
Impact de la Concentration de Marché sur les Tarifs Aériens : Une Analyse Économétrique des Routes Domestiques aux États-Unis (1997-2000)



Master 1 Analyse et Politique Economique
Parcours Data Scicene pour l'économie et l'entreprise

Par BALLOGOU Essi Carole Claudia

Prof : El Ouardighi Jahal

Plan

1. Introduction.....	3
2. Revue de la littérature	3
3. Cadre statistique.....	4
3.1 Données et sources	4
3.2 Etude descriptive	4
4. Résultats	4
4.1 Interprétation et commentaire	5
4.2 Test d'existence des effets fixes.....	7
4.3 Test d'Hausman.....	7
5. Conclusion	7
6. Références bibliographiques.....	8
7. Annexes.....	9

1. Introduction

Le transport aérien est un secteur complexe, soumis à de nombreuses variables qui influent directement sur les prix des billets pour un trajet donné. Parmi ces variables, nous avons non seulement les coûts liés à la distance du trajet (carburants, salaire de l'équipage...) mais aussi la concentration de marché (la part de marché) détenu par les compagnies sur un trajet donné. L'étude qui sera menée dans ce papier a pour but d'évaluer, à l'aide d'outils économétriques, l'impact de la part de marché des compagnies aériennes sur les tarifs aériens en tenant compte de la présence potentielle d'effets temporels et individuels non observés. On suppose les effets individuels (par route) pour capturer les caractéristiques propres et non observées de chaque route (comme les infrastructures aéroportuaires, la localisation des villes, ou la concurrence locale sur chaque itinéraire) et les effets temporels pour tenir compte des variations globales dans le temps comme des chocs économiques, des changements réglementaires ou des tendances du secteur aérien entre 1997 et 2000 aux USA.

Pour contrôler ces effets, nous incluons des **variables indicatrices** pour les années (y98, y99, y00), malgré le fait que la méthode « within » les éliminera. Ces variables restent néanmoins utiles dans l'estimation OLS et GLS, permettant une comparaison des résultats et une analyse complète des éventuelles tendances temporelles. La **variable dépendante** est le **log des tarifs aériens moyen** (*lfare*) d'un aller simple d'une ville A à destination d'une ville B. Concernant les **variables explicatives**, nous utilisons la **distance** (log-transformée, *ldist*) et son carré (*ldistsq*) pour capturer une éventuelle relation non linéaire entre la distance et les tarifs aériens. En effet, il est plausible que les tarifs augmentent plus rapidement sur des trajets courts (en raison des coûts fixes élevés par kilomètre) et moins rapidement sur des trajets longs. La variable logarithmique du **nombre de passagers** (*lpassen*) est incluse car, les avions pleins sont moins chers (par passager) que les avions vides, donc les coûts, et par conséquent les tarifs, pourraient être plus bas sur les routes très fréquentées. Et donc, le log du nombre moyen de passagers sur un trajet influence les tarifs. La **concentration de marché** (*concen*), qui est notre **variable d'intérêt**, mesure la domination de la plus grande compagnie aérienne sur une route donnée. Ceci permet de mesurer si une concentration élevée est associée à des tarifs plus élevés.

Ainsi, L'équation du modèle est formulée comme suit :

$$lfare_{it} = \beta_0 + \beta_1 concen_{it} + \beta_2 lpassen_{it} + \beta_3 ldist_{it} + \beta_4 ldistsq_{it} + \varepsilon_t + \mu_i + v_{it} , \\ i = 1, ..., 4596, t = 1, ..., 4$$

Où : β_0 c'est le log du tarif moyen en 1997, ε_t capture les effets temporels et représente nos variable indicatrices ($\alpha_1 y98 + \alpha_2 y99 + \alpha_3 y00$), μ_i capture l'effet spécifique à chaque route, et v_{it} l'erreur résiduelle.

NB : Afin de limiter les écarts et de limiter le problème d'hétérogénéité j'ai utilisé le logarithme

2. Revue de la littérature

Les déterminants des tarifs aériens ont été largement étudiés dans la littérature économique. Nos résultats rejoignent plusieurs conclusions établies par des travaux précédents. Tout d'abord, la relation positive entre la concentration du marché et les tarifs, observée dans notre étude, est en lien avec les travaux de Borenstein (1989) qui a montré que la réduction de la concurrence sur des routes spécifiques permet aux compagnies aériennes d'exercer un pouvoir de marché et d'augmenter leurs prix. Ensuite, la structure non linéaire de la relation entre la distance et les tarifs, avec des coûts fixes plus élevés pour les courts trajets, est cohérente avec les analyses de Morrison et Winston (1986). Enfin, nos résultats mettent en évidence le rôle prépondérant des caractéristiques spécifiques à chaque route, en accord avec les travaux de Dresner et Tretheway (1992) qui ont souligné l'importance des effets individuels par rapport aux variations temporelles globales.

3. Cadre statistique

3.1 Données et sources

Le jeu de données utilisé est ‘airfare’ de Wooldridge, avec **4596 observations**. Aucune donnée manquante, ce qui est essentiel en panel.

On a les variables suivantes :

- **Année** (4ans): 1997, 1998, 1999, 2000
- **Id** : identifiant de la route, c’est la liaison spécifique entre deux villes
- **lfare** : Logarithme du tarif moyen
- **concen** : Concentration de marché détenue par la plus grande firme, sur une route donnée
- **lpassen** : Logarithme du nombre moyen de passagers (lpassen)
- **ldist** : Log-distance , en miles
- **ldistsq** : $ldist^2$
- **y98** : = 1 si année = 1998, sinon 0
- **y99** : = 1 si année = 1999, sinon 0
- **y00** : = 1 si année = 2000, sinon 0

Wooldridge Source: R.C. Fair (1978), “A Theory of Extramarital Affairs,” Journal of Political Economy 86, 45-61, 1978.

3.2 Etude descriptive

La distribution de nos différentes variables semble légèrement asymétrique car la médiane de chaque est légèrement différent de la moyenne (Cf *Annexe, Tableau1*). **Le tarif moyen d’un simple allé**, pour une distance donnée constant dans le temps et spécifique à chaque compagnie, vaut environ **164 \$** [$=exp(5.10)$]. Dans les années 90 aux USA, le prix minimum d’un billet d’avion en moyenne, était d’environ 37\$ [$=exp(3.61)$] pour une distance d’environ 153km [$=95miles*1.6km$]. Le prix maximum, en moyenne, était d’environ 533\$ [$=exp(6.26)$] pour 4385km soit la distance entre New York et Los Angeles. Ces tarifs sont bien sûr moins élevés que maintenant à cause de l’inflation actuelle. La part de marché, semble suivre une loi normale car la médiane est juste légèrement au-dessus de la moyenne. Toutes les compagnies étudiées des grandes compagnies de ce fait on a un faible écart avec la moyenne.

4. Résultats

L’estimation du modèle est réalisée à l’aide des méthodes OLS, within et GLS, afin de comparer les résultats.

Tableau 2 : Comparaison des modèles OLS, Within et GLS			
	Dependent variable:		
	Logarithme des tarifs moyens aériens (lfare)		
	OLS	panel linear	
	OLS (1)	Within (2)	GLS (3)
Concentration de marché (concen)	0.310*** (0.030)	0.150*** (0.022)	0.153*** (0.021)
Log-passagers (passen)	-0.079*** (0.006)	-0.370*** (0.007)	-0.321*** (0.006)
Log-distance (ldist)	-0.972*** (0.126)		-1.189*** (0.351)
Log-distance au carré (ldistsq)	0.107*** (0.010)		0.119*** (0.026)
Année 1998 (y98)	0.023* (0.014)		0.029 (0.065)
Année 1999 (y99)	0.042*** (0.014)		0.056 (0.065)
Année 2000 (y00)	0.106*** (0.014)		0.125* (0.065)
Constant	7.010*** (0.416)		9.479*** (1.155)
Variance du terme d'erreur	0.108	0.005	0.006
Variance individuelle			0.229
Variance temporelle			0.002
Constante dans le modèle Within		7.228	
Observations	4,596	4,596	4,596
R ²	0.431	0.435	0.396

Note: *p<0.1; **p<0.05; ***p<0.01

Source : Construit par l'auteur, à partir des données de Wooldridge sur les tarifs aériens aux USA

4.1 Interprétation et commentaire

Le tableau 2 présente le modèle économétrique estimé sur un panel de donnée, par trois estimateurs (OLS, Within, GLS). L'estimateur OLS est le plus simple et ne tient pas compte de la structure stochastique du terme d'erreur c'est-à-dire des effets fixes. L'estimateur Within lui, élimine les effets fixes en contrant les variables autour de leur moyenne. L'estimateur GLS tient compte de la structure de la covariance des erreurs, en supposant qu'elles sont corrélées au sein d'un même individu et au cours du temps. Chaque estimateur présente donc des estimations différentes.

Estimateur OLS sur le Modèle :

Tous les coefficients sont significatifs. Puisque notre variable dépendante est sous forme logarithmique, c'est donc normal d'avoir des coefficients qui ne dépasse pas 1. Le modèle prédit **qu'une augmentation de la part de marché** d'une compagnie aérienne sur une route donnée fait **augmenter le tarif moyen d'environ 36%** [$\exp(0.310)-1 \approx 0.36$] toute chose égale par ailleurs. Le coefficient associé au log-passagers indique qu'une augmentation que **10% du nombre de passager** sur un trajet donné entraine **une baisse des prix des billets de 0.79%**. Ce qui n'est pas surprenant, sur les vols où il y a beaucoup de passagers, les prix ont tendance à diminuer. La relation non linéaire entre les tarifs et la distance qui est invariante dans le temps et spécifique pour chaque route identifier, nous montre que le **tarif moyen** d'un allé simple a tendance **à diminuer** avec les **courtes distances** mais pour les **longues distances**, les tarifs ont tendance **à augmenter**. Les coefficients associés aux variables indicatrices indiquent des différences significatives des tarifs par rapport à 1997. Les prix des billets

sont plus élevés en 98, 99 et 2000 par rapport à 97. En effets, **en 2000**, le tarif aérien moyen **a augmenté de 11%** [$\exp(0.106) - 1 \approx 0.11$]**de plus par rapport à 1997**. La constante (7.010) correspond au logarithme du tarif moyen en 1997, lorsque toutes les variables explicatives sont nulles.

Estimateur Within sur le Modèle:

La constante n'est pas estimée car la transformation Within (déviations par rapport à la moyenne) **élimine** les effets fixes individuels et donc **toute constante**. Et donc comme les variables indicatrice et la variable distance (*ldist*) sont invariante dans le temps elles disparaîtraient avec la transformation within. L'estimateur Within ne permet pas d'identifier les coefficients de ces variables-là. Toutefois, cette **constante** peut-être **calculer** en soustrayant la somme pondérée des moyennes des variables explicatives multipliées par leurs coefficients des moyennes de la variable dépendante (*lfare*). On trouve **une constante = 7.228** ce qui est supérieur à celui de l'OLS, prouvant qu'il y a effectivement des effets fixes.

Néanmoins, les coefficients des variables explicatives restant, sont significatifs, mais on observe des différences notables :

La concentration de marché (0.150) a un **impact réduit** par rapport au modèle OLS. L'augmentation de la part de marché une augmentation d'environ 16% [$\exp(0.150) - 1 \approx 0.16$] du tarif moyen or dans le modèle OLS c'était d'environ 36%. Les log-passagers (-0.370) ont **un impact beaucoup plus négatif**, ce qui peut refléter des variations intra-individuelles importantes.

Estimateur GLS sur le Modèle:

La transformation et l'estimateur GLS permettent d'identifier, contrairement à l'estimateur Within, les coefficients des variables constants dans le temps.

La **concentration de marché** (0.153) a **un impact légèrement supérieur** à celui du modèle Within, un impact positif et significatif **d'environ 16.5% sur le tarif moyen**. Les **log-passagers** (-0.321) et **log-distance** (-1.189) conservent des impacts similaires c'est-à-dire qu'ils ont tous les deux **un impact négatif sur les tarifs aériens** mais notons que l'impact de la distance sur les tarifs dépend de la distance en question. Comme dans le modèle OLS, le log-distance reste positif et donc on tombe sur la même conclusion ; les prix augmentent ou diminuent compte tenu de la distance à parcourir entre deux points. **Les variables indicatrices** (*y98*, *y99*, *y00*) **perdent leur significativité**, ce qui indique que les effets temporels sont moins importants par rapport aux effets individuels. Mais il semble que la variable *y00* ait un impact sur le log-tarif, avec un seuil de 10%, et donc l'effet temporelle semble ne pas être totalement absent.

Concernant la **variance de l'estimateur GLS**. La **variance individuelle** (0.229) est **beaucoup plus importante** que la **variance temporelle** (0.002), ce qui explique pourquoi **les effets individuels dominent**.

Les effets temporels (capturés par les variables indicatrices) semblent peu significatifs dans le modèle GLS, renforçant l'idée que les variations temporelles n'ont pas un rôle majeur.

En somme, Le modèle **OLS** ne prend pas en compte les effets fixes, ce qui peut biaiser les coefficients en présence d'hétérogénéité non observée. Le modèle **Within** corrige pour les effets fixes individuels, mais il élimine également la constante. Cela rend son interprétation plus précise mais limitée aux variations intra-individuelles. Le modèle **GLS** est une approche mixte qui tente de combiner les avantages des deux premiers. Cependant, il dépend de l'hypothèse que les effets aléatoires ne sont pas corrélés avec les variables explicatives, ce qui sera testé avec le test d'Hausman.

Le **R²** est légèrement plus élevé pour le modèle Within (0.435) que pour le modèle OLS (0.431), ce qui indique un meilleur ajustement aux données intra-individuelles. En effet, dans le modèle estimé par l'estimateur **Within**, **43.5% de la variabilité du log-tarif est expliqué par le modèle** alors qu'avec le modèle estimé par l'estimateur **OLS**, seul **43.1% de la variabilité de la variable dépendante est expliqué par le modèle**. Le modèle **GLS** qui

repose sur une pondération qui dépend des variances intra-groupe (individuel) et inter-groupe (temporelle), présente un R^2 plus faible (0.396), reflétant la pondération qui pénalise les variations temporelles.

4.2 Test d'existence des effets fixes

Le test d'existence des effets entre le modèle estimé par OLS et par Within, effectué est celui de Fisher.

- **Hypothèse nulle:** Les effets fixes (individuels ou temporels) ne sont pas significatifs, c'est-à-dire qu'il n'y a pas de différences systématiques entre les individus ni entre les périodes.
- **Hypothèse alternative :** Au moins un des effets fixes (individuels ou temporels) est significatif.

Le F -test (*cf annexe*) de Fisher donne $F = 64.18$, $p\text{-valeur} < 2,2 \cdot 10^{-16}$. La p -value est extrêmement petite, bien inférieure au seuil classique de 5%. Cela signifie que nous **rejetons fortement** l'hypothèse nulle. **Les effets fixes individuels et/ou temporels sont significatifs**, ce qui justifie l'utilisation d'un modèle avec transformation Within.

4.3 Test d'Hausman

Le test de Hausman permet de comparer le modèle **within** (effets fixes) avec le modèle **GLS** (effets aléatoires). Cela aide à déterminer quel cadre est le plus approprié pour nos données (effets fixes ou aléatoires).

- **H0 :** Le modèle d'effets aléatoires (GLS) est cohérent et approprié pour les données. **Les effets fixes ne sont pas corrélés avec les variables explicatives.**
- **H1 :** Le modèle d'effets aléatoires (GLS) est incohérent, et les effets fixes (within) sont préférables. **Les effets fixes sont corrélés avec les variables explicatives** (*hypothèse de Mundlak*)

La statistique **Chi² est égale à 195.29** (*cf annexe*) avec un $p\text{-valeur} < 2,2 \cdot 10^{-16}$. La p -valeur étant très faible, on rejette l'hypothèse nul et donc les **variables explicatives sont bien corrélées avec les effets fixes**. Le modèle d'effets fixes (**within**) est préféré au modèle d'effets aléatoires (**GLS**).

Cependant, en calculant (*cf annexes*) les effets fixes et les différentes corrélations pour examiner les corrélations entre les effets fixes individuels et temporels et nos variables explicatives afin de voir laquelle des variables est la plus corrélée avec les effets, on remarque que les **effets fixes individuels sont principalement corrélés** avec les variables **ldist**(0.506) et **ldistsq**(0.513), ce qui est cohérent avec l'idée que la distance entre les destinations peut refléter des caractéristiques spécifiques (infrastructures aéroportuaires, la localisation des villes, ou la concurrence locale sur chaque itinéraire...). Nos variables explicatives étant corrélées avec les effets individuels, on doit utiliser un estimateur autre que le Within, on utilise l'estimateur IV (variable instrumentale)

Les **effets fixes temporels** sont surtout capturés par l'indicatrice temporelle **y00**, confirmant que les variations temporelles globales **jouent un rôle limité**.

5. Conclusion

Cette étude a analysé les déterminants des tarifs aériens aux États-Unis entre 1997 et 2000. Les résultats obtenus par régression sur données de panel (OLS, effets fixes, GLS), montrent que **la concentration du marché exerce une influence positive significative sur les tarifs**. Les compagnies opérant sur des routes peu concurrentielles peuvent ainsi imposer des prix plus élevés. La distance, quant à elle, a un effet non linéaire sur les tarifs : les courts trajets sont plus coûteux en raison des coûts fixes, mais cet effet s'atténue pour les longues distances. De plus, les routes très fréquentées bénéficient d'économies d'échelle, ce qui se traduit par des tarifs plus bas. Les effets fixes individuels, capturant les caractéristiques propres à chaque route, expliquent une part importante de la variation des tarifs, soulignant l'hétérogénéité des marchés aériens. Bien que les variables temporelles aient été introduites, les résultats suggèrent que les effets individuels sont prédominants. Cependant, une limitation de l'étude réside dans la possible endogénéité des variables, comme la variable *distance*, qui pourraient être corrélées avec des facteurs non observés spécifiques à chaque route. Cette endogénéité n'a pas pu être traitée dans le cadre de cette analyse.

6. Références bibliographiques

- Bailey, E. E., & Panzar, J. C. (1981). The contestability of airline markets during the transition to deregulation. *Law and Contemporary Problems*, 44(1), 125-145.
- Borenstein, S. (1989). Hubs and high fares: Dominance and market power in the US airline industry. *The RAND Journal of Economics*, 20(3), 344-365.
- Dresner, M., & Tretheway, M. (1992). Modeling and testing the effect of market structure on price: The case of international air transport. *Journal of Transport Economics and Policy*, 26(2), 171-184.
- Morrison, S., & Winston, C. (1986). *The economic effects of airline deregulation*. The Brookings Institution.
- Wooldridge, J. M. (2010). *Économétrie : Une approche moderne*

7. Annexes

Tableau 1 : Statistiques descriptives des principales variables

Statistic	N	Mean	Median	St. Dev.	Min	Max
Log-tarif moyen aller simple (en dollars)	4,596	5.10	5.12	0.44	3.61	6.26
Part de marché (%)	4,596	0.61	0.60	0.20	0.16	1.00
Distance (en miles)	4,596	989.74	861	611.83	95	2,724

On utilisera 'log' pour réduire les écarts

1 mile \approx 1,609 kilomètres.

- Test d'existence des effet (*Fisher Test*)

F test for twoways effects

```
data: lfare ~ concen + lpassen + ldist + ldistsq + y98 + y99 + y00
F = 64.18, df1 = 1146, df2 = 3442, p-value < 2.2e-16
alternative hypothesis: significant effects
```

- Test d'indépendance (*Hausman*)

Hausman Test

```
data: lfare ~ concen + lpassen + ldist + ldistsq + y98 + y99 + y00
chisq = 195.29, df = 2, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

- Test de corrélation avec les effets fixes individuels

```
> print("Corrélations avec les effets fixes individuels")
[1] "Corrélations avec les effets fixes individuels :"
```

concen	lpassen	ldist
-0.2894698	0.4666698	0.5063631
ldistsq	y98	y99
0.5131606	0.0000000	0.0000000
y00		
0.0000000		

- Test de corrélation avec les effets fixes temporels

```
> print("Corrélations avec les effets fixes temporels :")
[1] "Corrélations avec les effets fixes temporels :"
```

	concen	lpassen
	-2.856904e-02	3.598416e-02
	ldist	ldistsq
	1.718574e-21	-2.539160e-21
	y98	y99
	-2.976153e-01	5.729871e-02
	y00	
	8.988107e-01	