

# Is Virtual Codesharing A Market Segmenting Mechanism Employed by Airlines?

Philip G. Gayle\*  
Kansas State University

August 30, 2006

## Abstract

It has been suggested that virtual codesharing is a mechanism used by airlines to segment passengers based on their price sensitivity. The objective of this paper is to test whether passengers' choice behavior is consistent with market segmentation being the primary motive for virtual codesharing. The findings fail to support the market segmentation motive.

*JEL Classification:* L13, L93, C1, C2

*Keywords:* Airline Codesharing, Discrete Choice, Bi-level Nested Logit

**Acknowledgement:** I thank Dong Li, Lance Bachmeier, Tian Xia, and two anonymous referees for their invaluable comments. I also thank Muharrem Onemli for excellent research assistance. All remaining errors are my own.

---

\*Department of Economics, 320 Waters Hall, Kansas State University, Manhattan, KS, 66506, (785) 532-4581, Fax:(785) 532-6919, email: gaylep@ksu.edu.

# 1 Introduction

A codeshare agreement allows an airline ("ticketing carrier") to market and sell seats on their partners' plane ("operating carrier") as if these seats were owned by the ticketing carrier. A common structure that these agreements follow is where the ticketing carrier is responsible for setting the price for the entire round trip, and compensating its partner for the segments of the trip for which they provided operating services. The operating carrier would also pay the ticketing carrier a commission for marketing services provided. For the majority of domestic codeshare itineraries, passengers remain on a single carriers' network for the entire trip even though the ticket for the trip may have been marketed and sold by a codeshare partner. Ito and Lee (2007) label such codeshare itineraries, "virtual codesharing."<sup>1</sup>

Why would a carrier sell some of its operating services in a market directly (pure online product), but contract his partner carriers to market and sell some of its operating services in the said market (virtual codeshare product)?<sup>2</sup> Ito and Lee (2007) argue that passengers that are members of an airline's frequent flyer program may view the airline's virtual codeshare product as an inferior substitute to its pure online product since virtual tickets often do not allow the frequent flyer to upgrade to first class even though the flights on the two itineraries (pure online and virtual) are the same. Further, by offering a branded (pure online) and a lower priced non-branded (virtual) product in the same market, a carrier is able to separate customers based on their price sensitivity. Ito and Lee (2007) liken an airline offering pure online and virtual codeshare products to a pharmaceutical firm offering branded and generic drugs.

The market segmentation motive for virtual codesharing at first glance seems a very convincing argument. However, if market segmentation is the primary motive for virtual codesharing, why is similar segmentation not achievable via varying levels of ticket restrictions for pure online products? This question suggests that at a minimum we need to test if passengers' choice behavior is consistent with market segmentation being the primary motive for virtual codesharing, which is the main objective of this paper.

---

<sup>1</sup> According to Ito and Lee (2007), approximately 85% of U.S. domestic codesharing itineraries are "virtual."

<sup>2</sup> More formal definitions of pure online and virtual codeshare products are given in section 2 of the paper.

## 2 Definitions

Following Gayle (2005), a market is defined as directional round-trip air travel between an origin and destination city. The assumption that markets are directional implies that a round-trip air travel from Denver to Washington, DC is a distinct market than round-trip air travel from Washington DC to Denver. Further, this directional assumption allows for the possibility that origin city characteristics may influence market demand.

A flight itinerary is defined as a specific sequence of airport stops in traveling from the origin to destination city. Products are defined as a unique combination of airline(s) and flight itinerary. The products explicitly included in the model are pure online and virtual codeshare products. A pure online product means that the same airline markets and operates all segments of a round trip.<sup>3</sup> However, for virtual codeshare products, the marketing and operating airlines differ.<sup>4</sup>

## 3 The Model

Air travel demand is modelled in a discrete choice framework. Specifically, I use a demand specification similar to Goldberg and Verboven (2001) which is based on McFadden's (1978) generalized extreme value (GEV) model. Potential passenger  $i$  in market  $t$  faces a choice between  $J_t + 1$  alternatives. There are  $J_t + 1$  alternatives because I allow passengers the option ( $j = 0$ ) not to choose one of the  $J_t$  differentiated air travel products considered in the empirical model. A passenger chooses the product that gives them the highest utility, that is

$$\underset{j \in \{0, \dots, J_t\}}{\text{Max}} \{U_{ijt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt} + \eta_{ijt}\}, \quad (1)$$

where  $U_{ijt}$  is the value of product  $j$  to passenger  $i$ ,  $x_{jt}$  is a vector of observed product characteristics (whether or not the origin is a hub for the carrier, the number of intermediate stops, the carrier's flight frequency out of the origin airport, etc.),  $p_{jt}$  is the price,  $\xi_{jt}$  is the level of unobserved product quality, and  $\eta_{ijt}$  is a mean zero random component of utility.

---

<sup>3</sup>For example, three separate pure online products are, (1) a non-stop round trip from Denver to Washington, DC marketed and operated by Delta Air Lines, (2) a round-trip from Denver to Washington, DC with one stop in Atlanta marketed and operated by Delta Air Lines, and (3) a non-stop round-trip from Denver to Washington, DC marketed and operated by United Air Lines. Note that all three products are in the same market.

<sup>4</sup>Three examples of virtual codeshare products are, (1) a non-stop round-trip from Denver to Washington, DC marketed by US Airways but operated by United Air Lines, (2) a non-stop round-trip from Denver to Washington, DC marketed by United Air Lines but operated by US Airways, and (3) a round-trip from Denver to Washington, DC with one stop in Atlanta marketed by Continental Air Lines but operated by Delta Air Lines.

Based on McFadden's (1978) generalized extreme value (GEV) model, where  $\eta_{ijt}$  has a multivariate extreme value distribution, the probability that passenger  $i$  chooses product  $j$  is given by

$$s_j = \frac{e^{\delta_j} G_j(e^{\delta_j}, \dots, e^{\delta_J})}{G(e^{\delta_j}, \dots, e^{\delta_J})}, \quad (2)$$

where  $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$ ,<sup>5</sup>  $G_j(e^{\delta_j}, \dots, e^{\delta_J})$  is the derivative of  $G(e^{\delta_j}, \dots, e^{\delta_J})$  with respect to  $e^{\delta_j}$ .  $s_j$  can also be interpreted as the market share of product  $j$ . The market index,  $t$ , is suppressed throughout only to avoid notational clutter. A few extra notational definitions will allow me to specify the functional form that I use for  $G(\cdot)$ .

Let  $g(r, k)$  be subgroup  $r$  of airline  $k$ 's products, where  $r = 1, 2$ .  $r = 1$  when airline  $k$  is the sole operating carrier but not the marketing carrier of the product, while  $r = 2$  when airline  $k$  is both the operating and marketing carrier of the product. In other words, subgroup  $g(1, k)$  contains all virtual codeshare products for which airline  $k$  is the sole operating carrier, while subgroup  $g(2, k)$  contains all of airline  $k$ 's pure online products. Similar to the bi-level nested logit model in Goldberg and Verboven (2001), I assume that  $G(\cdot)$  takes the following functional form,<sup>6</sup>

$$G(\cdot) = 1 + \sum_{k \in K} \left\{ \left[ \left( \sum_{j \in g(1, k)} e^{\frac{\delta_j}{\rho_b}} \right)^{\frac{\rho_b}{\rho_a}} + \left( \sum_{j \in g(2, k)} e^{\frac{\delta_j}{\rho_b}} \right)^{\frac{\rho_b}{\rho_a}} \right]^{\rho_a} \right\}. \quad (3)$$

Thus,  $k$  indexes both airlines and aggregate product groups, while  $g(r, k)$  indexes product subgroups.

Intuitively, products are first nested by operating carriers and then sub-nested by product type (virtual versus pure online).<sup>7</sup> Such a nesting structure is particularly convenient for explicit testing of whether passengers' choice behavior is consistent with market segmentation being the primary motive for virtual codesharing. For instance, the values of  $\rho_a$  and  $\rho_b$  tell us the extent to which passengers perceive the products as differentiated. In other words, these parameters capture the pattern of substitutability across products. First, to be consistent with random utility maximization, we must have  $0 < \rho_b \leq \rho_a \leq 1$ . Second, the closer either  $\rho$  is to 0, the greater the substitutability between products within groups and subgroups. In other words, if both  $\rho$  are close to 0 while still maintaining  $0 < \rho_b < \rho_a$ , then passengers view airline  $k$ 's products as closer substitutes for each other compared to the substitutability of these products across airlines.

<sup>5</sup>  $\delta_{jt}$  is the mean utility derived from product  $j$ .

<sup>6</sup> For an alternative specification of  $G(\cdot)$  see Bresnahan et al. (1997).

<sup>7</sup> I thank an anonymous referee for suggesting this concise summary description of the nesting structure.

Further, since  $\rho_b < \rho_a$ , passengers view airline  $k'$ 's virtual codeshare products as closer substitutes compared to the substitutability of these products with airline  $k'$ 's pure online products. Conversely, if  $\rho_b = \rho_a$ , then passengers do not perceive a distinction between airline  $k'$ 's virtual codeshare and pure online products even though these products are imperfect substitutes for other airlines' products. Thus, if market segmentation is the primary motive for virtual codesharing, we should have  $\rho_b < \rho_a$ . Last, in the limiting case where,  $\rho_b = \rho_a = 1$ , consumer do not perceive products as differentiated based on either airlines or virtual codeshare versus pure online.

## 4 Estimation

The parameters to be estimated are  $\beta$ ,  $\alpha$ ,  $\rho_a$ , and  $\rho_b$ . Following Berry(1994), the estimation strategy involves choosing parameter values such that *observed* product shares,  $S_j$ , are equal to *predicted* product shares,  $s_j$ , that is,

$$S_j = s_j(\delta, \rho_a, \rho_b). \quad (4)$$

Observed product shares are computed by  $S_j = \frac{q_j}{M}$ , where  $M$  is the size of the population in the origin city and  $q_j$  is the actual number of travel tickets sold for a particular itinerary-airline(s) combination called product  $j$ .

**Lemma 1:** The functional form of the right hand side of equation (4), which is based on equations (2) and (3), implies that:  $\delta_j = \ln(S_j) - \ln(S_0) - (1 - \rho_b) \ln(S_{j|g(r,k)}) - (1 - \rho_a) \ln(S_{g(r,k)|k})$ , where  $S_0$  is the observed share of the outside option,  $S_{j|g(r,k)}$  is the observed within subgroup share of product  $j$ , and  $S_{g(r,k)|k}$  is subgroup  $g(r, k)$  observed share in group  $k$ .

*Proof:* In an appendix available upon request.

Based on lemma 1, the equation to be estimated is

$$\ln(S_j) - \ln(S_0) = x_j\beta + \alpha p_j + (1 - \rho_b) \ln(S_{j|g(r,k)}) + (1 - \rho_a) \ln(S_{g(r,k)|k}) + \xi_j. \quad (5)$$

Since  $p_j$ ,  $S_{j|g(r,k)}$ , and  $S_{g(r,k)|k}$  are all endogenous, an instrumental variable estimator has to be used to achieve consistent estimates. Fortunately, the simple linear structure of equation (5) allows me to use two-stage least squares (2SLS).

## 5 Data

The data set is drawn from the Origin and Destination Survey (DB1B), which is a 10% sample of airline tickets from reporting carriers. The U.S. Bureau of Transportation Statistics publishes this database along with other transportation data via its TranStats web site.<sup>8</sup> For this research, I focus on the U.S. domestic market in the first quarter of 2005. The sample contains 1,914 products<sup>9</sup> spread across 56 markets.<sup>10</sup> Table 1 lists carriers in the data set. A list of the markets is available upon request.

**Table 1**  
**List of Operating Carriers in the Data Set**

Airline Code	Airline Name
AA	American Airlines
AS	Alaska Airlines
CO	Continental Air Lines Inc.
DL	Delta Air Lines Inc.
NW	Northwest Airlines Inc.
UA	United Air Lines Inc.
US	US Airways Inc.

Observed product characteristics are, "Price", "Hub\_opcarrier", "Hub\_tkcarrier", "Departures", "Departures  $\times$  Interstop", "Interstop", "Virtual", while airline, and market dummies are used to control for unobserved product and market fixed effects. "Price" is the mean fare of a given itinerary-airline(s) combination, "Hub\_opcarrier" is a dummy that takes the value 1 if the origin airport is a hub for the operating carrier and 0 otherwise, "Hub\_tkcarrier" is a dummy that takes the value 1 if the origin airport is a hub for the ticketing carrier and 0 otherwise, "Departures" is the average number of scheduled daily departures of the operating carrier from the origin airport throughout the previous year, "Interstop" is the number of intermediate stops on an itinerary, and "Virtual" is a dummy taking the value 1 if the product is virtually codeshared, and zero otherwise. These variables are right hand side regressors in equation (5).

---

<sup>8</sup>For detail on air travel data published by U.S. Bureau of Transportation go to <http://transtats.bts.gov/>

<sup>9</sup>Since products are defined as a unique combination of operating carrier and specific route, the data are aggregated to the airline-route level. For example, product shares are based on total number of passengers traveling a particular route on an airline while, "Price" is the average fare paid by these passengers.

<sup>10</sup>A detailed description of the sample is available upon request.

## 6 Results

The estimated coefficients are presented in table 2. Both ordinary least squares (OLS) and two-stage least squares (2SLS) estimates are reported for the purpose of illustrating the importance of instrumenting for endogenous variables. As expected, the size of estimated coefficients for the endogenous variables (Price,  $\ln(S_{j|g(r,k)})$ , and  $\ln(S_{g(r,k)|k})$ ) depend crucially on whether instruments are used for these variables.<sup>11</sup> The most striking example being the counter intuitive positive OLS coefficient on price. A Hausman exogeneity test displayed in the table also provides statistical support for the need to instrument for the endogenous variables.<sup>12</sup> As such, from this point on I focus on the 2SLS estimates.

---

<sup>11</sup>Instruments include, a product's itinerary distance, the squared deviation of a product's itinerary distance from the average itinerary distance of competing products offered by other airlines, mean itinerary distance in a subgroup, and mean itinerary distance in a group. Choice of instruments are motivated by, (1) supply theory which predicts that a product's price and market share are affected by the number and closeness of competing products in the market and, (2) the assumption that the marginal cost of servicing an itinerary (i.e. the airline's marginal cost of transporting a passenger) is a function of the itinerary distance. The intuition behind (2) is that an airline's marginal cost of transporting a passenger between a given origin and destination city is likely to be lower if the passenger chooses a nonstop flight compared to the case where the passenger chooses to use (for whatever reason) a circuitous route with multiple intermediate stops. As such, since the distance flown on the chosen itinerary is directly related to the route used in getting the passenger from the origin to destination city, I posit that itinerary distance is correlated with the marginal cost to the airline of transporting the passenger.

<sup>12</sup>In the first-stage regressions (not reported) where "Price",  $\ln(S_{j|g(r,k)})$  and  $\ln(S_{g(r,k)|k})$  are regressed on instruments, the  $R^2$  for these regressions are 0.35, 0.77 and 0.87 respectively. All these  $R^2$  measures are sufficiently high to reassure us that the instruments do explain variations in the endogenous variables.

**Table 2**  
**Model Estimates**

Variables	Ordinary Least Squares Estimates	Two-Stage Least Squares Estimates (1)	Two-Stage Least Squares Estimates (2)	Two-Stage Least Squares Estimates (3)
Price	0.0002 (0.0001)	-0.0017** (0.0008)	-0.0024** (0.0008)	-0.0017** (0.00077)
HUB_opcarrier	-0.3096** (0.125)	-0.1368 (0.1903)	0.230 (0.174)	-
HUB_tkcarrier	0.167* (0.098)	0.0864 (0.1480)	-0.048 (0.16)	-
Departures	0.0044** (0.0005)	0.0034** (0.0007)	-	0.001** (0.004)
Departures × Interstop	-0.0005** (0.0002)	-0.0013** (0.0003)	-	-
Interstop	-0.2203** (0.037)	-0.9150** (0.0900)	-1.112** (0.095)	-1.04** (0.091)
Virtual	2.467** (0.714)	-1.4055 (1.2424)	-1.838 (1.20)	-0.031 (1.206)
$\ln(S_{j g(r,k)})$	0.801** (0.011)	0.2755** (0.0641)	0.212** (0.063)	0.285** (0.063)
$\ln(S_{g(r,k) k})$	0.667** (0.028)	0.4233** (0.0944)	0.373** (0.101)	0.417** (0.095)
R <sup>2</sup>	0.90			
Hausman Exogeneity Test		Test Statistic =116.17 $\chi^2(3, 0.95) = 7.82$		
Wald test: $H_0 : 1 - \rho_a = 1 - \rho_b$		F(1, 1818)=2.17 5% critical F value=3.84		

Notes: The dependent variable is  $\ln(S_j) - \ln(S_0)$ . Standard errors are in parentheses.

\*\* indicates statistical significance at the 5% level, while \* indicates statistical significance at the 10% level. The regressions are estimated with a full set of airline and market dummies even though these coefficient estimates are not reported.

I report three separate model specifications of the 2SLS estimates primarily to illustrate robustness of the main qualitative results. However, the following discussion focusses on the first column of the 2SLS estimates. As expected, the coefficient on price is negative and statistically significant suggesting that passengers are less likely to choose a flight itinerary the higher is its price. Surprisingly, the origin airport being a hub for the operating or ticketing carrier seem not to have a statistically significant effect on passengers' choice of products. Airlines' strategic advantage at their hub airports is derived from their large presence at these airports [see Berry (1990)]. The statistical insignificance of the hub dummies may be due to them being a noisy measure of an airline's airport presence.<sup>13</sup> The strategic advantage that airport presence may confer on an airline is supported by the positive and statistically significant coefficient for the "Departures" variable. However, given that the coefficient on the interaction between "Departures" and "Interstop" is negative and statistically significant, this suggests that an airline's large airport presence is less effective the larger the number of intermediate stops contained in its itineraries. Finally, the neg-

<sup>13</sup> Alternatively, the poor performance of the hub variables could be due to potential correlation with the "Departure" variable. This motivated the model specification in the second column of the 2SLS estimates (I thank an anonymous referee for suggesting that I try this.). The qualitative results are robust to deletion of "Departures" from the model. The last column also reveals that qualitative results are robust to deletion of the "HUB" variables.



ative and statistically significant coefficient on "Interstop" suggests that an itinerary is less likely to be chosen the larger the number of intermediate stops it has.

As illustrated in equation (5), the coefficients on  $\ln(S_{j|g(r,k)})$  and  $\ln(S_{g(r,k)|k})$  are  $1 - \rho_b$  and  $1 - \rho_a$  respectively. First, based on the size of the coefficients relative to their respective standard errors, both coefficients are statistically different from zero and one. Therefore, consumers do perceive products as differentiated. However, the results from a Wald test shown in table 2, suggest that we cannot reject  $1 - \rho_b = 1 - \rho_a$  at conventional levels of significance. In other words, the results suggest that passengers do not perceive a distinction between an airline's virtual codeshare and pure online products, but they do perceive products as imperfect substitutes across airlines. The values of the substitution parameters therefore suggest that if an airline marginally increases the price of one of its virtual codeshare products, its passengers are equally likely to substitute towards one of its pure online as they would to one of its other virtual codeshare products in the said market. The statistical insignificance of the coefficient on "Virtual" is also consistent with the result that passengers do not perceive a distinction between pure online and virtual codeshare products. Together, these findings cast doubt on whether market segmentation is the primary motive for virtual codesharing.

## 7 Conclusion

The main objective of this paper is to test whether passengers' cross substitution pattern is consistent with market segmentation being the primary motive for virtual codesharing. The test is performed within a discrete choice econometric framework which allows me to isolate and study passengers' substitution patterns across any given airline's products. The results suggest that while passengers do perceive products as differentiated across airlines, they do not perceive an airline's pure online products as distinct from its virtual codeshare products. As such, if an airline marginally increases the price of one of its virtual codeshare products, its passengers are equally likely to substitute towards one of its pure online products as they would to one of its other virtual codeshare products in the said market. The findings therefore cast doubt on whether market segmentation is the primary motive for virtual codesharing.

Notwithstanding the findings above, it is possible that an airline's most loyal frequent flyers perceive a difference between virtual codeshare and pure online products, but this group of pas-

sengers is too small a fraction of the market to show up in the results.<sup>14</sup> As such, a limitation of this study is that there is no way of identifying passengers that belong to an airline's frequent flyer program, which is necessary for testing the market segmentation motive for this specific group of passengers. If detail passenger-specific data becomes available, this may be a fruitful issue for future research to explore.

Last, when codeshare partners allow their frequent flyer members to earn and/or redeem frequent flyer points on partner carriers, this increases the value of each partners' frequent flyer program to its customers.<sup>15</sup> As such, the formation of a partnership may serve to boost each partner's demand especially in instances where the partners' networks are non-overlapping [see Lederman (2003)].<sup>16</sup> However, the partners' decision to allow frequent flyer members to earn and redeem frequent flyer points on partner carriers effectively reduce each partners' market power over their loyal frequent flyers in markets which the partners compete [see Lederman (2003)]. A deeper understanding of this trade-off between increase demand and lower market power that a codeshare partnership may bring for each partner may well be an important piece of the puzzle in understanding the motivation for virtual codesharing.

---

<sup>14</sup>I thank an anonymous referee for pointing this out.

<sup>15</sup>I thank an anonymous referee for suggesting that this may be an alternative motivation for virtual codesharing.

<sup>16</sup>The airline's boost in demand may also stem from its ability to tap into its partner's loyal customer base which might not have been possible without the partnership [see Lederman (2003)].

## References

- [1] Berry, S. (1994), “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, Vol. 25, 242-262.
- [2] Berry, S., (1990). “Airport Presence as Product Differentiation,” *The American Economic Review*, Vol. 80, 394-399.
- [3] Bresnahan, T., S. Stern, and M. Trajtenberg, (1997), “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s,” *Rand Journal of Economics*, Vol. 28, 17-44.
- [4] Gayle, P. (2005), “Airline Code-share Alliances and their Competitive Effects,” *Manuscript, Kansas State University*.
- [5] Goldberg, P., and F. Verboven (2001), “The Evolution of Price Dispersion in the European Car Market,” *The Review of Economic Studies*, Vol. 68, No. 4, 811-848.
- [6] Ito, H., and D. Lee, (2007), “Domestic Codesharing, Alliances and Airfares in the U.S. Airline Industry,” forthcoming in *The Journal of Law and Economics*.
- [7] Lederman, Mara (2003), “Partnering with the Competition? Understanding Frequent Flyer Partnership between Competing Domestic Airlines,” *Manuscript, Massachusetts Institute of Technology*.
- [8] McFadden, D. (1978), “Modelling the Choice of Residential Location,” in A. Karlqvist et al. (eds.), *Spatial Interaction Theory and Planning Models* (Amsterdam: North-Holland).