

and

$$v_{F+f} = - \frac{R'_{f2}}{e - R'_{f1}} \sum_{k \leq F} v_k - \frac{R'_{f1}}{e - R'_{f1}} \sum_{F < k \leq 2F} v_k.$$

Summing over firms yields

$$\sum_{f \leq F} v_f = - \sum_{f \leq F} \frac{R'_{f1}}{e - R'_{f1}} \sum_{k \leq F} v_k - \sum_{f \leq F} \frac{R'_{f2}}{e - R'_{f1}} \sum_{F < k \leq 2F} v_k$$

and

$$\sum_{F < f \leq 2F} v_{F+f} = - \sum_{F < f \leq 2F} \frac{R'_{f2}}{e - R'_{f1}} \sum_{k \leq F} v_k - \sum_{F < f \leq 2F} \frac{R'_{f1}}{e - R'_{f1}} \sum_{F < k \leq 2F} v_k.$$

Using the symmetry properties of \mathbf{B} it is clear that eigenvectors display $v_f = v_{F+f}$ for all $f \leq F$. Hence the two previous equations reduce to

$$\sum_{f \leq 2F} v_f = - \sum_{f \leq 2F} \frac{R'_{f1} + R'_{f2}}{e - R'_{f1}} \sum_{k \leq 2F} v_k.$$

Eigenvalues e of $-\mathbf{B}$ thus are solutions to

$$G(e) \equiv - \sum_{f \leq 2F} \frac{R'_{f1} + R'_{f2}}{e - R'_{f1}} - 1 = 0$$

The function G is continuous decreasing for all $e \geq 0$, with $G(0) > 0 > -1 = G(+\infty)$.

There is consequently a unique $e \geq 0$ solution to $G(e) = 0$. This is the spectral radius.

Since G is decreasing, this eigenvalue is lower than 1 if and only if $G(1) < 0$,

$$-\sum_{f \leq 2F} \frac{R'_{f1} + R'_{f2}}{1 - R'_{f1}} - 1 < 0.$$

The result follows from $R'_{fm} = R'_{(F+f)m}$ for every $f \leq F$ and m . ■

In the presence of several related markets Proposition 4 highlights two destabilizing forces. The first one relates to the actual market size. It can be identified assuming that the demand in the other market is exogenously fixed, $R'_{f2}(\mathbf{q}^*) = 0$. The stability condition then reduces to $S(\mathbf{q}^*) < 1/2$: the threshold for the stability index $S(\mathbf{q}^*)$ above which the equilibrium is unstable should be adjusted downwards if stability is assessed from a sub-market of the relevant one. The observation of a large spread between the theoretical Nash equilibrium number of passengers while $S(\mathbf{q}^*)$ is below 1 in the route thus may signal that the route in fact is only a part of the relevant market. This may provide competition authorities with an indirect test for assessing the size of the relevant market that does not require the knowledge of cross-price elasticities.

The second destabilizing force relates to a contagion-like argument. Firms active in sub-market m need to predict the behavior of firms in the other sub-market. If $|R'_{f2}(\mathbf{q}^*)|$ is high, i.e., the decisions of firms in sub-market m are very sensitive to the expected production in the remaining sub-market, the behavior of firms in sub-market m becomes difficult to predict by the other firms in this same sub-market m since any of these firms may possibly depart significantly from Nash.

3 Empirical illustration to the airline industry

Our empirical application focuses on the U.S. domestic airline industry over the

period 2003-2016. We exploit data from the Bureau of Transportation Statistics to estimate both demand for airlines tickets and aircraft cost functions fitting the linear-quadratic setup developed in Section IV. This gives us all the information needed to recover the number of transported passengers in the Nash equilibrium as well as the stability index $S(\mathbf{q}^*)$ that governs the process of iterated elimination of dominated strategies around this equilibrium. Our main theoretical prediction is that a rise in the stability index is associated with a greater difficulty for every airlines to assess correctly capacity choices of its competitors. This difficulty should eventually lead to coordination failures reflected in a larger spread between Nash and actual production levels. We empirically assess for this theoretical prediction as well as the role of the driving factors for market instability derived from Propositions 2 and 3 suggesting that larger spreads should be observed in routes with high equilibrium traffic and with competing airlines having similar equilibrium market shares.

3.1 Data

We combine demographic and climate information with three publicly available databases released by the Bureau of Transportation Statistics of the U.S. Department of Transportation (DOT): the Air Carrier Financial Reports, the Air Carrier Statistics and the Airline Origin and Destination Survey (DB1B).

Schedule P-5.1 of the Air Carrier Financial Reports contains cost information, e.g., input prices, maintenance expenses, equipment depreciation and rental costs, and total operating expenses disaggregated by airlines \times aircraft type. Since airlines necessarily use a single aircraft on every direct flight, cost estimation will be based on non-stop direct flights referred to as segments. In the airline industry markets are routes that

Table 7: Descriptive statistics per carrier.

Variable	Mean	Std.Dev.	Min	Max
Aircraft costs (thousands of USD)	148,887	187,083	329	1,895,361
Passengers per carrier (thousands)	1,931.1	2,657.2	16.242	26,000
Nb of operated segments	196.41	112.96	17	604
Salary (thousands of USD)	22.59	6.51	8.19	46.34
Average fuel price per 1000 gallons (USD)	2,199.3	649.6	834.4	6,867.9
Nb of observations	3,575			

consist of the flights between two cities. This amounts to abstract from possible interactions between routes, e.g., through multimarket contacts in common endpoint cities or airports, or through airlines network strategies involving several routes. In view of the available data we shall consider segments, which are special instances of route: a segment is a route where endpoint cities are served by a single flight. For instance a route linking cities A and C with one stop in city B comprises two segments AB and BC. The financial information in Schedule P-5.1 can be matched with Air Carrier Statistics T-100 Domestic Segment that provides us with domestic non-stop segment data reported by U.S. air carriers, including origin and destination points, number of passengers carried, flight frequency, and route length.

The T-100 consists of more than three millions observations over the sample window 2003-2016. From this raw data we select segments with distance above 100 miles, with more than 10 passengers per flight, with at least eight departures and 600 passengers during every quarter. This yields a database that contains 22 carriers operating on 1,298 segments and carrying 80 percent of the passengers transported in the U.S. domestic

market. Some descriptive statistics are given in Table 7.⁴⁵

Demand is estimated from the Airline Origin and Destination Survey (DB1B) database over the same 2003-2016 period. The raw sample contains more than 4 million observations for every quarter. In order to match the supply side, which refers to non-stop segments, we focus on routes with a high enough proportion of direct flights. We have chosen a threshold that eliminates routes with less than 60 percent of direct non-stop flight tickets. As shown in Table 8 there remains 3,337 routes in the DB1B database. Compared to the original sample, our dataset tends to be biased toward routes with higher passenger traffic. We disregard routes with one airlines in a monopoly situation, to which our theoretical model does not apply. We only keep routes matching a segment existing in our final T-100 dataset: if, e.g., there is no way to link directly cities A and C, this route is deleted from the sample. Then we apply a cleaning process that mirrors the one applied to the T-100 dataset: we remove routes under 100 miles and get rid of the routes with few passengers. We also discard tickets with extreme reported prices in the bottom and top 5% quantiles of the price per mile distribution, and we remove routes observed during less than 12 years in our 14-year sample window. There eventually remain 379 routes.

To estimate demand we compute the average number of passengers booking a direct flight and the corresponding average fare for every quarter from 2003:1 to 2016:4 for each of these 379 routes.⁴⁶ The database is enlarged with the characteristics of origin and destination: temperature and population on the cities.⁴⁷ Descriptive statistics are presented in Table 9.

⁴⁵Prices and costs are adjusted using the transportation sector price index of the Bureau of Labor Statistics, available at <http://www.bls.gov/cpi/>.

⁴⁶Fares are adjusted using the transportation sector price index of the Bureau of Labor Statistics.

⁴⁷This information is obtained from ggweather.com and citypopulation.de.

Table 8: Direct and indirect routes

Variable	Number of routes	Mean	Std Dev	Min	Max
Original base	11141				
Passengers per route in 1000		18 146	65 839	0 010	2986 0
Share of direct tickets per route		0 46	0 39	0	1
“Direct routes” (share of direct flights >0 6)	3337				
Passengers per route in 1000		38 118	95 359	0 050	2986 0
Share of direct tickets per route		0 87	0 12	0 60	1
Routes (in the final sample)	379				
Passengers per route in 1000		265 9	263 8	9 9	2986 0
Share of direct tickets per route		0 90	0 09	0 60	1

*Variables are corrected by the consumer price index for transportation sector

Table 9: Descriptive statistics. Route level data.

Variable		Mean	Std Dev	Min	Max
PopL	Population (larger city) in million	6 419	4 552	0 928	18 663
PopS	Population (smaller city) in million	1 785	1 518	0 011	12 368
Prlev	Average ticket price on the route*, in U S dollars	182 8	74 6	23 9	586 8
Prlev2	Medium ticket price on the route*, in U S dollars	161 7	54 2	8 2	467 1
Prml	Average price per mile , in U S dollars	0 286	0 108	0 079	0 654
Dist	Distance between two cities, in 1000 km	0 783	0 557	0 013	2 918
PassD	Passengers per route in 1000	265 9	263 8	9 9	2986 0
Obs	Number	20,808			

*Variables are corrected by the consumer price index for transportation sector

3.2 Costs

The estimation of an aircraft cost function is based on the quadratic specification

$$C_{afst} = \frac{q_{afst}^2}{2\sigma_{afst}}, \quad (21)$$

which applies to a given flight operated by airlines f , using a type a aircraft during period t on segment s . Note that our dataset contains information about the number of passengers q_{afst} on a specific segment s , while we only observe the aggregate cost

$$C_{aft} = \sum_s n_{afst} C_{afst}, \quad (22)$$

where n_{afst} denotes the number of departures. We shall use this aggregate cost to estimate the parameter σ_{afst} that enters the stability index at the segment level. To this aim we express this parameter as

$$\frac{1}{2\sigma_{afst}} = \beta_0 \xi_{ft} \mu_s \nu_a, \quad (23)$$

where β_0 is a constant term, ξ_{ft} varies across firms and time periods, while μ_s and ν_a are segment and aircraft fixed effects, respectively. The unobserved characteristics captured by ξ_{ft} may be viewed as accounting for some airlines productive efficiency, managerial effort or marketing strategies, each of which plausibly varies over time. In (23) segment and aircraft fixed effects are restricted to be time invariant, but our final cost specification includes time fixed effects common to segments and aircrafts. Using (23) the aggregate cost C_{aft} given in (22) becomes

$$C_{aft} = \beta_0 \xi_{ft} \nu_a \sum_s \mu_s n_{afst} q_{afst}^2. \quad (24)$$

We argue that unobserved managerial efforts or marketing strategies should be correlated with the input prices that airlines succeed to achieve. We accordingly model the contribution ξ_{ft} specific to the airlines as a linear function of wages and fuel prices faced by airlines f during period t

$$\log \xi_{ft} = b \log \text{Wage}_{ft} + (1 - b) \log \text{PFuel}_{ft} + \xi_f + \text{Quarter}_t + \text{Year}_t, \quad (25)$$

where ξ_f is a carrier fixed effect, and Quarter_t and Year_t are time dummies. The coefficients of input prices sum to 1 due to the property of homogeneity of degree one in input prices.

As suggested in Table 7, our data contains 3,575 observations, which entail 1,298 different segments. As a direct identification of each individual segment fixed effects μ_s is not feasible, we assume that

$$\mu_s = d_0 + d_1 \text{Distance}_s + d_2 \text{Temperature}_s,$$

where Distance_s is the segment length measured as the geographical distance between two cities, and Temperature_s is the average temperature at the departure and arrival cities over the whole sample window.

Hence, the final cost expression to be estimated is

$$\begin{aligned} \log C_{aft} = & \log b_0 + b \log \text{Wage}_{ft} + (1 - b) \log \text{PFuel}_{ft} \\ & + \log \left[\sum_s (1 + d_1^* \text{Distance}_s + d_2^* \text{Temperature}_s) n_{afst} q_{afst}^2 \right] \\ & + \xi_f + \nu_g + \text{Quarter}_t + \text{Year}_t + \varepsilon_{aft}, \end{aligned} \tag{26}$$

where ε_{aft} is an error term. Note also that the constant $\log b_0$ is the sum of the two constants $\log \beta_0$ and $\log d_0$ that cannot be estimated separately. This also explains why the cost function involves the two normalized coefficients $d_1^* = d_1/d_0$ and $d_2^* = d_2/d_0$. The data displays a high correlation between the airlines and aircraft fixed effects ξ_f and ν_a that prevents us from considering both fixed effects. To keep track of both airlines and aircraft unobserved heterogeneity we have considered limited aircraft heterogeneity by allocating the 29 existing aircraft types in the database to 12 groups referring to model characteristics and producing companies.⁴⁸ Then, we have replaced the aircraft

⁴⁸ Aircrafts in the same cluster belong to the same generation of models and have similar size. For example, Boeing 737 300, Boeing 737 400 and Boeing 737 500 are allocated to the same cluster while next generation larger Boeing 737 800 and Boeing 737 900 are in another cluster. There remain small clusters with rare aircraft types like Aerospatiale/Aeritalia ATR 72 or Saab Fairchild 340/B.

fixed effect ν_a with a more aggregated aircraft group fixed effect ν_g in (26).

Table 10 reports the estimated parameters of the cost function in (26) through nonlinear least squares with Gauss-Newton adjustment algorithm.

The results go in the expected direction. Input prices have a positive significant impact on costs. Seasonality plays an important role with a higher cost during the first and the last quarter of the year, and bad weather conditions entail additional expenses. Airlines fixed effects are also highly significant, which shows that airlines unobserved characteristics matter. These unobserved characteristics cover items such as the productive efficiency of firms, the productivity of inputs, the managerial effort, or marketing strategies. We have reported results for three variants. Every variant displays a high fit of the data. The two variants in columns (1) and (2) exclude aircraft group fixed effects and the average temperature on the segment, respectively. The only change in the results relates to the impact of the length of the segment: we find that greater distance reduces cost in the absence of control for aircraft types. We recover the expected positive impact once aircraft fixed effects are introduced: this reflects that airlines indeed allocate specific aircraft types to shorter segments and that keeping the production level and the type of aircraft fixed, a carrier faces higher costs on larger segments. This result suggests the presence of economies of density, i.e., increasing production while keeping the size of the transportation network fixed leads to a less than proportional increase of the operating costs (Caves et al., 1984). Increasing the flows of passengers in the existing network allows airlines to operate bigger aircrafts on high-density segments and reduce their costs. Our estimation results also suggest that $d_2 \leq 0$; in other words, increasing production is more costly on segments with low temperatures as flight delays and airport congestion are more frequent in this case.

Table 10: Estimated cost function C_{aft}

	log C_{aft}		
	(1)	(2)	(3)
Constant (log b_0)	2.513*** (0.34)	1.127*** (0.31)	2.182*** (0.45)
log Wage (b)	0.558*** (0.08)	0.597*** (0.07)	0.602*** (0.07)
Distance (d_1^*)	0.025 (0.02)	0.709*** (0.37)	0.205* (0.11)
Temperature (d_2^*)	0.006*** (0.00)		0.006*** (0.00)
Quarter 2	0.170*** (0.01)	0.157*** (0.01)	0.164*** (0.01)
Quarter 3	0.199*** (0.01)	0.180*** (0.01)	0.193*** (0.01)
Quarter 4	0.096*** (0.01)	0.086*** (0.01)	0.089*** (0.01)
Aircraft group f.e. (ν_g)	No	Yes	Yes
Airlines f.e. (ξ_f)	Yes	Yes	Yes
Year f.e. (θ_{year})	Yes	Yes	Yes
Obs.	3,575	3,575	3,575
Log Likelihood	1010	453	407

*** significant at the 1% level,

** significant at the 5% level, * significant at the 10% level

3.3 Demand

We start from the linear demand specification (19),

$$Q_{st} = \gamma_{st}^0 + \gamma_{st} P_{st} + \zeta_s + \text{Quarter}_t + \text{Year}_t + \nu_{st}, \quad (27)$$

where P_{st} and Q_{st} are the average price level and the aggregate quantity of passengers transported in segment s during period t . Demand includes segment fixed effects ζ_s and times dummies. Prices and quantities are available in the DB1B dataset. The intercept γ_{st}^0 and the slope γ_{st} are two parameters to be estimated. We assume that they depend on route distance and population size in cities of segment endpoints, Pop_{st}^1 and Pop_{st}^2 with $\text{Pop}_{st}^1 \leq \text{Pop}_{st}^2$. We also assume that the slope may directly depend on time period, so we include quarter and year fixed effects Quarter_t and Year_t in the estimated expression of the slope γ_{st} . The final expressions of the intercept and the slope are

$$\gamma_{st}^0 = \alpha_0 + \alpha_1 \text{Pop}_{st}^1 + \alpha_2 \text{Pop}_{st}^2, \quad (28)$$

and

$$\gamma_{st} = \alpha_3 + \alpha_4 \text{Distance}_s + \alpha_5 \text{Pop}_{st}^1 + \alpha_6 \text{Pop}_{st}^2 + \text{Quarter}_t + \text{Year}_t. \quad (29)$$

Hence, the demand equation that we estimate is (27) with intercept and slope given by (28) and (29), respectively. In order to deal with the joint determination of the price P_{st} and the quantity Q_{st} , we implement several estimation strategies, i.e., we use as an instrument the price level from the previous period and input prices. The estimation results are quite similar in both cases.

Demand functions are estimated by ordinary least squares using panel data tech-

Table 11: Estimated demand function

	Nb of Passengers			
	(1)	(2)	(3)	(4)
Constant α_0	133.646*** (16.24)	207.210*** (16.58)	126.531*** (16.26)	196.328*** (16.62)
Pop ¹ ($\alpha_1(\alpha_2)$)	47.003*** (3.02)	22.234*** (3.70)	46.288*** (3.03)	17.078*** (3.82)
Price (α_3)	-0.505*** (0.02)	-0.986*** (0.04)	-0.583*** (0.02)	-1.060*** (0.04)
Price×Distance (α_4)		0.168*** (0.02)		0.148*** (0.053)
Price×Pop1 (α_5)		0.004** (0.00)		0.013*** (0.00)
Price×Pop2 (α_6)		0.051*** (0.01)		0.078*** (0.01)
Year f.e.	Yes	Yes	Yes	Yes
Quarter f.e.	Yes	Yes	Yes	Yes
R^2	0.176	0.204	0.170	0.198
Obs.	20808	20808	20808	20808
F test (route f.e.)	1003.8	982.9	1008.7	978.1

Fixed-effects (within) IV regression, *** significant on the 1% level,

** significant of the 5% level, * significant on the 10% level

(1), (2) mean price is employed, (3), (4) median price is employed

niques with route fixed effects.⁴⁹ Table 11 presents the results of four models applied to routes that involve more than 60 percent of direct flights. The price variable is the average ticket price in columns 1 and 2 of Table 11. In columns 3 and 4 we use the median price. In each case we report results for two different specifications. In columns 1 and 3 the price elasticity is assumed to be independent on route characteristics. The second specification in columns 2 and 4 corresponds to the specification described in (19), (28) and (29).

⁴⁹A simple specification test suggests using the fixed effect model instead of the random effect or the cross sectional ones.

The estimated coefficients are usually significant and go in the expected direction. First, the impact on demand of the size of the population in the cities of origin and destination is positive and significant. Second, the own price parameter is overall negative and significant, but the demand price elasticity also depends on the characteristics of the routes. In particular, a larger population at the origin or at the destination entails a more elastic demand: As larger cities are connected more systematically to alternative transport facilities, airlines face a stronger competition on the markets that involve a big hub at the origin or destination. Moreover, it is noticeable that demand is less elastic on larger segments, as intermodal competition is probably less intense in this case.

The contribution of demand to market stability captured by the index $S(\mathbf{q}^*)$ only relies on $\delta_{st} = -1/\gamma_{st}$ which can be easily recovered from the results reported in Table 11. In the sequel we shall use the specification in column 2. For a 1,000km distance segment linking cities with endpoint populations of 2 and 6 millions inhabitants, we have $\gamma_{st} = -0.986 + 0.168 \times 1 + 0.004 \times 2 + 0.051 \times 6$ plus quarter and year fixed effects.

4 Stability index

From the 1,298 segments used to estimate cost functions, we keep the 337 segments that match demand. Hence the following analysis is based on these 337 segments.

The sufficient statistics σ_{afst} for the contribution of market supply to stability of the Nash equilibrium in the iterative process of elimination of dominated forecast strategies is obtained by replacing the parameters β_0 , ξ_{ft} , μ_s and ν_a that appear in (23) with their estimated values reported in Table 10. The methodology used in the estimation of

cost functions to deal with the correlation between the airlines and aircraft fixed effects implies that the statistics σ_{afst} is recovered at the coarse level of aggregation of the aircraft group g rather than the fine level of the aircraft type a .⁵⁰ The results in Table 11 yield estimated slopes of the demand functions $\gamma_{st} = -1/\delta_{st}$ that provide us with the relevant summary statistics for the contribution of market demand to stability. We therefore have all the information that is needed to compute the slope of the best response function (17) and eventually the stability index $S(\mathbf{q}^*)$.

Table 12: Market fundamentals sufficient statistics for stability

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
sigma	0.002	0.28	0.515	0.762	0.965	5.935
delta	0.915	1.28	1.495	1.609	1.816	5.113
best-response slope	-0.482	-0.375	-0.309	-0.299	-0.231	-0.002
stability index	0.167	0.458	0.568	0.629	0.774	1.794

The marginal cost for transporting one additional passenger quickly increases with the total number of transported passengers, a consequence of the limited aircraft capacity. A small subset consisting of the largest airlines however succeeds to contain the increase in the marginal costs.⁵¹ Price sensitivity of demand (δ_{st}) is symmetrically

⁵⁰The estimate of the constant in (23) is set to the estimate of $b_0 = \beta_0 + d_0$ including the constant term introduced in the expression of the segment component μ_s since β_0 and d_0 cannot be estimated separately.

⁵¹We find that the average number of transported passengers is 3.6 times higher for observations in the last quartile

distributed and displays a high concentration around an average sensitivity of 1.609. That is, a 1 euro increase in the fare implies a $1/1.609 = 0.62$ decrease in the average number of passengers who would like to travel on the segment.

The greater homogeneity of demand behavior makes the distribution of the slope of the best-response (17) at the airlines \times segment \times quarter level mostly driven by cost heterogeneity. Table 12 shows that airlines are found to display significant inertia in their reaction to the production of others. The average slope equals -0.299 , i.e., an average airlines would only transport 30 less passengers when its competitors are expected to transport 100 additional passengers. The number of transported passengers can be adjusted within the limit of a given route transportation capacity, which can be considered as fixed over a quarter. The theory predicts that this inertia favors stability of the Nash equilibrium. However the property of uniqueness of the Nash equilibrium in the set of rationalizable outcomes relies on the sum (18) of the slopes of the best-response functions of the airlines active on the segment. Stability thus should be observed in routes with few active airlines. It may be otherwise more difficult to achieve: (18) suggests that market stability is lost with 4 identical competing airlines having the average best-response slope of -0.31 .

Table 12 finally provides information about the resulting stability indexes given in Proposition 1 for every segment \times period,

$$S_{st} = \sum_f \frac{\delta_{st}\sigma_{fst}}{1 + 3\delta_{st}\sigma_{fst}}. \quad (30)$$

The index is lower than 1 for most segments: it exceeds 1 for only 9 percent of the segment \times period observations. Although our sample window includes the 2007 crisis of the (σ_{fst}) distribution, i.e., for $\sigma_{fst} \geq 0.965$.

and the Great Recession in the following years we find that time only explains 5 percent of the variance of the stability index whereas segments contribute to 81 percent of this variance.

Table 13 provides us with a more detailed picture of the routes with a high stability index. It reports results of the regression

$$\log(S_{st}) = X'\beta + \text{period}_t + \varepsilon_{st},$$

where the right-hand side variables in X consist of exogenous route characteristics as well as potentially endogenous competition indicators. We find high stability indexes in segments linking two widely spaced populated cities. Once again, an interpretation in accordance with Proposition 4 may be that a high distance makes land-based modes of transport (by train or car) less substitutes to air transport. We have constructed a variable h-distance that captures the substitution possibilities between non-stop segments and longer flights. It equals the ratio between the average distance over all the routes linking two cities (with more than 60 percent of direct flights) and the shortest non-stop distance between these cities. A higher value of h-distance thus tends to reduce the willingness to substitute an indirect flight to a direct one, and thus to narrow the relevant market. In line with Proposition 4 we also find that a narrower market (captured by a higher h-distance or a higher share $\text{share}_{\text{direct}}$ of direct flights in the route) is associated with a higher stability index.

We use two standard measures of competition intensity to assess the theoretical insights derived from Propositions 2 and 3. First the stability index is higher in routes with a greater number of competitors. In addition, given the number of competitors market share asymmetry also favors higher indexes.

Table 13: Determinants of the stability index

	log(stabindex)
log(distance)	0.028*** (0.003)
log(maxpop)	0.070*** (0.003)
log(minpop)	0.100*** (0.002)
log(nairlines)	0.101*** (0.016)
log(herfindhal)	-0.685*** (0.017)
log(sharedirect)	1.089*** (0.021)
log(h distance + 1e-10)	0.004*** (0.000)
factor(cluster)2	-0.091*** (0.006)
factor(cluster)3	0.003 (0.007)
factor(cluster)4	-0.024*** (0.007)
factor(cluster)5	0.077*** (0.005)
factor(cluster)6	0.042*** (0.009)
factor(cluster)7	0.005 (0.007)
factor(cluster)8	-0.032*** (0.006)
factor(cluster)9	0.052*** (0.006)
factor(cluster)10	0.069*** (0.006)
Constant	-3.768*** (0.057)
Observations	7,422
R ²	0.897
Adjusted R ²	0.896
Residual Std. Error	0.123 (df = 7350)
F Statistic	899.472*** (df = 71; 7350)

Notes:

***Significant at the 1 percent level
 **Significant at the 5 percent level
 *Significant at the 10 percent level

We have tried to delineate the interaction network that drive higher stability indexes by clustering routes based on the competing airlines. Unlike the clustering relying on exogenous route characteristics that was used in the cost estimation, the decision to enter is likely to be correlated with components of the stability index. The clustering procedure only refers to the 4 airlines with the greatest market share in the whole sample: American Airlines (AA), Delta (DL), United (UA) and Southwest (WN). We associate to every segment \times period a 4-component vector whose f -th component is 1 if airlines f (among these 5 airlines) is active on the segment during this period, and 0 otherwise. Then, in a second stage, we group segments by minimizing the within group distance between the representative vectors assuming a given total number of groups equal to 10.⁵² The clusters and an associated measure of within cluster similarity are reported in Table 14. The highest within cluster similarity is normalized to 1. It obtains when all the observations in the cluster are identical. Most of the 10 clusters exhibit routes with the same active carriers; similarity within clusters 5, 9 and 10 is less clear. Since our sample does not comprise monopoly routes, cluster 1 consists of routes where the low-cost airlines Southwest (WN) occupies a quasi-monopoly market position by competing against airlines with low market share in the whole sample. Delta and United Airlines stand in a similar position in the segments of clusters 3 and 7, respectively. In our sample where segments are grouped into 10 clusters, these three clusters represent around one-third of the total number of segments but the distribution of observations across clusters is very skewed toward this group because these three large carriers are

⁵²We use Fanny partitioning cluster method developed by Kaufman and Rousseeuw (1990) which spreads out every observation into the various clusters according to some measure of similarity. The outcome of the clustering is a matrix whose entry ik gives the similarity between observation i and cluster k . Similarity is normalized to be positive and sum to 1 for every observation (route \times period). Perfect similarity between observation i and cluster k corresponds to the situation where entry ik is 1 while entries ik' are 0 for all $k' \neq k$. No similarity is found if every entry in the i th row equals $1/k$. An index for within cluster homogeneity obtains by averaging all the entries in the k th column, e.g., all the observations in group k are identical if and only if this average is 1. Table 14 provides us with the nearest crisp clustering where every route is entirely allocated to the cluster with the highest similarity.

active in ‘persistent’ segments observed during many periods. Most of the remaining segments involves some competition among large carriers.

Table 14: Airlines network from the main airlines

	cluster	WN	UA	AA	DL	within homogeneity	nb of segments	nb of observations
1	1	100	0	0	0	1	83	1,448
2	2	0	0	0	0	1	71	631
3	5	0	0	0	100	1	62	1,071
4	8	100	0	100	0	1	57	738
5	10	100	17	8	100	0.580	57	719
6	9	100	100	0	0	1	56	957
7	4	0	100	0	15	0.740	50	486
8	7	100	100	100	5	0.920	43	614
9	3	0	0	100	32	0.570	30	490
10	6	0	100	100	36	0.590	21	268

Table 13 delineates three groups of clusters: compared to the reference group formed by clusters 1, 3 and 7, we find one group with low stability indexes (consisting of clusters 2, 4 and 8) and a group with higher indexes (with clusters 5, 6, 9 and 10). Clusters 2, 4 and 8 consist of segments where either both American Airlines (AA) and Southwest (WN) compete one against each other, or neither AA nor WN are active. Symmetrically, only one among these two carrier is always active in segments of clusters 5, 6, 9 and 10.

Markets where Southwest is present and competes against larger airlines are associated with higher stability indexes. In other words, the presence of Southwest in competition games which also involve large traditional airlines complicates convergence

toward a Nash equilibrium. This is not a surprising result as the empirical literature focusing on airlines has documented that Southwest has the ability to trigger fierce reactions by its competitors: Morrison (2001) for instance estimates fares to be 11 to 13% lower when Southwest serves one end-point of a route and 33% when it serves both endpoints. Goolsbee and Syverson (2008) also sheds light on the fact that the powerful competitive effect of Southwest Airlines does not operate solely through Southwest's head-to-head competition with major carriers; it seems also that the threat of competing with Southwest is enough to induce substantial fare reductions from major carriers. These fare declines seem to be accompanied by significant increases in passenger traffic on the incumbents' threatened routes.

4.1 Assessing the Nash equilibrium

We want to assess whether the index computed in the theoretical Nash equilibrium in Proposition 1 can be regarded as a reliable predictor of the occurrence of the equilibrium in segment s during quarter t . To this purpose we now introduce the spread between the theoretical Nash and the actual number of transported passengers. The prediction to be tested involves a spread positively correlated with the indexes S_{st} over segments s and quarters t . Let \mathcal{F}_{st} be the set of airlines active in segment s during quarter t . The spread is measured by

$$\|\mathbf{q}_{st} - \mathbf{q}_{st}^*\| = \sqrt{\sum_{f \in \mathcal{F}_{st}} (q_{fst} - q_{fst}^*)^2}.$$

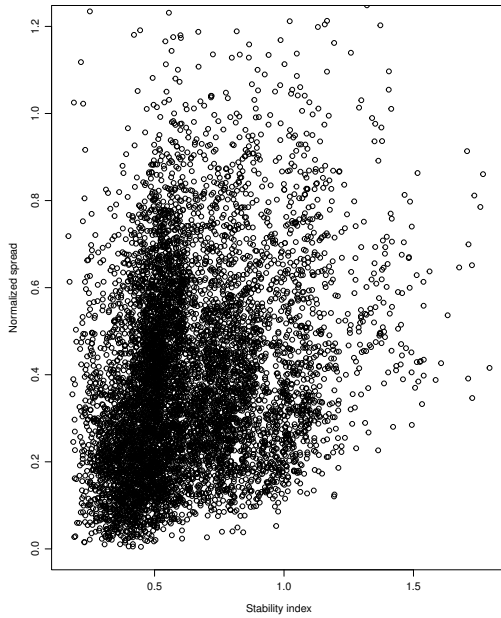


Figure 1: All points

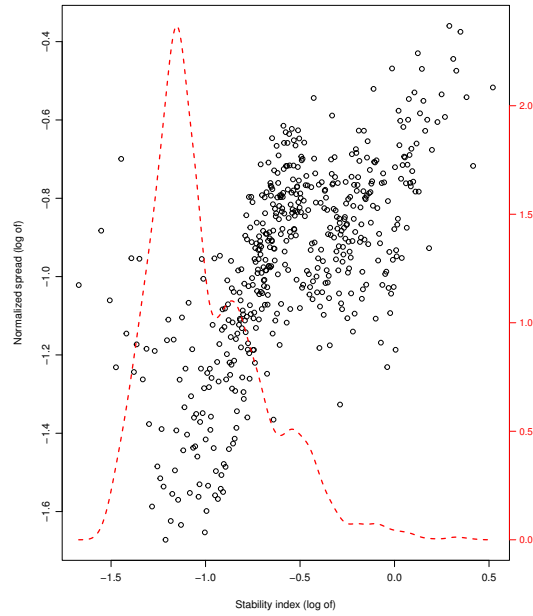


Figure 2: Density and average spread per range 2e-02 stability indexes

In order to avoid a bias with a higher spread in segments with intense traffic, it is normalized by the number of passengers

$$\|\mathbf{q}_{st}^*\| = \sqrt{\sum_{f \in \mathcal{F}_{st}} q_{fst}^{*2}}$$

Our normalized measure for the spread between Nash and actual production (at the segment \times period level) is therefore

$$\Delta_{st} = \frac{\|\mathbf{q}_{st} - \mathbf{q}_{st}^*\|}{\|\mathbf{q}_{st}^*\|}$$

Figures 1 and 2 suggest a positive correlation between the stability index and the spread Δ_{st} . However, unlike the prediction of Proposition 1 the spread is not 0 if the

index falls below 1. Instead the figure seems to display a continuous rise in the spread from 0 (where actual and Nash number of transported passengers coincide) when the index increases from 0. The spread seems to reach its highest level around 0.8, i.e., the departure from Nash represents 80% of the Nash production. Indeed Figure 1 possibly suggests the existence of different regimes: the one with the greatest number of observations obtains for stability indexes below 1/2 and seems to be associated with the largest departures from Nash outcomes while the remaining one has indexes between 1/2 and 1 and display some dampened Nash departures. In both regimes there is a high positive correlation between the value of the index and Nash departure. We estimate a model of spread depending on the route characteristics:

$$\log \Delta_{st} = b_0 + b_1 \log S_{st} + \text{Segment}_s + \text{Quarter}_t + \varepsilon_{st}$$

The results of this regression are reported in Table 15.

Table 15: Nash departure and stability index

	Nash spread $\log \Delta_{st}$		
	(1)	(2)	(3)
Stability index $\log S_{st}$ (b_1)	0.539*** (0.131)	0.574*** (0.145)	0.713*** (0.089)
Constant (b_0)	-0.829*** (0.077)	-0.731*** (0.117)	-0.560*** (0.085)
Quarter fixed effect	No	Yes	Yes
Segment fixed effect	No	No	Yes
Observations	7,422	7,422	7,422
R ²	0.096	0.117	0.596
Adjusted R ²	0.096	0.110	0.575
Residual Std. Error	0.632 (df = 7420)	0.627 (df = 7365)	0.433 (df = 7065)
F Statistic	788.090*** (df = 1; 7420)	17.459*** (df = 56; 7365)	29.230*** (df = 356; 7065)

Notes: ***Significant at the 1 percent level; **Significant at the 5 percent level; *Significant at the 10 percent level. We report robust standard errors at the segment level.

As expected from Figure 1 and 2 we find a positive significant relation between the theoretical stability index and departures from Nash outcomes: the index is shown to

be a reliable predictor of the occurrence of the Nash equilibrium. Important insights come from the fixed effects that are controlled for. In our preferred specification, which includes both segments and quarter fixed effects, a 1 percent increase in the index is associated with a 0.713 percent rise in the spread. An inspection of the R^2 of the various regressions shows that time is about irrelevant to explain the importance of departures from Nash, though the period includes the 2008 great recession. In fact unobserved heterogeneity in the segments explains a large part of the Nash departures.

5 Conclusion

This paper focuses on the Cournot oligopolistic competition the U.S. airline market. Inspired by the theoretical insights of Desgranges and Gauthier (2013) we develop a detailed theoretical model of competition on the airline routes. In a strategic game where firms compete against each other, the set of rationalizable strategies for each player entails all the best responses to the other's decisions. The theoretical literature has suggested that the uniqueness of the rationalizable outcomes coincides with the Nash equilibrium of the game. This paper proposes an empirical test of the existence of the uniqueness of the Nash equilibrium in a Cournot oligopoly.

Data of the airline industry allows us to apply this model to several markets at the same time, so we employ a number of databases in order to estimate in details the cost and demand functions. To differentiate the stability criterion between routes we estimate marginal costs depending on company, segment and aircraft characteristics. Estimation of cost equation is the main challenge of this paper, as there is no accessible expenses data for separate routes. Partly this problem is solved by nonlinear estimation

and employment of the aircraft costs instead of firm level data. Price elasticity of demand is also function of time and particular route characteristics. Compatible to the goal of the paper the computed stability criterion (SI) is specific for routes and periods, its value is as expected: more than 90% of the market are suppose to be in an equilibrium. The criterion is expectedly higher for more popular routes with a larger number of competitors.

We also identify the main determinants which prevent firms from reaching an equilibrium. We show in particular that local markets which include a higher number of competitors are the ones where the quantity produced is more volatile.

Part V

Demand estimation for airline passenger services

Abstract

This paper employs a rich Survey of International Air Travelers (US) to estimate price elasticities of demand for airline tickets and draw implications for market definition. The goal is to construct a model of airline ticket demand to measure how the price change influences the market share of this product. The main focus is on the elasticity between direct and indirect routes given that the consumers have different appreciations of the product characteristics. I estimate a multinomial logit which controls for the observable consumers characteristics and a mixed logit model in which households choose between groups of tickets with different characteristics such as the number of coupons, the reservation moment and the ticket price.

1 Introduction

The airline industry is characterised by an important price dispersion. For a seat on the same flight one passenger can pay several times more than another client. The huge price difference for the same flight is explained by a marketing policy of the airline companies, called yield management. Yield management aims at differentiating the airline service products in order to have the possibility to charge different types of consumers with different prices. In this paper I treat airline tickets as a heterogeneous product in order to shed light on the connection between prices and product characteristics.

This paper proposes an econometric analysis of consumer behaviour. I estimate

price elasticities using discrete choice models, that accounts for the actual choice of consumers. I estimate a relevant demand of the airline market and I shed some light on the consumers willingness to pay for one or another characteristic of airline ticket. As the demand shifters the model includes price, direct flight dummy and number of days before departure the ticket is bought.

Although discrete choice models are popular in the empirical industrial organisation literature focusing on price elasticities, airline ticket are usually not treated as a differentiated product. This is a potential drawback given that the moment of reservation, the service class or the number of coupon have a direct connection with consumers evaluation of these characteristics. In a discrete choice setup the consumer chooses between different types of products such as a direct flight served by one company or a ticket for an indirect flight bought from another airline. I estimate a logit model, which controls for observable heterogeneity between clients characteristics such as income, sex and age of the consumers and I also estimate a range of random coefficient logit models which allows for unobserved consumer heterogeneity⁵³. I find that some of the clients characteristics significantly influence the ticket choices and price sensitivity. The model allows estimating the substitution effects between different types of tickets.

I exploit a detailed set of microdata, which provides information about consumers characteristics. More specifically I employ a representative survey of 70000 international travelers of the U.S. of 2005 and 2007.

The airline industry has received an important attention in the competition literature since the industry got deregulated in the 70s. Nevertheless, the contribution to pricing policies and consumers behavior patterns on that market is slight. Mostly it is due

⁵³In different sources several models can be called “mixed logit model”. In this paper by “mixed logit” or “random coefficient logit”, I call a standart multinomial logit with introduced consumers heterogeneity. Consequently a model without consumer heterogeneity is called “multinomial logit”.

to the lack of detailed data and difficulty to model the yield management strategy. Most papers present either qualitative analysis (Smith and all (1992), Weatherford and Bodily (1992), Weatherford and Belobaba (2002)) or develop theoretical framework (Botimer and Belobaba (1999), Coughlan (1999), Dana (1999), Desiraju and Shugan (1999), Feng and Xiao (2000), Chen and all (2003)).

Estimates of price elasticities of demand using discrete choice models are commonly used for a lot of purposes, starting from the relevant product market definition to the prediction of the likely price effects of mergers. For example, Ivaldi and Verboven (2005), Grzybowski and Pereira (2007), Durand and Pesaresi (2007) employ the random coefficient discrete choice models to estimate the merger effect, Grzybowski and all (2014) estimate price elasticities and contribute to market definition for broadband internet. In this paper I analyze consumer choices and substitution patterns between product types. Lee and al. (2006) and Grzybowski and Pereira (2011) estimate price elasticities and switching costs for telecommunication industry. In both paper survey data is employed. There are several earlier papers devoted to discrete choice models such as Nevo (2000).

2 Econometric model

A discrete choice framework is a natural framework for analyzing consumer decision regarding different products. The tickets from point A to point B is differentiated by the number of characteristics: day of travel, airline providing services, number of coupons, ticket class, day of purchase. Each consumer choose between a set of discrete alternatives depending on her own characteristics and product attributes. The first step

in discrete choice model is to define an exhaustive and mutually exclusive choice set. Second step is to describe the consumers decision process. The estimated substitution patterns between alternative depend on the definition of the choice set and the decision-making process. The discrete choice methodology requires information on prices of all the types available to consumer at the moment of choice. As price differences significantly influence client's decisions, a matrix of own and cross elasticities between different product can be estimated for each client. Individual price elasticities are then aggregated to the whole set of clients. In this paper I want to specify a realistic decision making process that creates a flexible substitution patterns between product types. The simplest approach is to estimate the multinomial logit, which allow to estimate flexible aggregate substitution patterns when the clients characteristics are among choice determinants. However, this model has the property of independence of irrelevant alternatives (IIA), which may result in implausible proportional substitution patterns across alternatives if there is a lot of unobserved consumers heterogeneity. Based on these considerations, I estimate first a standard multinomial logit model with all the available clients characteristics to control for the observed heterogeneity of consumers. Then I subsequently estimate a mixed logit model which in addition to consumers characteristics controls for the unobserved heterogeneity across clients.

2.1 Consumers Utility

I employ a standard linear utility specification for consumers $i = 1, \dots, N$ over the different product types $j = 1, \dots, J$. Utility depends on product characteristics and on observable and unobservable household characteristics. So the utility of consumer i for

product j is given by:

$$U_{ij} = x_j' \beta_i - \alpha_i \log(p_{ij}) + \epsilon_{ij} \quad (31)$$

Here, x_j is a $J \times 1$ vector of the characteristics of product j and β_i is a $J \times 1$ vector of coefficients denoting the client specific valuations for the different characteristics. Furthermore p_{ij} denotes the price paid by consumer i for product j ⁵⁴, α_i is a random coefficient for the individual-specific valuation of price, ϵ_{ij} is a consumer specific valuation for product type, “logit error term”. It is identically and independently distributed across different products. The vector of coefficients β_i and α_i may depend on both observed consumers characteristics and unobserved heterogeneity. More specifically:

$$\begin{pmatrix} \beta_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \bar{\beta} \\ \bar{\alpha} \end{pmatrix} + \Pi D_i + \begin{pmatrix} \sigma_\beta \\ \sigma_\alpha \end{pmatrix} v_i, v_i \sim N(0, 1) \quad (32)$$

where $(\bar{\beta}, \bar{\alpha})$ refers to a $(J + 1) \times 1$ vector of mean valuations, D_i is a $d \times 1$ vector of observed consumers characteristics and ϕ is a $(J + 1) \times d$ matrix of parameters capturing the impact of consumers personal characteristics on the valuation of the product’s j attributes and price, v_i is a randomly drawn vector from the standard normal distribution capturing unobserved household heterogeneity regarding price and characteristics. And $\sigma = (\sigma_\beta, \sigma_\alpha)$ refers to a $(J + 1) \times 1$ vector of standard deviations around the mean valuations. In this empirical analysis the vector of observable characteristics D_i consists of consumers age, income and gender.

In the special case when vector σ being a vector of zeros there is no unobserved heterogeneity and the estimation can be conduct just by standard conditional logit model. More generally, when I control for unobserved heterogeneity between consumers

⁵⁴The price is used in logarithm terms as it allows to compare several routes with different price levels

mixed (or random coefficient) logit model is employed.

2.2 Choice probabilities

A consumer i chooses a product j if this choice maximizes her utility among all available alternatives, which means $U_{ij} = \max_{k \in C_i} U_{ik}$ where C_i is consumers available choice set. Therefore the probability that client i chooses product j is given by (with the given random coefficients β_i and α_i):

$$l_{i,j}(\beta_i, \alpha_i) = Pr \left(U_{ij} = \max_{k \in C_i} U_{ik} \right) = \frac{\exp(x'_j \beta_i - \alpha_i \log(p_{ij}))}{\sum_{k \in C_i} \exp(x'_k \beta_i - \alpha_i \log(p_{ik}))} \quad (33)$$

where the second part of the equation comes from the assumption of the logistic distribution for error term $\epsilon_{i,j}$.

If there is no unobserved consumers heterogeneity ($\sigma=0$), this is standard multinomial logit choice probability that enters the likelihood function. With unobserved heterogeneity, it is necessary to integrate the conditional choice probability $l_{ij}(\alpha_i, \beta_i)$ over the distribution of α_i and β_i

$$s_{ij} = \int_{\alpha} \int_{\beta} l_{ij}(\alpha, \beta) f(\alpha) f(\beta) d\alpha d\beta \quad (34)$$

where the distribution of α_i and β_i is specified earlier in (32) and it consists of an observable part and an unobservable part that is normally distributed. This is the mixed logit choice probability.

2.3 Estimation Strategy

The consumers probability of choosing an option can be expressed as $\prod s_{ij}^{y_{ij}}$, where $y_{ij} = 1$ if consumer i chose option j and $y_{ij} = 0$ otherwise. Assuming that each consumer's choice is independent of the choice of others, the probability that each client in the sample chooses the product which is observed, the log-likelihood function can be written as follows:

$$L(\theta) = \sum_i y_{ij} \sum_j \log(s_{ij}) \quad (35)$$

where θ is the vector of all the parameters to be estimated. As it is not possible to solve the integral (34) analytically I use simulation method. Following Train (2003) and Grzybowski and all. (2014) we take R Halton draws for the vector v_i from the standard normal distribution to obtain the average choice probability per consumer

$$\bar{s}_{ij} = \frac{1}{R} \frac{\exp(x'_j(\beta + \sigma_\beta v_i^r) - (\alpha + \sigma_\alpha v_i^r) \log(p_{ij}) + (x'_j, \log(p_{ij})) \Pi D_i)}{\sum_{k \in C_i} \exp(x'_k(\beta + \sigma_\beta v_i^r) - (\alpha + \sigma_\alpha v_i^r) \log(p_{ik}) + (x'_k, \log(p_{ik})) \Pi D_i)} \quad (36)$$

In the special case of no unobserved consumers heterogeneity ($\sigma=0$). the equation (36) reduces to the multinomial choice probability:

$$\bar{s}_{ij} = \frac{1}{R} \frac{\exp(x'_j \beta - \alpha p_{ij} + (x'_j, \log(p_{ij})) \Pi D_i)}{\sum_{k \in C_i} \exp(x'_k \beta - \alpha p_{ik} + (x'_k, \log(p_{ik})) \Pi D_i)} \quad (37)$$

The formula (36) substitute s_{ij} in equation (35) and maximization of the likelihood function L provide us with an estimated vector of parameters θ^{55} .

⁵⁵Train (2003) describes the detailed algorithm of multinomial and mixed logit estimations.

2.4 Price elasticities of demand

The multinomial and mixed logit models can be used to calculate the price elasticities of demand. I calculate the own-price elasticities between different products. Product type elasticity is a percentage change in demand for a product type in response to a 1% change in the price of this product (own price elasticity). The calculation of the price elasticities is following. The total market share of the product j is $s_j \equiv \sum_i s_{ij}/N$, where s_{ij} is the probability of the consumer i choose the product j and N is the number of clients. As in this case there are several product characteristics the elasticities are estimated for the product characteristic, not for the each type of product. The effect of the percentage price increase of product k on the level of individual household probability of choosing the product j is:

$$\frac{\delta s_{ij}}{\delta p_{ik}} p_{ik} = \frac{\delta s_{ij}}{\delta \log(p_{ik})} = \begin{cases} -\alpha_i s_{ij}(1 - s_{ij}) & \text{if } k=j \\ \alpha_i s_{ij} s_{ik} & \text{otherwise} \end{cases} \quad (38)$$

This is an individual semi-elasticity of demand for j with respect to the price of k . Summing over all the clients on the route, I get the aggregate semi-elasticity:

$$\frac{1}{N} \sum_i \frac{\delta s_{ij}}{\delta p_{ik}} p_{ik} = \begin{cases} \frac{1}{N} \sum_i (-\alpha_i) s_{ij}(1 - s_{ij}) & \text{if } k=j \\ \frac{1}{N} \sum_i \alpha_i s_{ij} s_{ik} & \text{otherwise} \end{cases} \quad (39)$$

This is the effect of a percentage price increase on the level of aggregate demand for technology j (semi-elasticity), while the aggregate elasticity given as:

$$\varepsilon_{jk} = \frac{1}{N} \left(\sum_i \frac{\delta s_{ij}}{\delta p_{ik}} p_{ik} \right) \frac{1}{s_j} = \begin{cases} \sum_i (-\alpha_i) s_{ij}(1 - s_{ij}) / \sum_i s_{ij} & \text{if } k=j \\ \sum_i \alpha_i s_{ij} s_{ik} / \sum_i s_{ij} & \text{otherwise} \end{cases} \quad (40)$$

To calculate the influence of particular characteristic we need to calculate the elas-

ticity taking into consideration all the products with this characteristic. The aggregate market share for the product of a characteristic δ is given by $s_\delta \equiv \sum_i \sum_{j \in \delta} s_{ij}/N$, N is the number of clients on the given route and δ is the group of all the products with the given characteristic. The effect of 1% price increase for products belonging δ on the level of the individual probability of choosing from the group δ is

$$\sum_{j \in \delta} \sum_{k \in \delta} \frac{\partial s_{ij}}{\partial p_{ik}} p_{ik} = -\alpha_i \sum_{k \in \delta} s_{ik} p_{ik} (1 - \sum_{j \in \delta} s_{ij}) \quad (41)$$

Summed for all the clients on the route ($s_\delta \equiv \sum_i \sum_{j \in \delta} s_{ij}/N$) the aggregate characteristic level semi-elasticity may be defined as:

$$\frac{1}{N} \sum_i \left[\sum_{j \in \delta} \sum_{k \in \delta} \frac{\partial s_{ij}}{\partial p_{ik}} p_{ik} \right] = \frac{1}{N} \sum_i \left[-\alpha_i \sum_{k \in \delta} s_{ik} p_{ik} (1 - \sum_{j \in \delta} s_{ij}) \right] \quad (42)$$

This figure corresponds to the effect of a joint percentage price increase for all the products of some characteristic on the level of aggregate demand for the product with this characteristic. So the aggregate group-level elasticity of demand for the group of products δ with respect with the joint percentage price increase may be defined as:

$$\varepsilon_\delta = \frac{1}{N} \sum_i \left[\sum_{j \in \delta} \sum_{k \in \delta} \frac{\partial s_{ij}}{\partial p_{ik}} p_{ik} \right] \frac{1}{s_\delta} = \frac{1}{N} \sum_i \left[-\alpha_i \sum_{k \in \delta} s_{ik} p_{ik} (1 - \sum_{j \in \delta} s_{ij}) \right] / \sum_i \sum_{j \in \delta} s_{ij} \quad (43)$$

3 Data

3.1 Survey of International Air Travelers

For the purpose of this paper I employ the Survey of International Air Travelers conducted by the American International Trade Administration Office of Travel and

Tourism Industries (2005, 2007). The survey covers the international trips originated from the US and provides the detailed information about each trip. The respondents gave the details about the international flight and if they have any connecting flights, the survey provides information about the starting points and principal destinations of their travel.

For the purpose of my research a number of variables are particularly important: airline company serving the international flight, origin-destination points, number of coupons (direct or indirect flight), departure date, moment of purchase and ticket price. As the ticket price varies a lot among consumers, only a few database can provide with a real price paid for a particular type of trip.

The survey provides also the personal characteristics of the consumers, which are employed as explanatory variables in the model: age of the respondent, income (11 income groups, included as continuous variable), gender (dummy variable equal to 1 if the respondent is female).

The previous articles devoted to the airline competition research employ other databases for the estimations. The commonly used is Destination and Origin Survey (DB1B, by Department of Transportation), a 10% random sample of the all tickets sold in the U.S.⁵⁶ This database is highly reliable but it does not include any details about the passengers or product characteristics. It also concerns the other databases used for competition analysis of the airline market before: Digest Statistics (ICAO), World Air Transport Statistics (IATA), etc.⁵⁷

As it was mentioned above the survey we employ has not been used for competition analysis before. It is can be explained by a number of disadvantages of the database.

⁵⁶see Borenstein (1989), Borenstein and Rose (1994), Dresner and Tretheway (1992), Evans and Kessides (1993), Brueckner and Spiller (1994). etc.

⁵⁷See Gagnepain and Marin (2006, 2010), Marin (1995), Oum and all (1996)

First of all, it is survey type data so there is a great number of missing information. Secondly, it is possible that the variables are not absolutely accurate. For example, not all the respondents remember the exact day of the ticket purchase, so they give an approximate number of days. The same problem can arise with the most important variable for our analysis, the ticket price. The respondent can make a mistake providing the information about the ticket price, so we don't possess the true value of the ticket price (P_t^{obs}); we have a price which can contain a possible measurement mistake ($P_t^{obs} = P_t^{real} + \varepsilon_t$). The third disadvantage of the Survey is a sample problem. We cannot be sure that this database is a perfect approximation of the real market. For example, it is not possible to obtain any plausible information about the volumes of production. Nevertheless, the similar databases are used for travel demand analysis.⁵⁸ We employ this database regardless all the disadvantages as only survey type of data can provide us with all the ticket details and consumers characteristics. The possible measurement problem is not so crucial for the analysis as both variables are introduced in the model as independent variables. To avoid the problem of the possible bias of sample on the real market we truncate our database only to the largest routes of the survey.

3.2 Data handling and descriptive statistics

A lot of work has been done in order to prepare the database for the necessary estimations. I start the analysis with the definition of a route. The principal characteristic of a route is the origin and destination points. The database provides codes of the origin

⁵⁸ Reece (2001), Divisekera (2003), Racello and all (2005), Laesser and Crouch (2006), Cortes Jimenez and Blake (2010).

and destination airports of an international direct flight originated from the U.S. But there are a lot of passengers traveling by indirect flights and we have to take them into consideration as well. During the survey respondents answered the questions about the starting and destination points of their journey. The database provides the origin airports which are situated on the territory of the United States. As far as destination point is concerned, I know only the city (town) or region of destination, not the airports. Therefore it is decided to use the cities or towns as the origin-destination points, as the lack of the information does not allow me to use the airports in my analysis. Moreover, this approach allows having more observations for every route that could make our results more consistent.⁵⁹

After the unification of the origin and destination information, I receive a database of a number of direct and indirect flights covering the whole journey of a passenger. The number of coupons vary from one (for direct flight) to three. In database there is only variable that indicates what airline served the direct international flight, so I have to assume that the same airline which is indicated for the direct flight served all the connecting flights. It is a reasonable assumption, as it is usually much cheaper to buy the whole trip from the same company than to buy the tickets separately from point to point.⁶⁰ Some of the flights were bought as a part of the package tour. For the further analysis I do not control for the number of coupons, the data is just split in two part: direct and indirect flights. It is connected with the fact that in case of multinomial

⁵⁹The both approaches are used in the literature. Some authors use the airports not cities as the origin destinations points (Borenstein (1989) , Morrison, Winston (1990), Kim, Singal (1993), Borestein, Rose (1994)). Others prefer cities pairs for the analysis of competition on the airline market (Berry (1990), Forbes and Lederman (2009), Barley, Pazar (1981), Dresner, Tretheway (1992), Evans, Kessides (1993), Brueckner, Spiller (1994), Sinclair (1995), Marin (1995), Oum, Park, Zhang (1996)). There is also a number of article with the control for both: cities and airports (Brueckner, Dyer, Spiller (1992), Ciliberto and Tamer (2009)).

⁶⁰Usually the tickets for indirect flights are less expensive than for direct flights by the same company and for the same route, however they cost more for the company. It also a yield management strategy allowing to distinguish consumers more sensitive for price.

logit I need to limit the number of choices⁶¹. Also the percentage of flights with more than two coupons is very small.

Then I have to delete a number of observations for several reason. First of all, as in any survey there is a lot of omissions when the interviewees failed to answer some questions, so we eliminated the observations without information for crucial variables, like price or reservation moment. Two more types of observations are also deleted: one way tickets and first class tickets. For both of them the airline companies employ special pricing strategy and these tickets present a very small percentage of the data. I try to introduce the business class tickets into model, but the model proves to be unstable. Therefore only the economy class tickets are used for analysis, which should not bias the results as business class and economy class travelers represent almost independent markets. Calculation of the descriptive statistics shows a huge dispersion of the price per kilometer, so in order to minimize the measurement mistake I delete the extreme values: 1% quantile for the highest values and 1% quantile for the smallest price variable (price per kilometer).

3.3 Choice set

A set of mutually exclusive and exhaustive choice alternatives is to be specified for each consumer. I specify the choice set as consisting of all the types of tickets available on a given route. There are some slight differences in types of products depending on specific route, therefore the choice set varies between routes. So as a choice option for the model I consider a direct or an indirect ticket by a particular company on a

⁶¹The number of observation in our dataset is small (less than 300 for the largest routes), and multinomial logit models are very sensitive to the number of choices, so I need to keep the number of choices to the minimum.

given route. I start the analysis the model estimations for the largest route in the dataset: route Los-Angeles, US - Taipei, Taiwan. Then I also estimate the models on the number of different routes. On each route the estimations are held independently, so the routes should also answer a number of conditions. They should have more than 80 observation per route and at least 3 choice options. There is also should be at least one airline which has both direct and indirect flights on this route. Under all these restrictions I end up with the 18 route ⁶². The introduction of the route characteristics in the model is meaningless, as a separate regression is estimated separately for each route.

Tables 16-17 present descriptive statistics of the main variables for the whole set. Even though there is only 18 routes they are quite different. All of them originate from the U.S., but they have various destinations points starting from cities in Mexico to Japan and South Korea. That gives us a huge dispersion in distance from 1.000 kilometers to 13.500 kilometers. Prices are various, which can be explained by the diversity of routes and their characteristics. The price per kilometer is also calculated (but it is not used in regressions).

The maximum period of the reservation is 450 days in the restricted base. In the original base this figure is much bigger (730 days, but less than 1% of the tickets reserved more than 450 days before departure). The reservations made more than eight months before departure are usually exceptional.

Most of the flights are direct, about 87%. In the original data set the part of direct flights is much less: in the unrestricted database this ration is 43% of direct flights. The larger the route the higher the number of direct options and share of the direct tickets.

⁶²The list of the routes with the statistics in Annex, Table 27

Table 16: Descriptive statistics

		Mean	Sd.dev	Min	Max
Price	Price per tike, US\$	1028	523	50	4500
Pr km	Price per kilometer	0.12	0.07	0.03	0.49
Distance	Distance between origin and destination cities, km	9203	2461	1161	13588
Res	The time before departure the reservation was made, days	37	44	0	450
Age	Consumer age, years	42	13	18	85
Nr	Number of observation per route	260	143	86	520
Nc	Number of choices per route	6	2	3	9
Obs.number		3173			

Table 17: Descriptive statistics (2)

Percent	
Direct	87%
Indirect	13%
Freq.	
Female	37%
Male	63%

4 Estimation results

4.1 Demand models for the route Los Angeles - Taipei

The multinomial logit is very sensitive to number of variables. As I do not have a lot of observations per route I include only the interaction of consumer characteristics with the option characteristics, but not the option variable itself. This approach also allows for the clearer interpretations of the estimation results. Continuous variables like price and reservation moment (number of days before departure the reservation is made) are used as logarithms.⁶³The results of the several multinomial logit models are in the Table 18. All the significant coefficients have the expected signs. Nevertheless the interactions with the most of the consumer characteristics are insignificant. On this route only the interaction of price and income has a significant influence on the market share of these product type. The price impact is strictly negative: higher the price, less likely the ticket is bought and this influence is decreasing with the growth of the consumers income. The moment of the reservation has also significant influence: consumers prefer to stay flexible and buy tickets closer to the departure date. It explains why the ticket prices usually increase closer the departure date. The binary variable of the direct flights is also strongly significant: the direct flights are more preferable by the consumers.

The second set of the equations are estimated with the mixed logit models, when σ is different from zero (see equation 32). The results are in the Table 19. I employ 100 Halton draws to estimate the standard deviations of the coefficients⁶⁴. These models

⁶³I try also to estimate the models with linear price and linear reservation moment, but in the most of the models they are insignificant. Moreover to measure the price influence in relative terms allows to compare the coefficient between routes, as the price level varies between routes.

⁶⁴In a number of articles has been shown that 100 Halton draws is enough for mixed logit models (Grzybowski and all. (2009), Nevo (2001), etc.)

Table 18: Multinomial logit model, market Los Angeles, US - Tapei, Taiwan

Variable	Model 1.1	Model 1.2	Model 1.3	Model 1.4	Model 1.5
log (price)	-1,81 0,36	-4,98 1,25	-5,01 1,26	-5,23 1,26	-5,12 0,74
log(res)	-1,43 0,11	-1,40 0,11	-1,40 0,11	-0,52 0,43	-1,44 0,14
direct	1,84 0,16	1,84 0,16	3,11 0,65	3,07 0,65	1,84 0,16
log(price)*income		0,54 0,10	0,54 0,10	0,58 0,11	0,54 0,10
log(price)*age		0,00 0,03	0,00 0,03	0,00 0,03	
log(price)*female		-0,15 0,63	-0,12 0,63	-0,15 0,63	
direct*income			0,02 0,06	0,02 0,06	
direct*age			-0,03 0,01	-0,03 0,01	
direct*female			-0,26 0,32	-0,24 0,32	
log(res)*income				-0,10 0,04	
log(res)*age				-0,01 0,01	
log(res)*female				0,09 0,21	0,11 0,21
LL	-675,42	-661,30	-658,40	-654,35	-661,18

Number of options on the market - 9

are estimated to test how the possibility of the consumers heterogeneity effects the coefficients of the product characteristics. The models predict large standard deviations for all coefficients estimated, nevertheless significant for price and reservation moment variables the standard deviations are strongly significant. Coefficients of the direct route dummy show no connection with the consumers heterogeneity. The results of mixed logit models are similar to the standard multinomial logit: price and number of days before departure have strongly negative influence, and direct flight dummy strongly positive. Among all the interactions of product attributes with consumers characteristics, as in the standard multinomial logit, only the interaction of income with price is significant: the price elasticity is decreasing with the growth of consumers income. Although the results seems to be similar to standard multinomial logit the absolute values of the coefficients for price and direct flights are larger than in standard multinomial logit. The effect of the number of days before departure is almost the same. That means that there some unobservable consumers heterogeneity which should be taken into account in order to get more accurate estimation of the price elasticities.

4.2 Price elasticities

The estimated price elasticities for all the options of the route Los Angeles - Tapei are presented in Table 20 and Table 21. With the price increase by 10 % for the direct ticket with China Airlines, the share of the direct tickets of this airline decrease by about 0.1. In case of the other options the influence is much less. While semi-elasticities are much smaller for the indirect options, the relative percentage change of the market share is much more impressive (Table 21). In standard multinomial models price-elasticities are very close between options, 1% change in prices leads to the 1-2% change in market share of this product type. In case of the mixed logit models the price elasticities are

Table 19: Mixed logit model, market Los Angeles, US - Tapei, Taiwan

Variable	Model 2.1	Model 2.2	Model 2.3	Model 2.4	Model 2.5
log (price)	-7,75 1,32	-13,91 0,64	-9,71 3,91	-6,25 3,19	-19,15 1,88
log(res)	-1,41 0,12	-6,09 1,12	-0,15 0,54	-0,74 1,46	-1,95 3,09
direct	2,36 0,30	5,36 0,80	4,11 1,09	3,40 3,63	5,09 0,81
log(price)*income			1,25 0,43	0,74 0,35	1,50 0,45
log(price)*age			-0,10 0,09	-0,01 0,05	
log(price)*female			-1,16 2,12	-0,03 1,27	
direct*income			0,02 0,12	0,05 0,29	
direct*age			-0,04 0,02	-0,04 0,06	
direct*female			-0,40 0,56	-0,02 1,43	
log(res)*income			-0,14 0,04	-0,12 0,15	
log(res)*age			-0,01 0,01	-0,01 0,03	-0,08 0,07
log(res)*female			0,02 0,26	-0,07 0,74	
sd.lpr	16,60	24,47	16,21	5,95	23,12
d sd.lpr	1,96	3,05	1,99	2,78	4,27
sd.lnres		13,54		5,81	13,01
d sd.lnres		1,83		2,28	2,34
sd.direct		0,12		7,95	
d sd.direct		4,57		5,96	
LL	-558,66	-473,40	-547,30	-628,44	-470,12

Number of options on the market - 9

Table 20: Semi-price elasticity of the options

Airline	Route	Model 1.1	Model 1.5	Model 2.1	Model 2.5
CI-China Airlines	indirect	0,09	0,11	0,38	0,06
CI-China Airlines	direct	0,29	0,34	1,17	0,97
JL-Japan Airlines	indirect	0,01	0,02	0,01	0,00
SQ-Singapore Airlines	indirect	0,03	0,02	0,03	0,01
SQ-Singapore Airlines	direct	0,21	0,25	0,72	0,22
MH-Malaysian Airlines	indirect	0,01	0,01	0,01	0,00
MH-Malaysian Airlines	direct	0,18	0,24	0,50	0,14
BR-EVA Airways	indirect	0,03	0,03	0,07	0,01
BR-EVA Airways	direct	0,30	0,34	0,95	0,89

much higher, especially for the indirect routes. An increase in price by 1 % lead to an increase of market share by 20-30% , although the market share of the indirect route is very small and this change does not lead to a considerable change on the market.

I also calculate an elasticity for a group options: change of direct share with the price change of all the direct options (see the equation 41 for calculation of the group level elasticity). Tables 22 and Table 23 presents the results. The price elasticity of mixed logit modes is several times higher than the price elasticity estimated with the model without control for consumers heterogeneity. The results of the estimation shows necessity to control for unobservable consumers heterogeneity, otherwise the model results can considerably underestimate the price elasticities on the market.

4.3 Demand models for all the routes⁶⁵

In my database I have a lot of routes available for the research so I test my model

⁶⁵The list of routes is in the Table 27, in Annex

Table 21: Price elasticities of the options

Airline	Route	Multinomial logit*		Mixed logit model*		Option share
		Model 1.1	Model 1.5	Model 2.1	Model 2.5	
CI-China Airlines	indirect	1,7	2,1	6,9	6,5	0,02
CI-China Airlines	direct	1,0	1,2	3,0	2,1	0,33
JL-Japan Airlines	indirect	1,8	2,6	7,3	9,6	0,01
SQ-Singapore Airlines	indirect	1,7	1,6	4,3	4,4	0,01
SQ-Singapore Airlines	direct	1,1	1,3	3,9	2,3	0,19
MH-Malaysian Airlines	indirect	1,8	1,9	6,8	8,6	0,01
MH-Malaysian Airlines	direct	1,3	1,7	4,7	3,1	0,13
BR-EVA Airways	indirect	1,7	1,8	5,9	2,5	0,05
BR-EVA Airways	direct	1,1	1,2	3,8	2,4	0,25

*Exact model specification see in the Tables 18 and 19

Table 22: Semi - price elasticities between direct-indirect tickets

	Model 1.1	Model 1.5	Model 2.1	Model 2.5
Semi-price elasticity	0,16	0,19	0,48	0,08

Table 23: Price elasticities between direct-indirect tickets

	Model 1.1	Model 1.5	Model 2.1	Model 2.5
Indirect routes	7,7	9,3	31,1	23,9
Direct routes	0,7	0,8	2,1	0,3

Table 24: Significance of variables in the multinomial logit models on all markets

	Model 1.1	Model 1.2	Model 1.3	Model 1.4	Model 1.5
log (price)	72%	50%	44%	44%	78%
log(res)	100%	100%	100%	22%	94%
direct	89%	89%	44%	33%	89%
log(price)*income		44%	44%	39%	39%
log(price)*age		6%	6%	6%	
log(price)*female		28%	28%	28%	
direct*income			0%	0%	
direct*age			11%	17%	
direct*female			0%	6%	
log(res)*income				28%	
log(res)*age				22%	
log(res)*female				44%	28%

Number of markets - 18

on a set of the suitable routes. Table 24 presents the summary of the results of all the estimated models for these routes. It shows in how many routes the variables are significant in several model specifications. On some route a range of the consumer characteristics product attributes interactions are significant, but mostly the models of the other routes are not different from the route Los Angeles - Taipei

Another set of estimation is made by mixed logit model (random coefficient model), table 25 presents all the models which are estimated. The results are similar the basic multinomial logit: only the interaction between consumers income and product price is significant on most of the routes.

Table 26 presents an average influence of the variables in the models. Average influence is calculated only if the variable on this route is significant. The standard deviations given in the table are the dispersion of all the significant coefficients in the models of all the routes. In case of multinomial logit the influence of the main variables (price, reservation moment and direct/indirect flights) is stable: on all the routes

Table 25: Significance of variables in the estimated mixed logit models on all markets

	Model 2.1	Model2.2	Model 2.3	Model 2.4	Model 2.5
log (price)	61%	61%	28%	22%	67%
log(res)	100%	72%	17%	17%	28%
direct	89%	50%	28%	0%	67%
log(price)*income			17%	22%	50%
log(price)*age			6%	11%	
log(price)*female			11%	11%	
direct*income			0%	0%	
direct*age			6%	0%	
direct*female			6%	0%	
log(res)*income			17%	0%	
log(res)*age			17%	11%	33%
log(res)*female			17%	11%	
sd.log(price)	78%	72%	61%	50%	78%
sd.direct		28%		0%	
sd.log(res)		56%		28%	61%

Number of markets - 18

the influence of the variables is very similar. The coefficients of the price variable are strongly negative: higher the price of the ticket less likely this ticket is bought. The binary variable of the direct flights is also strongly significant: the direct flights are more preferable by the consumers. The influence of the reservation moment is highly negative: all others things being equal consumers prefer to buy tickets closer to the departure data.

In the mixed logit models, the variables influence is less heterogeneous. The price influence varies considerably between the routes. The possible explanation is that the price levels on the routes are very diverse. The impact of the moment of reservation is also not very stable among routes, which also can be explained by the diversity of routes. The set of the carriers on the route, touristic or business destination, distance can all influence the consumers appreciation of the flexibility.

The introduction of the consumers characteristics into the model makes model less

Table 26: Average influence of the parameters in the models

	Model 1.1		Model 1.5		Model 2.2		Model 2.5	
	Coef	St.dev	Coef	St.dev	Coef	St.dev	Coef	St.dev
log (price)	-1,95*	(0,44)	-4,92*	(2,20)	-11,29	(8,71)	-14,01*	(8,07)
log(res)	-1,20*	(0,24)	-1,37*	(0,32)	-4,63	(3,94)	-2,68	(3,00)
direct	2,20*	(0,91)	2,22*	(0,92)	7,39*	(4,64)	4,17*	(1,91)
log(price)*income			0,68*	(0,22)			1,19	(0,90)
log(price)*age								
log(price)*female								
direct*income								
direct*age								
direct*female								
log(res)*income								
log(res)*age							-0,04	(0,09)
log(res)*female			0,42	(1,49)				
sd.log(price)					13,76	(25,77)	21,98	(14,38)
sd.log(res)					7,51	(5,91)	6,20	(4,12)
sd.direct					-3,17	(4,06)		

Number of markets - 18

robust. Even when on the most of the routes the variables are significant, their influence is different different routes. Therefore I cannot make any conclusion about the effect of the particular consumers characteristics on the ticket demand.

5 Conclusions

This paper provides demand estimations for airline passenger market which treated as a market with the product differentiation. The paper offers a way to estimate demand taking into consideration the differences in ticket characteristics. I apply the multinomial logit model for demand estimation, and extend the analysis with random coefficient model to achieve more accurate estimation of the price elasticities. For the purpose of the analysis I apply a Survey of the International Air Travelers, which provides detailed product characteristics as well as clients characteristics. Survey covers

the international flights from the USA to all over the world.

I start the analysis with the largest route in the sample: Los Angeles, US - Taipei, Taiwan. There are two characteristics defining the choice options: airline company and direct - indirect attribute. Multinomial logit allows to estimate how the product characteristics influence the shares of each option. All the coefficients in the estimated regressions have the expected signs. Price and buying ticket in advance have a strongly negative impact on market share, direct flights are on the contrary much more preferable by consumers. I include gender, age and income level of the consumers in the model to control for consumer heterogeneity. These consumers characteristics show no significant influence except for the interaction of the price and consumer income: clients with higher level of income care less about the price change (their price elasticity is smaller).

The random coefficient model shows the similar results, the coefficients have the same significance and signs as in the standard coefficient models. But the absolute values of the coefficients of the price variable and direct route dummy are much higher than in the standard multinomial model. Based on the estimated models I calculate the price elasticities for the options. The results shows that using the mixed logit models leads to significantly higher values of the elasticities.

I also test how the model works on the other routes in the survey. The models are estimated on a sample of largest route with enough options (18 route). On the most of the routes the signs of the coefficients and significance of the coefficients are the same, but the absolute values can be very diverse. On most of the routes consumers characteristics, except for income-price interaction, have no significant influence. With the control for consumers heterogeneity the difference in model between routes become more profound.

The results of all the estimated models show that the consumers heterogeneity plays an important role in estimation of demand and price elasticities for the airline passengers market. The number of available consumers characteristics like income, age, gender cannot explain it, therefore it is necessary to rely on simulations.

For the further research on this subject, it is necessary to obtain more extend database, which, first, provides more accurate market estimation and, second, supplies more observations to boost the quality of the model. Also in order to control for the possible endogeneity problem, it is necessary to look for the instrumental variables for price.

Bibliography

Basso, L. J. (2008). Airport deregulation: Effects on pricing and capacity. *International Journal of Industrial Organization*, 26(4), 1015-1031

Basu, K. (1992). A characterization of the class of rationalizable equilibria of oligopoly games. *Economics Letters*, 40(2), 187-191.

Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, 1007-1028.

Boguslaski, C., Ito, H., & Lee, D. (2004). Entry patterns in the southwest airlines route system. *Review of Industrial Organization*, 25(3), 317-350.

Borenstein, Severin (1989) Hubs and High Fares: Dominance and Market Power in the U.S., *The RAND Journal of Economics*, vol 20, No 3

Borenstein, S. and N. Rose (1994) Competition and Price Dispersion in the U.S. Airline Industry, *Journal of Political Economy*, vol. 102, No.4, pp.653 -683

Botimer, T.C. and P. P. Belobaba (1999) Airline Pricing and Fare Product Differentiation: A New Theoretical Framework, *The Journal of the Operational Research Society*, Vol. 50, No. 11, pp.1085-1097

Brander, J. A., & Zhang, A. (1990). Market conduct in the airline industry: an empirical investigation. *The RAND Journal of Economics*, 567-583

Brueckner, J. K. (2002). Airport congestion when carriers have market power. *American Economic Review*, 92(5), 1357-1375

Brueckner J., P. Spiller (1994) Economies of Traffic Density in the Deregulated Airline Industry, *Journal of Law and Economics*, Vol.37, No.2, pp.379-415

Caves, D. W., Christensen, L. R., & Tretheway, M. W. (1984). Economies of density versus economies of scale: why trunk and local service airline costs differ. *The RAND Journal of Economics*, 471-489

Chen, V., Günther D. and E. Johnson (2003) Solving for an Optimal Airline Yield Management Policy via Statistical Learning, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 52, No. 1, pp.19-30

-
- Clemons, E., Hann, I. and L. Hitt (2002) Price Dispersion and Differentiation in Online Travel: An Empirical Investigation, *Management Science*, Vol. 48, No. 4, pp.534-549
- Coughlan, J. (1999) Airline Overbooking in the Multi-Class Case, *The Journal of the Operational Research Society*, Vol. 50, No. 11, pp.1098-1103
- Dana, J. (1999) Using Yield Management to Shift Demand When the Peak Time is Unknown, *The RAND Journal of Economics*, Vol. 30, No. 3, pp.456-474
- Davis, P., & Garcés, E. (2009). Quantitative techniques for competition and antitrust analysis. *Princeton University Press*.
- Desgranges, G., & Gauthier, S. (2016). Rationalizability and efficiency in an asymmetric Cournot oligopoly. *International Journal of Industrial Organization*, 44, 163-176.
- Desiraju, R. and S. Shugan (1999) Strategic Service Pricing and Yield Management, *Journal of Marketing*, Vol. 63, No. 1, pp.44-56
- Dresner, M. and M. Tretheway (1992) Modeling and Testing the effects of Market Structure on Price, *Journal of Transport Economics and Policy*, Vol. XXVI, No2, pp.171-184
- Durand, B. and E. Pesaresi (2007) Empirical estimation of discrete choice model for filler calcium carbonates in the paper industry, *Merger Control*
- Evans, W. and I. Kessides (1993) Localised Market Power in the U.S. Airline Industry, *The Review of Economics And Statistics*, vol.75, No.1m, pp.66-75
- Feng, Y. and B. Xiao (2000) A Continuous-Time Yield Management Model with Multiple Prices and Reversible Price Changes, *Management Science*, Vol. 46, No. 5, pp.644-657
- Gagnepain, P. and P. Marin (2006) Regulation and Incentives in European Aviation, *Journal of Law and Economics*, vol. XLIX
- Gagnepain, P. and P. Marin (2010) The Effects of Airline Alliances: What Do the Aggregate Data Say? , *SERIEs*, Spanish Economic Association, vol.1 (3), pp.251-276
- Giaume, Stephanie and Sarah Guillou (2004) Price Discrimination and Concentration in European Airline Markets, *Journal of Air Transport Management*, Vol. 10, pp.305-310
- Goolsbee, A., & Syverson, C. (2008). How do incumbents respond to the threat of entry? Evidence from the major airlines. *The Quarterly journal of economics*, 123(4), 1611-1633.

Guesnerie, R. (1992). An exploration of the eductive justifications of the rational-expectations hypothesis. *The American Economic Review*, 1254-1278.

Grzybowski, L., Nitsche, R., Verboven, F., and L.Wiethaus (2014) Market definition for broadband internet in Slovakia – Are fixed and mobile technologies in the same market?, *Information Economics and Policy*, 28, pp.39-56

Grzybowski, L. and P. Pereira (2007) Simulation of Merger in mobile Telephony in Portugal, *Review of Industrial Organisation*, 31 (3), pp.205-220

Grzybowski, L. and P. Pereira (2011) Subscription choices and switching costs in mobile telephony, *Review of Industrial Organisation*, 38, pp.23-42

Ivaldi, M. and F.Verboven (2005) Quantifying the effects from horizontal mergers in European competition policy, *International Journal of Industrial Organization*, 23, pp.669-691

Leal de Matos, P. (2001) Yield Management for Privatised Air Traffic Control?, *The Journal of the Operational Research Society*, Vol. 52, No. 8, pp.888-895

Lee, J., Kim, Y., Lee, J.D. and Y. Park (2006) Estimating the extent of potential competition in the Korean mobile telecommunication market: switching costs and number portability, *International Journal of Industrial Organisation*, 24, pp.107-124

Marin, P. (1995) Competition in European Aviation: Pricing Policy and Market Structure, *The Journal of Industrial Economics*, Vol. 43, No.2 , pp.141-159

Morrison, S. A. (2001). Actual, adjacent, and potential competition estimating the full effect of Southwest Airlines. *Journal of Transport Economics and Policy (JTEP)*, 35(2), 239-256.

Moulin, H. (1984). Dominance solvability and Cournot stability. *Mathematical social sciences*, 7(1), 83-102.

Nevo, A (2000) “Mergers with differentiated products: The case of the ready-to-eat cereal industry, *Rand Journal of Economics*, 31, pp.395-421

Novshek, W. (1984). Finding all n-firm Cournot equilibria. *International Economic Review*, 61-70.

Smith, B., Leimkuhler, J. and R. Darrow (1992) Yield Management at American Airlines, *Interfaces*, Vol. 22, No. 1, pp.8-31

Peteraf, M. (1995) Sunk Costs, Contestability and Airline Monopoly Power, *Review of Industrial*

Organisation, Vol.10, pp.289-306

Stavins, (2001) Price Discrimination in the Airline Market: The Effect of Market Concentration, *The Review of Economics And Statistics*, vol.83, No.1,

Train, K., (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge

Zhang Anming, Park Jong-hun, Oum Tae Hoon (1996) The Effects of Airline Codesharing Agreements on Firm Conduct and International Air Fares, *Journal of Transport Economics and Policy*, vol.30, no. 2., pp.187-202

Table 27: List of routes with descriptive statistics

Origin	Destination	Price	Res	Direct	Income	Age	Female	Pr	km	Dist	Nc	Nr
Dulles, MD	London	842	38	0,95	7,2	46	0,44	0,14	5951	3	117	
	U.K.	438	44		2,9	14		0,07				
New York, NY	Athens	1024	70	0,92	6,2	43	0,54	0,13	7955	4	153	
	Greece	560	67		3,0	15		0,07				
New York, NY	Tokyo	1176	40	0,97	6,0	40	0,39	0,11	10859	5	189	
	Japan	399	40		2,9	12		0,04				
New York, NY	Moscow	927	31	0,73	5,0	41	0,46	0,12	7504	5	131	
	Russia	366	28		3,1	14		0,05				
New York, NY	Seoul	1400	32	0,90	5,6	41	0,42	0,13	11119	5	231	
	S. Korea	455	40		2,9	13		0,04				
New York, NY	Kiev	1050	48	0,64	3,8	44	0,52	0,14	7556	4	86	
	Ukraine	475	45		2,4	16		0,06				
New York, NY	Frankfurt	767	41	0,96	6,3	43	0,36	0,12	6207	4	167	
	Germany	341	43		2,8	13		0,06				
Los Angeles, CA	Taipei	956	24	0,90	5,3	43	0,39	0,09	10946	9	520	
	Taiwan	434	28		2,6	12		0,04				
Los Angeles, CA	Tokyo	914	39	0,95	5,5	41	0,38	0,10	8776	8	267	
	Japan	391	41		2,7	13		0,04				
Los Angeles, CA	Seoul	1234	33	0,66	5,2	42	0,39	0,13	9652	4	91	
	S. Korea	524	34		2,5	13		0,05				
Phoenix, AZ	Los Cabos	407	72	0,93	6,6	46	0,43	0,35	1161	3	90	
	Mexico	117	78		3,1	16		0,10				
Phoenix, AZ	Puerto Vallarta	407	67	0,95	6,1	48	0,39	0,26	1563	3	97	
	Mexico	139	69		2,9	15		0,09				
San Francisco, CA	London	899	57	0,93	6,3	43	0,36	0,10	8641	4	121	
	U.K.	418	50		3,2	15		0,05				
San Francisco, CA	Singapore	1596	50	0,43	7,3	40	0,23	0,12	13588	4	122	
	Singapore	650	67		2,9	10		0,05				
San Francisco, CA	Taipei	1099	27	0,93	6,6	42	0,32	0,11	10415	5	413	
	Taiwan	547	30		2,9	11		0,05				
San Francisco, CA	Tokyo	1032	35	0,97	6,4	42	0,33	0,13	8250	5	189	
	Japan	519	35		2,8	13		0,06				
San Francisco, CA	Seoul	1050	26	0,89	6,6	41	0,28	0,12	9111	7	246	
	S. Korea	441	28		3,1	12		0,05				
San Francisco, CA	Hong Kong	1180	37	0,71	6,1	41	0,31	0,11	11152	8	210	
	Hong Kong	689	41		2,8	12		0,06				