Using New Mode Choice Model Nesting Structures to Address Emerging Policy Questions: A Case Study of the Pittsburgh Central Business District

Zulqarnain H. Khattak 1,*,†, Mark J. Magalotti 2,†, John S. Miller 3,† and Michael D. Fontaine 3,†

1 Center for Transportation Studies, Department of Civil and Environmental Engineering, Thornton Hall D101, 351 McCormick Road, University of Virginia, Charlottesville, VA 22904, USA
2 Center for Sustainable Transportation Infrastructure, 706 Benedum Hall, University of Pittsburgh, Pittsburgh, PA 15213, USA; mjm25@pitt.edu
3 Virginia Transportation Research Council, 530 Edgemont Rd., Charlottesville, VA 22903, USA; John.Miller@VDOT.Virginia.gov (J.S.M.); Michael.Fontaine@VDOT.Virginia.gov (M.D.F.)

* Correspondence: zk6cq@virginia.edu; Tel.: +1-412-608-9346
† All authors contributed substantially to the research article.

Abstract: As transportation activities affect a region’s environmental quality, knowing why individuals prefer certain modes can help a region make judicious transportation investments. Using a nested logit model, this paper studies the behavior of commuters to downtown Pittsburgh who use auto, bus, light rail, walking, and biking. Although statistical measures influence the selection of a nesting structure, another criterion for model selection is the policy questions such models inform. Hence this paper demonstrates how an alternative model structure allows planners to consider new policy questions. For example, how might a change in parking fee affect greenhouse gas emission (GHGs)? The proposed model showed that a 5%, 10% and 15% increase in parking cost reduces GHGs by 7.3%, 9% and 13.2%, respectively, through increasing carpoolers’ mode share. Because the proposed model forecasts mode choices of certain groups of travelers with higher accuracy (compared to an older model that did not consider the model selection criteria presented here), the proposed model strengthens policymakers’ ability to consider environmental impacts of interest to the region (in this case, GHGs). The paper does not suggest that one nesting structure is always preferable; rather the nesting structure must be chosen with the policy considerations in mind.

Keywords: nested logit model; multinomial logit model; transport policy; travel behavior; econometrics; utility function; systematic term; random term; discrete mode choice; carpool; sustainable transport

1. Introduction

Although one may view a city as a geographical unit, a city itself is a “functional region” defined by a series of networks—trade, communications, and transportation [1]. Not only is the environmental health of the region influenced by actions taken [2], but the very definition of what constitutes a region rests in part on the connectedness of the transportation network. Recognizing that transportation impacts both the natural world (e.g., air, noise, and water quality) and the human environment (e.g., community cohesion, economic development, and property values) [3], it is appropriate to evaluate these impacts when considering potential changes to the transportation system. For example, in the U.S. alone, $4 trillion were forecast to be spent to accommodate new commercial and residential development over a quarter century—but more than a $400 billion reduction in these
costs could be achieved with more compact development, which certain modes of transportation support [4]. Failure to consider indirect impacts of transport investments, such as additional land consumption [1], increased barriers for social interaction [3] or vehicle emissions means that a large contributor to a community’s health is ignored. Because these effects may not be easily apparent prior to investment decisions, transportation “models” may be used to help forecast demand over medium to longer periods of time (e.g., 5 to 20 years) and through considering external forces (such as changes in a region’s employment), shorter term factors (e.g., fuel costs), and longer-term initiatives (e.g., increasing transit capacity) [5]. Although a model may take many forms, it is at its essence “a series of mathematical equations that are used to represent how people travel” [6]; in a model, the impacts of a variety of decisions (do nothing at location x, build a new freight hub at location y, and expand the hours of service for transit line z) can be examined before such decisions are implemented.

A model’s scope is not limited to actions dictated from a central authority; in fact, the fragmented decision-making structure of U.S.-based metropolitan planning [2] renders it naïve to assume such an authority can implement policies with no other institutional support. Gossling [7] and Feng [8] differentiate between such investments that entail a directive (also described as “command and control”; examples are law enforcement moving crowds after a sporting event or speed limits) and two other actions that do not require a centralized entity. These are: (1) investments that provide information only (e.g., changeable message signs indicating travel times on parallel routes thereby enabling travelers to reduce delay if they so choose); and (2) market-based approaches that are a combination of the former two categories (e.g., taxes on certain routes and subsidies for certain modes of travel). The role of the model is thus to provide information to decision-makers regarding the costs and benefits of potential investments, and this role can be served in a variety of planning contexts. For example, the model developed herein shows how work trips (relative to other trip purposes) tend to increase bus use more than they increase light rail use. While such a finding could, in limited situations, inform a command-and-control measure (such as the provision of bus-only lanes during peak periods), its relevance is more apt for “soft” policies [7] such as apps allowing travelers to compare this mode to the cost of other modes or market-based policies (e.g., a change in fares during peak periods).

2. The Niche for Discrete Mode Choice Models

Metropolitan residents have the option to choose from different travel modes multiple times throughout the day. Since the cumulative effect of these decisions affects the environment, understanding how travelers choose among these alternatives matters for policymakers. Some factors are specific to a given mode (e.g., travel time, cost, comfort, and reliability); others are specific to the traveler (income, awareness of alternatives, and the reason for the trip). As mode choice is a complex decision making process for each person, a large sample of individual choices are used to extract insights despite seemingly random variation [9]. For example, slower modes might be tolerated to a greater degree by commuters with less disposable income than by wealthier travelers.

Because an individual’s socioeconomic characteristics are a critical part of the decision process, and because some of these data can only be obtained through relatively expensive survey methods, transportation agencies have a vested interest in being able to mine historical survey-based data sets, such as travel diaries. However, new policy questions are emerging with the passage of time that may not have been emphasized just a few years ago when diaries were collected. Hence, there is a need to better understand how alternative model structures can be applied to answer such questions. This paper addresses that need by demonstrating how an alternative discrete choice model structure can answer one such question: how do changes in age and parking cost influence willingness to carpool? While carpooling itself is an old mode, the advent of a new supplier—the transportation network company (TNC)—has increased opportunities to share a vehicle and thus generated new interest in the factors that influence ridesharing. While some of this benefit results from improved
model accuracy, the need to develop a model to address a specific question—rather than presuming one model can answer all questions—is germane. The paper thus has three objectives:

- To determine if altering the nesting structure affects the ability of discrete choice models to address questions of interest.
- To demonstrate the impact of such a revised structure on forecast accuracy.
- To quantify how better models influence estimation of environmental impacts (e.g., how increased vehicle sharing may affect emissions).

The first bullet is significant because new nesting structures can incorporate new variables, and there is an increased understanding that the individual socioeconomic characteristics influence mode choice as much as transportation characteristics [10–12]. The second bullet is of interest because while much literature [13–16] has focused on modeling decisions, less frequently have models been validated for transferability to similar regions [2,17,18]. (This paper assesses the model’s accuracy using a different data set than that used to build the model, thereby helping to provide guidance on transferability elsewhere.) The third bullet is important because impacts on greenhouse gas emissions (GHGs) are increasingly considered in transport investment decisions [19,20]. Thus, assessment of how changes in certain mode shares, such as carpooling, affect GHGs is an integral component of future policy decisions. While the paper uses a data set for an urbanized section of Pittsburgh, Pennsylvania as a case study, the demonstration of the value of new modeling structures for answering new questions is relevant to other municipal locations seeking to maximize the benefit of existing discrete choice data sets.

An extensive amount of literature related to mode choice decisions provides a starting point for considering model refinements [10]. Less attention has been paid to forecast accuracy and using revised model structures to address policy questions of interest.

- Following the advocacy of econometric techniques for discrete choice modeling, some studies employed explicit modeling of latent psychological explanatory variables, heterogeneity, and latent [9,21]. Others have integrated two decisions via a combined mode and departure time choice model [22,23]. Two other studies [24,25] used multinomial logit models to study intercity travel behavior in Libya and choice behavior of concert participants at Taipei Arena, respectively. Wang et al. [26] developed binary and nested logit models to check the extent to which visitors’ individual attributes (e.g., income, travel frequency distance, and home delivery) influence shopping trip mode choices. Subbarao [27] developed a typology of trip chains based on the structure and activity of trips in a metropolitan region of India, and later proposed a nested logit model. Another study [28] used integrated choice and a latent variable model to examine the factors influencing school teenagers’ travel decisions. Gao et al. [29] reviewed how altering public transit networks can improve the performance of mode choice forecasting. Two recent studies [30,31] considered commuters’ willingness to carpool using cross nested model structures. Discrete choice models are not limited to mode selection but include highway safety [32], express delivery service [33], travelers’ willingness to pay for better quality information [34], and economic impacts of disruptions to entire industries [35].

- Forecast accuracy, especially the desire to make models transferable, has been a motivation for discrete choice models in particular [14]; Rossi and Bhat [36] summarize related efforts. Tellingly, however, the authors state that “There is no basis in the research for defining situations in which model parameters are clearly and definitively spatially transferable [36].”

- For the study’s third objective, some authors have developed models for policy questions of interest to non-modelers. Schlaich [37] showed that logit modeling can inform investments made in traveler information; at one location, variable message signs led to a maximum diversion rate of 30%. Kalaee et al. [13] showed that students not in neighborhood schools, students from families with high income, high school students, and female students are less likely to walk or
bike in comparison to other students. Veras and Wang [38] concluded that time factors instead of cost factors are more influential in choosing an electronic toll collection system. Islam et al. [15] showed how mode choice behavior of park and ride users was affected by transit vehicles and transfer time at stations using logit models. A hypothetical hurricane scenario presented to Miami residents showed that although special evacuation buses were the most likely mode choice, other modes could be favored such as taxi (wealthy evacuees) or regular bus (evacuees destined for a hotel) [14]. Incentives for carpooling (e.g., [39–41]) have also received attention; this paper however, contributes to these existing studies by seeking to explicitly consider the carpooling and vanpooling modes as discrete alternatives and comparing the proposed model to an existing model used in practice for forecast accuracy and validation in order to provide guidance on transferability to other regions.

3. Motivation for Considering an Alternative Nesting Structure

The Southwestern Pennsylvania Commission is the regional planning agency serving the Pittsburgh 10-county area [42]. Figure 1 shows the travel demand model developed for the region in 1989, and the structure informed certain policy questions that faced decision makers at that time. For example, the model differentiates between walking to transit and driving to transit. This enables one to evaluate how improvements to pedestrian facilities might influence transit use. One might compare the cost of such improvements to, say, the provision of better parking at such facilities, and thereby determine which initiative can increase transit patronage in the most cost-effective manner. The 2010 survey conducted by the Pittsburgh Downtown Partnership [43] was later used by the Southwestern Pennsylvania Commission to update some features of their model, however the model was not recalibrated and the structure was also left intact.

![Figure 1. Original Mode Choice Model Developed for Pittsburgh.](image-url)

With the passage of time since the development of the original model structure, new information has become available that suggests it may be productive to revise the model structure:

A primary reason is that new policy questions are emerging. One question concerns transportation network companies (TNCs), also known as “technology-enabled mobility service companies,” which provide on-demand transport services or ride hailing [44]. As the availability of TNCs increases, a better understanding of the factors that cause individuals to carpool generally may shed insight into the potential market for TNCs. A related set of questions is whether commuter vans are similar to carpools, are they a competitor to transit services (e.g., [45]), or are they complementary to such services (e.g., [46]). Another question is to what extent may carpooling be affected, if any, by an increase in driverless vehicles. In fact, TNCs have started testing such vehicles in the case study area and
elsewhere [47,48]. For these types of questions, a focus on two modes “carpool” and “vanpool” is of interest. By contrast, because it focused on whether certain managed lanes should allow only vehicles with at least two, three or four people in them, it was logical for the original model to differentiate among the carpool modes (e.g., two, three, or more than four occupants)—and to avoid a separate vanpool mode entirely. A related policy item is that what’s the role of environmental assessments in transportation investments; while this paper considers how GHGs may be affected by use of shared vehicles, decisions such as increased parking rates (to alter mode selection) or rezoning (to alter land development) also affect emissions [49].

A second reason for revising the structure is to increase model transferability. To some degree, demographic changes affect forecasts. For example, the number of persons age 65 and over is forecasted to increase by 78% in the U.S. between 2014 and 2040 [50]. Accordingly, the se socioeconomic variables (such as age, gender, and income) capture individual preference. (Adeel et al. [12] found that females are more likely than males to use personal auto due to comfort, Easton and Ferrari [11] found that boys more than girls prefer to walk or cycle to school, Kalaee et al. [13] found that preference for car travel is affected by income.) Neglecting these factors hampers forecast accuracy, and hence transferability. Of particular interest to this paper, a separate validation data set [18] is fundamental to model development.

4. Data Description

This study used a stated preference survey designed by the Pittsburgh Downtown Partnership [43] in 2009 and conducted in 2010. The emphasis of the survey was the transportation needs of Pittsburgh regional residents who worked in downtown. Data from the survey concerning factors that influence travel decisions were used to update some features of the original model developed in 1989 [51]. The survey was released to each employer’s human resources manager and was made available to individual workers in that company. The survey was distributed to residents using emails and hard copies. The workers and residents were provided with a response time of one month in order to receive most responses. In total, nine districts near the downtown area were selected to participate in the survey, as shown in Figure 2.

Figure 2. Map of Nine Districts included in the Downtown Pittsburgh Survey (Source: Pittsburgh Downtown Partnership Survey, base map © Google Earth [52]).
A total of 6513 surveys were completed during the survey period. The survey data included information on demographic and socioeconomic characteristics, travel purpose, commute distance, parking fee, and many other factors that might influence mode choice such as travel cost, travel time, number of workers in the household, and vehicle ownership. Respondents were asked to select between seven modes of transport (bus, light rail transit, car, walk, biking, carpool, and vanpool) based on various factors affecting their travel decisions. The descriptive statistics for the explanatory variables used to generate the model are provided in Table 1.

### Table 1. Descriptive statistics of explanatory Variables.

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic and Socioeconomic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator Variable expressing Marital Status (1 if married)</td>
<td>0.52</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Variable for Gender (1 if female)</td>
<td>0.62</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Household income ($K)</td>
<td>56</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>Variable for Race (1 if Hispanic)</td>
<td>0.21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Trip related characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable for distance from downtown (miles)</td>
<td>13</td>
<td>0.4</td>
<td>48</td>
</tr>
<tr>
<td>Variable for Purpose of trip (1 if work)</td>
<td>0.86</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Variable for Travel Time (sec)</td>
<td>2340</td>
<td>300</td>
<td>4500</td>
</tr>
<tr>
<td>Parking related characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable for Parking cost (1 if &gt;$30)</td>
<td>0.51</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Household characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of people age 45 and over living in the household</td>
<td>0.43</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

5. Methodology

The modeling of this discrete choice problem for travel modes involves the seven discrete outcomes shown in Figure 3. (Readers interested in modeling details should refer to Appendix A.)

Figure 3. Two-Level Nested Logit Model for Pittsburgh Traveler Mode Choice.

A nested logit model consisting of three final mode choice groups were public transportation modes (including modes available for general public), private transportation modes (including modes that are private property of travelers) and commuter pool modes (consisting of ridesharing options). The data for the 6513 travelers were then split into two groups: 70% of these data points (e.g., 4559)
were used for calibration—e.g., developing the model. The remaining 30% (e.g., 1954 data points) were used for validation and thus never used for calibration.

Multiple nesting structures were considered based on three criteria. First, the authors assessed the extent to which alternative structures demonstrated a competitive relationship among alternatives (within the nest) and hence were representative of true behavior. (For example, a nest that includes walking and biking may be appealing because of the common element of active transport; however, because the lengths of such trips differ substantially, another plausible nest is one that includes both bus and biking because of the common elements of not requiring auto ownership and comparable trip lengths). Second, because many plausible nesting structures exist, the likelihood ratio test (Appendix A) was used to test the significance of selected nest over the rejected one. A third criterion was whether the inclusive value parameters are reasonable.

6. Results and Discussion

6.1. Best Model

Table 2 shows the best model based on calibration from 70% of the data. For example, after evaluating multiple models, the best model includes a parameter of 4.87 for using the drive-alone car mode (if married), and the t-statistic for this parameter, since it is larger than 1.96, suggests this variable is significant. The values of the McFadden pseudo R-squared and adjusted R-squared for summary statistics are greater than 0.1, suggesting that the model provides good explanatory power. Almost all of the defined variables are statistically significant.

Table 2. Results for the Best Nested Logit Model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Specific Constant</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.8)</td>
</tr>
<tr>
<td>Variable expressing Marital Status (1 if married)</td>
<td>-</td>
<td>-</td>
<td>4.87</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Variable for Gender (1 if female)</td>
<td>-1.243</td>
<td>(−3.24)</td>
<td>-1.187</td>
<td>(−2.84)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Variable for household income</td>
<td>-2.326</td>
<td>(−2.85)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Variable for distance from downtown</td>
<td>4.163</td>
<td>(1.72)</td>
<td>3.267</td>
<td>(2.68)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of persons in the household age 45 and over</td>
<td>-</td>
<td>-</td>
<td>-5.745</td>
<td>(−4.32)</td>
<td>-3.459</td>
<td>(−1.73)</td>
<td>-4.568</td>
</tr>
<tr>
<td>Variable for trip purpose (1 if work)</td>
<td>4.894</td>
<td>(3.56)</td>
<td>2.432</td>
<td>(3.24)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Variable for Parking cost (1 if $&gt;30)</td>
<td>-</td>
<td>-</td>
<td>-6.638</td>
<td>(−3.45)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Travel Time</td>
<td>2.435</td>
<td>(2.45)</td>
<td>4.562</td>
<td>(3.67)</td>
<td>-5.631</td>
<td>(−1.89)</td>
<td>-4.353</td>
</tr>
<tr>
<td>Variable for Race (1 if Hispanic)</td>
<td>1.430</td>
<td>(1.80)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Goodness of Fit measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood LL(0)</td>
<td>−4604.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood at convergence LL(β)</td>
<td>−3160.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McFadden’s pseudo R-squared</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted pseudo R-squared</td>
<td>0.308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>2370</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusive Value parameter Public nest</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusive Value parameter Private nest</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusive Value parameter Commuter Pool nest</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
All three inclusive value parameters ($\theta_m$ in Appendix A) are below 1.0, justifying that the three nests are appropriate and indicating that the modes possess some degree of unobserved shared effects. The lowest value of 0.23 for commuter pool, for example, suggests that carpool and vanpool are substitutable for each other and the nesting structure has taken care of the unobserved shared effects that existed between the two modes. (The literature supports at least two different ways of viewing vanpools: as similar to “larger capacity transportation modes” such as buses [53] or as similar to carpools (e.g., [54]) analyzed carpools and vanpools together, with the common feature being their exemption from tolls). Placement of vanpool in a public transportation nest along with bus and light rail resulted in illogical inclusive parameter values, in contrast to the chosen model of Figure 3. For this particular data set, therefore, it appears appropriate to treat vanpooling as more similar to carpooling.

6.1.1. Variables with Intuitive Signs

Most variables have the expected signs. Household income shows a negative coefficient for bus, carpool, and vanpool, suggesting that higher incomes lead to a greater likelihood of driving alone. Travel time is also negative for walking and bicycling in the private nest: as travel time increases, travelers are more likely to substitute a private car for these non-motorized modes. Race also shows a positive coefficient for bus; however, the practical significance of this variable is lower for present-day work since Pittsburgh has a low population of Hispanics (around 2%). Distance from downtown shows a positive coefficient for bus and light rail, suggesting that travelers living farther from the downtown are more likely to choose these modes. Large parking costs, however, reduce the likelihood of using the car, as shown by this negative coefficient.

The signs in the model also can inform policy initiatives. Presently, the lack of street and garage parking during peak hours, for example, coupled with the negative coefficient for parking costs in the model, points to the use of park and ride services originating outside the Central Business District (CBD). Shared driverless vehicles could also be considered as a viable alternative in the future to eliminate the need for parking in the CBD—and thus knowing whether parking cost substantially affects mode split may help forecast the attractiveness of this alternative. (The impact of how individuals might react to an increase in parking cost is the focus of Section 7.)

6.1.2. Variables Requiring Additional Interpretation

For a few variables, however, the signs were different than that provided in some literature—and the reasons for differentiation require some additional interpretation. For example, these data showed that women were more likely to use public transportation than the private auto, which directly contradicts [55]. However, given that Belwal and Belwal [55] found that female travelers preferred private cars due to a “gender disadvantage in developing countries”, it appears that such a condition does not apply to Pittsburgh. The results also showed that married travelers are more likely to use a personal car than other choices and, while Table 2 alone does not provide a definitive reason, the fact that ‘family obligations’ was the second highest reason for individuals indicating they wanted a private car (Figure 4) suggests that these travelers may have needed the vehicle to pick up children from daycare or school. The results also suggest that people over the age of 45 are less likely to use active modes of transportation. The finding however, matches that of Schoner and Lindsey [56] who found that travelers age 40–49 were about two-thirds as likely to walk or bike as persons age 20–29.
6.2. Travel Mode Market Share

Figure 5 shows that the highest market share among the three upper level nests (e.g., $m_1$, $m_2$, and $m_3$) belongs to public transport (44.6%), while personal cars represent the single highest portion of trips (42.5%) among the seven lower level choices ($i_1, i_2,...i_7$). Carpool represents 6.5% of trips followed by walking (4%), bicycling (1.6%), and vanpool (0.8%). The forecasted and observed values revealed similar mode shares; for example, the observed mode shares for carpool was 7.5% (compared to 9% which was forecasted by the model). The mode splits shown in Figure 5 match those elsewhere that offer substantial public transportation options, although American Association of State Highway and Transportation Officials (AASHTO) [57] notes regional factors influence these shares.

6.3. Public Transport Attractiveness

Collectively, respondents indicated that the most important factors contributing to their decision to use the personal car was convenience (48.5%) and family obligations (14.3%). Only 9.7% said they needed their car while downtown and only 4.6% named privacy as a factor. Substantially different results were obtained by considering respondents who chose public transport: saving money was
the most common reason (48.5%), with other reasons cited less: convenience (14.3%), no car (10.6%), reduced stress (9.7%), and less pollution (4.2%). These observations have three implications for policies: first, for auto use, convenience (which may include control over departure and arrival times) is critical; second, cost has a much bigger impact for those choosing transit than those choosing the auto, and third, socioeconomic factors may play a sizeable role (given the 14.3% of persons choosing the private car cited family obligations).

7. Alternative Nesting to Consider Policy Questions of Interest

A question facing decision makers is what factors may affect the decision to carpool rather than drive alone—and what are the environmental consequences of intervention to influence these factors? This question matters as the “supply” of opportunities to share a vehicle may increase in the future. Although the focus herein is parking cost and persons age 45 and over as factors that may affect the decisions to carpool, the approach may be extended to other attributes that also affect willingness to carpool: road tolling, fuel pricing, education levels, job status, household income, and car ownership [39,58]. Parking is a good example both because of its influence on urban density and because it was a primary issue of interest to decision makers in the case study location. Persons age 45 and over was used because the original model omitted this variable and, with the passage of time, it has become clearer that persons age 45 and over is projected to differ substantially by 2040. Table 3 shows how an alternative model structure can help planners consider such emergent policy questions. The original structure supported a different policy question: should the occupancy requirement for managed lanes be 2, 3, or 4? That model suggested a 10% increase in parking cost would increase carpoolers’ mode share by about 1.5%. The new structure, however, suggests the same parking fee increase would raise the carpool mode share by 6.3%. Further, a 10% increase in travelers age 45 and over (a variable not included in the original model) increases the mode share of carpoolers by 3.7%. Because the proportion of persons age 45 and over is forecast to increase from 41% to 51% by 2040 [50], such socioeconomic variables may merit attention as new policy questions arise.

Table 3. Considering Alternative Nesting to Answer Policy Questions of Interest.

<table>
<thead>
<tr>
<th>Model</th>
<th>Scenario</th>
<th>Base Mode Share for Carpool</th>
<th>Scenarios Mode Share for Carpool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Model</td>
<td>10% increase in age 45 and over</td>
<td>12% *</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>10% increase in Parking cost</td>
<td>12% *</td>
<td>13.5% *</td>
</tr>
<tr>
<td>Revised Best Model</td>
<td>10% increase in age 45 and over</td>
<td>9.8%</td>
<td>13.5%</td>
</tr>
<tr>
<td></td>
<td>10% increase in Parking cost</td>
<td>9.8%</td>
<td>16.1%</td>
</tr>
</tbody>
</table>

* Sum of 2 occupant auto, 3 occupant auto, and 4 occupant auto.

The revised model structure can thus answer questions regarding the environmental impact of encouraged vehicle sharing through an increase in parking cost: how would a shift from a drive-alone vehicle to shared vehicles reduce GHGs? For this demonstration, the study further focused on the example of change in parking cost only. Given an average trip length of 11.2 miles to downtown Pittsburgh, average vehicle emissions of 8887 grams (g) of carbon dioxide per gallon, a ratio of carbon dioxide to total GHGs of 0.989, and an average of 21.6 miles/gallon [59], Table 4 shows amount of GHGs based on the 4559 total trips to the downtown area. After scaling this result to reflect annual emissions for downtown travelers (not just those sampled), decision-makers can compare this impact to potential emissions reductions from other policies such as recycling programs and better insulation for low-income housing [19].
Table 4. Assessment of How an Increase in Parking Prices affects Greenhouse Gas Emissions (GHG).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Result</th>
<th>Revised Model</th>
<th>Total Emissions g CO₂eq</th>
<th>Reduction in Emissions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Drive Alone</td>
<td>Carpool</td>
</tr>
<tr>
<td>Baseline</td>
<td>Trips</td>
<td></td>
<td>1922</td>
<td>447</td>
</tr>
<tr>
<td></td>
<td>Emissions (g CO₂eq)</td>
<td></td>
<td>8,953,225</td>
<td>1,041,359</td>
</tr>
<tr>
<td>5% increase in parking cost</td>
<td>Trips</td>
<td></td>
<td>1673</td>
<td>629</td>
</tr>
<tr>
<td></td>
<td>Emissions (g CO₂eq)</td>
<td></td>
<td>7,795,053</td>
<td>2,930,716</td>
</tr>
<tr>
<td>10% increase in parking cost</td>
<td>Trips</td>
<td></td>
<td>1589</td>
<td>734</td>
</tr>
<tr>
<td></td>
<td>Emissions (g CO₂eq)</td>
<td></td>
<td>7,403,670</td>
<td>3,419,945</td>
</tr>
<tr>
<td>15% increase in parking cost</td>
<td>Trips</td>
<td></td>
<td>1440</td>
<td>843</td>
</tr>
<tr>
<td></td>
<td>Emissions (g CO₂eq)</td>
<td></td>
<td>6,709,430</td>
<td>3,987,812</td>
</tr>
</tbody>
</table>

* (Trips)/2; b Shifts to other modes (bicycle, walk, and bus) are not presumed to generate emissions (e.g., while some additional persons use the bus, the increase is not large enough to require additional bus service); c Greenhouse Gas Emissions, measured as carbon dioxide equivalent grams (g CO₂eq).

Table 4 shows how models can aid in policy decisions regarding parking costs in urban areas based on the expected future consequences. Table 4 suggests that by increasing parking costs, travelers prefer shifting to shared carpools, thus reducing net GHGs. A 5% increase in parking cost suggests a reduction of 7.3%, a 10% increase in parking cost suggests 9% reduction and a 15% increase in parking cost suggests a 13.2% reduction in GHGs. These numbers provide the basis for selection of future policies on parking costs to maintain a sustainable environment in dense urban settings.

8. Model Transferability

The strength of discrete choice models is that they explain variation in behavior based on a wealth of individual characteristics [60] (income, gender, urban environment), making it feasible, at least in theory, to transfer such models to other locations [2,17,18] (although some “updating” [36] of the model parameters using local data will be advantageous) [32]. By contrast a model linked to a particular set of transportation analysis zones from whence the data were calibrated will likely not be transferable as no two locations will have the same zonal configuration [2]. McFadden [60] noted discrete choice models, because they do not aggregate data by zone, preserve the “detailed associations between individual circumstances and travel choices”. Thus, Rossi and Bhat [36] indicate that the likelihood of transferability is higher for person-level models than for zone-based models.

While a robust model can explain some variation in a given data set, the model’s practical value depends on whether it can be used for forecasting scenarios of interest. That is, changes in a particular mode share will ultimately translate into a set of behaviors that have short-term consequences (e.g., congestion on bicycle network) and longer-term effects (e.g., consumption of land for new development and the resultant water and sewer costs). Thus, for any model, one must ask how does it perform with a different data set not used to calibrate the model? Hence, the authors determined the accuracy of the revised model, defined as the difference between model forecasts and observations, by using the validation data set (which had never been used for calibration); this approach provides the magnitude of error [18]. For example, consider the 1,954 travelers not used to develop the revised model. Individual 1 chose to carpool. A perfect model would have forecasted the probability of carpooling as 1.0 for this individual, whereas the model forecasted carpooling with a probability of 0.749. The sum of the error terms (such as 0.251 for individual 1), divided by 1954, yielded an error of approximately 23.2% for carpooling for the revised model.

Table 5 shows the error for seven modes for the original model and the revised nesting structure, where lower values indicate more accurate forecasts. Table 5 suggests the revised nesting structure has higher forecast accuracy for six modes common to both models. For instance, the error for carpoolers reduces from 29.6% (from the original model) to 23.2% (for the revised model). For other regions, this exercise suggests that a revised nesting structure can be of value as the modes of interest change:
for instance, if one wants to differentiate between requiring two versus requiring three persons per vehicle, the original nesting structure is preferable. However, if one wants to differentiate between two non-motorized forms of transport (bike and walk) then the revised model is preferable. Table 5 is also encouraging in that for all six modes shown, the revised nesting structure reduced the forecast error. Furthermore, regions interested in transferring the model can also use their local samples to test the forecast accuracy and then update model parameters as suggested elsewhere [2,36].

Table 5. Forecast Error for Seven Modes of Interest.

<table>
<thead>
<tr>
<th>Nesting Structure</th>
<th>Statistical Measure</th>
<th>Bus</th>
<th>Light Rail</th>
<th>Car</th>
<th>Walk</th>
<th>Bicycle</th>
<th>Carpool</th>
<th>Vanpool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Model</td>
<td>Mean absolute deviation, **</td>
<td>0.287</td>
<td>NA</td>
<td>0.381</td>
<td>0.268</td>
<td>NA</td>
<td>0.296</td>
<td>NA</td>
</tr>
<tr>
<td>Revised Model with New Nesting</td>
<td>MAD = ( \frac{1}{n} \sum</td>
<td>0.236</td>
<td>0.210</td>
<td>0.271</td>
<td>0.267</td>
<td>0.230</td>
<td>0.232</td>
<td>0.235</td>
</tr>
</tbody>
</table>

\* \( \epsilon_i = \text{Observed Outcome} - \text{Predicted Outcome} \); \*\* \( n = \text{Number of Observations} \); \( a \) (Travelers using rail and buses for at least a section of their trip); \( b \) (sum of 2 occupant auto, 3 occupant auto, and 4 occupant auto in the original model).

Additionally, the error from Table 5 can be combined with the mode shares of Table 3 to provide a forecast range instead of a point value—thereby conveying a more realistic understanding of any model’s limitations. Table 3 shows that a 10% increase in persons age 45 and over increases the propensity of carpoolers by 3.7%. With an error of 23.2% for this mode (Table 5) and a 95% confidence level, the forecast increase is a range of 3.2% to 4.2%.

9. Conclusions and Application to Planning Practice

Using the over 6000 samples from a case study conducted in Pittsburgh, the paper reports on the development of a proposed model that incorporates, at its inception, both statistical criteria and policy questions.

The proposed model targets a specific question of interest: what factors lead to the decision to carpool rather than drive alone. The model shows that a 10% rise in parking cost increases carpoolers’ mode share by 6.3%—an increase that is four times the amount forecasted by the original nesting structure. Demographics are a critical component of the revised model: a 10% increase in persons age 45 and over, adds almost 4 percentage points to carpoolers’ mode share. Returning to the paper’s second objective, the proposed model appears to increase transferability, as it reduces aggregate forecast error by six percentage points. Finally, the model shows the value of what-if analyses: a 5%, 10% and 15% increase in parking costs reduces net GHGs by 7.3%, 9% and 13.2% through shifting travelers to shared carpools. This type of scenario visioning can facilitate the incorporation of environmental impacts into investment choices.

The paper’s demonstration of how changing the nesting structure affects both (1) the accuracy of the model and (2) the ability to answer certain policy questions of interest contributes to the existing literature. The authors therefore, recommend that new nesting structures should be considered along with statistical criterion during model calibration in order to tailor the model to a wide range of questions that may arise. Such joint consideration of both policy questions and statistical criteria at the same time may take some coordination between modelers and analysts. (For instance, the paper showed that the inclusion of socioeconomic variables in the model matters as much as transportation characteristics since the age distribution in the U.S. is projected in 2040 to differ substantially from what is seen today.)

The paper, therefore, does not suggest that one nesting structure is always preferable; rather the nesting structure should be chosen based on the policy questions at hand. The findings do show some benefits to recalibrating relatively recent data sets for that purpose. Further, because of the importance of age, income, gender, and marital status in the Pittsburgh model, there may be a benefit to collecting additional demographic data, especially with respect to trip chaining, as resources become available.
Acknowledgments: The authors acknowledge the effort of Pittsburgh downtown partnership in conducting the survey.

Author Contributions: All authors contributed substantially to the research article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Justification for the Use of the Inclusive Value Parameter as a Decision Criterion (Nested Logit Model)

To understand why the inclusive value parameter was relevant to model selection, it is appropriate to consider how this parameter influences the model structure. With discrete nominal choices (such as the seven modes $i_1, \ldots, i_7$ in Figure 3), utility functions link the choice outcome to the decision process and follow the form of Equation (A1) for mode $i$ and person $n$. Equation (A1) shows a systematic component $\beta X_{in}$ and a random component $\epsilon_{in}$. The systematic component contains attributes perceived by the modeler (e.g., $X_1$ might be age and $\beta_1$ would be the parameter for age) while the random component accounts for unobserved impacts such as preferences for traveler $n$. Because utility is not deterministic, one can only find the probability of individual $n$ choosing mode $i$. With independent and identical (IID) distribution of the random terms, the probability for such a choice is given by Equation (A2) [9].

\[
U_{in} = \beta X_{in} + \epsilon_{in} \quad \text{(A1)}
\]

\[
P_n(i) = \frac{e^{\beta X_{in}}}{\sum e^{\beta X_{in}}} \quad \text{(A2)}
\]

When the IID assumption is violated, as is the case with the alternatives of carpooling and vanpooling where errors are highly correlated, a nested logit model is preferable as shown in Figure 3. Such models eliminate the problem of unobserved correlation by nesting different alternatives [32]. After assuming the Gumbel distribution for the random term, Equation (A3) can be derived for calculation of the two level nested logit model [9] for person $n$, where $m$ refers to a choice at an upper level (e.g., selecting a particular nest) and $i$ refers to a choice at a lower level (e.g., selecting a particular alternative within a nest).

\[
P_n(i) = P_n(i/m) \times P_n(m) \quad \text{(A3)}
\]

Equation (A3) indicates that nested logit model can be expressed as a product of two probabilities: the marginal choice probability (e.g., the probability of choosing an upper level nest such as public transport) and a conditional choice probability (e.g., the probability of choosing a bus given that the traveler will use a public transport mode). Because there may still be some correlation among the upper level nests (e.g., there may be correlation between public transport and commuter pool due to random error present within lower levels of those nests), the marginal choice probability for $m = \text{public transport, private, or commuter pool}$ is given by Equation (A4).

\[
P_n(m) = \frac{e^{(V_m+\mu_m)}}{\sum_{m' \in M_n} e^{(V_{m'}+\mu_{m'})}} \quad \text{(A4)}
\]

$\mu_m$ represents the product of inclusive value parameter $\theta_m$ and $LS_{i,m}$. $LS_{i,m}$ is estimated from the log of sum of exponents of nested utilities under consideration i.e., $\ln \left[ \sum e^{V_{i,m}} \right]$. The conditional choice probability for the lower level choices ($i_1, \ldots, i_7$) in Figure 3 is given by Equation (A5) and the product of the two choice probabilities provides the required probability for the nested logit model.

\[
P_n(i/m) = \frac{e^{(V_i)\mu_d}}{\sum e^{(V_{i'})\mu_d}} \quad \text{(A5)}
\]

In Equation (A5), $\mu_d$ represents the inverse of inclusive value parameter $\theta_m$ whose appearance implies that all the parameters of the utility are scaled by a common value [10]. Logical values for
$\theta_m$ are between 0 and 1—a fact that proved critical for later evaluating the many potential nesting structures for this data set.

The actual implementation of Equations (A1)–(A5) was performed with the econometric package N-Logit 4.0. The survey results were converted into the proper format for the N-logit package, including the creation of dummy variables and skipping of missing observations to clean the data properly. Because highly correlated variables are problematic [32,61], variables with correlation above 0.85 were removed from the analysis. (In this data set, travel time and cost had a high correlation coefficient of 0.96.) Then, a multinomial logit model was tested, but due to correlation effects of certain alternatives, such as vanpool and carpool, the results were not satisfactory. As an alternative approach, multiple nested logit models were created and tested to develop the best possible model having lower Akaike Information Criteria (AIC) values. The likelihood ratio test was also used to check whether the selected model was superior to the rejected models. The results for each model were tested for their statistical significance, where $t$-statistics larger than 1.96 (corresponding to a $p$-value of 0.05 for a two-tailed test) were viewed as significant. Note that in Table 2, the log-likelihood at zero may appear relatively high (e.g., closer to zero) than expected because some individuals did not have all seven modal choices available to them. Note also that because the bus system is much larger than the light rail system (that runs in a very small portion of the city) [51], approximately four times as many travelers use bus than light rail.

References


