Journal of Choice Modelling 54 (2025) 100537

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Control Function Approach for Addressing Endogeneity in Transport Models: A Case Study on the London–Amsterdam Route

Thomas E. Guerrero B. [a], Nicolò Avogadro [b],*, Raúl Ramos [c]

[a] Department of Civil Engineering, Universidad Francisco de Paula Santander Seccional Ocaña, Sede El Algodonal Ocaña, Colombia
[b] Department of Management, Information and Production Engineering, University of Bergamo, Italy
[c] Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Santiago, Chile

A R T I C L E I N F O

Keywords:
Endogeneity
Discrete choice model
Control Function
High-speed rail
Air transport

A B S T R A C T

Endogeneity is a key empirical challenge in transportation modeling, which may lead to inconsistent estimates and biased policy decisions. This paper investigates the sources of endogeneity and focuses on tackling this issue for a discrete choice model analyzing the multimodal London–Amsterdam route, where air transport and high-speed rail (HSR) compete. Contrary to previous literature, we found no evidence of endogeneity in service frequency for the London–Amsterdam market. This could be attributed to market-specific features, such as feeding considerations, slot retention dynamics, and the congestion of the HSR network, which constrains capacity expansion opportunities. Conversely, we observed that fare introduced endogeneity into the model. To address this issue, we applied the control function approach and proposed two novel instruments: the fare for similar markets and the price of power sources. These instruments proved to be effective in correcting for endogeneity by increasing model performance. We also discuss the adverse impact of neglecting endogeneity and estimate price and frequency elasticities, ultimately demonstrating the significance of dealing with endogeneity in ensuring the reliability of results in transportation studies and appropriately informing policy decisions.

1. Introduction

High-speed rail (HSR) has become an affordable and sustainable alternative to air travel for domestic and cross-border trips worldwide (for recent reviews on the topic, see, e.g., Givoni and Dobruszkes, 2013; Li et al., 2019a; Zhang et al., 2019b). On one hand, this pattern was triggered by the spread of high-speed networks on a global scale. The length of high-speed networks in operation worldwide has increased rapidly, reaching an extension of approximately 58,000 km in 2021, and it is expected to increase by 20,000 km in the near future (UIC, 2022). On the other hand, the undeniable environmental impacts of air transport —which accounts for approximately 2.5% of global CO_2 emissions, increasing to 3.5% if non-CO_2 impacts are considered (Lee et al., 2021)— have provided the impetus for policymakers and industry stakeholders to devise mechanisms to reduce emissions. Among them, the fostering of a modal shift toward alternative transport modes, especially HSR, emerged as a promising solution for short and medium-haul routes characterized by high demand intensity, primarily due to the *greenness* of HSR compared to air transport on a per-seat basis (Avogadro and Redondi, 2023; Jiang et al., 2021).

Following this trend, researchers have increasingly investigated the drivers underpinning transport demand in markets where HSR competes with air carriers. The standard approach to these analyses is the use of linear econometric models or the adoption of

* Corresponding author.
E-mail addresses: teguerrerob@ufpso.edu.co (T.E. Guerrero B.), nicolo.avogadro@unibg.it (N. Avogadro), rramos4@uc.cl (R. Ramos).

https://doi.org/10.1016/j.jocm.2024.100537
Received 2 March 2024; Received in revised form 2 December 2024; Accepted 22 December 2024

gravity-based formulations, where the dependent variable is the aggregate demand or the market share of a single alternative (see, e.g., Gu and Wan, 2022; Li et al., 2019a,b; Mizutani and Sakai, 2021; Strauss et al., 2021). However, at the individual level, transport demand is not continuous but discrete. This is because single passengers choose the best alternative for their trip. Thus, empirical studies that use individual data rely extensively on discrete choice models (DCMs) to investigate passenger preferences.

Table 1 summarizes previous studies jointly investigating HSR and air transport demand using individual passenger choices and compares them with this paper. Most studies have only leveraged stated preferences (SP) surveys or combined them with revealed preferences (RP) data. This is because the former allows the modeler to study hypothetical scenarios and countervail possible correlations and lack of variability among alternatives' attributes. However, this approach suffers from potential response bias, since real and stated choices may diverge significantly (Cameron et al., 2011; Haghani et al., 2021; Jiang et al., 2022). Not surprisingly, from a methodological perspective, the majority of previous works estimated multinomial (MNL) or nested logit (NL) models, which are the most common DCMs applied in the transport literature, since they have closed-form expressions for the choice probability, making them computationally easy to estimate (Arteaga et al., 2022; Ortúzar and Willumsen, 2011; Molloy et al., 2021; Train, 2009). Regarding modal attributes, the fare, service frequency, travel time, and access and egress times are among the most investigated explanatory variables. However, other variables, such as comfort, reliability, and individual socioeconomic characteristics, have also been demonstrated to affect passenger choices.

Notably, previous studies leveraging DCMs have systematically disregarded possible correlations between the error term and the explanatory variables, which may lead to biased and inconsistent estimates of the model parameters (Danaf et al., 2020; Guerrero et al., 2022; Guevara, 2015; Sarrias, 2021; Watanabe and Maruyama, 2023). This so-called endogeneity problem has been extensively investigated in the case of linear models. For example, it is well known that the fare may be endogenous because higher consumption allows transport providers with market power to charge higher prices (Birolini et al., 2020; Pagoni and Psaraki-Kalouptsidi, 2016; Zhang et al., 2017). In this case, demand, which is the dependent variable, affects the price, leading to a problem known as simultaneous estimation, which causes endogeneity (Lurkin et al., 2017; Perera and Tan, 2019; Vigren, 2017; Zhang et al., 2019a). Moreover, demand may also affect frequency, so even this variable could introduce endogeneity problems (Birolini et al., 2020; Yang et al., 2020; Zhang et al., 2019a). Another source of endogeneity is the correlation of ticket price with the unobservable service quality characteristics or passengers' subjective attitudes, which can affect demand, an effect known as omitted variable bias (Berry and Jia, 2010; Fernández-Antolín et al., 2016; Fu et al., 2015; Gama, 2017; Hsiao and Hansen, 2011; Li et al., 2020; Palma et al., 2016).

When endogeneity is neglected, it can result in inconsistent estimates of model parameters and elasticities. More importantly, these biased estimates may lead to suboptimal (if not inadequate) policy decisions when DCM outcomes are used in the planning and social evaluation of transportation projects. Therefore, devising methods to tackle this issue in DCMs and accurately quantify the effects stemming from endogeneity is a crucial priority in promoting effective planning of transportation systems (Guerrero et al., 2021).

Based on these gaps, this paper contributes to the transport literature in three ways. First, contrary to the prior literature, this study evaluates potential sources of endogeneity in a DCM analysis of a multimodal corridor featuring competition between HSR and air transport. To do so, we apply the Control Function (CF) approach. The reason behind this choice is threefold: (i) its use is largely supported by the literature, which recognizes this method as a suitable approach to tackling endogeneity in DCMs (Guerrero et al., 2021; Guevara and Ben-Akiva, 2012; Guevara, 2015); (ii) its adoption is facilitated by the ease and speed of estimation using readily available software and the possibility of estimating models using RP data without requiring additional information from individuals, such as perception or attitude indicators; and (iii) it yields consistent parameters, even considering the potential trade-off with efficiency due to the estimation process (Guevara, 2015). We applied the CF approach to analyze one of the latest additions to the European HSR network: the Eurostar connection between Amsterdam and London. The HSR alternative entered this highly dense route in 2018, gaining a market share of about 10% against a broad set of air alternatives, including traditional and low-cost carriers, connecting Amsterdam with various airports within the London airport system (Avogadro et al., 2023). To analyze passenger preferences in this market, we leveraged a set of RP data from the International Passenger Survey conducted by the UK Office for National Statistics (ONS), which interviews passengers traveling to and from the United Kingdom by air or the Channel tunnel.

The CF approach has a major shortcoming: finding valid instruments is challenging, since they must be correlated with the endogenous variable (relevance condition) and must be independent of the error term in the DCM (exogeneity condition). Thus, the second contribution of this paper is the proposal of relatively easy-to-obtain valid instruments for correcting fare endogeneity in markets where HSR competes with air carriers. In this case, we propose the fare observed in similar markets and the fuel price of each alternative (e.g., oil or electricity) as instruments. Although we proved endogeneity in the fare when estimating the MNL, the CF method considering the proposed instruments was capable of correcting the estimates. Moreover, by applying the likelihood ratio test, we proved that the corrected model overperformed the endogenous model, and also quantified the impacts of endogeneity on price and frequency elasticities.

The third contribution of our paper concerns the possibility of endogeneity in frequency. While this problem has been somewhat argued in the transportation literature, we found no proof of endogeneity concerning this service attribute in the London–Amsterdam route after checking various instruments. This is likely due to the peculiarities of the specific market, which limit the possibility for airlines and HSR operators to modify service frequency in response to changes in demand.

The remainder of the paper is structured as follows. Section 2 introduces the CF approach used to solve endogeneity in DCMs and describes how to test instrument validity. Section 3 presents the data, while Section 4 discusses the empirical results. Lastly, Section 5 summarizes the findings and suggests avenues for further research.

Table 1
Summary of studies jointly analyzing HSR and air transport demand using individual observations and DCMs.

Paper	Country or Region	Market	Survey[a]	Model[b]	Transport alternative attributes							Endogeneity Correction
					Fare	Frequency	Travel time	Access/egress time	Expected delay (Reliability)	Comfort	Others	
Hensher (1997)	Australia	Sydney Canberra	SP	HEVL	✓	✓	✓	✓				
González-Savignat (2004)	Spain	Madrid Barcelona	SP	MNL	✓	✓	✓	✓				
Park and Ha (2006)	Korea	Seul Busan Mokpo	SP	MNL	✓	✓		✓				
Ortúzar and Simonetti (2008)	Chile	Santiago Concepción	RP-SP	NL	✓	✓	✓		✓	✓		
Román et al. (2010)	Spain	Madrid Zaragoza Barcelona	RP-SP	NL	✓	✓	✓	✓	✓	✓		
Cascetta et al. (2011)	Italy	Rome Naples	RP	MNL, NL	✓		✓	✓			✓	
Behrens and Pels (2012)	Europe	London Paris	RP	MNL-ML	✓	✓	✓	✓	✓			
Pagliara et al. (2012)	Spain	Madrid Barcelona	RP-SP	MNL-ML	✓	✓	✓	✓	✓		✓	
Martín et al. (2014)	Spain	Madrid Barcelona	RP-SP	NL	✓	✓	✓	✓	✓	✓		
Bergantino and Madio (2020)	Italy	Bari Brindisi Rome	SP	BL	✓	✓						
Hong and Najmi (2022)	USA	Dallas Houston	SP	MNL	✓	✓	✓					
This paper	Europe	London Amsterdam	RP	MNL-CF	✓	✓	✓	✓	✓			✓

[a] Stated Preferences (SP), Revealed Preferences (RP).
[b] Heteroscedastic Extreme Value Logit (HEVL), Multinomial Logit (MNL), Nested Logit (NL), Mixed Logit (ML), Binomial Logit (BL), Multinomial Logit with Control Functions (MNL-CF).

2. Modeling framework

Endogeneity is an unavoidable challenge in econometric modeling and stems from potential sources such as omitted variables, measurement or specification inaccuracies, simultaneous determination, and self-selection (Guevara, 2015). To mitigate this issue for DCMs, we apply the CF technique, which is recognized as a practical approach to cope with endogeneity issues (Danaf et al., 2023; Petrin and Train, 2010; Train, 2009; Wooldridge, 2015).

2.1. Control function approach

To introduce the CF approach, let us assume a DCM with endogeneity resulting from the omission of a certain variable q, correlated with an observable variable X. Let V_{in} denote the utility obtained by individual n when choosing alternative i:

$$V_{in} = ASC_i + \beta_y Y_{in} + \beta_x X_{in} + \underbrace{\beta_q q_{in} + e_{in}}_{\varepsilon_{in}}, \tag{1}$$

where ASC_i is an alternative specific constant, Y_{in} is a set of known (measurable) and exogenous attributes, X_{in} is the possible endogenous variable, β_y, β_x, and β_q are parameters to be estimated, and e_{in} is an exogenous error term. Then, assuming that q_{in} is an unobservable variable, the specification proposed by the modeler is as follows:

$$V_{in} = ASC_i + \beta_y Y_{in} + \beta_x X_{in} + \varepsilon_{in}, \tag{2}$$

where the new error term, ε_{in}, contains both e_{in} and q_{in}.

The variable X_{in} is endogenous because it is correlated with ε_{in} through q_{in} in Eq. (2). The correlation arises because in Eq. (3), the variable X_{in} depends on q_{in} as follows:

$$X_{in} = \gamma_0 + \gamma_{z_1} z_{1,in} + \gamma_{z_2} z_{2,in} + \gamma_y Y_{in} + \underbrace{\gamma_q q_{in} + \phi_{in}}_{\delta_{in}}, \tag{3}$$

where $z_{1,in}$ and $z_{2,in}$ are the so-called instrumental variables (IVs or instruments), and ϕ_{in} is an exogenous error term. For illustration, we assume that the endogenous variable X_{in} is correlated with the exogenous variables Y_{in}. However, these variables can also be independent without affecting the validity of the CF method. Moreover, we can rewrite Eq. (3) as follows:

$$X_{in} = \gamma_0 + \gamma_{z_1} z_{1,in} + \gamma_{z_2} z_{2,in} + \gamma_y Y_{in} + \delta_{in}, \tag{4}$$

where the error term δ_{in} contains both ϕ_{in} and q_{in}. By construction, the IVs are correlated with X_{in} through Eq. (4) but are independent of the modeler error term ε_{in}. Then, it can be shown that a DCM corrected by endogeneity has the following functional form (Guevara and Ben-Akiva, 2012):

$$V_{in} = \widetilde{ASC}_i + \widetilde{\beta}_y Y_{in} + \widetilde{\beta}_x X_{in} + \beta_\delta \widetilde{\delta}_{in} + \widetilde{e}_{in}, \tag{5}$$

where $\widetilde{\delta}_{in}$ is a proper estimator of δ_{in}. The intuition is that $\widetilde{\delta}_{in}$ captures the part of the endogenous variable X_{in} correlated with the error term ε_{in}. Therefore, if the instruments $z_{1,in}$ and $z_{2,in}$ are truly exogenous and correlated with the endogenous variable, $\widetilde{\delta}_{in}$ should be added to the utility V_{in} to control the endogeneity problem.

The practical implementation of the CF approach follows two main stages. The first is to find an estimator for δ_{in}, which can be computed as the residual of the ordinary least squares (OLS) regression of X_{in} on the instruments ($z_{1,in}$, $z_{2,in}$) and the exogenous variables (Y_{in}). The second stage is to estimate the DCM by considering $\widetilde{\delta}_{in}$, X_{in}, and Y_{in} as explanatory variables to obtain consistent estimators $\widetilde{\beta}_x$ for the parameter β_x (Guevara and Ben-Akiva, 2012).

Sequential estimation implies efficiency loss, which, in any case, would not occur when the error terms ϵ_{in} and δ_{in} are homoscedastic and are not autocorrelated (Rivers and Vuong, 1988). Estimating standard errors is also more complicated since they cannot be directly obtained from the Fisher information matrix. Therefore, to make inferences, the variance–covariance matrix must be determined using nonparametric methods, such as bootstrapping (Petrin and Train, 2003) or the approach proposed by Karaca-Mandic and Train (2003). Despite these concerns, the sequential estimation of the CF approach proved to be more robust to model misspecification and less computationally expensive (Guevara, 2015).

To verify the hypothesis that a given variable is endogenous in the uncorrected model, we follow the Rivers and Vuong (1988) method by evaluating the significance of the endogenous variable's residual in the second stage of the CF approach. If the residual is significant, the model exhibits endogeneity in that variable. Moreover, since the endogenous model described in Eq. (2) is a restricted version of the correct one proposed in Eq. (5), to evaluate the goodness of fit of these two models and, thus, determine whether the CF approach effectively addresses endogeneity issues, the likelihood ratio test (LR) can be used (Ortúzar and Willumsen, 2011). This test is based on the likelihood function, which measures the probability that a given model generated the observed data. The test statistic for the LR test is defined as follows:

$$LR = -2(l(\Theta)^E - l(\Theta)^{CF}) \sim \chi^2_r, \tag{6}$$

where $l(\Theta)^E$ and $l(\Theta)^{CF}$ are the log-likelihood of the endogenous and corrected model, respectively. Under the null hypothesis, the LR statistic defined above is asymptotically chi-square distributed with r degrees of freedom, where r corresponds to the number of linear restrictions required to transform the more generic model into its restricted version. The LR test has several advantages, such as not requiring strong assumptions about the distribution of errors and its suitability to be applied to a broad range of models. Notwithstanding, it also has some limitations, including the sensitivity to small sample sizes and dependence on proper model specification. In our case, for instance, given that the term $\beta_\delta \widetilde{\delta}_{in}$ is zero, the LR test can be used to compare the models in Eqs. (2) and (5).

2.2. Instruments for endogeneity correction

Prior studies have grouped potential IVs for transportation applications into four main categories (Guerrero et al., 2021; Hotle et al., 2015; Mumbower et al., 2014):

- Cost-shifting instruments (Casey, 1989), with the aim of explaining cost variations across geographic areas or product characteristics;
- Stern-type measures (Stern, 1996) of competition and market power, which focus on the number of products in the market and the time since a product (or firm) was introduced;
- Hausman-type price instruments, which rely on the prices of the same product in other geographic contexts (Hausman et al., 1994; Hausman, 1996);
- BLP-type measures of non-price characteristics of other products, introduced by Berry et al. (1995) and based on the average non-price attributes of other products.

To be considered valid instruments and, thus, solve endogeneity issues, the IVs required for the implementation of the CF method in a DCM need to satisfy two requirements (Rivers and Vuong, 1988; Villas-Boas and Winer, 1999): (i) they must be correlated with the endogenous variable (relevance condition), and (ii) they must be independent of the error term in the DCM (exogeneity condition). Identifying appropriate IVs for practical applications has been a consistent challenge and the subject of controversy (Bresnahan, 1997). However, compliance with these conditions can be verified by specific tests.[1]

Instrument relevance has typically been investigated by evaluating the correlation between the instrument and the endogenous variable. This aspect, initially overlooked, has since been extensively analyzed for linear models (Staiger and Stock, 1997). Stock and Yogo (2005) formalized the analysis of the weak instruments problem in linear regression, determining critical values for identifying weak instruments based on two unambiguous criteria: relative bias and size distortion in the Wald test (Wald, 1943). These tests draw on the F statistic, which tests whether the coefficients of the instruments are zero in the first-stage regression of the CF approach.

[1] Furthermore, for the CF model to be identifiable, at least as many IVs as endogenous variables are required (Guevara and Ben-Akiva, 2012).

Transposing the F-test to evaluate instrument strength in DCMs has then become a common practice and is widely recognized to provide a good proxy of instrument relevance (Guevara, 2024; Guevara and Navarro, 2015). Only recently, Frazier et al. (2020) highlighted that the F-test approach might be unsuitable for DCMs, instead proposing an innovative solution for detecting weak instruments in DCMs. Although this approach appears to be more suitable for DCMs, its implementation is quite complex. Therefore, in this study, in line with current practice, we evaluate the strength of the instrument using the F-test. Nevertheless, we acknowledge this as a limitation of the current work, which can be addressed in future research.

Instrument exogeneity is more challenging to demonstrate because the independence of IVs from the error term —which is not observed— needs to be verified. This requirement is typically assessed using overidentification tests that rely on having more instruments than endogenous variables. Various tests can be used to establish the exogeneity of instruments in DCMs, including the Amemiya-Lee-Newey (ALN) test of overidentifying restrictions, the Refutability (REF) test, and the Hausman (HAU) test (Amemiya, 1978; Guevara, 2010; Hausman, 1978).[2] Recently, Guevara (2018) proposed a novel overidentification test for instrument exogeneity in DCMs —the modified refutability test (S_{mREF})— and evaluated its performance relative to other tests using a binary choice Monte Carlo experiment. The results demonstrated that the S_{mREF} test offers a more straightforward application and is recommended due to its larger power, smaller size distortion, and greater robustness compared to the other tests. Accordingly, in this study, we evaluate instrument exogeneity by applying the S_{mREF} test.

The S_{mREF} test implies adding all the instruments (in our case, $z_{1,in}$ and $z_{2,in}$) as additional variables within the utility function and considering the parameters $\widehat{ASC_i}$, $\widetilde{\beta}_x$, $\widetilde{\beta}_y$, and β_δ to be fixed (as estimated in the corrected model in Eq. (5)) and estimating the parameters $\widetilde{\beta}_{z_1}$ and $\widetilde{\beta}_{z_2}$ while obtaining the associated log-likelihood $l(\Theta)^{CFz}$. The model specification in this case is as follows:

$$V_{in} = \widehat{ASC_i} + \widetilde{\beta}_y Y_{in} + \widetilde{\beta}_x X_{in} + \beta_\delta \widetilde{\delta}_{in} + \widetilde{\beta}_{z_1} z_{1,in} + \widetilde{\beta}_{z_2} z_{2,in} + \widetilde{e}_{in}. \tag{7}$$

Consistent with this formulation, the statistic of the S_{mREF} test for exogeneity is as follows:

$$S_{mREF} = -2(l(\Theta)^{CF} - l(\Theta)^{CFz}) \sim \chi^2_r, \tag{8}$$

where $l(\Theta)^{CF}$ is the log-likelihood of the corrected model obtained in Eq. (5), χ^2_r represents the value of a chi-square distribution with r degrees of freedom, and r equals the degrees of overidentification of the model. The null hypothesis for the S_{mREF} test states that the instruments are exogenous. Conversely, the alternative hypothesis suggests that at least one of the IVs is endogenous. Thus, if S_{mREF} is lower than the critical value of χ^2_r at the required significance level, the instruments can be considered exogenous and independent from the error in the DCM.

3. Data

Data on passenger choices and the characteristics of available transport alternatives are needed to estimate the endogenous and corrected DCM for the Amsterdam–London market.

We used data on modal passenger choices between Amsterdam and London gathered from the International Passenger Survey (IPS), a continuous survey conducted by the UK Office for National Statistics (ONS), which interviews a random sample of about 750,000 passengers per year traveling to and from the United Kingdom by air, sea, or the Channel tunnel. The information collected includes the respondents' sociodemographic characteristics (such as age and gender) and details of the travel arrangements, the departure and arrival travel facilities, and the fare paid. Furthermore, for UK residents, the main county of residence is collected, and for visitors leaving the United Kingdom, the town they visited during their stay is noted.

For our analysis, we selected observations of passengers traveling via London travel facilities (airports and railway stations) to Amsterdam and *vice versa* between 2015 and 2019. Furthermore, observations of air passengers who claimed to travel through London or Amsterdam airports for an interconnecting flight were excluded. The sample assembled according to these specifications was composed of 5199 observations. Table 2 reports the number of observations per year and the travel alternative selected. Overall, low-cost carriers (i.e., easyJet and Vueling) held a market share ranging between 53% in 2015 and 31% in 2019. Full-service carriers served about 47% of passengers in 2015, increasing to about 57% in 2019. Lastly, the Eurostar HSR connection, which began operations in the second quarter of 2018, has gained a relatively limited market share of about 10%.

The choice set faced by passengers in each period, namely the various travel alternatives, was reconstructed by considering the availability and characteristics of the transport modes between London and Amsterdam. The choice set contained at least seven and a maximum of ten travel alternatives. All supply characteristics of the aviation and HSR connections were collected from various secondary sources with monthly granularity. The average weekly frequencies and onboard travel times were collected from the OAG Schedule Analyzer database for the air alternative, while official Eurostar timetables were used for the HSR alternative. Expected delays were calculated based on CAA[3] and Eurostar data.

We calculated the access/egress time to reach each HSR station and airport in the United Kingdom based on the county of residence (UK residents) or the last town visited (foreign visitors) indicated in the IPS survey.[4] In regard to Amsterdam, due to the absence of specific information in the questionnaire about the passenger departure zone or destination, we did not include the access/egress time in the Netherlands. Lastly, the average fare per alternative and period was computed by considering the primary IPS data source. Descriptive statistics for the explanatory variables used are reported in Table A.6 in Appendix A.

[2] For a more comprehensive review on the topic, please refer to Guevara (2018).

[3] Civil Aviation Authority (CAA) UK flight punctuality data.

[4] To compute the access and egress times, we leveraged the routing services Rome2Rio and Openrouteservice.

Table 2
Number of observations per year and chosen alternative (market shares are in brackets).

Year	Alternative									
	AIR 1 LCY-AMS BA	AIR 2 LCY-AMS KL	AIR 3 LGW-AMS BA	AIR 4 LGW-AMS U2	AIR 5 LHR-AMS BA	AIR 6 LHR-AMS KL	AIR 7 LTN-AMS U2	AIR 8 LTN-AMS VY	AIR 9 STN-AMS U2	HSR QQS-ZYA EUR
2015	33 (2.6)		80 (6.3)	261 (20.6)	231 (18.2)	250 (19.7)	286 (22.5)		129 (10.2)	
2016	60 (5.1)		75 (6.4)	223 (18.9)	194 (16.4)	235 (19.9)	181 (15.3)	38 (3.2)	174 (14.7)	
2017	51 (4.9)	38 (3.6)	38 (3.6)	120 (11.4)	172 (16.4)	262 (25)	244 (23.3)	63 (6)	61 (5.8)	
2018	40 (4.7)	25 (2.9)	33 (3.8)	123 (14.3)	129 (15)	216 (25.1)	179 (20.8)	44 (5.1)	34 (4)	36 (4.2)
2019	30 (3.6)	46 (5.5)	32 (3.8)	73 (8.7)	115 (13.7)	254 (30.2)	139 (16.5)	30 (3.6)	16 (1.9)	106 (12.6)

AMS: Amsterdam Schiphol; LCY: London City; LGW: London Gatwick; LTN: London Luton; QQS: London St Pancras; STN: London Stansted; ZYA: Amsterdam Centraal; BA: British Airways; EUR: Eurostar; KL: KLM; U2: easyJet; VY: Vueling Airlines.

4. Results

4.1. Sources of endogeneity

Prior research has reported three primary instances in which the condition of exogeneity becomes violated and, therefore, endogeneity may occur in econometric models: omission of variables, errors-in-variables, and simultaneous causality (Wooldridge, 2010). When considering transportation applications, it is worth taking into account the two main variables that can introduce endogeneity: fare and frequency (Yang et al., 2020; Zhang et al., 2017).

Fare endogeneity can occur due to both simultaneous causality and omission of variables. The former arises because higher consumption (and more urgent transportation needs) allows transport providers to increase their market power and charge higher prices (D'Alfonso et al., 2023). Thus, demand, which is the dependent variable, can affect the price and cause biased estimates (Pagoni and Psaraki-Kalouptsidi, 2016; Zhang et al., 2017). Furthermore, the fare may be correlated with unobservable (thus, not modeled) quality features of transport modes or passengers' subjective attitudes and perceptions, which may affect demand. In these cases, the fare variable also internalizes the effect of this omitted variable(s), resulting in potentially inconsistent estimates of model parameters (Hsiao and Hansen, 2011; Li et al., 2020; Berry and Jia, 2010).

Simultaneous causality also determines frequency endogeneity. This phenomenon arises due to the existence of supply–demand interactions (Hsu and Wen, 2003). Naturally, companies tend to adapt the frequency offered and, more generally, their supply based on the level of demand. However, the overall level of supply in a market also affects overall demand. In addition to the exogenous strength of the socioeconomic attraction factors between the two regions, the quality of available transportation options constitutes a potential driver in contributing to the stimulation of additional flows or, conversely, restrict passenger traffic (Adler et al., 2018; Boonekamp et al., 2018). Importantly, these phenomena (endogeneity in fare and frequency) can occur simultaneously, thereby increasing the complexity of proposing solutions to correct them.

Interestingly, our investigation into the sources of (potential) endogeneity in the London–Amsterdam market revealed no evidence of endogeneity in the frequency. We estimated models assuming that the frequency is also an endogenous variable using two additional instruments. The first IV corresponds to the monthly number of passengers in the London–Amsterdam market, since companies tend to adapt the frequency to changes of demand for air and HSR transport. The second instrument is the weekly frequency in a different market, specifically the London–Paris market. This is a Hausman-type instrument, which relies on the characteristics of the same product in other geographic contexts than the one under investigation (Hausman et al., 1994; Hausman, 1996). We verified the validity of the instruments and the hypothesis of service frequency endogeneity using the set of tests presented in Section 2.[5] While we concluded that the monthly number of passengers on the London–Amsterdam market is not a valid instrument because it did not meet the exogeneity requirement, the weekly frequency for the London–Paris market resulted in being a valid instrument. However, in this case, the weekly frequency residual was not significant, indicating the absence of service frequency endogeneity (Rivers and Vuong, 1988). Furthermore, the likelihood ratio test showed that the model aimed at correcting for frequency and fare endogeneity did not outperform the model aimed at correcting only fare endogeneity.

The reason behind the lack of this phenomenon in our case needs to be traced back to the specific characteristics of the market under consideration. Amsterdam airport as well as airports in London (e.g., London City and Heathrow) use slot regulation to control airport congestion.[6] Although there are no property rights, grandfather rights are used to regulate slot allocation. The grandfather rule determines slot assignment priorities and limits the entrance of new competitors and increases in the frequency of existing airlines. If a carrier has used some slots at least 80% of the time during a season, it is entitled to use the same slot in the next corresponding period; otherwise, the slots become free and may be assigned to new carriers (D'Alfonso et al., 2023). This *use-it-or-lose-it* dynamic may induce slot hoarding and inefficient use of slots, with behaviors such as the so-called *slot babysitting*, in which

[5] The test results are available in Appendix B.

[6] In the European Union, the Slot Allocation Regulation (EC Regulation 95/93, as amended by Regulation 793/2004) defines the mandatory rules for coordinated airports (EC, 2004).

airlines only use them for the minimum amount required to retain their grandfather rights (Ball et al., 2018; Dempsey, 2001). This results in few carriers holding many slots and potentially operating several flights simply to comply with the *use-it-or-lose-it* rule (Madas and Zografos, 2008, 2006). Furthermore, given the specific market under consideration and the presence of hubs, frequencies are largely determined by network considerations (Wei and Hansen, 2005).

As a result, the London–Amsterdam air market is not only highly concentrated but also highly rigid. This significantly reduces the possibility for airlines to increase capacity (frequency) to accommodate higher demand and prevents a reduction in frequency in a certain period of the year due to the feeding dynamic and the threat of losing slots. Similarly, HSR capacity expansion opportunities are severely undermined by the congestion of the core European HSR network. From a theoretical point of view, the above evidence supports the absence of frequency endogeneity resulting from the existence of supply–demand interactions. Consequently, we focus below on investigating instruments aimed at correcting fare endogeneity.

4.2. Empirical results

To analyze passenger choices in the London–Amsterdam market, we first estimated an MNL model in which each alternative was a single travel option (i.e., the combination of the departure/arrival travel facility and the carrier, as shown in Table 2). We modeled passenger choice based on onboard travel time, access/egress time, the logarithm of weekly frequency, expected delay, and fare, which are the variables that usually contribute to consumer choice in this context (see, e.g., Pagliara et al., 2012; Martín et al., 2014). Service frequency was included in logarithmic form for two reasons (Hansen, 1990). First, to account for the expected decreasing marginal utility of frequency. Second, since a route alternative is considered an aggregation of individual flights or trains, the logarithmic form is considered the most suitable for a characteristic that captures the size of an aggregate alternative. In addition to relevant supply attributes that affect passenger choice, we also included a set of alternative specific constants (ASCs) to model the effect of characteristics not explicitly included in the utility formulation (e.g., comfort, safety, and individual preference in the market under study; Ortúzar and Willumsen, 2011).[7]

We opted for this straightforward formulation to avoid all other possible sources of biased or inconsistent parameters, such as identification issues, poor specification, or problems in the optimization procedure, since the MNL has proven to produce robust and consistent parameter estimators. Since the omission of sociodemographic variables from the analysis may introduce an apparent endogeneity issue that could be mitigated if they are properly accounted for, we also tested different model specifications incorporating the sociodemographic variables available in our sample (i.e., gender and age) as systematic taste variations in our utility function. Notably, the results demonstrated that the model still exhibits fare endogeneity even when controlling for sociodemographic attributes, indicating that their omission is not the source of this endogeneity. Accordingly, the only source of biased parameters in the model was fare endogeneity due to omitted variables (other than sociodemographic attributes) or reverse causality. For simplicity, we excluded sociodemographic variables from the base model presented below, as they do not provide additional insights into the main findings of this paper. Nevertheless, the best-performing models (endogenous and corrected) that incorporate sociodemographic variables are presented in Appendix C.

The results of the model without sociodemographic variables, denoted as the endogenous or uncorrected model, are presented on the left-hand side of Table 3. Except for the fare parameter, the parameters of the endogenous model showed correct signs and were statistically significant at 99%. Overall, the model confirmed the sensitivities toward the different travel characteristics reported in previous transport mode choice studies (e.g., Avogadro et al., 2023; Behrens and Pels, 2012; Wardman et al., 2016). Thus, travel time, access time, and expected delay negatively influenced passenger utility, while a higher frequency increased passenger attitudes toward all travel alternatives. Additionally, all ASCs had negative signs, indicating an inherent preference for HSR. This preference is likely due to greater comfort onboard and the absence of stringent luggage size requirements (Avogadro and Redondi, 2023).

However, the positive sign of the fare parameter was evidence of endogeneity in the model. The positive coefficient implies that the passenger's utility increases with the cost incurred when traveling, which is not in line with basic economic theory and typical passenger behavior (Birolini et al., 2020; Pagoni and Psaraki-Kalouptsidi, 2016; Perera and Tan, 2019). Since the fare was significant in the endogenous model and was considered a highly relevant policy variable, its positive sign was considered a major modeling problem that needed to be adequately addressed (Ortúzar and Willumsen, 2011). Furthermore, due to the presence of endogeneity, not only the fare parameter can be biased, but other coefficients can also be biased in magnitude.

To correct for fare endogeneity, we propose two instruments: (i) the average fares observed for trips occurring during the same month and year but in a similar market (i.e., London–Paris) by mode and (ii) the average price of the power source of each alternative (i.e., oil or electricity) for the month and year when the trip occurred. Given the nature of the instruments, the former can be considered a Hausman-type instrument, while the latter is a cost-shifting-type instrument. Theoretically, the two proposed IVs are correlated with the endogenous variable fare and do not confound with market share. In other words, they can affect the endogenous variable through aggregate travel demand but not the individual passenger's travel utility and associated unobservable service attributes. Fuel cost has been considered a valid instrument, as it is correlated with the ticket price and not confounded with market share (Birolini et al., 2020; Gama, 2017; Hsiao and Hansen, 2011). Similarly, Mumbower et al. (2014) and Lurkin et al. (2017) used average prices for other markets as an effective instrument to control for the effects of endogeneity in modeling airline price-demand elasticity.

[7] In a simultaneous but independent work, de Grange et al. (2024) demonstrated that the CF approach could lead to biased estimations of ASCs, which could, in turn, affect the estimates of elasticities and marginal effects. However, we did not test this hypothesis in this paper for two reasons. First, de Grange et al. (2024) also showed that the CF approach did not introduce bias in the parameters of the explanatory variables. Second, comparing our results with those obtained using the innovative approach proposed by de Grange et al. (2024) was beyond the scope of this paper.

Table 3
Endogenous and corrected DCM for the multimodal London–Amsterdam market.

Variable	Alternative	Endogenous model		Corrected model	
		Coefficient[a]	Std. Error	Coefficient[a]	Std. Error[b]
ASC	AIR 1	−11.230	0.8463	−8.2560	0.8675
	AIR 2	−10.960	0.8202	−8.0765	0.8396
	AIR 3	−8.8570	0.7765	−6.3023	0.7930
	AIR 4	−8.1550	0.7815	−5.6479	0.7986
	AIR 5	−9.2730	0.7876	−6.3260	0.8118
	AIR 6	−9.1810	0.7974	−6.1407	0.8228
	AIR 7	−8.6980	0.8252	−6.2176	0.8399
	AIR 8	−9.1370	0.7994	−6.6753	0.8135
	AIR 9	−9.3210	0.8356	−6.5675	0.8528
	HSR	(base)	–	(base)	–
Onboard travel time	All	−0.0638	0.0053	−0.0457	0.0054
Access/egress time	All	−0.0529	0.0006	−0.0532	0.0007
Log (weekly frequency)	All	0.6787	0.0562	0.6541	0.0591
Expected delay	All	−0.0140	0.0033	−0.0114	0.0033
Fare	All	**0.0026**	**0.0003**	**−0.0101**	**0.0010**
Fare's residual	All	–	–	0.0147	0.0010
Sample size		*5199*		*5199*	
Log-likelihood		*−8921.164*		*−8892.846*	

[a] All parameters are significant at the 99% level.
[b] Standard errors determined using bootstrap.

The first IV (i.e., average fare for the London–Paris market) was calculated in the same way as the fare for the London–Amsterdam market, that is, by leveraging the fares reported in the IPS observations of passengers traveling between London and Paris. Regarding the second IV, the average oil price, which is highly correlated with the price of jet fuel, was collected from the EIA (2023), while the electricity price, which is the energy source of the HSR, was obtained from Eurostat (2023).

The right-hand side of Table 3 reports the model corrected for endogeneity using the CF approach. As mentioned in Section 2.1, the standard errors could not be directly inferred from the Fisher information matrix, since we estimated the corrected parameters in two stages. Therefore, we determined the standard errors of the corrected model using the bootstrap approach.[8] The corrected model provides an additional estimate compared to the endogenous model: that corresponding to the fare residual derived from the first stage of the CF method.

Before discussing the estimated parameters, let us verify the hypothesis of fare endogeneity in the uncorrected model and the compliance of the proposed IVs with the exogeneity and relevance requirements. Considering the former, fare endogeneity can be proved following the Rivers and Vuong (1988) method by evaluating the significance of the fare residuals in the second stage of the CF approach (see the right-hand side of Table 3). In our case, the fare residual is significant in the second stage of the CF approach; thus, there is evidence that the model exhibits endogeneity due to the fare variable. Second, we evaluated the validity of the proposed IVs by checking their relevance and exogeneity conditions using the tests explained in Section 2.2. We summarize the results of these tests in Table 4.

Concerning the relevance condition, we considered the recommendation of Guevara and Navarro (2015) to check whether the F-test was greater than 10 in the first-stage regression. This condition held in our model, which confirmed that our instruments (energy source price and London–Paris fares) were correlated with the fare in the London–Amsterdam market. The exogeneity condition of the instruments was confirmed by the S_{mREF} test, which is the most recommended test in the literature (Guevara, 2018). Since the model included one endogenous variable and two instruments, the degree of overidentification for this test was equal to one. In this case, the critical value of the test Chi-square ($\chi^2_{r=1}$) was 3.84, above the value of the S_{mREF} test for our model. Therefore, we can conclude that all our instruments were exogenous, meaning that they were independent of the error term and only affected passenger choice through the value of the fare.

Finally, to evaluate the goodness of fit of the two models (endogenous and corrected, shown in Table 3), we used the likelihood ratio (LR) test and applied Eq. (6) for $l(\theta^E) = -8921.164$ and $l(\theta^{CF}) = -8892.846$. Due to the restrictions on the fare residuals, we

[8] Each bootstrapping sub-sample was built by randomly selecting 80% of the observations from the original sample of 5199 individuals. We arbitrarily set the number of bootstrapping sub-samples to 1000, which was large enough to obtain consistent estimations of the standard errors and confidence intervals.

Table 4
Relevance, exogeneity, goodness of fit, and fare endogeneity tests for the corrected model.

Property	Test	Value	Threshold	
Instruments' relevance	F-test	11.85	> 10	✓
Instruments' exogeneity	S_{mREF}	0.204	< 3.84	✓
Model's goodness of fit	LR	56.64	> 3.84	✓
Fare endogeneity	Fare residual t-test	14.70	> 1.96	✓

compared the LR statistic to the critical value for one degree of freedom at the 95% level ($\chi^2_{r=1} = 3.84$). Since the LR value exceeded the critical value, we confidently rejected the null hypothesis and concluded that the corrected model outperformed the restricted one.

Looking at the coefficients of the corrected model (right-hand side of Table 3), we observed that the CF method allowed us to obtain the appropriate sign of the fare parameter and, therefore, to correct for the endogeneity in this variable. This again confirmed that the fare correlated with the error term of the endogenous model. Interestingly, the ASCs in the corrected model were significantly lower than those in the endogenous model. This suggests that a possible cause of fare endogeneity was the omission of one characteristic of the HSR, which affects individual choice and is correlated with fare. In the endogenous model, the omission of this characteristic was captured both in the ASCs and the fare parameter, while in the corrected model it was captured in the residual parameter of the fare. However, the magnitude of the parameter associated with the weekly frequency did not change significantly compared to the endogenous model, confirming that endogeneity in the London–Amsterdam market only affects the fare. Lastly, it is worth noting that when employing the CF approach, the variances of the corrected model estimators tend to be larger than those of the endogenous model, leading to wider confidence intervals in the CF approach (Guerrero et al., 2021).

Overall, the main results of our analysis can be summarized as follows. First, we did not find evidence of frequency endogeneity in the London–Amsterdam bimodal corridor. This may occur due to market-specific conditions, such as slot regulation in the corresponding airports and capacity restrictions in the HSR network. Second, we found strong evidence of endogeneity affecting transport fares for the market under consideration. Theoretical and empirical evidence suggests that the source of endogeneity could be the omission of relevant HSR attributes and the interdependence of transport demand and ticket prices. Third, we successfully applied the CF method using two IVs: the price of oil or electricity and the fares in a competing market. These instruments, extensively applied in linear models, have proven relevant and exogenous, allowing us to successfully correct for fare endogeneity in the proposed DCM.

To further corroborate that the CF approach did correct the endogeneity problem in the fare parameter, let us focus on the magnitude of the effect of the fare on passenger choice by analyzing the aggregate demand elasticities of the corrected model.[9] Elasticities are frequently used in transport projects, as they model how demand responds to changes in relevant variables (Ortúzar and Willumsen, 2011). Understanding price, frequency, income, and cross-elasticities helps planners anticipate shifts in transportation needs and provides useful insights into optimizing fare structures, ensuring cost-effective infrastructure development, evaluating environmental policy impact, and designing service frequency. Ultimately, by grasping the elasticities, decision-makers can make informed choices regarding the sustainability of transport projects and the effectiveness of infrastructure planning. Table 5 presents the price and frequency elasticities for the multimodal corridor between London and Amsterdam. Taking into account the price elasticity, we observed values below one for all travel alternatives, demonstrating an inelastic demand for both HSR and airlines. This implies that an increase of one percent in the ticket price will result in a demand reduction that is less than proportional. This aligns substantially with previous results on HSR and air competition that leverage DCMs (see the bottom part of Table 5). Similarly, we also observed a low elasticity of demand with respect to the frequency, which is also consistent with previous findings and a very saturated market between the two cities (Givoni and Rietveld, 2009). As expected, alternatives with higher connection frequencies (e.g., LGW-AMS-U2) exhibit lower frequency elasticity due to the decreasing marginal utility of frequency resulting from its inclusion in logarithmic form.

5. Conclusions

Endogeneity is a key challenge in transportation modeling, yet it remains largely overlooked in DCMs. It can result in biased estimates and suboptimal, if not inadequate, policy decisions. This paper investigated sources of endogeneity in a multimodal corridor featuring competition between HSR and air transport and proposed novel IVs to correct for endogeneity. Specifically, we focused on the London–Amsterdam market, which recently observed the introduction of one of the latest additions to the European HSR network. Building on prior literature, we investigated potential endogeneity induced by frequency and fare.

[9] We computed aggregate price and frequency elasticities using sample enumeration, weighing the individual elasticities by the choice probabilities of each alternative (see, e.g., Ramos et al., 2017).

Table 5
Mean and confidence intervals for price and frequency elasticities.

Alternative		Price elasticity		Frequency elasticity	
		Mean	Confidence interval[a]	Mean	Confidence interval[a]
LCY-AMS-BA	AIR1	−0.85	[−0.98, −0.71]	0.60	[0.60, 0.71]
LCY-AMS-KL	AIR2	−0.92	[−1.06, −0.78]	0.60	[0.52, 0.70]
LGW-AMS-BA	AIR3	−0.71	[−0.82, −0.60]	0.59	[0.51, 0.69]
LGW-AMS-U2	AIR4	−0.53	[−0.61, −0.45]	0.46	[0.40, 0.54]
LHR-AMS-BA	AIR5	−0.90	[−1.04, −0.76]	0.53	[0.45, 0.61]
LHR-AMS-KL	AIR6	−0.84	[−0.96, −0.71]	0.47	[0.40, 0.55]
LTN-AMS-U2	AIR7	−0.41	[−0.48, −0.35]	0.46	[0.39, 0.53]
LTN-AMS-VY	AIR8	−0.60	[−0.69, −0.51]	0.61	[0.52, 0.70]
STN-AMS-U2	AIR9	−0.59	[−0.68, −0.51]	0.51	[0.44, 0.60]
QQS-ZYA-EUR	HSR	−0.74	[−0.85, −0.62]	0.56	[0.48, 0.65]
Study	Mode	Mean	(Min, Max)	Mean	(Min, Max)
Martín et al. (2014)	AIR	−0.83			
	HSR	−0.60			
Behrens and Pels (2012)	AIR	−1.43	(−4.37, −0.48)	0.82	(0.6, 1.03)
	HSR	−0.44	(−0.64, −0.13)	0.24	(0.13, 0.25)
Román et al. (2010)	HSR	−0.64	(−0.72, −0.55)		

[a] [2.5th and 97.5th percentile].

First, we showed that frequency could not be endogenous in the specific context of the London–Amsterdam multimodal route. From a theoretical perspective, this could be traced back to certain market features —such as the existence of feeding dynamic and slot retention considerations on the air transport side and congestion on the HSR side— that reduce the existence of supply–demand interactions and, thus, dampen potential endogeneity. We corroborated this thesis by empirically testing different potential instruments that failed to demonstrate frequency endogeneity. This result adds to the literature on frequency endogeneity by proving that particular market configurations may limit the occurrence of this problem.

Second, we identified the fare variable as a potential troublemaker due to simultaneous causality and the omission of specific factors that typically affect this variable. This also emerged from the fare coefficient, resulting in our baseline model (endogenous) being inconsistent with basic economic theory principles. To tackle fare endogeneity, we proposed the average fare observed for trips in a similar market (i.e., London–Paris) and the average price of the power source of each alternative as promising IVs. These variables were demonstrated to satisfy the relevance and exogeneity conditions, supporting their use in correcting for endogeneity. By applying the LR test, we proved that the corrected model outperformed the original model (endogenous). Beyond fixing the inconsistency of the fare coefficient, the endogeneity correction provided more precise estimates for other factors under investigation. This allowed us to suitably compute price and frequency elasticities and uncover variations in estimated parameters when neglecting endogeneity, thus demonstrating the significance of dealing with endogeneity to ensure greater reliability of results in transportation studies and appropriately supporting policy decisions.

The current research paves the way for further research directions. First, a promising topic for future research would be to investigate the implications of travel purpose (business vs. leisure) in this type of market when the model is affected by endogeneity and the effectiveness of the proposed IVs in this context. Second, it would be worthwhile to adapt the model to allow for disentangling mode-specific sensitivity toward alternative attributes (e.g., travel time and frequency). Third, given that a relevant challenge to apply the CF approach lies in identifying suitable instruments that fulfill the exogeneity and relevance criteria, another research avenue might be to investigate how the nature of the instrument (demand- or supply-related) affects the correction for endogeneity. Finally, in light of the recent paper by de Grange et al. (2024) suggesting that ASC estimates using the CF approach could be biased, we believe that comparing our model results with those obtained using that of de Grange et al. (2024) would be another interesting avenue for future research.

CRediT authorship contribution statement

Thomas E. Guerrero B.: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Nicolò Avogadro:** Conceptualization, Data curation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Raúl Ramos:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

R. Ramos gratefully acknowledges financial support from ANID BECAS/DOCTORADO NACIONAL 21180398 and the Erasmus+ programme of the European Union.

Table A.6
Descriptive statistics for the explanatory variables.

	Air[a]		HSR		Total	
	Mean	SD	Mean	SD	Mean	SD
Onboard travel time	74.1	5.5	224.7	2.4	79.0	27.1
Access/egress time	58.9	36.3	44.3	45.8	58.4	36.7
Weekly frequency	38.2	18.6	15.0	3.0	37.4	18.8
Log (weekly frequency)	3.5	0.6	2.7	0.2	3.5	0.7
Expected delay	5.5	4.6	0.9	0.0	5.4	4.6
Fare	85.3	41.2	86.1	22.4	85.4	40.8
Sample size	*43,071*		*1,431*		*44,502*	

[a] Alternatives from 1 to 9 in Table 2.

Table B.7
Relevance, exogeneity, goodness of fit, and frequency endogeneity tests compared with a corrected model assuming endogeneity in the frequency.

Property	Test	MNP-LA			WF-LP		
		Value	Threshold		Value	Threshold	
Instrument's relevance	F-test	218.10	> 10	✓	208.80	> 10	✓
Instrument's exogeneity	S_{mREF}	19.08	< 3.84	✗	1.52	< 3.84	✓
Model's goodness of fit	LR	−19.29	> 3.84	✗	−4.77	> 3.84	✗
Weekly frequency endogeneity	WFR t-test	0.71	> 1.96	✗	1.20	> 1.96	✗

MNP-LA: Monthly number of passengers in the London–Amsterdam market.
WF-LP: Weekly frequency for the London-Paris market.
WFR t-test: Weekly Frequency residual t-test.

Appendix A. Descriptive statistics

See Table A.6.

Appendix B. Weekly frequency endogeneity tests

See Table B.7.

Appendix C. Models including sociodemographic variables

In Table C.8, we present the endogenous and corrected models when controlling for sociodemographic attributes as taste variations in the utility function. Specifically, we modeled the influence of sociodemographic variables observed in our sample (i.e., *age* and *gender*) as taste variations affecting onboard travel time and weekly frequency. *Age* is defined as a dummy variable equal to 1 if the respondent is 35 years old or younger, and otherwise 0. *Gender* is a dummy variable equal to 1 if the respondent is male, and otherwise 0. reports the results of the endogenous and corrected models based on these specifications. Notably, including sociodemographic variables does not correct the sign of the fare parameter in the endogenous model. Additionally, fare residual is significant in the second stage of the CF approach. These outcomes suggest that the source of endogeneity is not the omission of sociodemographic variables but rather the influence of demand on pricing.

Data availability

The authors do not have permission to share data.

Table C.8
Endogenous and corrected DCM for the multimodal London–Amsterdam market with sociodemographic variables as systematic taste variations.

Variable	Alternative	Endogenous model		Corrected model	
		Coefficient	Std. Error	Coefficient	Std. Error[b]
ASC	AIR 1	−9.5330[a]	0.7970	−8.1375[a]	0.7693
	AIR 2	−8.5940[a]	0.7733	−7.1678[a]	0.7513
	AIR 3	−9.7160[a]	0.7773	−8.0248[a]	0.7558
	AIR 4	−9.6600[a]	0.7859	−7.9116[a]	0.7652
	AIR 5	−9.1340[a]	0.8137	−7.7271[a]	0.7892
	AIR 6	−11.660[a]	0.8294	−9.9613[a]	0.8036

(*continued on next page*)

Table C.8 (*continued*).

Variable	Alternative	Endogenous model		Corrected model	
		Coefficient	Std. Error	Coefficient	Std. Error[b]
	AIR 7	−9.7380[a]	0.8272	−8.1719[a]	0.8011
	AIR 8	−9.2380[a]	0.7994	−7.7856[a]	0.7442
	AIR 9	−11.400[a]	0.7682	−9.7451[a]	0.7800
	HSR	(base)	–	(base)	–
Onboard travel time	All	−0.0619[a]	0.0053	−0.0515[a]	0.0051
Onboard travel time - Age	All	−0.0057[a]	0.0006	−0.0057[a]	0.0054
Onboard travel time - Gender	All	−0.0028[a]	0.0008	−0.0027[a]	0.0054
Access/egress time	All	−0.0528[a]	0.0006	−0.0530[a]	0.0006
Log (weekly frequency)	All	1.0510[a]	0.0680	1.0412[a]	0.0686
Log (weekly frequency) - Age	All	−0.5866[a]	0.0340	−0.6246[a]	0.0006
Log (weekly frequency) - Gender	All	−0.0040	0.0270	0.0213	0.0008
Expected delay	All	−0.0142[a]	0.0030	−0.0129[a]	0.0030
Fare	**All**	**0.0022**[a]	**0.0003**	**−0.0054**[a]	**0.0009**
Fare's residual	All	–	–	0.0091[a]	0.0010
Sample size		*5199*		*5199*	
Log-likelihood		*−8871.029*		*−8858.445*	

[a] Significant at the 99% level.

[b] Standard errors determined using bootstrap.

References

Adler, N., Njoya, E.T., Volta, N., 2018. The multi-airline p-hub median problem applied to the African aviation market. Transp. Res. A 107, 187–202. http://dx.doi.org/10.1016/j.tra.2017.11.011.

Amemiya, T., 1978. The estimation of a simultaneous equation generalized probit model. Econometrica 1193–1205.

Arteaga, C., Park, J., Beeramoole, P.B., Paz, A., 2022. xlogit: An open-source Python package for GPU-accelerated estimation of Mixed Logit models. J. Choice Model. 42, 100339. http://dx.doi.org/10.1016/j.jocm.2021.100339.

Avogadro, N., Pels, E., Redondi, R., 2023. Policy impacts on the propensity to travel by HSR in the Amsterdam – London market. Socio-Econ. Plan. Sci. 87, 101585. http://dx.doi.org/10.1016/J.SEPS.2023.101585.

Avogadro, N., Redondi, R., 2023. Diverted and induced demand: Evidence from the London-Paris passenger market. Res. Transp. Econ. 100, 101304. http://dx.doi.org/10.1016/J.RETREC.2023.101304.

Ball, M.O., Berardino, F., Hansen, M., 2018. The use of auctions for allocating airport access rights. Transp. Res. A 114, 186–202. http://dx.doi.org/10.1016/j.tra.2017.09.026.

Behrens, C., Pels, E., 2012. Intermodal competition in the London-Paris passenger market: High-Speed Rail and air transport. J. Urban Econ. 71 (3), 278–288. http://dx.doi.org/10.1016/j.jue.2011.12.005.

Bergantino, A.S., Madio, L., 2020. Intermodal competition and substitution. HSR versus air transport: Understanding the socio-economic determinants of modal choice. http://dx.doi.org/10.1016/j.retrec.2020.100823.

Berry, S., Jia, P., 2010. Tracing the Woes: An empirical analysis of the airline industry. Am. Econ. J. Microecon. 2 (3), 1–43. http://dx.doi.org/10.1257/mic.2.3.1.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 841–890. http://dx.doi.org/10.2307/2171802.

Birolini, S., Cattaneo, M., Malighetti, P., Morlotti, C., 2020. Integrated origin-based demand modeling for air transportation. Transp. Res. E 142, 102050. http://dx.doi.org/10.1016/j.tre.2020.102050.

Boonekamp, T., Zuidberg, J., Burghouwt, G., 2018. Determinants of air travel demand: The role of low-cost carriers, ethnic links and aviation-dependent employment. Transp. Res. A 112, 18–28. http://dx.doi.org/10.1016/j.tra.2018.01.004.

Bresnahan, T.F., 1997. The apple-cinnamon cheerios war: Valuing new goods, identifying market power, and economic measurement. In: The Economics of New Goods. The University of Chicago Press, Chicago/London, URL: https://web.stanford.edu/~tbres/Unpublished_Papers/hausman%20recomment.pdf.

Cameron, T.A., DeShazo, J., Johnson, E.H., 2011. Scenario adjustment in stated preference research. J. Choice Model. 4 (1), 9–43. http://dx.doi.org/10.1016/S1755-5345(13)70017-4.

Cascetta, E., Papola, A., Pagliara, F., Marzano, V., 2011. Analysis of mobility impacts of the high speed Rome-Naples rail link using withinday dynamic mode service choice models. J. Transp. Geogr. 19 (4), 635–643. http://dx.doi.org/10.1016/J.JTRANGEO.2010.07.001.

Casey, B., 1989. Early retirement: the problems of "instrument substitution" and "cost shifting" and their implications for restructuring the process of retirement. In: Redefining the Process of Retirement: An International Perspective. Springer, pp. 133–150. http://dx.doi.org/10.1007/978-3-642-74513-3_8.

D'Alfonso, T., Gregori, M., Silva, H.E., Basso, L.J., 2023. The economics of airports' pricing. In: Tirachini, A., Hörcher, D., Verhoef, E.T. (Eds.), Handbook on Transport Pricing and Financing. Edward Elgar Publishing, pp. 207–228. http://dx.doi.org/10.4337/9781800375550.00018.

Danaf, M., Guevara, A., Atasoy, B., Ben-Akiva, M., 2020. Endogeneity in adaptive choice contexts: Choice-based recommender systems and adaptive stated preferences surveys. J. Choice Model. 34, 100200. http://dx.doi.org/10.1016/j.jocm.2019.100200.

Danaf, M., Guevara, C.A., Ben-Akiva, M., 2023. A control-function correction for endogeneity in random coefficients models: The case of choice-based recommender systems. J. Choice Model. 46, 100399. http://dx.doi.org/10.1016/j.jocm.2022.100399.

de Grange, L., González, F., Marechal, M., Troncoso, R., 2024. Estimating multinomial logit models with endogenous variables: Control function versus two adapted approaches. Transp. Res. A 183, 104068. http://dx.doi.org/10.1016/j.tra.2024.104068.

Dempsey, P.S., 2001. Airport landing slots: barriers to entry and impediments to competition. Air Space Law 26 (1), http://dx.doi.org/10.54648/332045.

EC, 2004. European Council Regulation No 793/2004 of April 2004 amending Council Regulation No 95/93 on common rules for the allocation of slots at Community airports. Off. J. Eur. Union L 138 (2004), 50–60.

EIA, 2023. Europe Brent Spot Price FOB (Dollars per Barrel). URL: https://www.eia.gov/dnav/pet/hist/rbrteM.htm.

Eurostat, 2023. Electricity price statistics. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics.

Fernández-Antolín, A., Guevara, C.A., De Lapparent, M., Bierlaire, M., 2016. Correcting for endogeneity due to omitted attitudes: Empirical assessment of a modified MIS method using RP mode choice data. J. Choice Model. 20, 1–15. http://dx.doi.org/10.1016/j.jocm.2016.09.001.

Frazier, D.T., Renault, E., Zhang, L., Zhao, X., 2020. Weak identification in discrete choice models. arXiv preprint arXiv:2011.06753. http://dx.doi.org/10.48550/arXiv.2011.06753.

Fu, X., Lei, Z., Wang, K., Yan, J., 2015. Low cost carrier competition and route entry in an emerging but regulated aviation market – The case of China. In: Selected papers from the Air Transport Research Society World Conference, Bergamo (Italy) and World Conference on Transport Research, Rio de Janeiro (Brazil), 2013. Transp. Res. A 79, 3–16. http://dx.doi.org/10.1016/j.tra.2015.03.020.

Gama, A., 2017. Own and cross-price elasticities of demand for domestic flights and intercity trains in the U.S. Transp. Res. D 54, 360–371. http://dx.doi.org/10.1016/j.trd.2017.06.010.

Givoni, M., Dobruszkes, F., 2013. A Review of Ex-Post Evidence for Mode Substitution and Induced Demand Following the Introduction of High-Speed Rail, vol. 33. Routledge, pp. 720–742. http://dx.doi.org/10.1080/01441647.2013.853707, http://dx.doi.org/10.1080/01441647.2013.853707.

Givoni, M., Rietveld, P., 2009. Airline's choice of aircraft size–explanations and implications. Transp. Res. A 43 (5), 500–510. http://dx.doi.org/10.1016/j.tra.2009.01.001.

González-Savignat, M., 2004. Competition in air transport: The case of the high speed train. J. Transp. Econ. Policy 38 (1), 77–108, URL: www.jstor.org/stable/20173046Copy.

Gu, H., Wan, Y., 2022. Airline reactions to high-speed rail entry: Rail quality and market structure. Transp. Res. A 165, 511–532. http://dx.doi.org/10.1016/j.tra.2022.09.016.

Guerrero, T.E., Guevara, C.A., Cherchi, E., Ortúzar, J.d., 2021. Addressing endogeneity in strategic urban mode choice models. Transportation 48 (4), 2081–2102. http://dx.doi.org/10.1007/S11116-020-10122-Y.

Guerrero, T.E., Guevara, C.A., Cherchi, E., Ortúzar, J.d., 2022. Characterizing the impact of discrete indicators to correct for endogeneity in discrete choice models. J. Choice Model. 42, 100342. http://dx.doi.org/10.1016/J.JOCM.2021.100342.

Guevara, C.A., 2010. Endogeneity and Sampling of Alternatives in Spatial Choice Models (Ph.D. thesis). Massachusetts Institute of Technology.

Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transp. Res. A 82, 240–254. http://dx.doi.org/10.1016/j.tra.2015.10.005.

Guevara, C.A., 2018. Overidentification tests for the exogeneity of instruments in discrete choice models. Transp. Res. B 114, 241–253. http://dx.doi.org/10.1016/j.trb.2018.05.020.

Guevara, C.A., 2024. Endogeneity in discrete choice models. In: Handbook of Choice Modelling. Edward Elgar Publishing, pp. 668–692. http://dx.doi.org/10.4337/9781800375635.00032.

Guevara, C.A., Ben-Akiva, M.E., 2012. Change of scale and forecasting with the control-function method in logit models. Transp. Sci. 46 (3), 425–437. http://dx.doi.org/10.1287/trsc.1110.0404.

Guevara, C., Navarro, P., 2015. Detection of weak instruments when correcting for endogeneity in binary logit models. In: Proceedings 14th International Conference on Travel Behaviour Research (IATBR), Windsor, UK.

Haghani, M., Bliemer, M.C., Rose, J.M., Oppewal, H., Lancsar, E., 2021. Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. J. Choice Model. 41, 100309. http://dx.doi.org/10.1016/j.jocm.2021.100309.

Hansen, M., 1990. Airline competition in a hub-dominated environment: An application of noncooperative game theory. Transp. Res. B 24 (1), 27–43. http://dx.doi.org/10.1016/0191-2615(90)90030-3.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica: J. Econ. Soc. 1251–1271.

Hausman, J.A., 1996. Valuation of new goods under perfect and imperfect competition. In: Bresnahan, T.F., Gordon, R.J. (Eds.), The Economics of New Goods. University of Chicago Press, pp. 207–248.

Hausman, J., Leonard, G., Zona, J.D., 1994. Competitive analysis with differenciated products. Ann. Econ. Stat. 159–180. http://dx.doi.org/10.2307/20075951.

Hensher, D.A., 1997. A practical approach to identifying the market potential for high speed rail: A case study in the Sydney-Canberra corridor. Transp. Res. A 31 (6), 431–446. http://dx.doi.org/10.1016/S0965-8564(97)00001-3.

Hong, S.-J., Najmi, H., 2022. Impact of High-speed rail on air travel demand between Dallas and Houston applying Monte Carlo simulation. J. Air Transp. Manag. 102, 102222. http://dx.doi.org/10.1016/j.jairtraman.2022.102222.

Hotle, S.L., Castillo, M., Garrow, L.A., Higgins, M.J., 2015. The impact of advance purchase deadlines on airline consumers' search and purchase behaviors. Transp. Res. A 82, 1–16. http://dx.doi.org/10.1016/j.tra.2015.09.001.

Hsiao, C.-Y., Hansen, M., 2011. A passenger demand model for air transportation in a hub-and-spoke network. Transp. Res. E 47 (6), 1112–1125. http://dx.doi.org/10.1016/j.tre.2011.05.012.

Hsu, C.-I., Wen, Y.-H., 2003. Determining flight frequencies on an airline network with demand–supply interactions. Transp. Res. E 39 (6), 417–441. http://dx.doi.org/10.1016/S1366-5545(02)00060-1.

Jiang, Q., Penn, J., Hu, W., 2022. Real payment priming to reduce potential hypothetical bias. J. Choice Model. 45, 100383. http://dx.doi.org/10.1016/j.jocm.2022.100383.

Jiang, C., Wan, Y., Yang, H., Zhang, A., 2021. Impacts of high-speed rail projects on CO2 emissions due to modal interactions: A review. Transp. Res. D 100, 103081. http://dx.doi.org/10.1016/j.trd.2021.103081.

Karaca-Mandic, P., Train, K., 2003. Standard error correction in two-stage estimation with nested samples. Econom. J. 6 (2), 401–407. http://dx.doi.org/10.1111/1368-423X.t01-1-00115.

Lee, D.S., Fahey, D.W., Skowron, A., Allen, M.R., Burkhardt, U., Chen, Q., Doherty, S.J., Freeman, S., Forster, P.M., Fuglestvedt, J., et al., 2021. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. Atmos. Environ. 244, 117834. http://dx.doi.org/10.1016/j.atmosenv.2020.117834.

Li, H., Strauss, J., Lu, L., 2019a. The impact of high-speed rail on civil aviation in China. Transp. Policy 74, 187–200. http://dx.doi.org/10.1016/J.TRANPOL.2018.11.015.

Li, H., Wang, K., Yu, K., Zhang, A., 2020. Are conventional train passengers underserved after entry of high-speed rail?-Evidence from Chinese intercity markets. Transp. Policy 95, 1–9. http://dx.doi.org/10.1016/j.tranpol.2020.05.017.

Li, Y., Yang, B., Cui, Q., 2019b. The effects of high-speed rail on air passenger transport in China. Appl. Econ. Lett. 26 (9), 745–749. http://dx.doi.org/10.1080/13504851.2018.1494798.

Lurkin, V., Garrow, L.A., Higgins, M.J., Newman, J.P., Schyns, M., 2017. Accounting for price endogeneity in airline itinerary choice models: An application to Continental U.S. markets. Transp. Res. A 100, 228–246. http://dx.doi.org/10.1016/j.tra.2017.04.007.

Madas, M.A., Zografos, K.G., 2006. Airport slot allocation: From instruments to strategies. J. Air Transp. Manag. 12 (2), 53–62. http://dx.doi.org/10.1016/j.jairtraman.2005.08.001.

Madas, M.A., Zografos, K.G., 2008. Airport capacity vs. demand: mismatch or mismanagement? Transp. Res. A 42 (1), 203–226. http://dx.doi.org/10.1016/j.tra.2007.08.002.

Martín, J.C., Román, C., García-Palomares, J.C., Gutiérrez, J., 2014. Spatial analysis of the competitiveness of the high-speed train and air transport: The role of access to terminals in the Madrid–Barcelona corridor. Transp. Res. A 69, 392–408. http://dx.doi.org/10.1016/j.tra.2014.09.010.

Mizutani, J., Sakai, H., 2021. Which is a stronger competitor, High Speed Rail, or Low Cost Carrier, to Full Service Carrier? – Effects of HSR network extension and LCC entry on FSC's airfare in Japan. J. Air Transp. Manag. 90, 101965. http://dx.doi.org/10.1016/j.jairtraman.2020.101965.

Molloy, J., Becker, F., Schmid, B., Axhausen, K.W., 2021. Mixl: An open-source r package for estimating complex choice models on large datasets. J. Choice Model. 39, 100284. http://dx.doi.org/10.1016/j.jocm.2021.100284.

Mumbower, S., Garrow, L.A., Higgins, M.J., 2014. Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. Transp. Res. A 66, 196–212. http://dx.doi.org/10.1016/j.tra.2014.05.003.

Ortúzar, J.d., Simonetti, C., 2008. Modelling the demand for medium distance air travel with the mixed data estimation method. J. Air Transp. Manag. 14 (6), 297–303. http://dx.doi.org/10.1016/j.jairtraman.2008.08.002.

Ortúzar, J.d., Willumsen, L.G., 2011. Modelling Transport. Wiley, p. 586. http://dx.doi.org/10.1002/9781119993308.

Pagliara, F., Vassallo, J.M., Román, C., 2012. High-Speed Rail versus Air Transportation, vol. 2289. SAGE Publications Sage CA, Los Angeles, CA, pp. 10–17. http://dx.doi.org/10.3141/2289-02, https://doi-org.ezproxy.uninorte.edu.co/10.3141/2289-02.

Pagoni, I., Psaraki-Kalouptsidi, V., 2016. The impact of carbon emission fees on passenger demand and air fares: A game theoretic approach. J. Air Transp. Manag. 55, 41–51. http://dx.doi.org/10.1016/j.jairtraman.2016.04.004.

Palma, D., de Dios Ortúzar, J., Rizzi, L.I., Guevara, C.A., Casaubon, G., Ma, H., 2016. Modelling choice when price is a cue for quality: a case study with Chinese consumers. J. Choice Model. 19, 24–39. http://dx.doi.org/10.1016/j.jocm.2016.06.002.

Park, Y., Ha, H.-K., 2006. Analysis of the impact of high-speed railroad service on air transport demand. In: Selected papers from the 8th ATRS Conference. Transp. Res. E 42 (2), 95–104. http://dx.doi.org/10.1016/j.tre.2005.09.003.

Perera, S., Tan, D., 2019. In search of the ''Right Price'' for air travel: First steps towards estimating granular price-demand elasticity. Transp. Res. A 130, 557–569. http://dx.doi.org/10.1016/j.tra.2019.09.013.

Petrin, A., Train, K., 2003. Omitted Product Attributes in Discrete Choice Models. Working Paper Series 9452, National Bureau of Economic Research, http://dx.doi.org/10.3386/w9452.

Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. J. Mark. Res. 47 (1), 3–13. http://dx.doi.org/10.1509/jmkr.47.1.3.

Ramos, R., Cantillo, V., Arellana, J., Sarmiento, I., 2017. From restricting the use of cars by license plate numbers to congestion charging: Analysis for Medellin, Colombia. Transp. Policy 60, 119–130. http://dx.doi.org/10.1016/j.tranpol.2017.09.012.

Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. J. Econometrics 39 (3), 347–366. http://dx.doi.org/10.1016/0304-4076(88)90063-2.

Román, C., Espino, R., Martín, J.C., 2010. Analyzing competition between the high speed train and alternative modes. The case of the Madrid-Zaragoza-Barcelona Corridor. J. Choice Model. 3 (1), 84–108. http://dx.doi.org/10.1016/S1755-5345(13)70030-7.

Sarrias, M., 2021. A two recursive equation model to correct for endogeneity in latent class binary probit models. J. Choice Model. 40, 100301. http://dx.doi.org/10.1016/j.jocm.2021.100301.

Staiger, D.O., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586. http://dx.doi.org/10.2307/2171753.

Stern, S., 1996. Market Definition and the Returns to Innovation: Substitution Patterns in Pharmaceutical Markets. Working Paper, Sloan School of Management, Massachusetts Institute of Technology.

Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D.W., Stock, J.H. (Eds.), Identification and Inference for Econometric Models, Essays in Honour of Thomas Rothenberg. Cambridge University Press, New York, pp. 80–108.

Strauss, J., Li, H., Cui, J., 2021. High-speed Rail's impact on airline demand and air carbon emissions in China. Transp. Policy 109, 85–97. http://dx.doi.org/10.1016/j.tranpol.2021.05.019, URL: https://www.sciencedirect.com/science/article/pii/S0967070X21001633.

Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511805271.

UIC, 2022. High-Speed Rail Atlas. Technical Report, International Union of Railways.

Vigren, A., 2017. Competition in Swedish passenger railway: Entry in an open access market and its effect on prices. Econ. Transp. 11–12, 49–59. http://dx.doi.org/10.1016/j.ecotra.2017.10.005.

Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Manag. Sci. 45 (10), 1324–1338. http://dx.doi.org/10.1287/mnsc.45.10.1324.

Wald, A., 1943. Tests F-statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54, 426–482.

Wardman, M., Chintakayala, V.P.K., de Jong, G., 2016. Values of travel time in Europe: Review and meta-analysis. Transp. Res. A 94, 93–111. http://dx.doi.org/10.1016/j.tra.2016.08.019.

Watanabe, H., Maruyama, T., 2023. A Bayesian instrumental variable model for multinomial choice with correlated alternatives. J. Choice Model. 46, 100400. http://dx.doi.org/10.1016/j.jocm.2023.100400.

Wei, W., Hansen, M., 2005. Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets. Transp. Res. E 41 (4), 315–327. http://dx.doi.org/10.1016/j.tre.2004.06.002.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Wooldridge, J.M., 2015. Control function methods in applied econometrics. J. Hum. Resour. 50 (2), 420–445. http://dx.doi.org/10.3368/jhr.50.2.420.

Yang, H., Ma, W., Wang, Q., Wang, K., Zhang, Y., 2020. Welfare implications for air passengers in China in the era of high-speed rail. Transp. Policy 95, A1–A13. http://dx.doi.org/10.1016/j.tranpol.2020.04.007.

Zhang, R., Johnson, D., Zhao, W., Nash, C., 2019a. Competition of airline and high-speed rail in terms of price and frequency: Empirical study from China. Transp. Policy 78, 8–18. http://dx.doi.org/10.1016/j.tranpol.2019.03.008.

Zhang, A., Wan, Y., Yang, H., 2019b. Impacts of high-speed rail on airlines, airports and regional economies: A survey of recent research. Transp. Policy 81, A1–A19. http://dx.doi.org/10.1016/J.TRANPOL.2019.06.010.

Zhang, Q., Yang, H., Wang, Q., 2017. Impact of high-speed rail on China's Big Three airlines. Transp. Res. A 98, 77–85. http://dx.doi.org/10.1016/J.TRA.2017.02.005.