Transportation Research Part A 100 (2017) 228–246

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Accounting for price endogeneity in airline itinerary choice models: An application to Continental U.S. markets

Virginie Lurkin [a,1], Laurie A. Garrow [b,*], Matthew J. Higgins [c,d], Jeffrey P. Newman [b,2], Michael Schyns [a]

[a] University of Liege, HEC-Management School, 4000 Liege, Belgium
[b] Georgia Institute of Technology, School of Civil and Environmental Engineering, 790 Atlantic Drive, Atlanta, GA 30332-0355, United States
[c] Georgia Institute of Technology, Ernest Scheller Jr. College of Business, 800 West Peachtree NW, Atlanta, GA 30308, United States
[d] National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 25 January 2016
Accepted 3 April 2017

Keywords:
Air travel behavior
Itinerary choice
Price elasticity
Price endogeneity
Time of day preference

ABSTRACT

Network planning models, which forecast the profitability of airline schedules, support many critical decisions, including equipment purchase decisions. Network planning models include an itinerary choice model that is used to allocate air total demand in a city pair to different itineraries. Multinomial logit (MNL) models are commonly used in practice and capture how individuals make trade-offs among different itinerary attributes; however, none that we are aware of account for price endogeneity. This study formulates an itinerary choice model that is consistent with those used by industry and corrects for price endogeneity using a control function that uses several types of instrumental variables. We estimate our model using a database of more than 10 million passenger trips provided by the Airlines Reporting Corporation. Results based on Continental U.S. markets for May 2013 departures show that models that fail to account for price endogeneity overestimate customers' value of time and result in biased price estimates and incorrect pricing recommendations. The size and comprehensiveness of our database allows us to estimate highly refined departure time of day preference curves that account for distance, direction of travel, number of time zones traversed, departure day of week and itinerary type (outbound, inbound or one-way). These time of day preference curves can be used by airlines, researchers, and government organizations in the evaluation of different policies such as congestion pricing.

1. Introduction and motivation

Network planning models, which are used to forecast the profitability of airline schedules, support many important long- and intermediate-term decisions. For example, they aid airlines in performing merger and acquisition scenarios, route schedule analysis, code-share scenarios, minimum connection time studies, price-elasticity studies, hub location and hub buildup studies, and equipment purchasing decisions (Garrow et al., 2010).

* Corresponding author.
E-mail addresses: virginie.lurkin@epfl.ch (V. Lurkin), laurie.garrow@ce.gatech.edu (L.A. Garrow), matt.higgins@scheller.gatech.edu (M.J. Higgins), jnewman@camsys.com (J.P. Newman), M.Schyns@ulg.ac.be (M. Schyns).
[1] Current address: École Polytechnique Fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland.
[2] Current address: Cambridge Systematics, Inc., 115 South LaSalle Street, Suite 200, Chicago, IL 60603, United States.

http://dx.doi.org/10.1016/j.tra.2017.04.007
0965-8564/

Network planning models forecast schedule profitability by determining the number of passengers who travel in an origin destination (OD) pair, allocating these passengers to specific itineraries, and calculating expected costs and revenues. The passenger allocation model is often referred to as an itinerary choice model because it represents how individuals make choices among itineraries. Many airlines use discrete choice models to capture how individuals make trade-offs among different itinerary characteristics, *e.g.*, departure times, elapsed times, the number of connections, equipment types, carriers, and prices (see (Garrow et al., 2010; Jacobs et al., 2012) for reviews of itinerary choice models used in practice and (Coldren et al., 2003; Koppelman et al., 2008) for specific studies conducted for United Airlines and Boeing, respectively).

However, to the best of our knowledge, none of the itinerary choice models used in practice account for price endogeneity. Price endogeneity occurs when prices are influenced by demand, *i.e.*, higher prices are observed when demand is high and lower prices are observed when demand is low. Failure to correct for price endogeneity is critical, as it will result in biased estimates and incorrect profitability calculations. Recent work has focused attention on the importance of accounting for endogeneity in demand studies. For example Guevara (2015) notes that "endogeneity often arises in discrete-choice models, precluding the consistent estimation of the model parameters, but is habitually neglected in practical applications." Guevara (2015) provides several examples from the mode choice, residential location, and intercity travel demand literatures that provide evidence of endogeneity due to omission of attributes and reviews approaches researchers have been using to account for this endogeneity. These studies include those by Wardman and Whelan (2011) and Tirachini et al. (2013) for mode choice applications; Guevara and Ben-Akiva (2006, 2012) for residential location applications; and, Mumbower et al. (2014) for intercity applications.

Our prior work in air travel demand modeling has found strong evidence of price endogeneity. In Mumbower et al. (2014) we model flight-level price elasticities in four markets using linear regression models and find striking differences in price elasticity estimates between a model that ignores and a model that accounts for price endogeneity. The model that ignores price endogeneity produces inelastic results (−0.58) whereas the model that accounts for price endogeneity using a two-stage least squares (2SLS) approach produces elastic (−1.32) results. In Hotle et al. (2015) we investigate the impact of airlines' advance purchase deadlines on individuals' online search and purchase behaviors for 60 markets. Our model, which is also based on a 2SLS method, finds strong evidence of price endogeneity.

This paper builds on prior research by showing how to correct for price endogeneity for an itinerary choice model that is consistent with those used by industry. Unlike our previous applications, our model incudes "all" Continental U.S. markets and is based on discrete choice versus linear regression methods. Specifically, we follow the approach of Coldren et al. (2003) described for United Airlines and use a multinomial logit (MNL) to model itinerary choice for Continental U.S. markets. Results demonstrate the importance of accounting for price endogeneity; failure to do so results in value of time estimates that are too high, biased price estimates, and incorrect pricing recommendations. The results are intuitive, and validation tests indicate that the corrected model outperforms the uncorrected specification.

Our study is distinct from the majority of prior studies reported in the literature in that we use a large database of individual tickets from multiple carriers for our analysis. Specifically, we estimate our model using an analysis database of 10 million passenger trips provided by the Airlines Reporting Corporation (ARC). We are uniquely positioned to examine the potential of using the ARC ticketing database for itinerary choice modeling applications as we are able to work with detailed price data whereas airlines cannot due to anti-trust regulations. Our paper contributes to the literature in three key ways. First, we demonstrate the ability to use the ARC ticketing database (in spite of its limitations) to replicate itinerary choice models representative of those used in practice. Second, we find a valid set of instruments to correct for price endogeneity for Continental U.S. markets. Third, due to the size of our analysis database, we are able to estimate detailed departure time of day preference curves that are segmented by distance, direction of travel, number of time zones travelled, day of week, and itinerary type (outbound, inbound or one-way). To the best of our knowledge, these curves represent the most refined publicly-available estimates of airline passengers' time of day preferences.

The remaining sections are organized as follows. Section 2 describes the data processing assumption we used to create our analysis database and the variables used in our study. Section 3 presents our methodology, with a particular focus on how we addressed price endogeneity. Empirical results are presented in Section 4. We conclude by highlighting how our model contributes to the literature and offering directions for future research, many of which are based on the data limitations commonly faced by industry when estimating discrete choice models for itinerary choice applications.

2. Data

This section describes the data and variables we used, explains the process we used to generate choice sets, and assesses the representativeness of our analysis database.

2.1. Airlines Reporting Corporation ticketing database

The Airlines Reporting Corporation (ARC) is a ticketing clearinghouse that maintains financial transactions for all tickets purchased through travel agencies worldwide. This includes both online (*e.g.*, Expedia) and brick-and-mortar agencies. Some carriers, most notably Southwest, are under-represented in the database because the majority of their ticket sales are through direct sales channels (*e.g.*, southwest.com) that are not reported to ARC.