Assessing the Performance of Berkson-Theil Method on Multiple Choice Sets and Aggregated Choice Data

Weibo Li*, Maria Kamargianni, Philip Krammer, Lynnette Dray, Andreas Schafer

University College London
UCL Energy Institute, Central House, 14 Upper Woburn Place, WC1H 0NN, London, UK

*Corresponding Author
Email: weibo.li.10@ucl.ac.uk

ABSTRACT

One type of input data in choice modeling is that there are multiple or even a large number of choice sets and only an aggregated demand for each choice is available. Such data can usually be derived from official sources or big data platforms representing a high level of aggregation. An example could be when developing an air travel itinerary choice model at regional or national level, such that each different Origin-Destination pair will be a unique choice set with a number of itineraries in it as the choices. As an alternative approach to logistic regression based on maximum likelihood estimation (MLE), Berkson-Theil method that uses least squares is rarely being remembered. However, the method has significant practical advantages in terms of handling a large number of choice sets and software compatibility when dealing with this particular data type. As a result, this paper offers a leading research that assesses the performance of Berkson-Theil method in such a data case by comparing the model estimation results of Berkson-Theil method based on ordinary least squares (OLS) to a logistic regression based on MLE and also testing the predictive powers of the two methods. The comparisons reveal that the two methods can offer similar model estimation results; however, the results of logistic regression can lead to more accurate predictions. Heteroskedasticity is discovered in the end implying that the choice of OLS could be a cause for the lower predictive power of Berkson-Theil method. Overall, the findings suggest that Berkson-Theil method can be an effective approach in dealing with aggregated choice data with multiple choice sets and it may perform better if heteroskedasticity can be captured using weighted least squares (WLS) as the estimation technique instead. Choice modelers in air transportation and other domains that often deal with big data could therefore make use of this method given its significant practical advantages.

Keywords: Berkson-Theil method, Multiple choice sets, Aggregated data, Itinerary choice, Maximum likelihood, Least squares

1. INTRODUCTION

In choice behavior modeling, three types of input data are often involved: 1) There is a single choice set and individuals' choices are observed; 2) There is a single choice set and only an aggregated demand for each choice is available; 3) There are multiple or even a large number of choice sets and only an aggregated demand for each choice is available. Type 1 data is usually collected by surveys containing either revealed preference or stated preference choice behavior information. Type 2 and 3 data are usually derived from official sources or big data platforms representing a high level of aggregation.

Different methodologies may be chosen to analyze different types of data. Logistic regression using maximum likelihood estimation (MLE) is the most commonly used method to study choice behavior given type 1 data. Other rarely seen approaches such as linear probability models and non-linear models based on least squares estimation can also be applied though they have neither practical nor theoretical advantages over MLE with this data type (Ben-Akiva and Lerman, 1985).

For type 2 data, apart from logistic regression, one of the non-linear models relying on least squares named as Berkson-Theil method (i.e. developed by Berkson (1953) for binary choice and extended to multiple choices by Theil (1969)) were adopted by some researchers (Carrier and Weatherford, 2014; Schafer, 2015) due to the computational simplicity it offers when having aggregated data form and large data size (Ben-Akiva and Lerman, 1985).

Finally, there are three sub-cases in dealing with type 3 data since the capability of software becomes a critical concern. The first one is when the number of choice sets is small and as a result logistic regression can still be applied on most of the choice modeling software by introducing loops. The next case is that when the number of choice sets is large (i.e. thousands and more), to our knowledge on the existing modeling tools, logistic regression can only work on a unique software, GAUSS due to its strong looping facilities (Aptech Systems, Inc., 2003). Nevertheless, GAUSS has not been widely used to analyze type 3 data due to its complicated coding requirement and relatively costly price. Hence, most of the studies that have dealt with this data type limited their scope to a small number of choice sets (Ghobrial and Soliman, 1992; Ghobrial and Kanafani, 1995; Atasoy and Bierlaire, 2012; Busquets et al., 2016) or had to pick a small sample among a large number of choice sets due to software limit (Weidner, 1996). The last case is that as an alternative approach to logistic regression, Berkson-Theil method can work with any statistical software and any data size without having to reduce the number of choice sets. So far, the only application of Berkson-Theil method on type 3 data is found in Hsiao and Hansen (2011).

Briefly speaking, Berkson-Theil method transforms a logit choice model using MLE to a least squares regression model. Detailed transformation procedures are presented in the model specification section (Section 4). It has great potential to be adopted when dealing with type 3 data given its significant practical advantages in terms of handling a large number of choice sets and software compatibility. However, as a largely forgotten method, it has not been widely applied and its reliability when dealing with type 3 data has not been explored.

This paper aims to assess the performance of Berkson-Theil method on analyzing type 3 data. The study compares the model estimation results of Berkson-Theil method based on ordinary least squares (OLS) estimation to a logistic regression based on MLE and also tests the predictive powers of the two methods. Stata (StataCorp., 2011) and PythonBiogeme (Bierlaire, 2016) are used to develop the models respectively.

This work is expected to be the first research that compares the two methods when the data has multiple choice sets and is in aggregated form. If close results can be found, then it has great implication to the choice analysis based on big data as Berkson-Theil method can be a computationally simplified as well as a reliable alternative to the traditional logistic regression.

The structure of the paper is as follows. By taking air transportation industry as an example, section 2 reviews the current status of choice modeling studies based on the three data types. Section 3 provides information on the data that will be analyzed in this study. Model specifications by using Berkson-Theil method and logistic regression and the comparisons between their results are given in section 4 and 5. Section 6 concludes the paper.

2. LITERATURE REVIEW

Many studies have employed disaggregated data to study various choice topics on air transportation. Data from revealed preference surveys was commonly used to investigate the choices of airlines (Proussaloglou and Koppelman, 1995; Ishii et al., 2008), airports (Harvey, 1987; Hansen, 1995; Windle and Dresner, 1995; Pels et al., 2001; Pels et al., 2003; Hess and Polak, 2005; Pathomsiri and Haghani, 2005) and itineraries (Nassiri and Rezaei, 2012). Stated preference data was also gathered in a few cases, such as Nason (1981) on airline choices, Hess et al. (2007) and Loo (2008) on airport choices, Adler et al. (2005), Warburg et al. (2006) and Theis et al. (2006) on itinerary choices. Yoo and Ashford (1996) adopted both revealed and stated preference data to examine airline choices. Nevertheless, all of these studies used logistic regression to study choice behaviors, although some only developed multinomial logit models and some extended the models to more advanced levels by including nested logit and mixed logit models.

However, when the available choice data is in grouped form such that each choice is only associated with an aggregated demand, some researchers switched to Berkson-Theil method as an alternative to logistic regression. For instance, by having the market share of each alternative in the choice set, Carrier and Weatherford (2014) estimated separate models for four origin-destination pairs based on least squares to study air passenger choice behavior towards different services (i.e. combination of path and class). Similar application was seen in Schafer (2015) which used Berkson-Theil method to analyze the choices among three transportation modes, light-duty vehicles, public surface modes and aircrafts given an aggregated long-term travel data in the US. The practical advantage of Berkson-Theil method when dealing with aggregated data is explained in Ben-Akiva and Lerman (1985).

The situation becomes further complicated if the estimation is based on aggregated data as well as multiple choice sets. Weidner (1996) attempted to investigate the itinerary choices among the observed 7,405 city pairs (i.e. each city pair is a choice set) using logistic regression. However, due to the limit of the software Alogit (ALOGIT, 1990), the author had to select a sample of 380 city pairs to enter into the model. Some other studies also analyzed a limited number of city pairs (i.e. 100 in Ghobrial and Soliman (1992); 62 in Ghobrial and Kanafani (1995); 24 in Atasoy and Bierlaire (2012); 15 in Busquets et al. (2016)) though they did not explicitly say if such decisions were affected by software capabilities. Overall, it showed that the commonly chosen tools such as PythonBiogeme and Alogit etc. could work well with a small number of choice sets. Nevertheless, so far applying logistic regression on a large number of choice sets (i.e. thousands and more) was only observed with using GAUSS, a software that is relatively complicated to use (Cameron, 2001). As a result, only one group of researchers have adopted this tool to study the full choice sets (Coldren et al., 2003; Coldren and Koppelman, 2005; Koppelman et al., 2008). An alternative approach to using logistic regression via GAUSS was seen in Hsiao and Hansen (2011), in which the authors modeled a very large number of choice sets (i.e. 213,917 city pairs) by applying Berkson-Theil method.

At last, Goulias and Kitamura (1993) compared the results generated by Berkson-Theil method to binomial logistic regression using a travel survey data at individual level (i.e. type 1 data). They found that the logistic regression outperforms Berkson-Theil method in their binary data case. However, in another comparison by Carrier and Weatherford (2014) involving type 2 data, the findings showed that Berkson-Theil method produced close results to multinomial logistic regression, though only in terms of the coefficient values and they did not test the prediction performance of the two methods. Apart from these two studies, to our knowledge no other work exists to compare the performance of the two estimation methods especially when having type 3 data to which Berkson-Theil method is mostly needed given its significant practical advantages as explained in section 1.

3. DATA

This paper analyzes a type 3 dataset. A 2015 North America air travel itinerary choice dataset involving a large number of Origin-Destination pairs (OD pairs) is formed by sourcing a variety of information from Sabre's database (Sabre Travel Network, 2016). In this study, the OD pairs refer to the city pairs within North America and an itinerary is defined by a chain of selections involving the origin airport, the destination airport and any connection airport(s) in between.

In the dataset, the choice made for each itinerary is measured by an aggregated passenger demand (i.e. the annual number of passengers) and each itinerary is associated with a number of attributes such as journey time and cost etc. An explanation of all these variables will be provided when specifying the models in the next section. Besides, the original dataset is split by cabin classes so that each itinerary has two sets of the above information with one for the economy class and one for higher classes. Thus, in order to have a single observation for each itinerary, the two sets of information are summed up by using the aggregated passenger demand for each cabin class as a weighting factor.

The dataset also contains a large number of low-demand itineraries. Two selection criteria are therefore applied in order to exclude these less important itineraries while retaining the representativeness of the sample. The annual number of 52 passengers (i.e. the median value) is determined as the minimum demand level of an itinerary and any itineraries that have a lower demand are dropped out. If a city pair has a huge set of itineraries, it is only allowed to keep the top 9 itineraries ranked by passenger demand and all the rest itineraries are excluded. The city pairs that have only 1 itinerary (i.e. no alternative options) are also dropped out at this stage. Eventually, the final complete dataset includes 20,538 city pairs and 102,166 itineraries in total and still captures 95% of the total passenger demand in the original dataset.

Nevertheless, since the logistic regression model will be developed using PythonBiogeme, the complete dataset needs to be broken into sub-datasets each with a small number of choice sets due to the limit of the software as explained earlier. As a result, 5 sub-datasets each with 300 city pairs are formed and will be studied by the two modeling methods. The first dataset contains the top 300 city pairs in the complete dataset based on a rank by the total passenger demand in a city pair (i.e. combining the passenger demand for each itinerary). The second dataset corresponds to the 2nd top 300 city pairs and so on so forth. The descriptive statistics of the 5 sub-datasets are presented in Table 1. The "Top 300" dataset has a much larger range of city pair passenger demand compared to the other datasets.

Table 1: Key characteristics of the datasets

	Complete	**Top 300**	**2nd Top 300**	**3rd Top 300**	**4th Top 300**	**5th Top 300**
No. of city pairs	20,538	300	300	300	300	300
No. of itineraries	102,166	2,145	2,096	2,043	2,116	2,167
Maximum city pair passenger demand	3,469,634	3,469,634	284,567	155,165	105,845	75,750
Minimum city pair passenger demand	106	285,002	156,450	105,983	75,760	57,626

4. MODEL SPECIFICATIONS

The most commonly used approach to study a choice making problem is via developing a logistic regression model estimated under MLE (Ben-Akiva and Lerman, 1985). The model is based on the random utility theory such that a choice maker *n* 's probability of choosing an alternative *i* in a choice set $m = \{1, 2, 3, \ldots, j\}$ is specifically by

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in m} e^{V_{jn}}} \quad (1)$$

where *V* is the estimated utility that a choice maker perceives on an alternative so that in theory, a choice maker will choose the alternative that is associated with his/her highest perceived utility and *V* can be estimated by a function of

$$V_{in} = \sum_{k=1}^{K} \beta_{ik} X_{ink} \quad (2)$$

where X is the explanatory variable and β is the estimated parameter.

As a result, this research uses PythonBiogeme (Bierlaire, 2016) to develop a multinomial logistic regression (MNL) model for each of the 5 datasets. However, in each MNL model, there will be only one generic utility function (i.e. β_k instead of β_{ik} in equation (1) and n is removed from both equations) across choice makers (i.e. passengers), alternatives (i.e. itineraries) and choice sets (i.e. city pairs) given the particular data type in this research.

In contrast to the standard logistic regression approach, Berkson (1953) proposed an alternative solution by transforming a binomial logit regression (BL) model to a regression model that can be estimated under least squares. Such transformation theory was further developed by Theil (1969) to cover the case of more than two choices (i.e. an MNL model).

Considering the itinerary choice problem in this research, Berkson-Theil method defines the following expression for itinerary i in a city pair:

$$\mathrm{Ln}(\mathrm{P_i}) - \mathrm{Ln}(\mathrm{P_j}) \text{ or } \mathrm{Ln}\left(\frac{\mathrm{P_i}}{\mathrm{P_j}}\right) \tag{3}$$

where itinerary j is selected as a base itinerary in this specific city pair. Since the selection is arbitrary (Oum, 1979), the itinerary that has the maximum passenger demand is selected as the base in each city pair. The probabilities of choosing itineraries i and j can be calculated from the available data by taking the ratio of itinerary's passenger demand over the total passenger demand in the city pair. According to equation (1), the expression (3) is equal to $\mathrm{V_i} - \mathrm{V_j}$, the result of which can be calculated by using equation (2).

In summary, the functional form of Berkson-Theil method looks like

$$\mathrm{Ln}\left(\frac{\mathrm{P_i}}{\mathrm{P_j}}\right) = \mathrm{V_i} - \mathrm{V_j} = \sum_{\mathrm{k}=1}^{\mathrm{K}} \beta_{\mathrm{k}} \left(\mathrm{X_{ik}} - \mathrm{X_{jk}}\right) \tag{4}$$

from which the parameter β can be estimated by a least squares regression model and in this research, OLS is chosen as the estimation technique for applying Berkson-Theil method. The models are estimated in Stata (StataCorp., 2011).

Eventually, four explanatory variables are included in the models (Table 2). The number of legs for each itinerary captures the complexity of flight transfers such that a value of 1 indicates a direct flight and higher values refer to more transfers required in this itinerary. Journey fare comes from the average fare paid by passengers on an itinerary in the year of 2015 including taxes. Journey time is the total journey time (i.e. from the takeoff of first flight to the landing of final flight) including connection time. Finally, the frequency measure for an itinerary represents the number of possible flight combinations in the year of 2015 by all airlines on this itinerary. All variables are continuous and are only modeled in a linear form which is simply aimed for a result comparison between the two modeling methods.

Table 2: The variables

Variables	**Units**
Number of legs	no. of times
Journey fare	USD
Journey time	minute
Itinerary frequency	no. of times

Some other variables that may potentially affect itinerary choices are not included due to various considerations. For instance, the annual average delay of an itinerary and the average access times from city centers to both origin and destination airports are found only having minor impacts on passengers' itinerary choices and thus dropped out. The airport commercial revenue, which could be an indicator that may affect the choice of airport, is also excluded since such information is only available for approximately a half of the airports in the complete dataset.

5. MODEL ESTIMATION RESULTS AND PREDICTIONS

The model estimation results by applying Berkson-Theil method and MNL are presented in Table 3.

To compare the results, first of all, the impact signs of the factors are consistent across the two methods in all five models. Specifically, an increase in number of legs, journey fare or journey time would significantly decrease the probability of choosing an itinerary, whereas itinerary frequency is the only factor that is positively associated with the choice of an itinerary. Next, in each model the coefficient values between the two methods are similar but also slightly different. Such differences could be more intuitively revealed by examining the value of time (VOT) measurement, which equals the ratio of journey time's impact over journey fare's impact and indicates the amount of extra cost (i.e. USD in this case) that a traveler is willing to pay in order to exchange for one unit saving in journey time (i.e. one minute in this case). Table 4 shows the comparisons of VOT. In general, logistic regression generates higher VOT than those estimated via Berkson-Theil method in all five models. Particularly, the "Top 300" model estimated by logistic regression gives a much higher VOT of 4.43 USD per minute than the rest cases. One possible cause could be that the "Top 300" model includes some extremely popular city pairs and itineraries (i.e. in terms of passenger demand as shown in Table 1) so that the willingness to pay for time saving might be stronger.

Table 3: The comparisons of model estimation results

		Models				
		Top 300	**2^{nd} Top 300**	**3^{rd} Top 300**	**4^{th} Top 300**	**5^{th} Top 300**
Berkson-Theil Method (OLS)	Number of legs	- 2.5120	- 2.3797	- 1.9445	- 1.7371	- 1.6489
	Journey fare	- 0.0047	- 0.0056	- 0.0056	- 0.0045	- 0.0048
	Journey time	- 0.0029	- 0.0020	- 0.0035	- 0.0021	- 0.0027
	Itinerary frequency	0.0003	0.0003	0.0004	0.0005	0.0006
	Adj r-squared	0.678	0.556	0.497	0.400	0.392

Multinomial Logistic Regression (MLE)	Number of legs	- 2.1800	- 2.8600	- 2.7900	- 2.4500	- 2.3900
	Journey fare	- 0.0023	- 0.0046	- 0.0034	- 0.0016	- 0.0030
	Journey time	- 0.0102	- 0.0043	- 0.0030	- 0.0028	- 0.0027
	Itinerary frequency	0.0004	0.0004	0.0005	0.0005	0.0005
	Adj rho-squared	0.348	0.490	0.463	0.362	0.340
Note: all coefficients are significant at 99% level						

Table 4: The comparisons of value of time

		Models				
		Top 300	**2nd Top 300**	**3rd Top 300**	**4th Top 300**	**5th Top 300**
Berkson-Theil Method (OLS)	VOT	0.62	0.36	0.63	0.47	0.56
Multinomial Logistic Regression (MLE)	VOT	4.43	0.93	0.88	1.75	0.90

Given the similar but slightly different estimation results between the two methods, their predictive powers need to be assessed in order to reveal if Berkson-Theil method can predict the passenger demand for each itinerary as good as logistic regression. The same datasets are used to compare the predicted passenger demand by the two methods and the observed passenger demand. The model estimation results in table 3 are applied as the inputs for equation (2), so that the utility and the probability of choosing each itinerary within each city pair can be estimated. As a result, the predicted passenger demand for each itinerary can be obtained using the observed total passenger demand of a city pair. The coefficient of determination (R^2), which implies the closeness between the predicted passenger demand and the observed passenger demand, is calculated for each of the two methods in the five models.

The comparisons of R^2 show some differences between the predictive powers of the two methods (Table 5). In all five models, logistic regression has better performance in predicting the demand for itineraries, though this method gives a slightly lower R^2 (i.e. 0.725) in one model comparing to the 0.8+ values in the other four models. Such distinctive result in the “Top 300” model might be due to the large range of city pair passenger demand in this dataset (see Table 1), which is the same potential cause for the high VOT as observed above in this model by using logistic regression. In contrast, Berkson-Theil method seems to be less sensitive to such influence as it offers close R^2 across different models.

Table 5: The comparisons of predictive powers

		Models				
		Top 300	**2nd Top 300**	**3rd Top 300**	**4th Top 300**	**5th Top 300**
Berkson-Theil Method (OLS)	R^2	0.621	0.631	0.632	0.612	0.609
Multinomial Logistic Regression (MLE)	R^2	0.725	0.851	0.877	0.845	0.856

Overall, although Berkson-Theil method can predict the passenger demand with good R^2 values (i.e. all above 0.6), logistic regression is still better in terms of providing even more accurate predictions. A potential cause for the lower predictive power of Berkson-Theil method could be the violation of the homoskedasticity assumption for OLS adopted by this research. By recalling from literature, when using Berkson-Theil method to study a type 2 data, Carrier and Weatherford (2014) suspected that the error term could be heteroskedastic when the data is in aggregated form and they recommended using weighted least squares (WLS) estimation instead of OLS. Thus, based on their suspicion and our prediction results, this paper explores the presence of heteroskedasticity by examining the residual distributions.

For the five models estimated by Berkson-Theil method, Figures 1 to 5 are produced accordingly to display the relationships between the residuals (i.e. the differences between the observed demand and the predicted demand) and the predicted passenger demand. Since in general, the residuals become larger for higher passenger demand in all five models, the assumption of homoskedasticity can be seen as being violated. The finding implies that WLS, which can better capture heteroskedasticity, could be a more appropriate estimation technique than the OLS used in this study and might further increase the predictive power of Berkson-Theil method.

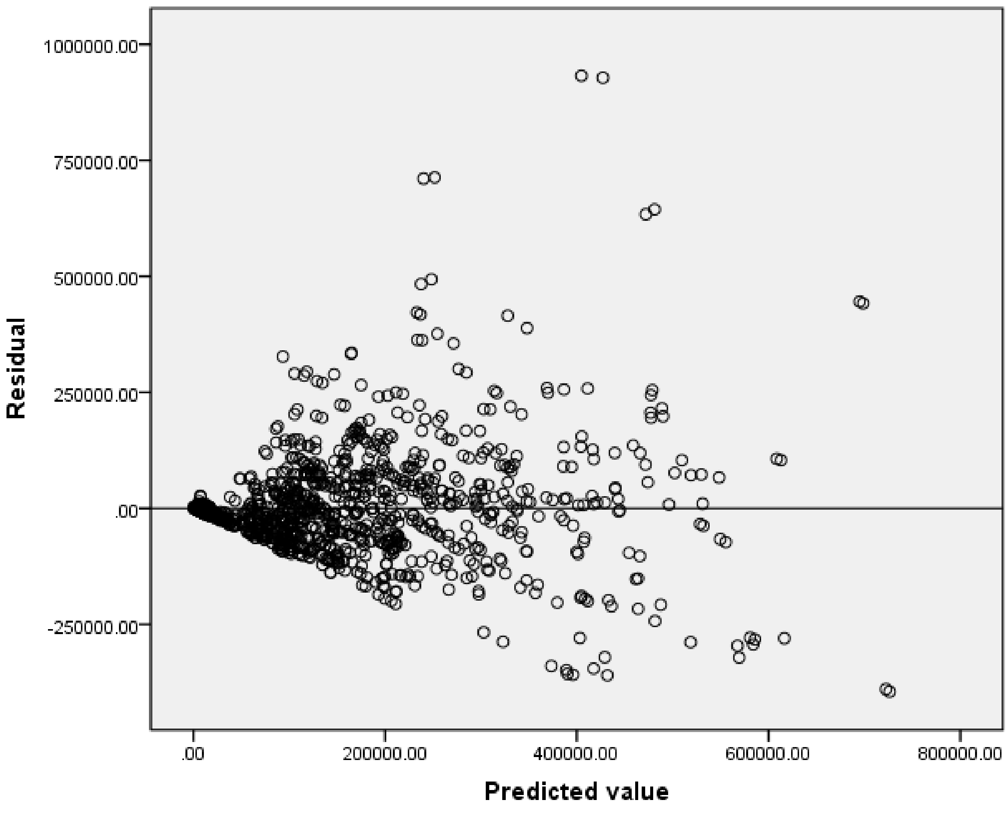

Figure 1: Residual plot for applying Berkson-Theil method on the “Top 300” model

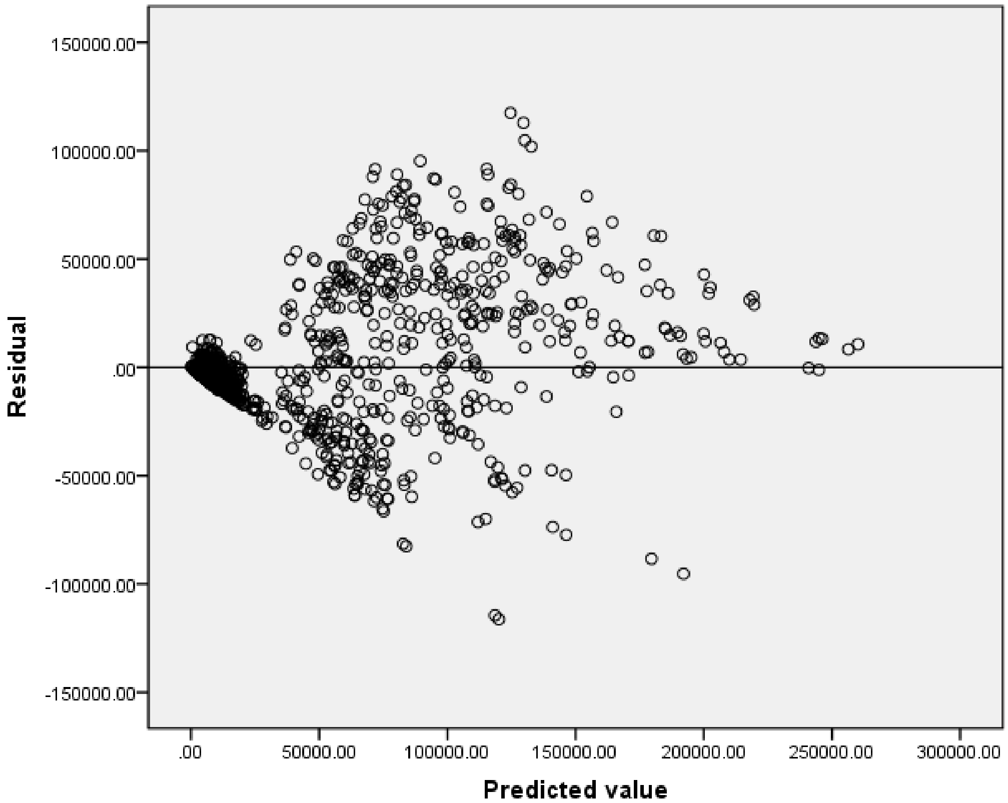

Figure 2: Residual plot for applying Berkson-Theil method on the "2nd Top 300" model

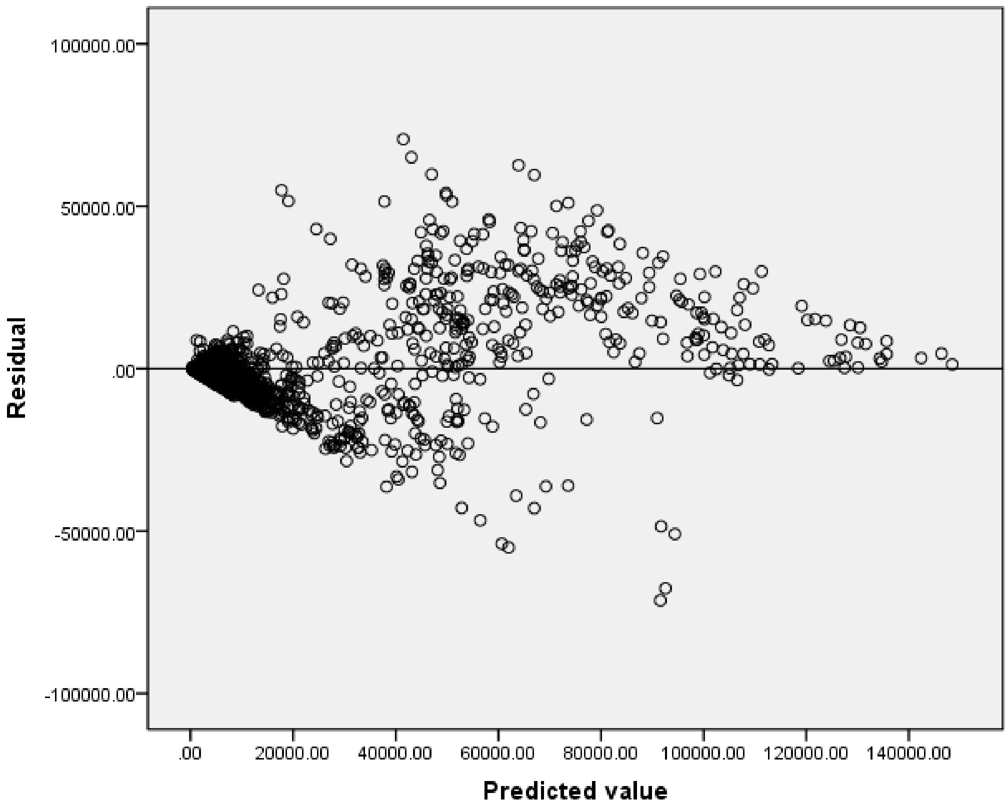

Figure 3: Residual plot for applying Berkson-Theil method on the "3rd Top 300" model

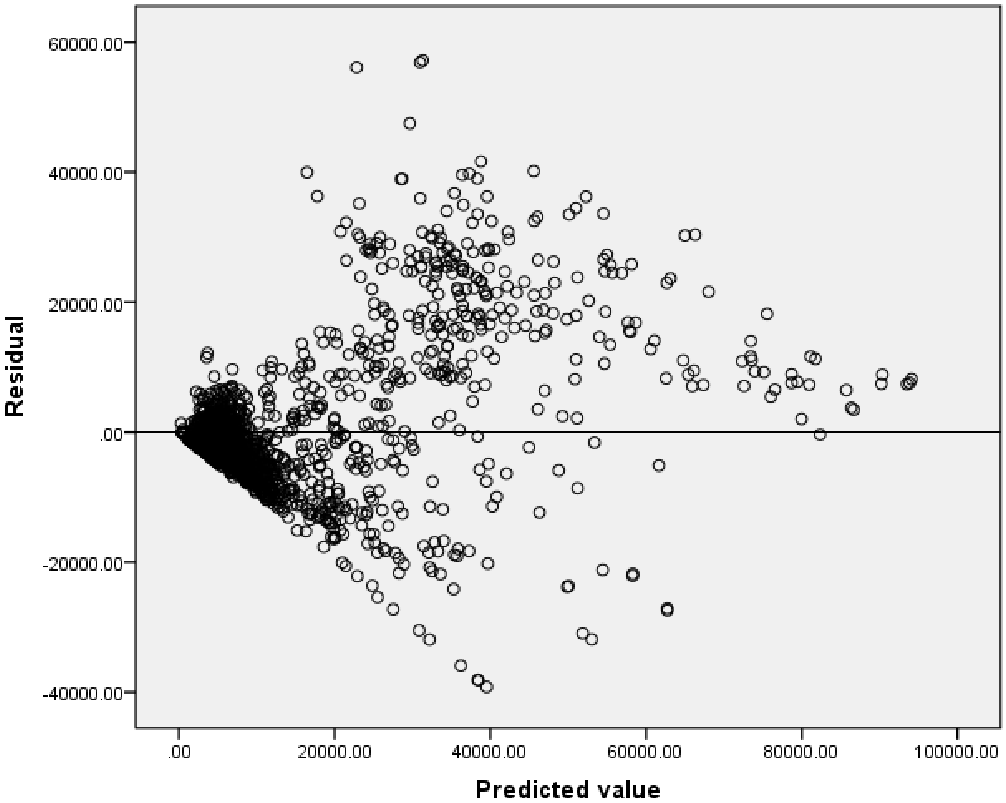

Figure 4: Residual plot for applying Berkson-Theil method on the “4^{th} Top 300” model

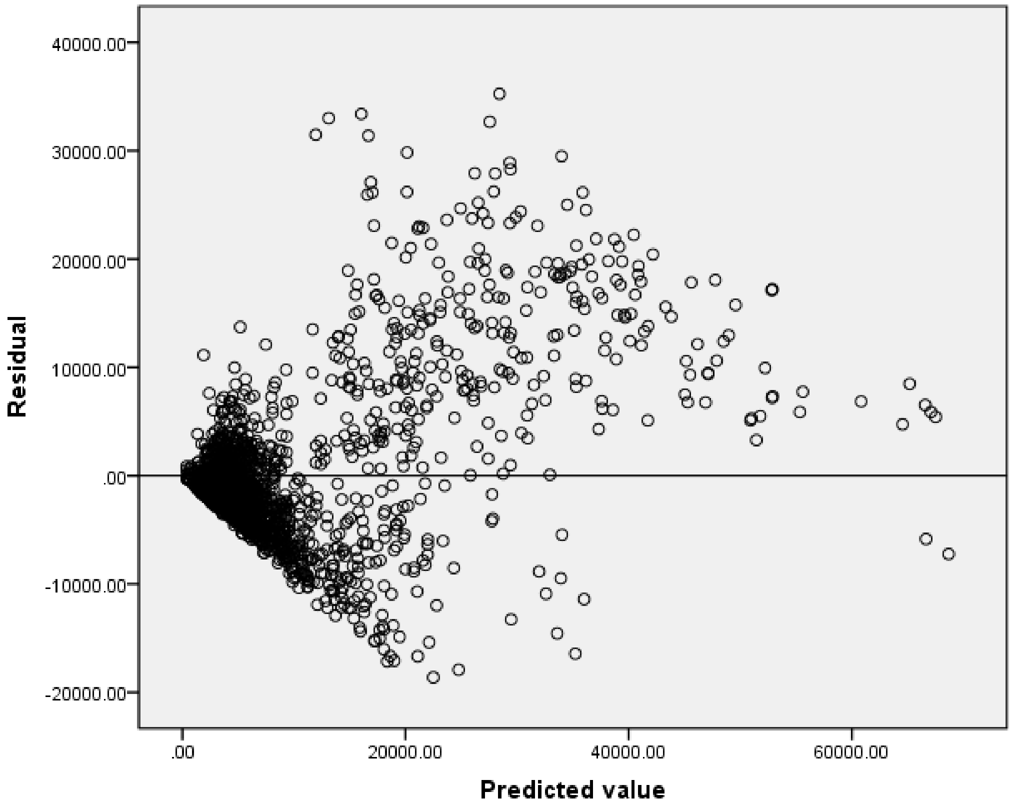

Figure 5: Residual plot for applying Berkson-Theil method on the “5^{th} Top 300” model

6. CONCLUSION

This research compares the model estimation results and the predictive powers between the two choice analysis techniques, Berkson-Theil method and logistic regression, in order to assess the reliability of using Berkson-Theil method to analyze type 3 data (i.e. aggregated choice data and multiple choice sets). The comparisons reveal that the two methods can offer similar model estimation results; however, the results of logistic regression can lead to more accurate predictions. In the end, evidence shows the presence of heteroskedasticity, which, if not being captured, could be a potential cause for the lower predictive power of Berkson-Theil method based on OLS estimation.

To our knowledge, this is the first study that evaluates the performance of Berkson-Theil method on the type 3 choice data which is more difficult to be analyzed (especially when the number of choice sets is large) by directly applying logistic regression on most of the software based on MLE. The findings suggest that Berkson-Theil method can be a good alternative approach to logistic regression; however, it may perform even better if heteroskedasticity can be captured using WLS estimation. Overall, choice modelers in air transportation and other domains that often deal with big data could make use of Berkson-Theil method given its significant practical advantages in terms of handling a large number of choice sets and software compatibility.

Nevertheless, future research is still demanded to provide more robust evidence on the performance of Berkson-Theil method. WLS needs to be applied in real practice to reveal to what extent the predictive power can be increased comparing to the current results based on OLS. Due to software limit, this study only investigates a relatively small number of choice sets (i.e. 300 in each model). However, by using more sophisticated tools such as GAUSS, it is possible to study a much larger dataset (e.g. the one with 20,538 city pairs in this research) in order to offer direct evidence showing the performance of Berkson-Theil method when dealing with type 3 data with a large number of choice sets.

REFERENCES

Adler, T. Falzarano, C. Spitz, G. (2005) Modeling service trade-offs in air itinerary choices. Transportation Research Record: Journal of the Transportation Research Board, (1915), 20-26.

ALOGIT. (1990) Version 3.1. Hague Consulting Group, May 1990.

Aptech Systems, Inc. (2003) GAUSS. Aptech Systems, Maple Valley.

Atasoy, B. Bierlaire, M. (2012) An air itinerary choice model based on a mixed RP. SP dataset. Technical Report TRANSP-OR 120426, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Ben-Akiva, M. Lerman, S. (1985) Discrete choice analysis: theory and application to travel demand, Vol.9. MIT press.

Berkson, J. (1953) A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function. J. Amer. Star. Assn. 48: 565-599.

Bierlaire, M. (2016) PythonBiogeme: a short introduction. Report TRANSP-OR 160706 ,Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Busquets, J.G. Evans, A.D. Alonso, E. (2016) Predicting Aggregate Air Itinerary Shares Using Discrete Choice Modeling. In 16th AIAA Aviation Technology, Integration, and Operations Conference (p. 4076).

Cameron, C. (2001) Introduction to GAUSS on SSDS Unix machines. Available at: http://cameron.econ.ucdavis.edu/gauss/gauss1.html

Carrier, E. Weatherford, L. (2014) Estimating a model of airline passenger choice using grouped booking data and least squares regression. Journal of Revenue and Pricing Management, 13(5), 347-353.

Coldren, G.M. Koppelman, F.S. (2005) Modeling the competition among air-travel itinerary shares: GEV model development. Transportation Research Part A: Policy and Practice, 39(4), 345-365.

Coldren, G.M. Koppelman, F.S. Kasturirangan, K. Mukherjee, A. (2003) Modeling aggregate air-travel itinerary shares: logit model development at a major US airline. Journal of Air Transport Management, 9(6), 361-369.

Ghobrial, A. Kanafani, A. (1995) Future of airline hubbed networks: some policy implications. Journal of Transportation Engineering, 121(2), 124-134.

Ghobrial, A. Soliman, S.Y. (1992) An assessment of some factors influencing the competitive strategies of airlines in domestic markets. International Journal of Transport Economics/Rivista internazionale di economia dei trasporti, 247-258.

Goulias, K.G. Kitamura, R. (1993) Analysis of binary choice frequencies with limit cases: Comparison of alternative estimation methods and application to weekly household mode choice. Transportation Research Part B: Methodological, 27(1), 65-78.

Hansen, M. (1995) Positive feedback model of multiple-airport systems. Journal of transportation engineering, 121(6), 453-460.

Harvey, G. (1987) Airport choice in a multiple airport region. Transportation Research Part A: General, 21(6), 439-449.

Hess, S. Polak, J.W. (2005) Mixed logit modelling of airport choice in multi-airport regions. Journal of Air Transport Management, 11(2), 59-68.

Hess, S. Adler, T. Polak, J.W. (2007) Modelling airport and airline choice behaviour with the use of stated preference survey data. Transportation Research Part E: Logistics and Transportation Review, 43(3), 221-233.

Hsiao, C.Y. Hansen, M. (2011) A passenger demand model for air transportation in a hub-and-spoke network. Transportation Research Part E: Logistics and Transportation Review, 47(6), 1112-1125.

Ishii, J. Jun, S. Van Dender, K. (2009) Air travel choices in multi-airport markets. Journal of Urban Economics, 65(2), 216-227.

Koppelman, F.S. Coldren, G.M. Parker, R.A. (2008) Schedule delay impacts on air-travel itinerary demand. Transportation Research Part B: Methodological, 42(3), 263-273.

Loo, B.P. (2008) Passengers' airport choice within multi-airport regions (MARs): some insights from a stated preference survey at Hong Kong International Airport. Journal of Transport Geography, 16(2), 117-125.

Nason, S.D. (1981) The airline preference problem: an application of disaggregate logit. In AGIFORS Proceedings.

Nassiri, H. Rezaei, A. (2012) Air itinerary choice in a low-frequency market: A decision rule approach. Journal of Air Transport Management, 18(1), 34-37.

Oum, T.H. (1979) A warning on the use of linear logit models in transport mode choice studies. The Bell journal of economics, 374-388.

Pathomsiri, S. Haghani, A. (2005) Taste variations in airport choice models. Transportation Research

Record: Journal of the Transportation Research Board, (1915), 27-35.

Pels, E. Nijkamp, P. Rietveld, P. (2001) Airport and airline choice in a multiple airport region: an empirical analysis for the San Francisco Bay Area. Regional Studies, 35(1), 1-9.

Pels, E. Nijkamp, P. Rietveld, P. (2003) Access to and competition between airports: a case study for the San Francisco Bay area. Transportation Research Part A: Policy and Practice, 37(1), 71-83.

Proussaloglou, K. Koppelman, F. (1995) Air carrier demand. Transportation, 22(4), 371-388.

Sabre Travel Network. (2016) Sabre GDS. Sabre Travel Network, Southlake.

Schäfer, A.W. (2015) Long-term trends in domestic US passenger travel: the past 110 years and the next 90. Transportation, 1-18.

StataCorp. (2011) Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Theil, H. (1969) A Multinomial Extension of the Linear Logit Model. Int. Econ. Rev. 10: 251-259.

Theis, G. Adler, T. Clarke, J.P. Ben-Akiva, M. (2006) Risk aversion to short connections in airline itinerary choice. Transportation Research Record: Journal of the Transportation Research Board, (1951), 28-36.

Warburg, V. Bhat, C. Adler, T. (2006) Modeling demographic and unobserved heterogeneity in air passengers' sensitivity to service attributes in itinerary choice. Transportation Research Record: Journal of the Transportation Research Board, (1951), 7-16.

Weidner, T. (1996) Hubbing in US air transportation system: Economic approach. Transportation Research Record: Journal of the Transportation Research Board, (1562), 28-37.

Windle, R. Dresner, M. (1995) Airport choice in multiple-airport regions. Journal of Transportation Engineering, 121(4), 332-337.

Yoo, K. Ashford, N. (1996) Carrier choices of air passengers in pacific rim: Using comparative analysis and complementary interpretation of revealed preference and stated preference data. Transportation Research Record: Journal of the Transportation Research Board, (1562), 1-7.