12-1-2009

Development of a Multi-Class Bicyclist Route Choice Model Using Revealed Preference Data

Joseph Broach
Portland State University

John Gliebe
Portland State University

Jennifer Dill
Portland State University, jdill@pdx.edu

Citation Details

Joseph Broach, John Gliebe, Jennifer Dill, "Development of a multi-class bicyclist route choice model using revealed preference data." Presented at the 12th International Conference on Travel Behavior Research, Jaipur, India, December 2009.

Development of a Multi-class Bicyclist Route Choice Model Using Revealed Preference Data

By Joseph Broach, John Gliebe* and Jennifer Dill

Portland State University
School of Urban Studies and Planning
P.O. Box 751—USP
Portland, OR 97207-0751
U.S.A.
Telephone: +01-503-725-4016
Email: gliebej@pdx.edu

* Corresponding author

Paper submitted to the 12th International Conference on Travel Behavior Research
Jaipur, India
December 13-18, 2009

Abstract

Existing regional travel forecasting systems are not typically set up to forecast usage of bicycle infrastructure and are insensitive to bicyclists' route preferences in general. We collected revealed preference, GPS data on 162 bicyclists over the course of several days and coded the resulting trips to a highly detailed bicycle network model. We then use these data to estimate bicyclist route choice models. As part of this research, we developed a sophisticated choice set generation algorithm based on multiple permutations of labeled path attributes, which seems to out-perform comparable implementations of other route choice set generation algorithms. The model was formulated as a Path-Size Logit model to account for overlapping route alternatives. The estimation results show compelling intuitive elasticities for route choice attributes, including the effects of distance and delay; avoiding high-volumes of vehicular traffic, stops and turns, and elevation gain; and preferences for certain bike infrastructure types, particularly at bridge crossings and off-street paths. Estimation results also support segmentation by commute versus non-commute trip types, but are less clear when it comes to gender. The final model will be implemented as part of the regional travel forecasting system for Portland, Oregon, U.S.A.

Key words: bicycle route choice, GPS data, choice set generation, path size logit

1. Introduction

Non-motorized travel options have been largely ignored in regional transportation planning studies in the U.S., where decisions on more resource-intensive investments in highway and transit facilities have been of primary concern. Recently, however, policy-maker interest in sustainable transportation systems and healthier lifestyles has shifted some of the decision-making focus to bicycling and walking and the extent to which the urban travel environment supports these alternative modes.

Travel forecasting models, the “workhorse” analytical tool for regional transportation decision making, have been well-developed to reflect the attributes important to predicting traveler responses to changes in the level of service of highway systems and transit systems, but are startlingly unrealistic in their representation of the pedestrian and bicycling environment. Focusing on bicycling, the practice implemented in all known operational travel forecasting models used in North America has been to assume that riders choose the minimum-distance path between origins and destinations using a fixed travel speed, usually without consideration of network attributes. Congestion effects and other travel environment attributes, such as elevation and the presence of dedicated bike lanes, separated bikeways, or "bike boulevards" are not considered. Moreover, extant models do not differentiate between classes of bicycle riders, nor do they segment the travel market in ways that could identify the viability of bicycling as a mode alternative, other than applying a maximum trip distance criterion.

In Portland, Oregon, in 2007, we collected detailed survey data revealing the actual paths taken by 164 bicyclists over the course of several days, using global positioning system (GPS) tracking devices. The data have been mapped to transportation network facilities, creating an enhanced digital bicycle network map files showing facility types, bike lanes and off-road trails. The GPS data reveal not only spatial paths, but also time-of-day readings at each starting and stopping points and elevation changes. Each participant also provided detailed socio-economic, behavior, and attitude data.

In this paper, we present the results of our development a bicyclist route choice model that will be applied in a regional travel forecasting framework for Oregon Metro, the metropolitan planning organization for the Portland region and a regionally-elected governing body. Metro Council is keenly interested in the capability of the modeling tool to project use of bicycle infrastructure investment alternatives and to derive economic welfare measures from such analysis. The benefits to regional modeling are more accurate estimates of bicyclists’ travel paths and costs, which affect upstream destination and mode choices and enable analysts to answer a more complex set of questions related to urban form and investments in bicycle facilities than existing regional models currently support. To our knowledge this is the first bicycle route choice model to be developed from revealed preference data that were generated through GPS methods, and the first such model to be applied to planning practice in a major metropolitan region in North America.

In the remainder of this paper we review the existing literature on bicycle route choice modeling; describe the GPS and network data we used in model development. We then describe important analytical elements of the work, namely the development of network route choice set algorithms and the adoption of a path-size logit model formulations. We then describe model specification and estimation results, including a final segmentations of decision makers by trip context and by gender. Next, we discuss how this route choice model fits into and will be used in the larger Metro regional forecasting system. Finally, we conclude this paper with an assessment of what we believe to be the salient contributions of this research and suggest possible avenues for further inquiry.

2. Existing Cyclist Route Choice Literature

Most existing work on cyclist route choice consists of small, targeted studies focusing on only a few variables. Sener et al. (2009) provide a comprehensive review of published work. The primary data collection strategies used to date have been binary choice stated preference surveys and recalled paths. Only one, unpublished study known to the authors has used cyclists' observed route choices, although the data used had not been originally collected for this purpose and therefore had some shortcomings (Menghini et al. 2009).

2.1 Stated Preference Studies

Stated preference studies have dominated the literature due to several appealing characteristics. Detailed travel network data are unnecessary. There is no need to solve the formidable problem of generating alternative routes. Model specification and estimation are also simpler due to the clean data and limited number of alternatives. From a policy perspective, the usual advantage of stated preferences for testing rare or nonexistent features also applies.

There are drawbacks to stated preference data for cyclist route choice. The usual technique in these studies is to show respondents a sequence of side-by-side comparisons from which a binary choice is made (see, for example, Krizek 2006, Hunt & Abraham 2007, Tilahun et al. 2007). It is difficult to know how well a participant can map these textual, or occasionally pictorial, representations to her preferences for real facilities. Many salient features of a route are sure to be missing on a piece of paper or computer screen. Also, although the choice set is in a sense controlled, it seems likely that respondents have in mind their own usual routes as points of comparison. Strategic bias is a possibility if participants think responses might influence policy outcomes. None of this is to say stated preference studies are not useful, only that their advantages in execution involve tradeoffs.

Landis et al. (1997) conducted an interesting variation on the typical stated preference method. Participants actually rode predefined alternative routes before evaluating each. There still may be a problem assuming cyclists can evaluate an unknown route in the same way they do a familiar one, but it does promise greater realism.

The stated preference work most comparable with our research are two similar studies of route choice based on web-based surveys (Stinson & Bhat 2003, Sener et al. 2009). Cyclists were provided with a base route and several alternatives with carefully designed characteristics. Mixed Multinomial Logit models were estimated using the stated preference data along with personal characteristics of participants. Taking into account specific data and modeling differences, we found the results to generally agree with our own. More specific comparisons are provided in the model estimation section of this paper.

2.2 Revealed Preference Studies

A handful of revealed preference studies have been undertaken on this topic, but most are limited studies that do not estimate a full route choice model. Most commonly, cyclists have been asked to recall routes. The routes are then compared with pre-selected routes based on shortest paths or other definitions of optimal paths (Aultman-Hall et al. 1997, McDonald and Burns 2001). These studies have the advantage of using actual routes and network data. The ability of cyclists to accurately recall routes is a question, but it may be quite accurate for habitual routes like commute trips. The larger shortcoming of these studies are the limited choice sets and lack of compensatory models.

Menghini et al. (2009) analyzed GPS data for travelers in Zurich, Switzerland, estimating a full route choice model similar to ours. Time and spatial data on 636 chosen routes was extracted from a larger, pooled database of trips. Because these were general GPS data records, not specifically targeting bicyclists, the actual travel mode had to be inferred algorithmically. Personal and trip purpose information were not available. Network data were also limited, and different types of bike facilities were not distinguished.

In addition, choice sets were generated using an exhaustive search algorithm and randomly selecting 20 alternate routes for each trip. The average non-chosen alternative route was 2.4 times the distance of the mean chosen route, with average maximum grades more than twice the mean chosen route. In our sample, routes that circuitous appear unlikely to be considered; however, the network topology may be different given the hilly terrain of Zurich, and it is unclear whether trips were for recreational or utilitarian purposes.

A Path Size Logit (PSL) specification was used similar to the one described below in Section 4. Estimation results showed some agreement with those presented here, but it is difficult to speculate whether the differences relate to data, choice set generation, or the route choice context and cyclist populations. The relatively high rho-squared model fit values, coupled with few significant parameters other than distance, and insignificant path size parameters suggest that the generated choice set may contain too few reasonable alternatives, biasing the parameter estimates.

3. Data Description

This research relies heavily on GPS data collected from collected during March through November 2007, from 162 bicyclists recruited from throughout the Portland metropolitan area. This research also relies heavily on accurate geographic information system (GIS) mapping of an urban street network and off-street bike paths, as well as compilation of attend attribute information regarding facility types, daily vehicular traffic volumes, and elevations. Tying this all together, coding the GPS observations to individual link traversals and compiling link traversals into coherent trip patterns was a painstaking but critical element of this work, which sought to provide the most accurate portrayal possible of the route attributes actually experienced by the survey respondents during their rides. A full report describing the GPS data collection methods and the processes used to prepare the data for our research may be found in the report by Dill and Gliebe (2008). A summary of the important features of the data for the purposes of route choice modeling is provided below.

3.1 Person attributes

GPS participants were outfitted with small hand-held devices, which they clipped onto their bicycles. They were instructed to enter both weather and trip attribute information and to record the beginning and end of a trip, defined by reaching a particular destination. They also completed demographic and attitudinal surveys.

The participants in this study were primarily regular bicyclists, who agreed to participate in the GPS portion of the study following an initial set of telephone interviews. Although regular cyclists are more likely to be male (80 percent according to the phone survey), we were able to recruit a GPS sample composed of 44 percent females. Among all respondents, 89 percent were between the ages of 25 and 64. Compared with the phone survey of bicyclists used to screen and recruit them, the GPS participants were slightly older, were more likely to have a college degree, had higher incomes, and were more likely to have full-time jobs. They were also more likely live in a two-person household, and only 7 percent lived in a household without a car. The phone survey participants had a demographic comparable to the general population.

While participating in the study, GPS respondents made an average of 1.6 bicycle trips per day. Most participants (77 percent) made an average of two or few bicycle trips per day while they had the GPS device. Participants rode an average of 6.2 miles per day. The median bicycle trip distance was 2.8 miles. The vast majority of the bicycle travel recorded by the participants was for utilitarian purposes. Only five percent of the trips were purely for exercise. Aside from riding back home, riding to work was the most frequent trip purpose (25 percent of all trips). About 18 percent of the trips were for shopping, dining out, or other personal business, and 12 percent were for social/recreation purposes (such as going to the movies, the gym, or visiting friends).

3.2 GPS survey records

When the bicyclists were riding for utilitarian purposes, they rode mainly on facilities with bicycle infrastructure. For over half (52 percent) of the miles bicycled on bicycle-only utilitarian trips were made on facilities with bicycle infrastructure, including lanes, separate paths, or bicycle boulevards. Over one-quarter of the mileage (28 percent) occurred on streets (arterials or minor streets) with bike lanes. An equal share (28 percent) of the mileage occurred on minor streets without bike lanes. These are typically low traffic volume, residential streets. Therefore, only 19 percent of the travel was on streets that would be expected to have high volumes of motor vehicle traffic and no separate facility for a bicycle. When asked about their route choices and preferences for utilitarian trips, participants placed highest importance on minimizing distance and avoiding streets with lots of vehicle traffic. Riding on a street with a bicycle lane was usually ranked third in importance, followed by reducing waiting time at stop lights and signs.

Bicycling for exercise purposes followed a different pattern. Exercise trips typically did not include a well-defined destination, tending to be structured as loops, thereby making shortest path comparisons nearly impossible. For such trips, the path followed itself is the "destination," thus a different decision paradigm is clearly being followed. Exercise trips were therefore not used to estimate route choice model parameters.

The average difference in distance between the actual bicycle trip and the shortest path between the same origin and destination was 0.95 miles, though the median was 0.27 miles. The difference between the shortest path and the observed route increases with trip distance. Looking only at the trips 10 miles or shorter in distance, the median difference between the observed route and the shortest path was just under a quarter mile (0.24 miles). This represents about an extra 1.5 minutes of travel, given the average speed on the trips.

Comparing the facilities used for the observed trips to the shortest paths reveals some preferences in facility type. Bicyclists spent a higher share of their miles on facilities with bicycle infrastructure and on low traffic streets than the shortest paths predicted. In particular, they rode 14 percent of their miles on bike facilities, compared with 6 percent of the miles for the shortest paths, a difference of eight percentage points. Arterials and highways that do not have bike lanes represented 19 percent of the bicyclists’ miles, compared with 36 percent of the shortest path miles. This also indicates that the major streets without bike lanes are often part of the shortest path between two points.

3.3 Network structure and attributes

Bicycle infrastructure in the Portland region includes about 550 miles of bike lanes on streets, 130 miles of separate bike paths, and 30 miles of "bicycle boulevards." Bicycle boulevards are low-traffic residential streets, usually running parallel to a major road, that use traffic calming features to give priority to bicycles over motor vehicles. For example, barriers at some intersections force cars to turn while bikes can continue on a through path. Traffic signals allow bikes traveling on the boulevard to cross busy streets safely. The routes are signed and usually connect with other bicycle infrastructure, including lanes and bridge crossings.

The network model developed for this research included more than 90,000 undirected links and 70,000 nodes. This network was constructed to include the various types of bike infrastructure discussed above, including off-street multi-use paths and low-volume residential streets. The bike network did not include facilities that would be off-limits to bicycle travel, such as controlled access highways and freeways.

The City of Portland provided interpolated average daily traffic volumes for nearly all streets in the study area. In addition, we coded elevation data to network nodes, enabling us to calculate elevation changes across a link.

4. Model Formulation

4.1 Overlapping Alternatives

Overlapping routes presumably have correlated errors. This violates the multinomial logit (MNL) model assumption of independently distributed errors across alternatives. From a statistical point of view, an MNL route choice model will tend to assign counter-intuitively high probabilities to routes that share common network links. From a behavioral point of view, we might say that the MNL considers overlapping routes distinct alternatives; whereas, cyclists may consider such routes jointly as minor variants of a single alternative.

There are two options to overcome the overlapping routes problem (Frejinger & Bierlaire 2007). A correction factor may be applied to partially adjust the utilities for overlap and the MNL model retained. Alternatively, more complex model forms may be specified that allow for correlated errors, including the multinomial probit model, mixed logit models, and closed-form members of the generalized extreme value (GEV) class of models.

For simplicity, we chose to retain the underlying MNL specification, and the route utilities were corrected using the path-size factor. The rationale for this assumption of model form was that we recognized the need to be able to apply the model for prediction across a very complex, detailed network. This requirement made the specifications of overlapping route calculations and nest memberships needed for the various probit, mixed logit, and GEV models seem somewhat intractable over such a large computational space.

A path size factor is calculated directly from route alternatives and network data, but avoids direct calculation of correlations across alternatives. The general form for the *j* alternatives in choice set C_n is specified as:

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \left(\frac{L_i}{L_j} \right)^{\gamma} \delta_{aj}} \tag{1}$$

where Γ_i are the links in alternative *i*, l_a is the length of link *a*, L_i is the length of alternative *i*, and δ_{aj} equals 1 if *j* includes link *a* (Frejinger and Bierlaire 2007). The parameter γ is a positive scaling term meant to penalize very long routes in a choice set. Fixing or estimating $\gamma>0$ has been shown empirically to improve route choice model fit (Hoogendoorn-Lanser et al. 2005, Bekhor et al. 2006, Prato & Bekhor 2006, Prato & Bekhor 2007); however, it has recently been shown that $\gamma>0$ can result in questionable utility corrections and illogical path probabilities (Frejinger and Bierlaire 2007). In addition, the choice set generation method described in the next section makes it unlikely that improbably long alternative paths will be included in our analysis. For these reasons, the path-size correction factor in equation 1 is used with $\gamma=0$, essentially dropping the long-path correction factor and yielding the basic Path Size Logit (PSL) model (Ben-Akiva & Bierlaire 1999).

While relatively simple, the PSL model has been shown to perform well relative to more complex model forms such as the cross-nested logit (CNL), although existing comparisons were performed with the generalized PS factor including $\gamma>0$ (Bekhor et al. 2006, Prato and Bekhor 2006, Prato and Bekhor 2007). Although nested logit models should outperform the PSL specification, they are limited in real network applications due to the huge number of parameters that would have to be estimated to exploit their full flexibility (Bekhor et al. 2006, Frejinger and Bierlaire 2007). Some promising work has been presented recently on using sub-network components as an improvement to the PSL which maintains much of its estimation simplicity (Frejinger and Bierlaire 2007). The method has not yet been applied to a real network problem but merits further research attention.

The remainder of this paper presents results obtained from the following specification of the Path Size Logit probability:

$$\Pr(i \mid C_n) = \frac{\exp^{V_{in} + \ln(PS_{in})}}{\sum_{j \in C_n} \exp^{V_{jn} + \ln(PS_{jn})}} \qquad (2)$$

where *PS* is the path size factor from equation (1) with $\gamma=0$. Since *PS* will always fall between 0 and 1, ln(*PS*) will be negative, consistent with a utility reduction proportional to the degree of overlap.

4.2 Other Issues

The original 162 participants were reduced to a usable estimation set for the purposes of route choice modeling, following a battery of data preparation exercises that eliminated riders who rode purely for exercise without a specific destination endpoint. Our estimation dataset therefore includes observations on 154 participants over multiple trips. It is likely that an individual's series of route choices are correlated to some extent. The inclusion of multiple trip purposes and the generally short one to two week periods of observation probably limit such correlation. That is, two trips for different purposes by the same individual are likely to be less correlated than, say, two commute trips. Furthermore, a random investigation of commute trip sequences, which we might expect to be the most regular, showed substantial route choice variation across trips. It did not seem as though these cyclists were "locked in" to a usual route. For simplicity, trips were assumed to be independent and pooled for analysis. An obvious future extension would be to consider different specifications including individual-specific effects.

5. Choice Set Generation

Generating the set of alternative route considered for each trip was the most difficult and time-consuming part of our analysis. The size and density of the Portland bicycle travel network increased the task's complexity, with nearly 90,000 undirected links and almost 70,000 nodes. In addition, the lack of existing revealed preference cyclist route choice studies demanded a careful rethinking of existing generation techniques. In particular, the common algorithms based on travel time and street hierarchy were not directly applicable. After experimenting with three common choice set generation methods, we chose a modified method of route labeling. Routes were chosen by maximizing individual criteria, subject to a flexible, calibrated distance constraint.

5.1 Existing Techniques

Most route choice set generation methods rely on a repeated shortest-path search that modifies either the network attributes or the search function at each iteration. Bovy (2009) provides a recent review. *K-shortest* paths is the technique with the longest tradition in route choice modeling. The shortest path is found, usually based on distance or travel time, and at each iteration either a link of the previous shortest path is removed (link-elimination) or the selected links are penalized (link-penalty). Simulated shortest paths has been proposed as a variation in which link costs are drawn from a specified probability distribution at each iteration (Ramming 2001, Bekhor et al. 2006). The route labeling approach takes a somewhat different approach (Ben-Akiva et al. 1984, Bekhor et al 2006). Instead of modifying network attributes, the shortest path search function is modified at each step to focus on a different attribute or related set of attributes. For instance, one "label" might seek to minimize congested travel time, while another might maximize use of interstates and highways. Distance must be included as a cost in each label function to prevent the algorithm returning unreasonably circuitous paths. Branch and bound search is an exhaustive search that finds all possible routes between a given origin and destination subject to an analyst-specified guidance function. Branch and bound has shown promising results but has been applied only to networks 50 times smaller and less dense than the Portland bike network (Prato & Bekhor 2006). It was not considered for our study.

We initially experimented with *K-shortest* paths, simulated shortest paths, and route labeling choice set generation methods. Multiple variations of each were tried. The resulting choice sets were not satisfactory. In auto route choice models, travel time variation will typically be much greater than variation in distance for parallel routes. For instance, a major arterial is likely to be considerably quicker than a parallel minor street, even if the distance on each is similar. The result does not necessarily hold for bicycle travel. Distance and travel time are practically identical, since speed limits, roadway design, and intersection delays have a much smaller impact on a cyclist's average speed. Combined with our dense street grid, this posed problems for *K-shortest* paths and simulated shortest paths, which tended to choose routes which were only minor variations of one another and with unreasonable numbers of turns.

There also was no straightforward method to ensure the existing algorithms include a reasonable number of alternatives using the bicycle network. Imposing turn penalties or arbitrarily reducing the cost on bike network links improved the results somewhat, but the selected routes still lacked coherency. Route labeling produced more reasonable routes; however, because each label produces only a single route, resulting choice sets often lacked a reasonable number of alternatives and sufficient variation on certain attributes. Out of 1,464 trips, 51 trips had no distinct alternative, and 50 more had only a single alternative, even when the distance weights were relaxed in each label. Poor initial estimation results confirmed the unrealistic nature of the choice sets.

5.2 A Modified, Calibrated Labeling Technique

Building on the relative success of the route labeling method, we modified the existing technique to include a parameterized label functions. Labels shown in below in Table 1 were chosen based on existing research findings and available network data. Most of the labels were selected from our respondents' answers to survey questions about important route choice factors. Maximizing bike lanes, signed routes, and paths as well as minimizing hills, traffic, and intersection delays were at least somewhat important to a majority of participants. The adjacent land-use variables employment and share commercial were included to proxy for driveway access, on-street parking, and parking turnover. These attributes were thought to be important, but we were unable to measure them directly.

For each label, we attempt to generate multiple alternatives by applying a cost function which balances the label attribute with the shortest-distance path. The algorithm starts by initially assigning all the weight to the shortest-distance criterion and thereby finding the shortest distance path. Then, the algorithm incrementally decreases this weight in small increments (e.g., 0.1) and thereby gradually placing more weight on the label attribute. With each decrease in the weight on distance, there is the possibility of finding a new least-cost path that attempts to minimize (or maximize) use of links with the attribute value. When a new least-cost path is found, it is added to the choice set. The process is constrained by setting a lower bound on the minimum value of the cost weight, such that the choice set will include only those paths that deviate from the minimum-distance path by the amounts observed in our data. This is carried out for each label category.

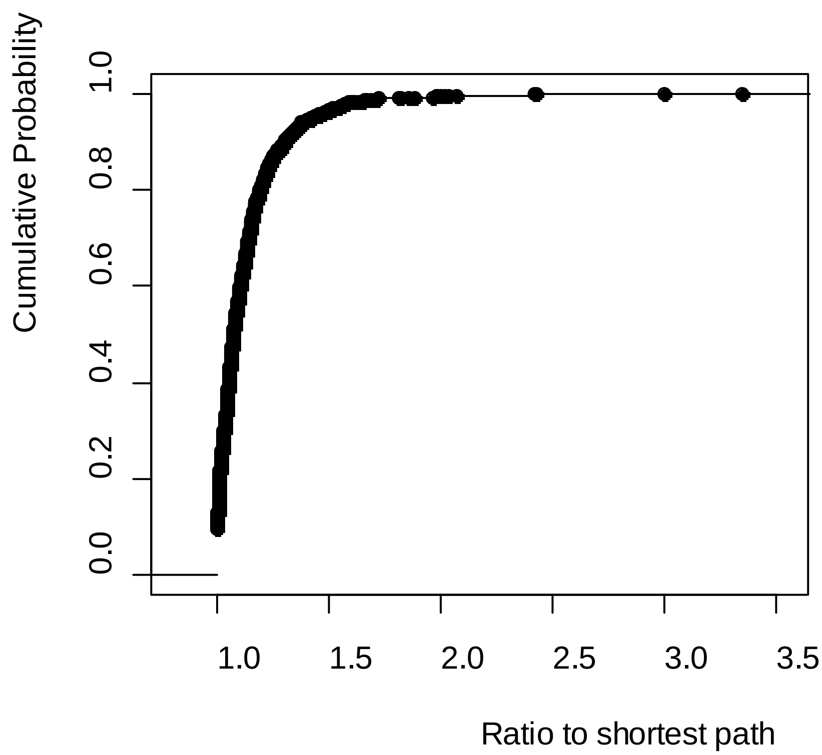
TABLE 1. Attribute Labels

Label	x_i	min(β)
Maximize all bike facilities	link length*(1-bike facility dummy)	0.2
Maximize on-street bike lanes	link length*(1-bike lane dummy)	0.3
Maximize improved and unimproved bike routes	link length*(1-bike route dummy)	0.2
Maximize improved bike route	link length*(1-improved bike route dummy)	0.2
Maximize off-street bike paths	link length*(1-bike path dummy)	0.3
Minimize upslope	(upslope/90th percentile observed travel upslope)*length	0.1
Minimize traffic volume	(AADT/95th percentile AADT)*length	0.1
Minimize stop signs and traffic signals	(stop dummy + signal dummy)*length	0.1
Minimize turns*	left turn dummy*100m + right turn dummy*50m	0.5
Minimize adjacent employment density†	(emp. density/99th percentile emp. density)*length	0.1
Minimize adjacent commercial land use†	commercial share*length	0.1

*turn penalties calculated from observed travel time estimation, converted to distance
†thought to proxy for on-street parking turnover and driveway access frequency

A step-by-step outline of our algorithm is included in an appendix at the end of this paper. This algorithm generates multiple routes for each label and produces different distance weights for each label category, which vary as a function of the way the label attribute is measured. The weights themselves have no meaning other than to prevent the inclusion of alternatives in the choice set with unrealistically high path lengths relative to the shortest-distance path. The range of the distance weight parameters was calibrated to the distribution of observed deviations from shortest path over the entire sample, shown in Figure 1.

As may be seen, cyclists are quite sensitive to detours. Half the trips were less than 10 percent longer than the shortest path, and over 95 percent of observed routes were no more than 50 percent longer. The behavioral rationale for our choice set selection method is that cyclists consider routes that deviate significantly from the shortest path only if those routes are attractive based on optimizing the value of some positive attribute label. The maximum willingness to deviate from the shortest path is a joint function of a route's performance on a given attribute and an unobserved deviation from shortest path preference function. An individual's preference function is assumed to be related to the observed willingness to deviate from shortest paths over the entire sample. If separate data were available, those could be used to generate the deviation preference distribution instead of the estimation sample.



5.3 Choice Set Generation Results

Replication of observed routes is by far the most common measure of choice set generation performance. Results are usually given as the percentage of observed routes "covered" by generated routes given overlap thresholds of 70 to 100 percent (Ben-Akiva et al. 1984, Ramming 2001, Bekhor et al. 2006, Prato & Bekhor 2006, Prato & Bekhor 2007). Prato and Bekhor (2006, 2007) generalized the overlap measure with a "consistency" index. While reproducing observed routes is a necessary condition for reproducing a consideration set, it is not a sufficient one. After all, it is the composition of the full choice set that matters, not only that the observed route is included. For example, a complete enumeration scheme would always maximize observed route overlap, but the resulting choice sets would be of poor quality in most cases.

Although our choice set selection method was not calibrated to maximize observed route replication, Table 2 shows that it generally outperformed existing techniques. For the overlap comparison, we implemented algorithm runs that generated roughly equivalent numbers of alternatives to make the comparisons more meaningful, with the exception of the single-route labeling method. The other three methods generate additional routes, at a decreasing rate, as the number of iterations is increased.

TABLE 2 Generated Choice Set Statistics

		% observed routes replicated at least:					
Method	**Alternatives**	**100%**	**90%**	**80%**	**70%**	**Captives***	**Runtime†**
Proposed	27,626	22.5	29.4	42.3	54.6	15	17h 14m
K-shortest paths	28,089	20.0	24.5	35.8	47.4	0	1h 57m
Simulation	28,029	18.2	21.3	27.9	37.0	0	4h 46m
Labeled routes	10,501	20.4	24.6	35.4	47.1	54	1h 10m

*cases with no alternative (N=1,464)

†runtime on 2.4GHz Intel Core 2 Duo; includes calibration runs for proposed method

Three results in Table 2 merit further attention. First, the single labeled routes technique is clearly the most computationally efficient, suggesting that the idea of attractive routes is useful. The single labeled routes, however, fails to find an alternative for 54 trips. When the label functions were adjusted to reduce the number of captive routes, the quality of choice sets degraded. Second, the proposed method outperforms the other methods on observed route replication, although the improvement is not huge. It also results in a small number of captive routes; however, inspection showed that these were all either very short routes or routes in sparse portions of the network. In such cases travelers may reasonably be captive to the observed route. Third, the computation time for the proposed method is much longer. About half of the time is spent calibrating the algorithm to the observed deviation distribution. In fairness, multiple runs were required to calibrate the other algorithm's parameters as well, but they are not included here to be consistent with existing results. Because the choice set only needs to be generated once, the long runtime is not viewed as unreasonably costly.

Finally, Figure 2 shows the distribution of choice set size for each method. The tendency of *K-shortest* and simulated shortest paths to choose numerous minor variations in our network results in relatively homogeneous choice set sizes. Both labeled route methods result in smoother distributions, which we think are more reasonable. Longer trips and trips in dense parts of the network with many attractive alternatives are likely to lead to more options being considered by cyclists. Subsequent estimation results in this paper use the modified, calibrated labeled routes choice sets.

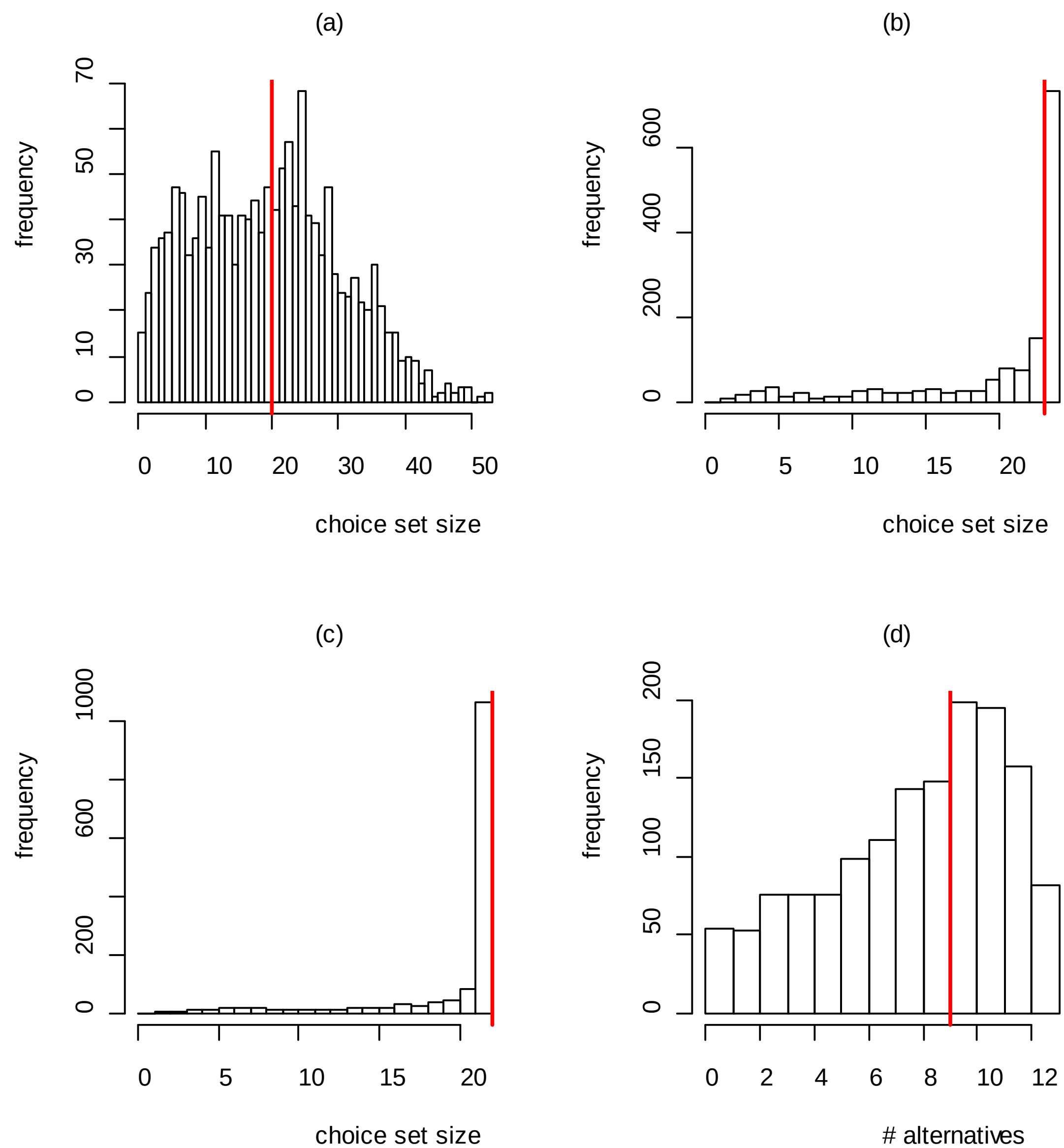


Figure 2. Choice set size distributions for (a) proposed, calibrated labeled routes, (b) *K-shortest* paths (*K*=23, shortest path added if not found), (c) Simulated shortest paths (20 draws, shortest path added if not drawn), and (d) single-permutation labeled routes. Heavy vertical lines show median.

6. Base Model Specification

Four categories of variables were considered for specifying the path-size logit route choice model:

- Delay/effort
- Navigation
- Perceived safety
- Designated bicycle facilities
- Adjacent land-use

Table 3 provides variable descriptions. Full estimation results are presented in Table 4. All model estimations were performed with the BIOGEME software package (Bierlaire 2003, 2008). Parameter estimate magnitudes cannot be directly compared in the multinomial logit model. Figure 3 provides a comparison of probability elasticities.

Aggregate choice elasticities require care to calculate meaningfully in the route choice context. Many network attributes appear relatively rarely in alternatives (e.g. bicycle facilities, major unsignalized crossings, highest traffic volume category, etc.). Simply calculating aggregate choice elasticity either at sample means or by sample enumeration leads to a downward bias for such attributes, since the majority of calculations will be zero. Since route choice alternatives are unlabeled (i.e. are not the same across cases), aggregate shares are not meaningful. What interests us is sensitivity to attributes when they are present. For these reasons, the following formula was used to calculate aggregate choice elasticities:

$$\eta_X^P = \sum_n \left(\frac{\sum_{j \in n} \beta_X X_j (1 - \hat{P}_j) \delta_{X,j}}{\sum_{j \in n} \delta_{X,j}} \right) \Big/ n \tag{3}$$

for attribute X, choice situation *n*, alternative *j*, estimated coefficient β, and estimated probability $\hat{P}$. The variable $\delta_{X,j}$ is 1 when variable X is present (i.e. non-zero over a path) and 0 otherwise. The basic form of sample enumeration follows Louviere et al. (2000) with the additional delta variable adjusting for zero-attribute bias and the denominator sum to account for multiple alternatives per case in the route choice context. Louviere at al. (2000) advise weighting cases by an alternative's choice probability, but of course this makes no sense in the route choice context, since alternatives are unlabeled.

For the two bridge dummy variables, sample enumerated arc elasticities were calculated by switching the dummy variable for trips using a bridge. The usual arc elasticity formula for probability changes was applied to bridge trips and the sample enumerated similarly to equation 1 (Louviere et al. 2000):

$$\eta_X^P = \sum_n \left(\frac{\sum_{j \in n} (\hat{P}_j' - \hat{P}_j)/(X_j' - X_j) \Big/ (\hat{P}_j' + \hat{P}_j)/(X_j' + X_j)}{j} \right) \Big/ n \tag{4}$$

for the n cases which include a river crossing. $X_j^{'}$ is set to 1 if X_j is 0 and vice versa. Elasticities in both cases presented here are best interpreted as the elasticity of probability to attribute X when present. Since distance enters in log form, elasticities with regard to distance are calculated directly from the estimated utility function as $B_X * (1 - \overline{P})$ where $\overline{P}$ is the sample mean probability (Ben-Akiva & Lerman 1985).

6.1 Delay and effort

Network variables in this category were thought to mainly increase a cyclist's travel time and effort. In addition to distance, variables measuring upslope and intersection delay were included. Since alternative trip lengths varied from 0.2 to 45 kilometers (mean 7.2 km), care was taken to ensure meaningful interpretations over the range of distances and also to guard against heteroskedasticity.

Log distance was chosen over simple distance, piecewise distance functions and polynomial formulations for its superior model fit and plausible behavior over the distance range. Log distance implies that a given unit of distance has lower disutility as trip length increases, with the percentage change in distance yielding a constant utility change. There are two complementary explanations for this. First, cyclists may simply compare routes in proportional terms, such that 100 meters on a 1 kilometer trip is equivalent to 1000 meters over 10 kilometers. Second, it seems likely that most cyclists would have a harder time distinguishing small distance differences on long trips, reducing their sensitivity to a given distance change. As can be seen in Figure 3, distance has the largest estimated impact on route choice probability with a 1 percent change in distance expected to decrease a route's choice probability by about 5 percent (not 5 percentage points) at the sample means.

The choice of log distance has implications for the form of other variables in the model. Point variables, such as intersection features, if entered as counts in the log distance model would imply counterintuitive results over the range of trip distances. A count variable would have a marginal utility equal to a fixed percentage change in distance. For instance, if a stop sign were estimated to cost 50 meters of distance for a 1 kilometer trip, the same stop sign would be worth 500 meters on a 10 kilometer trip. It seems unlikely that the same feature would be valued so differently. For this reason, point features were entered per unit distance. This provides the natural interpretation that the value of a point feature such as a stop sign is fixed relative to distance.

Intersection attributes include signalization, through, and cross volumes. Along with movement (through, left, or right) this yields a large number of potential categories. The categories presented are the result of many model iterations. Piecewise linear specifications were chosen both due to improved model fit and easier incorporation into the regional travel model, where precise estimates of local street volumes would be difficult. Entering right turn movements at traffic signals and busy intersections separately significantly improved model fit. This seems logical, since right turns may allow cyclists to avoid many intersection delays.

Significant intersection factors included stop signs, traffic signals, and the interaction of signalization with movement, through volume, and cross volume as shown in Table 4 and Figure 3. Crossing unsignalized intersections with high cross traffic volumes had the greatest estimated disutility among intersection variables, suggesting that cyclists are willing to go about 13 percent out of their way to avoid an unsignalized major street crossing on an average length alternative (7.2 km). Unsignalized left turns across heavy traffic volumes also had pronounced estimated effects on probability, implying a willingness to trade off a 9 percent increase in distance to avoid a difficult left turn on an average length trip.

A number of forms were tested for terrain. Differences in elevation gain and loss were highly correlated across alternatives, and thus gain alone was chosen to measure the degree of hilliness along a route. Average slope was estimated to have a relatively large influence on route choice decisions. A 1 percent increase in average upslope (e.g. a change from 1 to 1.01 meters per hundred) is expected to decrease selection probability by about 1.3 percent. Terrain, along with turn frequency, were the most important single factors after distance. Relative to distance, the estimation implies cyclists would be willing to travel roughly 27 percent farther to avoid an additional 1 percentage point average upslope.

Results regarding intersection delay factors are generally consistent with stated preference models; however, terrain appears to be considerably more important in a stated preference setting (Stinson & Bhat 2003, Sener et al. 2009). Stinson and Bhat (2003) and Sener et al. (2009) both found significant, positive effects of moderate hills on route choice, but we could not reproduce these findings with our data. It may be that cyclists prefer the idea of a more varied route in a stated choice experiment, but in real world situations are more sensitive to time and effort costs.

6.2 Navigation

The contiguity of a route was measured by the number of turns onto different streets or trails per kilometer. As well as imposing delays, which was captured in the intersection variables, turns impose a navigation cost on a cyclist. The sequence of turns and street names must be remembered. As Figure 3 reveals, the number of turns was second in importance only to distance. A 1.0 percent increase in the number of turns per kilometer results in an expected 1.3 percent decrease in route choice probability. On an average length trip, a cyclist would be willing to go about 1.5 percent out of her way to avoid an additional turn. These findings are consistent with measurements of “continuity” in stated preference models (Stinson & Bhat 2003, Sener et al. 2009).

6.3 Perceived Safety

The perceived safety or pleasantness of a route was not measured directly but instead proxied by traffic volumes along the route. Several formulations were tested. In the end, the proportion of travel with each of three volume categories was selected. The volume categories roughly align with the street classes secondary arterial, primary arterial, and highway, respectively. The proportion of travel on streets with less than 10,000 annual average daily traffic (AADT) comprise the reference category.

Because bike lanes provide a separate travel lane for cyclists, it was hypothesized that they would mitigate the effects of increasing traffic volume. Model fit failed to reject the hypothesis that bike lanes fully offset the effects of traffic volume in each category. It is somewhat surprising that, on a street with bike lanes, there is no discernable difference at different traffic volumes. Three explanations are offered. First, it may be that bike lanes tend to be designed to higher standards as traffic volume increases. Second, competing for space may constitute the major disutility with increasing traffic volume—speed and noise playing only minor roles. Third, higher traffic thoroughfares may be less likely to have on-street parking (for which data were not available for this study) and thus provide more of a buffer to the right of traffic. Stated preference work has also found that bike lanes completely offset major street traffic volume (Stinson & Bhat 2003).

Figure 3 shows that the estimated effect of traffic volume is similar in magnitude to intersection features. On streets with no bike lane, a 1 percent increase in the proportion of travel with 10-20, 20-30, and 30-plus thousand AADT is expected to reduce selection probability by 0.17, 0.45, and 1.0 percent, respectively, when such facilities are present. Stinson and Bhat (2003) and Sener et al. (2009) found similar relative effects of increasing traffic volume in stated preference work.

6.4 Bicycle Facilities

Designated bicycle facilities in the Portland region include bike boulevards (described in Table 3), on-street bike lanes, off-street bike paths, and signed, unimproved bike routes. Bicycle facilities are best viewed as bundles of other attributes. For instance, an off-street bike path has zero traffic volume, relatively few stops and major street crossings, and is generally contiguous. Despite controlling for the range of variables already mentioned, bike boulevards and bike paths were estimated to have significant residual effects on route choice.

We speculate that the residual value of bike facilities may be due to three additional factors not measured in the study. First, cyclists may simply be more knowledgeable about designated bike routes because they are signed and also printed on widely used bike maps. Second, there may be a perceived increase in safety due to higher numbers of other cyclists using a bike facility. Third, pavement quality may be higher on designated bike facilities than on competing routes.

Estimates of the residual value of bike facilities are provided in Table 4 and Figure 3. The elasticities of probability are fairly small, suggesting most of the value of those facilities is captured in the other model variables. A 1.0 percent increase in the proportion of bicycle boulevards and off-street paths is expected to increase selection probability by 0.16 and 0.249 percent, respectively. It is worth emphasizing that the entire bundle of bike facility attributes likely will have a much greater impact than the residual value suggests.

Similar to Stinson and Bhat (2003), we found significantly stronger effects for bicycle facilities on bridges. Our model implies bicyclists would be willing to go 34 percent out of their way to use a bridge with an improved, separated bicycle facility and 19.5 percent out of the way to find a bridge with a bike lane. Our results are only slightly lower than the sensitivities implied by the stated preference work of Stinson & Bhat (2003).

6.5 Land-use and Missing Variables

Adjacent employment, retail employment, and commercial zoning were tested during model development. It was hypothesized that these variables might capture unobserved variables such as driveway access, on-street parking, parking turnover, and the general busyness along a route segment. None of the land-use variables performed well in the model, and they are not included in the final specification.

We speculated that several factors might have led to the poor performance of adjacent land-use variables. First, land-use is strongly correlated with traffic volumes. Second, the measurement of employment variables was available only at the traffic analysis zone (TAZ), and several assumptions had to be made in order to join this data to the street link level. Third, it was unclear how to handle density; for example, to what degree does employment ten floors up affect street-level activity. Thus, while the land-use data available for this study was not helpful to understand route choice decisions, future studies with better data should not ignore land-use out of hand. A few other attributes found to be potentially important in stated preference work were unavailable to us. These include on-street parking, shoulder width, pavement condition, and traffic speed (Stinson & Bhat 2003, Sener et al. 2009).

6.6 Path-size Parameter

The path-size parameter estimate's positive coefficient is consistent with theory. It is significantly different from 1.0, which would be the expected value if the path-size parameter captures only the statistical error introduced by the IIA property of the MNL model. It has been suggested that the path-size parameter should not be arbitrarily fixed to 1.0, since it may have a meaningful behavioral interpretation (Frejinger and Bierlaire 2007).

In our case, estimating the parameter significantly improves model fit. Fixing the path-size parameter to 1.0 has the effect of reducing the magnitude of the distance coefficient while leaving the other parameters more or less unchanged. Since generated alternatives tend to cluster around the shortest-path, the greater than expected path-size correction may indicate unobserved disutility factors along shortest-path corridors. One plausible explanation is that many shortest paths in Portland involve a handful of busy, diagonal arterials that cut across the otherwise regular grid. These streets have generally poor riding environments which may not be fully captured by our observed attributes.

Another interpretation is that cyclists in our sample are less likely to distinguish between overlapping routes than statistically expected. That is, cyclists may consider two routes that overlap for just 25 percent of their lengths to be more similar than the physical overlap suggests. Perhaps they tend to share particularly unpleasant segments such as the diagonal arterials mentioned in the previous paragraph. A multi-modal route choice study found that trip "legs" rather than distance may sometimes be a better overlap measure (Hoogendoorn-Lanser et al. 2005).

Table 3. Variable Descriptions			
Variable	Description	Mean	Present in proportion alts (29,090 total)
Bridge w/ bike lane	bridge with on-street bike lane	dummy variable	0.05
Bridge w/ sep. facility	bridge with improved, separated bike facility	dummy variable	0.22
Avg upslope (m/100m)	average gross gain per 100m	1.01	1.00
Distance (km)	distance of route in kilometers	7.21	1.00
Path size (0-1, 1=unique)	path size (see section 4 for formula)	0.31	1.00
Left turn, unsig., AADT 10-20k (/km)	left turn without traffic signal and parallel traffic volume 10,000-20,000 per day	0.07	0.36
Left turn, unsig., AADT 20k+ (/km)	left turn without traffic signal and parallel traffic volume 20,000+ per day	0.03	0.18
Prop. bike boulevard	proportion of route on designated bicycle boulevard (signed bike route on low traffic volume streets with traffic calming, diversion, and enhanced right of way)	0.10	0.53
Prop. bike path	proportion of route on off-street, regional bike path (i.e. not minor park paths, sidewalks, etc.)	0.04	0.41
Prop. AADT 10-20k w/o bike lane	proportion of route on streets with traffic volume 10,000-20,000 per day without a bike lane	0.08	0.73
Prop. AADT 20-30k w/o bike lane	proportion of route on streets with traffic volume 20,000-30,000 per day without a bike lane	0.04	0.46
Prop. AADT 30k+ w/o bike lane	proportion of route on streets with traffic volume 30,000+ per day without a bike lane	0.02	0.26
Traffic signal exc. right turns (/km)	left turns and straight movements through traffic signals per kilometer	1.14	0.90
Stop signs (/km)	turns or straight movements through stop signs per kilometer	1.93	0.95
Turns (/km)	left and right turns per kilometer	2.26	1.00
Unsig. cross AADT 10-20k exc. right turns (/km)	left turns and through movements at unsignalized intersections with cross traffic volume 10,000-20,000 per day	0.26	0.72
Unsig. cross AADT 10k+ right turns (/km)	right turns at unsignalized intersections with cross traffic volume 10,000+ per day	0.10	0.44
Unsig. cross AADT 20k+ exc. right turns (/km)	left turns and through movements at unsignalized intersections with cross traffic volume 20,000+ per day	0.10	0.52

Table 4. Route Choice Model Estimation Results		
Variable	Est. Coeff.	Rob. t-stat
Bridge w/ bike lane	1.26	4.78
Bridge w/ sep. facility	2.44	9.93
Gain (m/100m)	-1.4	-9.6
Ln(distance)	-5.81	-10.91
Ln(path size)	1.72	18.83
Left turn, unsig., AADT 10-20k (/km)	-1.29	-3.15
Left turn, unsig., AADT 20k+ (/km)	-2.95	-4.19
Prop. bike boulevard	0.895	5.31
Prop. bike path	1.77	5.11
Prop. AADT 10-20k w/o bike lane	-1.56	-4.79
Prop. AADT 20-30k w/o bike lane	-5.29	-6.27
Prop. AADT 30k+ w/o bike lane	-14.6	-6.02
Traffic signal exc. right turns (/km)	-0.237	-4.21
Stop signs (/km)	-0.127	-3.54
Turns (/km)	-0.589	-14.12
Unsig. cross AADT 10-20k exc. right turns (/km)	-0.804	-5.12
Unsig. cross AADT 10k+ right turns (/km)	-0.508	-1.98
Unsig. cross AADT 20k+ exc. right turns (/km)	-3.86	-9.63
Number of observations	1,449	
Null log-likelihood	-4058.7	
Final log-likelihood	-3123.1	
Rho-square	0.231	
Prediction success rate	0.330	

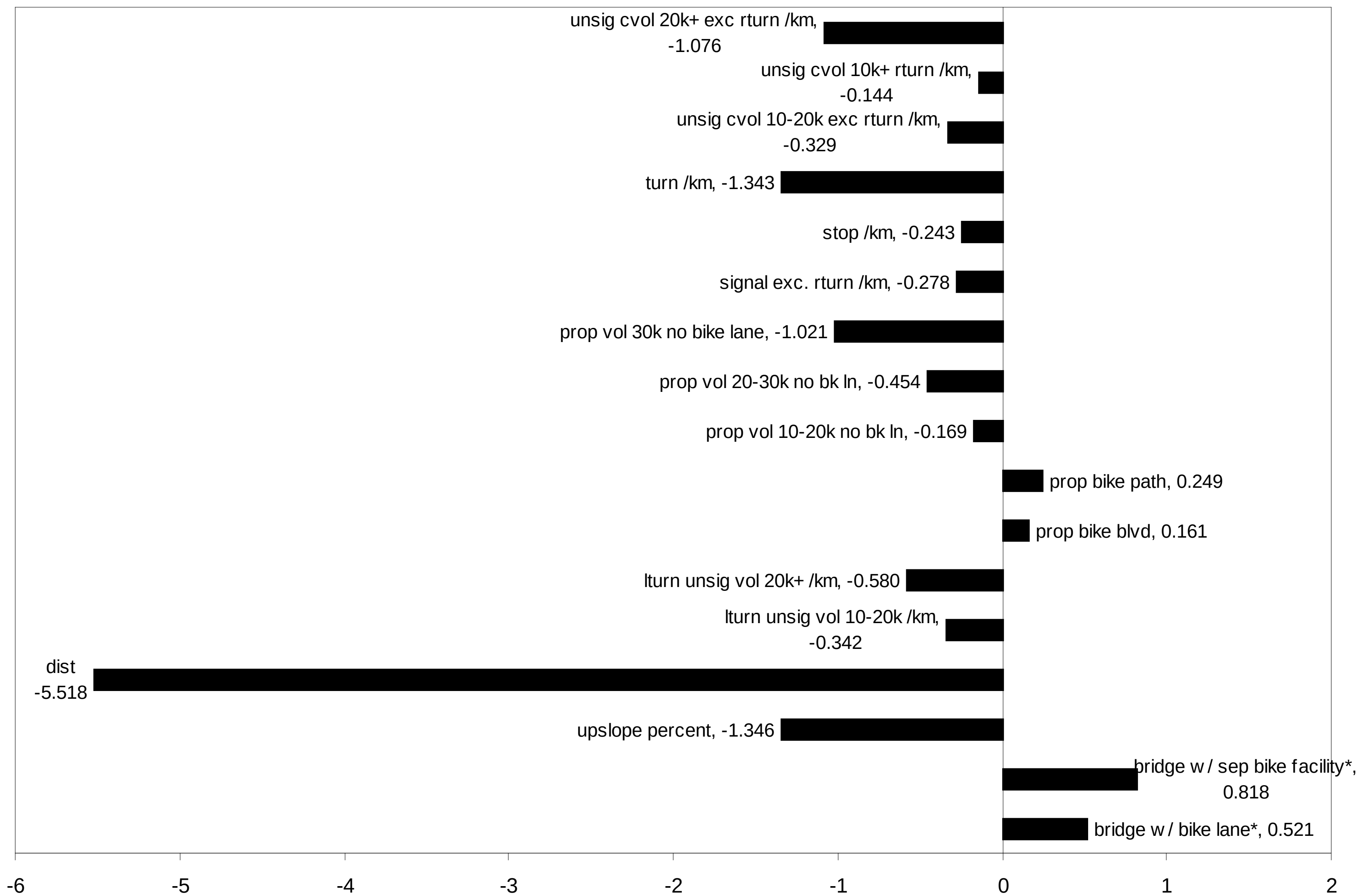


Figure 3. Aggregate choice elasticities (see text for calculation method)

7. Trip and Individual Segmentation Models

In addition to the base model, a number of different segmentation schemes were specified. Presented here are two fully segmented route choice models based on trip purpose and gender. In general, results were consistent with the base model; however, commuters were significantly more averse to distance, delays, and traffic.

7.1 Commute Trips

Commute trips are recurring trips, often with arrival deadlines at the work end. In contrast, other trips may be more irregular with more flexible arrival times. Commute trips in our sample are longer trips (about 25 percent longer, on average) and also more likely to occur during peak hours. Table 5 and Figure 4 show the model results and elasticities at sample means for commute and non-commute trip segments.

Commuting cyclists in our sample are considerably more sensitive to delay factors—distance, upslope, and stop signs and signals—than non-commuters. This finding is consistent with the arrival deadline hypothesis. Commuters also show somewhat more aversion to busy streets, perhaps because of peak traffic volumes. Overall model fit is better for the commute segment, although the non-commute model predicts more accurately. Unobserved factors would seem to be more important for non-commute travel. For both travel demand modeling and policy decisions, our results suggest that direct routes with minimal delays will be more significantly attractive to bicycle commuters.

7.2 Gender

A set of gender-specific models were specified. Table 6 and Figure 5 present the results. Most striking is the similarity between segments in the gender models. The largest apparent difference are the intersection variables for the busiest intersections. Female cyclists were considerably more sensitive to unprotected left turns across heavy traffic and somewhat more sensitive to heavy crossing volumes. Female cyclists in our sample were also somewhat more sensitive to distance than male cyclists. In fact, the results probably understate the difference, since male cyclists were more frequently commuters (40 percent versus 32 percent of trips), and commuters showed a stronger sensitivity to distance. Finally, the female segment showed slightly higher preference for bicycle facilities, particularly bike boulevards and both bridge facility types.

In our sample of mostly experienced cyclists, male and female cyclists do not show systematically different route choice preferences. While some of the small differences here merit further study, it appears that gender may be largely irrelevant for route choice among experienced cyclists.

Table 5. Commute/Non-commute model estimation results		
Variable	Est. Coeff. Commute	Est. Coeff. Non-Commute
Bridge w/ bike lane	1.41	1.04
Bridge w/ sep. facility	2.99	2.43
Gain (m/100m)	-2.20	-1.17
Ln(distance)	-10.7	-4.73
Ln(path size)	2.10	1.67
Left turn, unsig., AADT 10-20k (/km)		-0.85
Left turn, unsig., AADT 10k+ (/km)	-3.70	
Left turn, unsig., AADT 20k+ (/km)		-3.61
Prop. bike boulevard	0.98	0.86
Prop. bike path	1.84	1.62
Prop. AADT 10-20k w/o bike lane	-3.32	-1.29
Prop. AADT 20-30k w/o bike lane	-8.28	-4.53
Prop. AADT 30k+ w/o bike lane	-28.1	-10.5
Traffic signal exc. right turns (/km)	-0.54	-0.16
Stop signs (/km)	-0.28	-0.11
Turns (/km)	-0.75	-0.57
Unsig. cross AADT 10-20k exc. right turns (/km)		-0.93
Unsig. cross AADT 20k+ exc. right turns (/km)	-5.04	-3.41
Number of observations	433	1016
Null log-likelihood	-1343.5	-2715.2
Final log-likelihood	-933.3	-2085.2
Rho-square	0.305	0.232
Prediction success rate	0.296	0.342

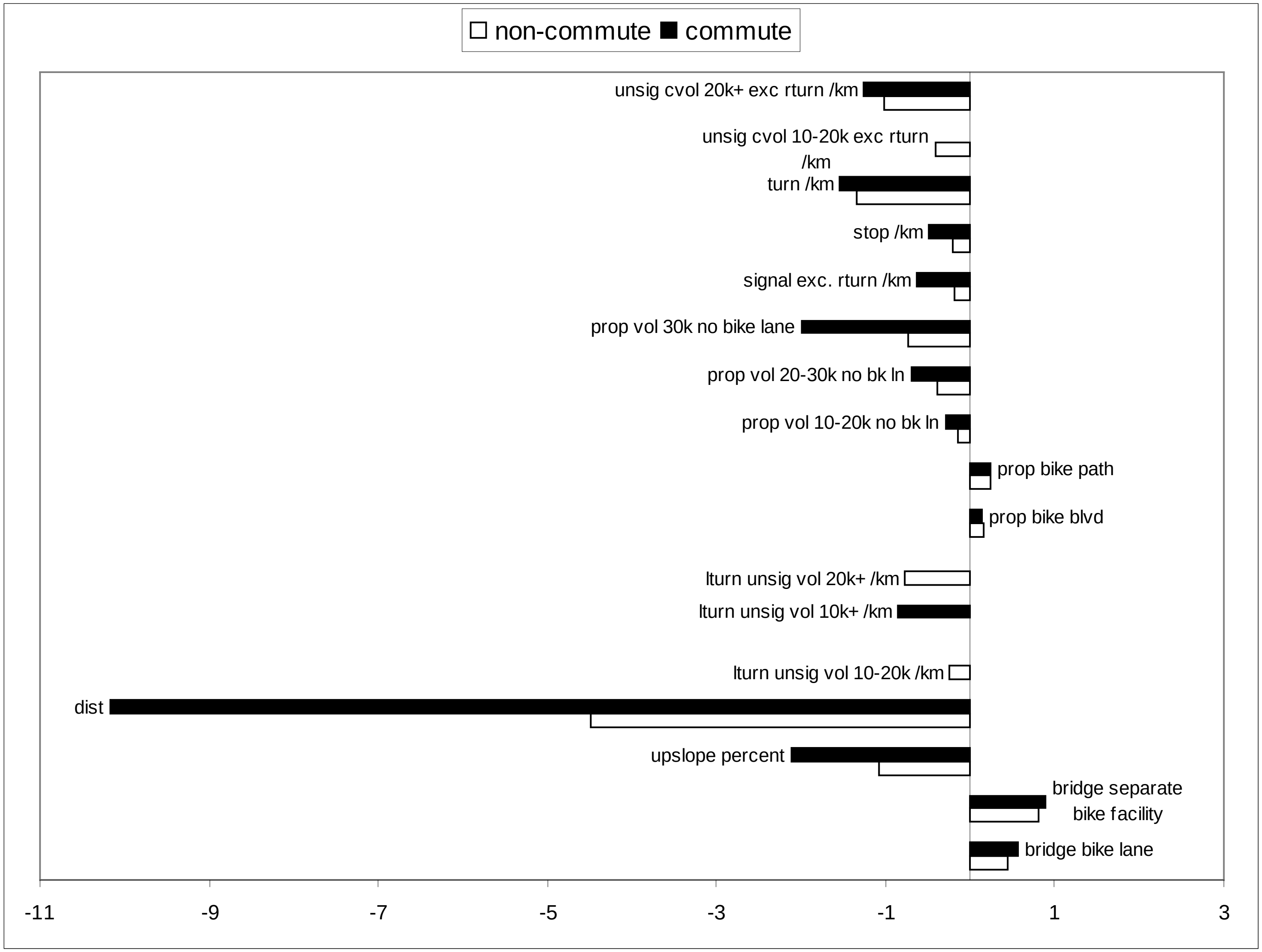


Figure 4. Commute/Non-commute elasticities (see text for calculation method)

Table 6. Female/Male model estimation results		
Variable	Est. Coeff. Female	Est. Coeff. Male
Bridge w/ bike lane	2.84	1.27
Bridge w/ sep. facility	4.53	2.08
Gain (m/100m)	-1.32	-1.37
Ln(distance)	-6.13	-5.69
Ln(path size)	1.57	1.87
Left turn, unsig., AADT 10-20k (/km)	-1.25	
Left turn, unsig., AADT 10k+ (/km)		-1.35
Left turn, unsig., AADT 20k+ (/km)	-9.04	
Prop. bike boulevard	1.16	0.725
Prop. bike path	1.91	1.6
Prop. AADT 10-20k w/o bike lane	-1.39	-1.83
Prop. AADT 20-30k w/o bike lane	-5.41	-5.22
Prop. AADT 30k+ w/o bike lane	-13.9	-14.5
Traffic signal exc. right turns (/km)	-0.208	-0.245
Stop signs (/km)	-0.0709*	-0.196
Turns (/km)	-0.557	-0.635
Unsig. cross AADT 10-20k exc. right turns (/km)	-0.676	-0.838
Unsig. cross AADT 20k+ exc. right turns (/km)	-3.89	-3.86
Number of observations	652	797
Null log-likelihood	-1784.3	-2274.4
Final log-likelihood	-1351.1	-1689.1
Rho-square	0.243	0.257
Prediction success rate	0.359	0.338
* not significant at 5% level		

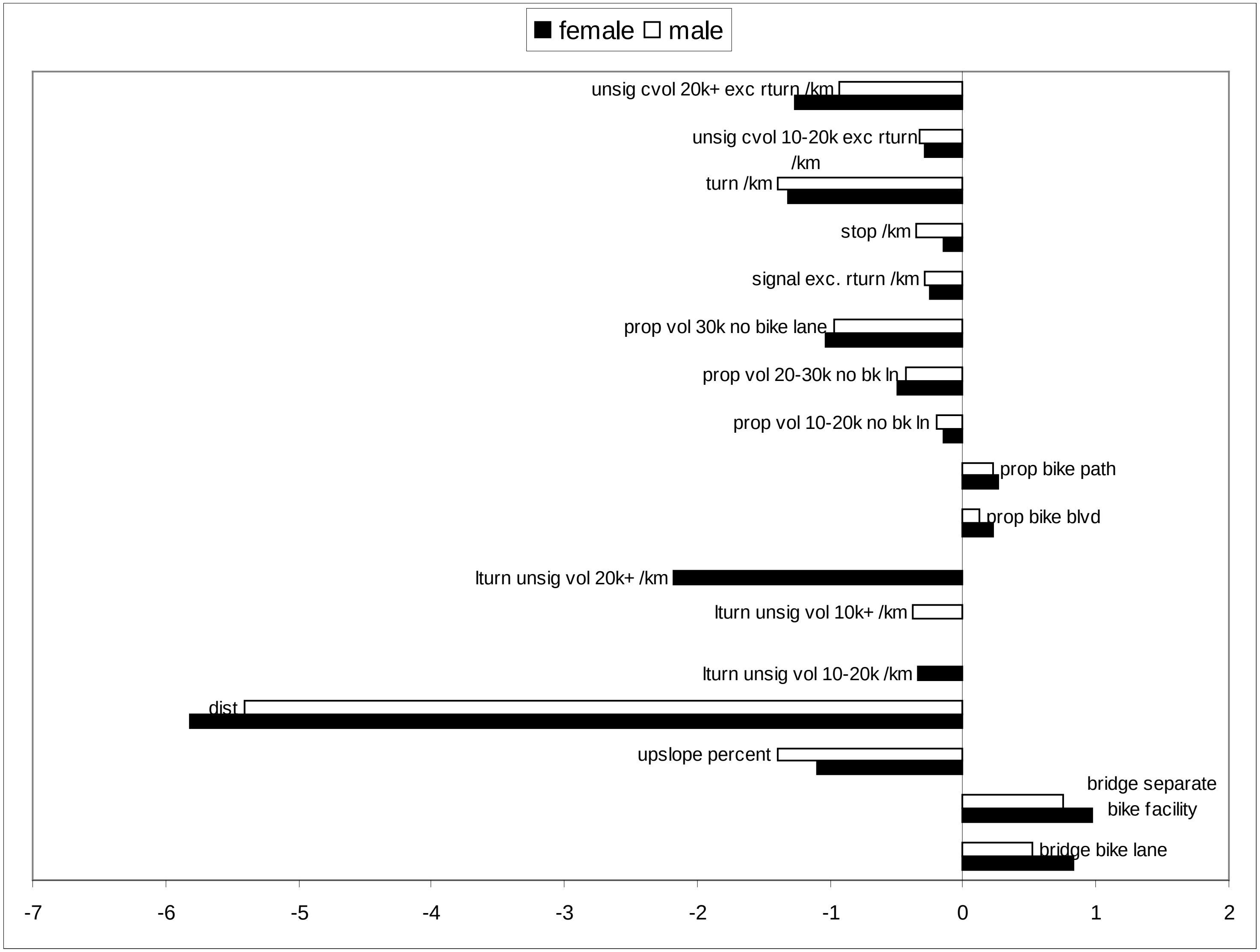


Figure 5. Female/Male elasticities (see text for calculation details)

8. Implementation

The bicyclist route choice model described in this paper will be implemented as part of the Metro regional travel forecasting system. The bicycle route choice and the bicycle network model, hereafter "the bike model," will be independent components of the system and will be utilized in a phased implementation plan. The bicycle network includes nearly all of the local streets in the region as well as off-street bike trails and is therefore much finer-grained in its detail than the Metro highway network.

In terms of bicyclist user classes, our analysis revealed some significant differences by trip context; therefore, the Metro model implementation will maintain market segmentation by work-commute and non-work-commute purposes. Gender-based segmentation showed little differences between groups and is not a dimension of segmentation maintained by Metro for its current suite of trip-based models.

As a first step towards integration, the bike model will be used to generate representative zone-based bicyclists' generalized cost path information, otherwise known as "skims," which will then be used to re-estimate the Metro mode choice model. Metro maintains a fairly detailed 2032-zone system, and has provided representative starting points for bicycle trips within each zone. The bike model skims will be expressed in terms of zone-to-zone generalized costs (dis-utilities) computed directly from the route choice model and calculated as a probability-weighted-average of competing paths; distances, calculated as a probability-weighted-average of competing paths; and travel times, computed using a special bicyclist travel time function applied to generalized least cost paths and weighted probabilistically similar to distance.

The bicyclist route travel time function is a linear model of travel times applied to the generalized least cost path determined through the route choice model. It was estimated from the same GPS data used to develop the route choice model and includes variables such as distance, slope, intersection and turning movement variables, and through- and crossing-traffic volume effects.

In the second phase, full implementation, the newly specified mode choice model will be in place. Trips that the model predicts as "bike" through the mode choice process will be assigned to the bike network in a single-pass stochastic assignment. This will enable Metro to project dispersion among reasonable competing paths through the bike network and usage levels of any existing or proposed bike facilities along those paths. In addition, analysts will be able to infer changes in bicycle rider welfare for competing investment scenarios by comparing changes in utility; however, this will be initially limited to changes to bike infrastructure itself or any connectivity improvements that result from changes to the highway network.

For this second phase, it is important to note that this is a static assignment without feedback. Bicycle facilities are not considered to be capacity constrained. Projected volumes from the auto assignment portion of the model system will not be fed back to the bike model network. This limitation is mainly due to the computational burden of iterating between the auto assignment, bike model and demand-side model components, but also because implementation of such a feedback mechanism requires further study of its dynamics. The value of feedback to the current variable specification would be to affect route utility with respect to the disutility of through- and crossing-traffic volumes, stratified by volume groups. It is planned to provide this full feedback to the bike model in a longer-term third phase, such that the disutility of alternative bike routes is affected by assigned auto volumes. In addition to the added computational time, this would require reconsideration of the volume group designations to avoid potential "cliff" effects that could produce oscillation in the system.

9. Conclusions

The regional bicycle model presented in this paper is novel in its use of GPS revealed preference data, collected specifically for the purposes of tracking bicyclists' route choice behavior. Coupled with a highly detailed bicycle network model and careful coding of observed paths, the model's data backplane is considered to be highly accurate in its portrayal of revealed choices and the path attributes faced by respondents. The trips represented in this model were primarily utilitarian in nature, at least a quarter of which were for work purposes.

The GPS survey was supported and corroborated by respondent demographic and attitude information. We used this supporting data to develop descriptive profiles of respondents preferences for route choice attributes and used this to inform the development of a novel choice set selection algorithm. Labeled attributes important to the respondents were balanced against observed deviations from shortest distance paths to create choice set alternatives that permitted multiple permutations from each label category, subject to a maximum deviation from the shortest path. The results of this choice set generation algorithm indicated a superior performance over comparable implementations of *K*-shortest paths, simulation methods, and one-permutation labeled approaches.

The route choice model was formulated as a Path-Size Logit model, with a path-size correction factor to account for overlapping alternatives based on network link lengths. Utility functions were specified to account for path distance, in log form, and other attributes relative to total path distance. In doing so, the model avoids problems of heteroskedasticity and intuitively accounts for perception variance over trip length. The most significant results indicate that bicyclists are most sensitive to total path length, but also avoiding turns across heavily-traveled arterials and high-traffic-volume through streets without separate bike facilities. Minimizing elevation gain, stops and turns in general were also significant.

Whereas bike infrastructure may be viewed as providing a bundle of attributes, including avoidance traffic volumes and minimizing stops and turns, our estimation results show strong residual preferences for the provision of separated bikeways along bridge crossings, followed by bike lanes on bridges, off-street multi-use paths, and bike boulevards. The results also indicate that striped bike lanes serve to completely offset the disutility of higher traffic volumes, but do not have a residual value beyond that. Our results are remarkably consistent with some recent stated preference work done by others.

In terms of market segmentation, our fully segmented models did not show an appreciable difference between males and females among these experienced cyclists, except for a stronger female preference for bike infrastructure on bridge crossings and to avoid left turns at un-signalized high-volume intersections.

There were more significant differences, however, by trip purpose. Commuter choice elasticities were much more sensitive to total trip distance, avoidance of high-volume streets without bike lanes, and avoiding elevation gain, compared with non-commute trips. Additional segmentation by income and other socioeconomic attributes was not supported by the data.

The final route choice model is being implemented as part of the larger Portland regional travel forecasting system. It will be used to inform the regional mode choice mode with better estimates of bicyclists' generalized cost of travel and will also be used to assign bicycle trips to the bike network to project facility usage. Areas for future research include full integration with the Metro area travel model such that there is feedback between the highway assignment steps and the bike model, allowing the bicycle route choices to vary in response.

Acknowledgments
The authors greatly appreciate the support of the legions of student research assistants, who helped to prepare the data used in this research. In particular, Daniel Costantino and Matt Waddell were instrumental transforming the GPS outputs into coherent representations of bicycle trips. The Active Living Research program of the Robert Wood Johnson Foundation provided the funding for the first part of this project, including much of the data collection. Subsequent funding was also provided by the Oregon Transportation Research and Education Consortium (OTREC) with in-kind matching support provided by Oregon Metro.

References

Aultman-Hall, L., F.L. Hall, and B.B. Baetz. (1998). Analysis of bicycle commuter routes using Geographic Information Systems: implications for bicycle planning. *Transportation Research Record*, 1578, 102-110.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models , *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2008). An introduction to BIOGEME Version 1.7, http://biogeme.epfl.ch.

Bekhor, S., M.E. Ben-Akiva, and M. Scott Ramming. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1), 235-247.

Ben-Akiva, M.E., M.J. Bergman, A.J. Daly, and R. Ramaswamy. (1984). Modelling inter urban route choice behavior. In *Proc., Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, Delft, The Netherlands, 299-330.

Ben-Akiva, M.E., and M. Bierlaire. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of Transportation Science*, Kluwer, Dordrecht, The Netherlands, 5-33.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete choice analysis:theory and application to travel demand*. Cambridge, MA: The MIT Press.

Bovy, P.H.L. (2009). On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews*, 29(1), 43-68.

Dill, J. L. and J. P. Gliebe. (2008) Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice: Final Report, OTREC-RR-08-03. Oregon Transportation Research and Education Consortium.

Frejinger, E. and M. Bierlaire. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 41, 363-378.

Hoogendoorn-Lanser, S. R. van Ness, and P. Bovy. (2005). Path size modeling in multimodal route choice analysis. *Transportation Research Record*, 1921, 27-34.

Howard, C. and E.K. Burns. (2001). Cycling to work in Phoenix: route choice, travel behavior, and commuter characteristics. *Transportation Research Record*, 1773, 39-46.

Hunt, J. D., and J.E. Abraham. (2007). Influences on bicycle use. *Transportation*, 34(4), 453-470.

Krizek, K. J.(2006). Two approaches to valuing some of bicycle facilities' presumed benefits. *Journal of the American Planning Association*, 72(3), 309-319.

Landis, B.W., V.R. Vattikutti, and M. Brannick. (1997). Real-time human perceptions: towards a bicycle level of service. *Transortation. Research Record*. 1578, 119–126.

Louviere, J.J., D.A. Hensher, and J.D. Swait. (2000). *Stated choice methods: analysis and applications*. Cambridge UP.

Menghini, G., N. Carrasco, N. Schüssler, and K.W. Axhausen. (2008). Route choice of cyclists in Zurich: GPS-based discrete choice models. ETH: Swiss Federal Institute of Technology, Zurich, Working Paper No. 544, January 2009, Retrieved June 1, 2009, http://www.ivt.ethz.ch/vpl/publications/reports/ab544.pdf.

Prato, C.G., and S. Bekhor. (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record*, 1985, 19-28.

Prato, C.G., and S. Bekhor. (2007). Modeling route choice behavior: how relevant is the composition of the choice set? *Transportation Research Record*, 2003, 64-73.

Ramming, S. (2001). *Network Knowledge and Route Choice*. PhD thesis. Massachusetts Institute of Technology, Cambridge, Massachusetts.

Sener, I.N., N. Eluru, and C.R. Bhat. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36, 511-539.

Stinson, M.A., and C.R. Bhat. (2003). Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation Research Record*, 1828, 107-115.

Tilahun, N.Y., D.M. Levinson, and K.J. Krizek. (2007). Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A*, 41(4), 287-301.

Appendix: Steps in Choice Set Generation Method

In outline form, our route choice set generation method was implemented as follows:

- Identify shortest paths for each origin-destination pair.
- Define a set of attribute labels (Table 1).
- For each label, specify a label cost function of form $L_i = \beta * c_i + (1-\beta) * x_i$, where L_i is the label value of link *i*, β is a weighting parameter between 0 and 1, c_i is the base cost (distance or time) of link *i*, and x_i is an attribute cast as a disutility. For example, x_i might be the distance link *i* traverses *without* a bike facility.
- Set an initial minimum value for β and a step size that specifies how much β will *decrease* with each iteration. At the limits, β=0 returns the path that minimizes attribute *x* regardless of distance, and β=1 returns the shortest path.
- Starting with β = 1 – step size, minimize the label cost function, and with each iteration decrease β by the step size until β<min(β). The combination of step size and min(β) determine the maximum number of unique routes each label generates. For example, with min(β)=0.1=step size, a maximum of nine routes would be returned.
- Calibrate min(β) by fitting the generated shortest path deviation distribution to the observed distribution. Optimizing a fit statistic such as minimizing the Kolmogorov-Smirnov (K-S) test statistic could be implemented; however, we fit by eye using quantile-quantile (Q-Q) plots. It was convenient to first fit the observed route deviations to a known distribution to minimize the effect of outliers and different sample sizes.
- Repeat the process for each label. Combine generated alternatives, observed routes, and, if desired, shortest paths. Duplicate and, if desired, highly overlapping routes should be filtered.

Specific calibration and filtering were implemented using the following steps:

- Perform initial label run with min(β)=0.5 and step size=0.1.
- Calibrate to fitted observed distribution of shortest path deviations. The calibration does not need to be overly precise. Using Q-Q plots like those in Figure 1, each label could be fitted in one to four iterations by changing min(β) in 0.1 increments. Table 1 includes calibrated min(β) values.
- Step size was left at the initial value of 0.1, which seemed to strike an acceptable balance among choice set size, route variation, and computation time.
- Generated labels were combined, shortest paths and observed paths were added if not generated, and routes overlapping more than 90 percent with others were removed.