

预测通道客运分担率的 MNL 模型特性变量选取

王江涛, 马 驹

(西南交通大学 交通运输学院, 四川 成都 610031)

摘要: 基于非集计离散选择模型理论, 结合通道内各种运输方式特性, 给出了在运用 MNL 模型预测通道客运分担率时特性变量的选取原则。分析了影响旅客出行方式选择的出行时间、安全性、舒适性、方便性、费用等因素间的关系以及各因素作为特性变量的适应性, 认为在运用 MNL 模型预测通道客运分担率时应选取出行时间、费用收入比和发车频率作为特性变量, 并给出了各特性变量的取值标定方法。结合旅客出行特征调查数据, 进行了实证分析和各个特性变量灵敏度分析, 得出制定客运专线运营策略时应主要考虑出行时间、费用和出行者收入变化影响的结论。

关键词: 效用理论; MNL 模型; 特性变量; 客运分担率

中图分类号: U293.1⁺3

文献标志码: A

文章编号: 1674-0696(2010)06-0947-04

MNL Model Characteristic Variables Selection on Forecasting Corridor Mode Split

WANG Jiang-tao, MA Si

(School of Traffic and Transportation, Southwest Jiaotong University, Chengdu 610031, Sichuan, China)

Abstract: Based on the theory of disaggregate discrete choice and characteristics of different transportation mode in transportation corridor, the selection principle of utility function explanation variables of MNL model was determined. According to the selection principle, relationship among travel mode split influencing factors such as travel time, safety, comfort, convenience, cost and adaptability of these factors as characteristic variables were analyzed in the paper; travel time, departure frequency and relative cost (cost divide by income) should be considered as characteristic variables when forecasting transportation corridor mode split by MNL model; meanwhile, corresponding calibration method was given. According to travel mode split data, empirical analysis and sensibility of the characteristic variables analysis were carried out by MNL model. The parameters indicated that the influence of travel time, cost and change of income should be taken into consideration while making operation strategies of passenger dedicate line.

Key words: utility theory; MNL model; characteristic variables; mode split

客运分担率是指某种运输方式或线路在同方向各种运输方式或线路中所承担的客运量比例, 是旅客在各种运输方式之间选择的结果, 它表明各种运输方式在通道客运市场所占有的份额。在规划建设客运专线时, 需分析客运专线与通道内高速公路、民航的竞争关系, 估计不同运输方式的客运分担率。

MNL 模型是比较成熟的分担率估计方法, 但是运用 MNL 模型估计通道内各种运输方式的客运分担率时, 特性变量的选取方法及其值标定不统一, 如: 文献 [1]、[2] 选取进出站时间、候车时间、车上时间、总费用作为特性变量; 文献 [3] 选取旅行时间、费用和发车频率; 文献 [4] 选取时间和费用作为特性变量; 文献 [5] 选取运行速度, 单位里程运价和

发班频率作为特性变量; 文献 [6] 中选取用人均花费成本表示的各种运输方式的经济性指标, 用乘客利用该运输方式的旅行时间(不包括市内交通耗费的时间)表示的快速性指标和用市区内交通走行时间与候车时间之和表示的方便性指标作为特性变量。

笔者根据效用理论和通道内各种运输方式特性, 给出了在运用 MNL 模型预测通道客运分担率时应选取出行时间、费用收入比、发车频率作为特性变量, 同时给出了相应的标定方法; 结合西兰通道旅客出行特征调查数据, 选取上述特性变量进行了实证分析和灵敏度分析, 得出在客运专线的运营策略的制定时应主要考虑出行时间, 费用和出行者收入变

化的影响。

1 理论基础

多项模型(Multinomial Logit Model, MNL)是常用的非集计模型之一,表达形式如式(1):

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

式中: P_{in} 为出行者 n 选择第 i 种运输方式的概率; V_{in} 为出行者 n 选择方案 i 的效用函数固定项; A_n 为通道内可供旅客选择的运输方式种类的集合。其中效用函数为:

$$V_{in} = \theta X_{in} = \sum_{k=1}^l \theta_k X_{ink} \quad (i \in A_n) \quad (2)$$

式中: $\theta = (\theta_1, \dots, \theta_l)$ 为未知的参数向量; $X_{in} = (X_{in1}, \dots, X_{ink}, \dots, X_{inl})^T$ 是出行者 n 的选择方案 i 的影响因素向量 $(X_{in1}, \dots, X_{ink}, \dots, X_{inl})$, 其中 $X_{ink} = 1$ 表示的是常数项, 由该运输方式的技术经济特征确定, 因此只标定效用系数 $\theta_1, \dots, \theta_l$ 就可以预测出客运分担率。

2 特性变量的选取

2.1 特性变量选取原则相关问题

1) 效用函数中的特性变量有不同的形式, 可以是各种运输方式的特性或出行者的特性, 也可能是描述运输方式特征与出行者特征之间相互关系的变量。

2) 旅客出行调查数据中特性变量的标定值必须因起讫点的不同而不同, 否则将导致效用函数的参数无法估计。

3) 模型中不能包含线性相关的变量, 否则变量参数无法估计。当模型中包含线性相关变量时解决方法见文献[7]。

4) 特性变量之间应相互独立。

5) 在标定特性变量的值时不必进行量纲的统一。下证:

MNL 模型中 $\theta = (\theta_1, \dots, \theta_l)$ 的值是式(3)得出[8]:

$$\frac{\partial L}{\partial \theta_k} = \sum_{n=1}^N \sum_{i \in A_n} X_{ink} \left(\delta_{in} - \frac{e^{\theta X_{in}}}{\sum_{j \in A_n} e^{\theta X_{jn}}} \right) = 0, \quad (3)$$

$$\delta_{in} = \begin{cases} 1, & \text{当出行者 } n \text{ 选 } i \text{ 时} \\ 0, & \text{当出行者 } n \text{ 不选 } i \text{ 时} \end{cases}$$

对 $X_i = (X_{i1}, \dots, X_{ki}, \dots, X_{li})^T$ 进行统一量纲得 $\alpha_i^T X_i$, 代入式(3)得:

$$\sum_{n=1}^N \sum_{i \in A_n} X_{ink} \left(\delta_{in} - \frac{e^{\theta_1 \alpha_i^T X_{in}}}{\sum_{j \in A_n} e^{\theta_1 \alpha_j^T X_{jn}}} \right) = 0 \quad (4)$$

令 $\theta = \theta_1 \alpha_i$ 即可得到式(3)。由此可得对特性变量标定时进行量纲的统一只会影响效用函数的固定项 $\theta = (\theta_1, \dots, \theta_l)$ 的解, 而不会对客运分担率的结果产生影响。

6) 选择肢固有常数项变量和个人特性变量的总数取 A_n 集合中选择肢的数减去 1, 即为 $n-1$ 。

7) 与选择肢明显无关的个人特性变量或选择肢的固有变量只可用于其相应的选择肢。

2.2 特性变量的选取分析

特性变量 X_{ink} 可以分为选择肢特性变量和个人特性变量, 前者又可以分为选择肢固有常数项、选择肢固有变量和共同变量。

影响旅客出行方式选择的主要因素有时间(速度)、票价、方便性、拥挤度、舒适性和安全性等。

2.2.1 安全性

各种运输方式的安全性由各种运输方式的安全率等指标来表示, 不因起讫点的不同而不同, 安全性不宜作为特性变量出现在效用函数中。

2.2.2 方便性

方便性通常以间接旅行时间度量。间接旅行时间是指旅客前往乘车点的走行时间与候车时间。对于不同运输方式其间接旅行时间的标定值不因起讫点的不同而不同, 因此, 方便性不宜单独作为一个特性变量进入效用函数, 解决方法如下:

出行时间一般由式(5)标定:

$$T_{OD}^i = \frac{L_{OD}^i}{V_i} + 2t_{\text{到站}}^i + t_{\text{上车}}^i + t_{\text{下车}}^i + t_{\text{候车}}^i \quad (5)$$

式中: $t_{\text{到站}}^i$ 为选择第 i 种运输方式的旅客到达、离开车站的平均时间; $t_{\text{上车}}^i$ 、 $t_{\text{下车}}^i$ 、 $t_{\text{候车}}^i$ 分别为选择第 i 种运输方式的旅客上车、下车、候车的平均时间; L_{OD}^i 为 OD 之间的距离; V_i 为第 i 种运输方式速度。

从式(5)可以看出标定出行时间时已考虑方便性。那么基于式(5)标定出行时间时, 效用函数特性变量可不再选取方便性。

2.2.3 舒适性

现有舒适性的标定方法有以下 3 种: ①用旅客恢复疲劳所需时间来反映, 而恢复疲劳所需时间与乘车时间和乘车环境相关, 由此得出舒适性与出行时间不独立, 舒适性不能与出行时间同时出现在特性变量中, 否则由各种运输方式出行时间的差异导致分担率结果出现误差; ②按票价乘以一个系数得出, 结合出行时间和费用可得舒适性与费用不独立; ③用人均占有面积和振动频率来描述, 这样出现与上述安全性相同的问题, 因此基于上述标定的舒适

性不能出现在特性变量中。

2.2.4 发车频率

发车频率的值按各种运输方式不同起讫点实际的日发车频数来标定。

2.2.5 出行时间和费用

出行时间的标定见式(5)。费用的标定公式为各种运输方式的票价率乘以各城市间距离来确定旅客出行费用,用式(6)表示:

$$C_{od}^i = \beta_i L_{od}^i \tag{6}$$

式中: C_{od}^i 为某一 OD 对之间第 i 种运输方式的费用; β_i 为第 i 种运输方式的票价率。

结合式(5)、式(6)可得,对于同一种运输方式其出行时间和费用高度相关,但是出行时间和式(6)标定的费用不能同时作为效用函数的特性变量,笔者认为在标定费用时应以实际票价为其标定值,而由于各种运输方式内部又有不同等级的票价,因此在问卷调查时应包含旅客选择了哪种等级票价。例如:西安到兰州既有线开行普通旅客列车和快速旅客列车,且其票价有硬座、软座、硬卧和软卧之分,若在进行问卷调查时旅客 A 选择既有铁路作为其出行选择,同时调查 A 会选择哪种列车的哪种等级的票价。这就可以解决费用的标定时与出行时间高度相关。

2.2.6 收入

基于效用理论,收入为描述出行者特性的一个重要因素,因此收入应作为个人特性变量出现在特性变量中,同时必须满足 2.1 中第 6)、7)条所述。

2.2.7 费用收入比

依据 2.1 中第 1)条,考虑到出行费用对不同收入出行者的影响是相对的,可选取费用/月收入代替费用、收入变量。

基于上述分析,在运用 MNL 模型计算通道内各种运输方式的客运分担率时,可选取出行时间、发车频率、费用、收入(或费用收入比)作为特性变量。

3 实例分析

3.1 选取时间、发车频率、费用、收入作为特性变量

结合西兰通道旅客出行特征调查数据,选取出行时间、发车频率、费用、收入作为特性变量,使用 TransCad 软件计算,输出的数据统计报告如表 1。

表 1 数据统计分析报告

Tab.1 Analysis report of parameters			
变量	E	标准误差	t 检验值
客运专线	-0.698 708	0.147 688	-4.730 972
常数项 高速公路	-0.670 753	0.158 984	-4.219 012
航空	-4.030 526	0.382 330	-10.542 008

(续表 1)

变量	E	标准误差	t 检验值
客运专线	0.026 067	0.008 800	2.962 249
收入 高速公路	0.022 359	0.009 749	2.293 608
航空	0.042 072	0.010 366	4.058 480
出行时间/h	-0.351 621	0.035 511	-9.901 669
出行费用/元	0.041 579	0.014 143	2.939 907
发车频率	-0.001 700	0.000 932	-1.824 880
LL(zero)		-3 311.857 229	
LL(end)		-2 393.214 306	
-2 [LL(zero) - LL(end)]		1 837.285 845	
McFadden 系数		0.277 380	
修正极大似然比		0.274 662	

表 1 中出行费用的参数应该是负值,发车频率的参数应为正值,而估计出的参数值不符合,选取费用单独作为该模型的特性变量在该例中不可取。

3.2 选取时间、发车频率、费用收入比作为特性变量

结合 2.2.7 条,选取出行时间、发车频率、费用收入比作为特性变量,使用 TransCad 软件计算,输出的数据统计报告如表 2。

表 2 数据统计分析报告

Tab.2 Analysis report of parameters			
变量	估计值	标准误差	t 检验值
客运专线	-0.076 524	0.105 425	-0.725 859
常数项 高速公路	0.081 992	0.084 718	0.967 813
航空	-1.203 365	0.186 310	-6.458 927
出行时间/h	-0.566 073	0.024 620	-22.992 065
费用收入比	-0.080 002	0.006 724	-11.898 769
发车频率	0.004 898	0.002 592	1.889 521
LL(zero)		-3 311.857 229	
LL(end)		-2 327.954 787	
-2 [LL(zero) - LL(end)]		1 967.804 884	
McFadden 系数		0.297 085	
修正极大似然比		0.295 273	

分析表 2 得,各特性变量 t 统计量的绝对值都大于 1.0,上述特性变量对该模型具有显著的解释能力;McFadden 系数在 0.2~0.4 之间^[9],为非常好的结果。

将表 2 中对应的参数值,代入式(2)得通道内 4 种运输方式的效用函数:

$$\begin{cases} V_t = -0.566\,073 \times T + 0.080\,002 \times C - 0.004\,898F \\ V_{hstr} = -0.566\,073 \times T + 0.080\,002 \times C - 0.004\,898F - 0.076\,524 \\ V_{fw} = -0.566\,073 \times T + 0.080\,002 \times C - 0.004\,898F + 0.081\,992 \\ V_{air} = -0.566\,073 \times T + 0.080\,002 \times C - 0.004\,898F - 1.203\,365 \end{cases} \tag{7}$$

式中: $V_t, V_{hstr}, V_{fw}, V_{air}$ 分别表示既有铁路、客专、高速公路、航空等运输方式的效用值; $T、C、F$ 代表出行时间、费用收入比、发车频率。

通道客运分担率结果如表 3。

表 3 西兰(西安—兰州)通道客运分担率预测结果
Tab.3 Prediction results of Xi'an ~ lanzhou transportation corridors mode split

	出行距离/ km	各种运输方式分担率			
		既有铁路	客运专线	高速公路	航空
西安—宝鸡	173	0.2 265	0.3 709	0.3 599	0.0 427
宝鸡—兰州	503	0.1 402	0.5 427	0.1 761	0.1 410
西安—兰州	676	0.1 013	0.6 160	0.1 126	0.1 702

分析表 3 表明,客运专线将吸引大量的旅客,其分担率随距离的增加显著增加,印证了客运专线的优势是中长途旅客运输。西兰客运专线的修建必将极大缓解陇海线的客运压力。

表 4 客运专线各属性敏感度分析结果
Tab.4 Analysis of parameters sensitivity of passenger dedicate line

	$T_{hstr}(-1\%)$	$T_{hstr}(+1\%)$	$C_{hstr}(-1\%)$	$C_{hstr}(+1\%)$	$F_{hstr}(-1\%)$	$F_{hstr}(+1\%)$
$E(\text{train})$	-3.600 29	3.620 458	-0.917 17	0.919 774	0.014 34	-0.014 35
$E(\text{hstr})$	2.244 565	-2.257 14	0.571 803	-0.573 42	-0.008 94	0.008 94
$E(\text{fw})$	-3.600 29	3.620 458	-0.917 17	0.919 774	0.014 34	-0.014 35
$E(\text{air})$	-3.600 29	3.620 458	-0.917 17	0.919 774	0.014 34	-0.014 35

分析表 4 得,发车频率(F_{hstr})的需求弹性值比出行时间(T_{hstr})、费用收入比(C_{hstr})的需求弹性值小很多,说明出行者在选择客运专线时对出行耗时间和费用收入比的变化更敏感,而发车频率的变动和对出行需求的影响相对较小。因此在客运专线的运营策略的制定时应主要考虑出行时间,费用和出行者收入变化的影响。

4 结 论

在运用 MNL 模型预测通道客运分担率时,由于舒适性、方便性等因素与出行者个人主观认知有关,难以准确标定,现有标定方法不能满足特性变量的独立性,笔者认为应选取出行时间、发车频率和费用收入比作为特性变量,能更有效的反映个人出行选择意愿,运用 MNL 模型计算通道客运分担率的结果更为准确。

运用 MNL 模型时也可以选取年龄和性别作为个人特性变量,但它们对模型的解释作用不大,所以一般情况不予考虑,即使选取它们作为特性变量,其 t 检验值也不能满足要求。在计算时把上述变量的标定值基本统一在一个数量级上,这样可以减小预测误差。

3.3 灵敏度分析

当某种运输方式的运行速度、票价和发车频率等供给属性变动时,其他运输方式的需求受到影响。第 i 种出行方式的需求弹性定义为^[3]:

$$E_{x_{jm}}^{p(i)} = - \frac{\Delta P_k(j) x_{jm} \theta_m}{\Delta x_{jm}} \tag{8}$$

式中: $\Delta P_k(j)$ 为第 j 种出行方式被选择概率的变化幅度; x_{jm} 为第 j 种运输方式的第 m 个属性值; Δx_{jm} 为该属性值的变化幅度。当 $i = j$ 时,计算得到的为直接弹性;当 $i \neq j$ 时,计算得到的为交叉弹性。

利用式(8)分析客运专线的各属性变化 1% 时出行需求的影响,其计算结果如表 4。

参考文献:

[1] Hensher D A, Bradley M. Using stated response data to enrich revealed preference discrete choice models [J]. Marketing Letter, 1993, 4(2) : 139 - 152.

[2] 张戎,吴晓磊,张天然. 基于 RP /SP 融合数据的沪杭客运通道公铁客流分担率研究 [J]. 铁道学报,2008,30(3) : 7 - 13.

[3] Park Y, Ha H K. Analysis of the impact of high - speed railroad service on air transport demand [J]. Transportation Research: Part E, 2006, 42(2) : 95 - 104.

[4] 李军,朱顺应,王红,等. 基于 RP_SP 联合数据的非集计模型应用研究 [J]. 铁道科学与工程学报,2007,4(1) : 87 - 90.

[5] 杨志振,王璐,蒋永雷. 高速客运通道上多种运输方式的分担率分析 [J]. 大连海事大学学报,2008,34(4) : 154 - 157.

[6] 聂伟,韩彪. 都市圈内运输通道客流分担率推算模型探讨 [J]. 铁道运输与经济,2008,30(6) : 53 - 60.

[7] 金安. LOGIT 模型参数估计方法研究 [J]. 交通运输系统工程与信息,2004,4(1) : 71 - 75.

[8] Ben-Akiva M, Lerman S R. Discrete choice analysis: theory and application to travel demand [M]. London: The MIT Press Cambridge, 1985.

[9] Daniel M. A method of simulated moments for estimation of discrete response models without numerical integration [J]. Econometrica, 1989, 57(5) : 995 - 1026.