Statistical Analysis of Transit User Preferences Including In-Vehicle Crowding and Service Reliability

Peter Vovsha
Parsons Brinckerhoff
1 Penn Plaza, 3rd Floor
New York, NY 10119
Phone: (212) 465 5511
Email: vovsha@pbworld.com

Marcelo G. Simas Oliveira
GeoStats LP
530 Means St NW Suite 310
Atlanta, Georgia 30318
Phone: (404) 588 1004
Email: moliveira@geostats.com

William Davidson
Parsons Brinckerhoff
303 Second St. #700N
San Francisco, CA, 94107
Phone: (415) 243 4601
Email: davidson@pbworld.com

Chaushie Chu
Los-Angeles County Metropolitan Transit Authority
1 Gateway Plaza
Los Angeles, CA 90012
Phone: (213) 922 3059
Email: chuc@metro.net

Robert Farley
Los-Angeles County Metropolitan Transit Authority
1 Gateway Plaza
Los Angeles, CA 90012
Phone: (213) 922 2532
Email: farleyr@metro.net

Michael Mitchell
530 Means St NW Suite 310
Atlanta, Georgia 30318
Phone: (404) 588 1004
GeoStats LP
Email: mmitchell@geostats.com

Gaurav Vyas
Parsons Brinckerhoff
1 Penn Plaza, 3rd Floor
New York, NY 10119
Phone: (212) 465 5730
Email: vyas@pbworld.com

Paper size: 6,950 words + 1 table (×250) + 7 figures (×250) = 8,950 words
Submitted for presentation at the 92nd Annual Meeting of the Transportation Research Board and Publication in the Transportation Research Records

August 1, 2013

1
Abstract
The paper describes recent experience with the application of an innovative web-only Stated Preference (SP) survey in the Los Angeles Region. The survey data was used to model the effects of crowding and on-time performance on transit demand. This SP survey was developed as a 100% self-complete web survey with most relevant details about a recently completed transit trip are collected in the opening questions, with all SP game scenarios generated dynamically. The survey featured online geocoding of the trip origin and destination, integration with Metro’s trip planner for downloading candidate itineraries, spatial processing using GIS data, as well as real-time parsing of modeled level-of-service variables as part of the SP game generation. The SP games were generated pivoting off the reported trip where the respondent was prompted to trade-off the crowding level or service reliability for travel time and/or cost.

Based on the collected survey sample that included 2,609 valid complete responses and over 30,000 generated choice observations, a discrete choice model for transit modes and routes was estimated. The estimation results confirmed that crowding perceived by transit users as a negative factor expressed as a perceptual weight on in-vehicle time that can reach a value of 1.6 or higher for extreme crowding levels. Additionally, crowding effects proved to be differential by trips purpose, person age, and transit mode. With respect to transit reliability effects, different model specifications were explored with such variables as amount of delay and probability of delay.

Keywords: Stated Preference Survey, Transit Crowding, Transit Service Reliability

Word Count: 6,750 words + 1 table × 250 words + 7 figures × 250 words = 8,750 words
Introduction
The paper describes recent experience with the application of an innovative web-only SP survey in the Los Angeles Region. The survey data was used to model the effects of crowding and on-time performance on transit demand, with input data on these attributes collected from Metro riders participated in the survey. This survey was developed as a 100% self-complete web survey with most relevant details about a recently completed transit trip collected in the opening questions, and with SP game scenarios generated dynamically. The survey featured online geocoding of trip origin and destination, integration with Metro’s trip planner for downloading candidate itineraries, spatial processing using GIS data provided by Metro, as well as real-time parsing of modeled Level-of-Service (LOS) variables as part of the SP game generation. The SP games were generated pivoting off the reported trip where the respondent was prompted to trade-off either crowding conditions or reliability for travel time and/or cost.

Inclusion of crowding and reliability in the survey allowed for estimation of the corresponding user preferences that were subsequently incorporated in the travel model of the Los-Angeles County Metropolitan Transit Authority (LACMTA). Most applied travel models still utilize simplified unconstrained transit assignment procedures. This simplification results in two particular problems. First, ridership greater than total line capacity is allowed that is obviously an unrealistic outcome. Secondly, inconvenience and discomfort in crowded transit vehicles (in particular, standing instead of seating) is ignored despite that this factor strongly affects transit route choice, mode choice, and other travel choices. The current research was intended to incorporate both factors in an operational travel model in a consistent non-duplicative way. Quantification of transit reliability is another new area with only a few published works. There are several approaches to measure travel time reliability of a transit service of which schedule adherence at boarding stop (expressed in extra wait time at boarding) proved to be the most “implementable” at the current stage of research.

Transit Capacity Constraining and Crowding in Travel Models
Most applied travel models still utilize simplified unconstrained transit assignment procedures. This simplification results in two particular problems. First, ridership greater than total line capacity is allowed that is obviously an unrealistic outcome. Secondly, inconvenience and discomfort in crowded transit vehicles (in particular, standing instead of seating) is ignored despite that this factor strongly affects transit route choice, mode choice and other travel choices. The current research was intended to incorporate both factors in an operational travel model in a consistent non-duplicative way.

The first related aspect is to ensure feasibility of transit ridership forecast for each line and segment with respect to the total capacity constraint. This means that in a case where the transit volume exceeds total segment capacity a penalty should be applied until the feasible solution is reached. A feasible solution might not exist especially if a restricted transit assignment framework with a fixed transit table is employed (i.e. the riders of overcrowded lines can only switch to some other lines). It is normally a better chance to find a feasible solution if a mode choice framework is also employed (i.e. the riders of overcrowded lines can also switch to alternative modes). In terms of behavioral realism, the most
appealing method to address infeasible volumes is to increase transit wait times at the corresponding boarding stations, i.e. use effective headways rather than schedule-based headways [Cepeda et al, 2005]. This is based on the assumption that the riders will not always be able to board the first-arriving vehicle and will have to wait for the next vehicle. Effective headways is general is difficult to observe in reality. Thus, the form of the effective headway function is derived based on theoretical considerations and evaluated by the aggregate outcome of model application.

The second related aspect is to take into account crowding in the vehicle as a negative factor in the user perception of transit service quality. From this standpoint, not only exceeding of the total vehicle capacity but also exceeding the seated capacity (or even approaching it) should be penalized since standing is generally perceived by transit users as a very strong negative factor. Also, in a crowded vehicle, seating passengers experience inconvenience in finding a seat and getting off the vehicle. Crowding, however, should not be penalized in the same way as exceeding the total capacity since it is still a feasible observed situation. In terms of behavioral realism, the probability of having a seat should be reflected in the perceived in-vehicle time weight. This factor should be incorporated in the transit assignment and mode choice model. Penalizing in-vehicle-time in crowding vehicles in transit assignment is algorithmically similar to applying volume-delay functions in highway assignment [Spiess, 1993; Maunsell, 2004; Abraham & Kavanagh, 1992; Davidson et al, 2011; Barquin, 2011]. This perceived weight should be estimated statistically which was one of the main purposes of the current research.

There were successful examples of applying both effective headways and in-vehicle time crowding weights in one model equilibrium framework [Barquin, et al, 2010]. We also applied two functions (effective headway and crowding in-vehicle time weight) in parallel. However, since only the crowding function is subject to statistical estimation we will be focusing on this component in the current paper.

Transit Service Reliability Measures

Quantification of transit reliability is a new area with only a few published works. There are three main approaches to measure travel time reliability of a transit service [Batley & Ibanez, 2009; Li et al, 2010]:

1. Schedule adherence at boarding stop (expressed in extra wait time at boarding). This measure of reliability can be incorporated in transit assignment and mode choice as an additional wait time component that relates to transit mode, service frequency, cumulative distance or number of stops from the beginning of the line, etc.

2. Impact of congestion (expressed in extra in-vehicle time). This component can also be incorporated in transit assignment and mode choice as a link in-vehicle time function of transit mode, congestion level, and level of separation from the traffic (mixed, bus lane, guided).

3. Combined lateness at destination versus planned arrival time similar to the way how reliability is measured for auto trips, this measure incorporates the first two. Incorporation of this measure in transit assignment requires the first two measures to be implemented.
The concept of (transit) travel time reliability can be extended beyond just extra time components at boarding stops and line segments. Eventually, lateness at the final destination vs. the preferred arrival time should be measured. However, this approach is beyond the scope of the current paper. Also, one can speculate that contrary to the auto time reliability where arrival on time at the final destination is all that matters, for transit, extra wait at the boarding stop can have a specific negative impact on the user perception even if subsequently the bus catches up with the schedule. The suggested extra time components associated with vehicle lateness at the boarding stop and in terms of runtime vs. the schedule can be presented in the SP survey along with the average travel time for each alternative. This way we can estimate user perception of transit service reliability relative to the average travel time.

Some recent estimates for transit travel time reliability can be found in [Bates, et al, 2001; Holander, 2006; Bhat & Sardesai, 2006; Asensio & Matas, 2008; Batley & Ibanez, 2009]. Different studies used very different frameworks for data collection as well as different choice model formulations. Hence, these studies are not easily comparable and reported Values of Reliability (VOR) relate to different definitions of reliability measures. In particular, it is common for a VOR based on the lateness measure to by significantly higher that a VOR based on a standard deviation or other measure of travel time distribution. One minute of being late versus the schedule is much more onerous that just one minute of travel time variability (that also includes cases of being early).

**Survey Design**

The SP survey design was based on the central concept of retrieving an actual transit trip with a full level of detail along with the necessary person and household characteristics with a subsequent construction of several games where the reported trip would be compared to generated realistic alternatives.

The choice of languages was provided in the opening screen which presented text and launch controls in both languages side by side. Once a language was selected an intro screen was presented which described the survey and invited participants to provide an email address in order to be included in a raffle (the survey incentive). Following the promo question the origin and destination geocodes were obtained along with an approximate trip start time. This information was then passed on to Metro’s online trip planner, which also includes other transit systems in the region, in order to generate itineraries for the participant to select. This greatly reduced the amount of data that had to be input by respondents. The survey instrument also allowed participants to manually build itineraries for cases when matching itineraries were not provided.

After an itinerary was selected the user was asked to confirm the travel time of the legs in the selected itinerary (defaults were downloaded from the trip planner) as well as provide information about how late the vehicles were and their crowding levels. Once an actual transit trip was obtained the survey entered the game generation and presentation phase. The generated alternatives were designed to be different from the actual trip itinerary in terms of travel time, cost, crowding, and/or reliability in order to generate the necessary trade-offs between these variables and capture crowding and reliability impacts in units of travel time and/or cost.
The SP survey goes through the following steps, some of which involve respondent responses and other ones represent internal calculations invisible to the respondent:

1. Welcome screen that explains:
   1.1. Choice of language (English, Spanish)
   1.2. Purpose of SP
   1.3. Promotion rules (entering a raffle upon completion)
2. Retrieval of the actual (pivot) trip information through the following steps:
   2.1. Initial input screen:
      2.1.1. Origin, destination, purpose at origin, purpose at destination
      2.1.2. Number and sequence of transit modes and lines
      2.1.3. For each line boarding stop, alighting stop
      2.1.4. Access to the 1st boarding stop, egress from the last alighting stop
      2.1.5. Trip departure time, day of week
      2.1.6. Actual crowding level 1-7 (categories explained below)
      2.1.7. Actual reliability level by 3 reliability measures:
         2.1.7.1. Bus late/early arrival at boarding stop
         2.1.7.2. Bus in-vehicle time delays
         2.1.7.3. Lateness at final destination (last alighting stop)
   2.2. Calculation of derivative LOS variables:
      2.2.1. Travel time components for all trip legs
      2.2.2. Fare
   2.3. Screen for verification/amendment of derived LOS variables
3. Person and household characteristics:
   3.1. Person characteristics
   3.2. Household characteristics
   3.3. Relevant trip circumstances
      3.3.1. Travel alone or jointly and with whom
      3.3.2. Availability of a car for driving or carpooling as passenger
      3.3.3. How the fare was paid, reimbursement by employer
4. Identification of available alternative transit modes/routes and calculation of LOS variables for them based on the reported trip origin, destination, and departure time:
   4.1. If car/carpool was available or used for access or egress the following paths are analyzed:
      4.1.1. Walk to transit or used access/egress car/carpool option in combination with
      4.1.2. Local Bus (LB), Rapid Bus (RB), Express Bus (EB), Transit Way (TW), BRT (BR), Urban/LRT (UR), Commuter Rail (CR)
   4.2. If car/carpool was not available the following paths are analyzed:
      4.2.1. Walk to transit only in combination with
      4.2.2. LB, RB, EB, TW, BR, LR, CR
   4.3. Realistic feasible paths used in SP meet the following criteria:
      4.3.1. Walk access & egress 1 mile or shorter
      4.3.2. Drive access or egress 5 miles or shorter
4.3.3. In-vehicle time by main mode 5 min or longer
4.3.4. Not more than 2 transfers between transit lines
4.3.5. Ratio of total in-vehicle time to number of boardings 5 min or greater

4.4. Realistic feasible paths are ranked by total travel time and compared to the reported pivot trip:

4.4.1. The best alternative is identified (BestSameModeAlt) by the same group of modes (LB/RB/EB vs. TW/BR/UR vs. CR) and access/egress sub-mode (walk, P&R, K&R)

4.4.2. The best alternative is identified (BestDiffModeAlt) by a different group of modes (LB/RB/EB vs. TW/BR/UR vs. CR) but the same access/egress sub-mode (walk, P&R, K&R)

4.4.3. The best alternative by the same group of modes (LB/RB/EB vs. TW/BR/UR vs. CR) and different access/egress sub-mode (walk, P&R, K&R) – replace 4.4.2 only if faster than 4.4.2 by at least 10 min

4.4.4. Both best alternatives are normalized if the total travel time is longer than the pivot alternative by 5 min

5. Screen explaining SP games

6. 9 SP games with trinary choices that have to be ranked by the respondent:

6.1. (Transit route choice) 3 choices between transit options of the same mode group

6.1.1. Pivot that stays the same across all games of this type and fare is assumed the same

6.1.2. Improved alternative (BestSameModeAlt) where at least one attribute (crowding, reliability, total travel time) is better than the pivot and at least one attribute (crowding, reliability, total travel time) is worse than the pivot

6.1.3. The same improved alternative where the same attribute (crowding, reliability, total travel time) is better than 6.1.2 and the same attribute (crowding, reliability, total travel time) is worse than 6.1.2

6.2. (Transit mode choice w/o switching to auto) 3 choices between transit options of different mode groups

6.2.1. Pivot that stays the same across all games of this type

6.2.2. Improved alternative (BestDiffModeAlt) where at least one attribute (crowding, reliability, total travel time) is better than the pivot and at least one attribute (crowding, reliability, total travel time, fare) is worse than the pivot

6.2.3. The same improved alternative where the same attribute (crowding, reliability, total travel time) is better than 6.2.2 and the same attribute (crowding, reliability, total travel time, fare) is worse than 6.2.2

6.3. (Mode choice w/switching to auto) 3 choices between transit modes (applied only for those who have auto driver or carpooling option available):

6.3.1. Pivot that stays the same across all games of this type with one attribute (crowding or reliability) getting progressively worse

6.3.2. Improved alternative (BestDiffModeAlt) where at least one attribute (crowding, reliability, total travel time) is better than the pivot and at least one attribute (crowding, reliability, total travel time, fare) is worse than the pivot

6.3.3. Auto driver or carpooling (whichever reported as available) from origin to destination

6.4. Concluding screen with possibility to list previous screens and revise
Web Survey Design

In order to give the survey website legitimacy it incorporated visual elements of Metro’s public websites. A custom uniform resource locator (URL) that resembled that of Metro’s public website was also obtained, which made it easy for potential participants to remember the URL after seeing it advertised.

Online geocoding and place searches were implemented using version 3 of the Google Maps Application Programming Interface (API) and the Google Local Search API. The geocoding question also included help information in the form of geocoding and search tips which could be made visible by the user. Transit itineraries were downloaded directly from TripMaster, Metro’s regional trip planner, using web services. In order to generate as many alternatives as possible a local instance of Open Trip Planner was setup at GeoStats using Metro's GTFS feed information. The starting and ending stops of the Metro portion of the base itinerary obtained from TripMaster were identified and then used to invoke Open Trip Planner with small time changes in order to increase the number of alternate itineraries shown to participants. The selected itinerary elements were then loaded in a question roster so that it could be recalled by the game generation questions in the survey.

Figure 1: Example of the Geocoding Question

2 Documentation and downloads available at https://github.com/openplans/OpenTripPlanner.
Presentation of Vehicle Crowding Levels and Service Reliability in SP

Crowding in the transit vehicle was quantified by 7 verbal categories associated with probability of having a seat and inability to board when crowding reaches an extreme level as shown in Table 1.

Table 1: Definition of Crowding Levels

<table>
<thead>
<tr>
<th>Crowding level</th>
<th>Probability of having a seat</th>
<th>Verbal description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100% (5 out of 5 trips)</td>
<td>Not crowded</td>
</tr>
<tr>
<td>2</td>
<td>80% (4 out of 5 trips)</td>
<td>Slightly crowded</td>
</tr>
<tr>
<td>3</td>
<td>60% (3 out of 5 trips)</td>
<td>Somewhat crowded</td>
</tr>
<tr>
<td>4</td>
<td>40% (2 out of 5 trips)</td>
<td>Crowded</td>
</tr>
<tr>
<td>5</td>
<td>20% (1 out of 5 trips)</td>
<td>Very crowded</td>
</tr>
<tr>
<td>6</td>
<td>0% (0 out of 5 trips)</td>
<td>Extremely crowded</td>
</tr>
<tr>
<td>7</td>
<td>0% (0 out of 5 trips)</td>
<td>1 out of 5 trips unable to board</td>
</tr>
</tbody>
</table>

In addition to these verbal presentations the question types showing alternatives used graphical representations of vehicle crowding levels, which were selected based on the vehicle type used by the main mode of the presented alternative. An example of the choices screen is presented in Figure 2.

Figure 2: Example of the Alternative Selection Question
Service reliability was presented in terms of the amount of time the vehicles were expected to be late on average followed by the frequency of the time vehicles were late, presented both in terms of the number of time out of five occurrences and as a percentage. Animation was used to fade-in the choices the first time a game scenario was presented.

**Collected SP Sample**

The collected sample after some cleaning and discarding incomplete responses included 2,609 completed person interviews. Depending on the auto availability each respondent participated either in 6 SP games (all compared different transit alternatives if auto was not available) or 9 SP games (6 of them bound to transit and another 3 included an auto alternative if it was available). Each game included 3 alternatives (one of them corresponding to the actually implemented transit trip) that were ranked. This allowed for “explosion” of the choice observations where each game produced two observations. The first one corresponds to the 1st ranked alternative chosen over the 2nd and 3rd ranked alternatives. The second one corresponds to the 2nd ranked alternative chosen over the 3rd ranked alternative. Overall, more than 30,000 observations usable for the model estimation were generated.

Person distributions along most important dimensions like gender, age, income, and student status are presented in Figure 3. Overall, the sample looked very reasonable across all dimensions. It was balanced with respect to gender and surprisingly uniform across a wide range of age groups. Distribution by household income was somewhat biased towards lower income groups but it is quite reasonable for transit users (recall that the survey participant requested to enter an actually implemented transit trip). Finally, we had a reasonably high percentage of students (high school and college) but not a dominant one.

*Figure 3: Person Distribution in the Collected Survey Sample*
Distributions of the reported transit trip along the most important dimensions (purpose at destination, mode, and length) are presented in Figure 4. In general we obtained a reasonable and representative distribution by trip purposes where work constituted about 60% the trip destinations while the remaining 40% of trips were evenly distributed across different non-work purposes. It is interesting to note that the absolute majority of the reported trips were outbound trips originated from home although the survey design allowed for an inbound trip (return back home) equally. In terms of transit modes, the reported trips reflected the actual proportion in the region with majority of trips made by different bus modes but also with quite a significant share of LRT and commuter rail (“MetroLink”) trips. Finally, a reasonable trip length distribution was obtained with respect to in-vehicle time with a mix of short and long trips. Work trips are logically biased towards longer distances than non-work trips.

Figure 4: Distribution of Reported Transit Trips in the Collected Survey Sample
Additional important aspect relates to the reported non-conventional attributes of interest such as crowding levels and reliability measures that are presented in Figure 5. The survey respondents reported a wide range of crowding levels with a substantial number of cases where vehicles were very crowded, extremely crowded, and even impossible to board. This is a realistic description of the situation in LA. The reported reliability levels were heavily biased towards small amounts of delay (5 min) that happen with high frequency (80-100%). There were some reported cases with greater delays and/or more partial frequencies but relatively few. This in general is a realistic portraying of the LA urban transit system that operates mostly with high frequency services; hence delays are frequent but most of them are of a small amount. In the process of generation of SP alternatives, some more extreme cases were pivoted off the reported levels of reliability.

Figure 5: Reported Crowding Levels and Reliability

Overall, we believe that we obtained an adequate dataset for the analysis that was rich enough for the purpose of the current research in term of variety of persons and reported transit trips. Further variation and trade-offs between different transit LOS attributes including in-vehicle crowding and service reliability were introduced by generation the SP alternatives that were compared by the respondents to their actual trips.
Mode and Transit Route Choice Model

The main vehicle for the subsequent statistical analysis was a choice model that was formulated in correspondence with the SP design. By the SP design, the respondents chose between different transit routes of the same mode (primarily games 1-3), different transit modes (primarily games 4-6), and different transit modes and available auto option (games 7-9). This model can be referred to as “mode and transit route choice”. The choice set consisted either of three alternatives (the entire SP game when the 1st ranked alternatives was chosen over the 2nd and 3rd ranked alternatives) or two alternatives (the 2nd ranked alternative chosen over the 3rd ranked alternative). The model structure used was a simple MNL with the following variables used in the utility functions:

- 8 mode-specific constants for auto driver, auto passenger, local bus, express bus, rapid bus, transit way, BRT, LRT, and commuter rail; the first one set to zero as the reference,
- 3 ordering constants to account for the uneven left-to-right perception of the presented alternatives (left alternative that is the actual one, middle alternative that is hypothetical, and right alternative that is hypothetical); the first constant is set to zero as the reference,
- Transit service frequency (for the main line),
- Wait time (for the main line)
- Transit in-vehicle time by mode,
- Transit access time to the main line by mode (walk, bike, drive, driven, other transit by mode),
- Transit egress time from the main line by mode (walk, bike, drive, driven, other transit by mode),
- Number of transfers between transit lines,
- Transfer time (wait and walk),
- Crowding level on the main line,
- Amount of delay (late arrival of the transit vehicle) at the boarding of the main line,
- Frequency of occurrence of amount of delay at boarding,
- Transit fare
- Auto time (for auto trips),
- Amount of delay for auto trips (late arrival at the destination),
- Auto parking cost

In the model estimation, many of the time and cost coefficients were segmented by person type, mode and other relevant characteristics. For example, the in-vehicle time coefficient was segmented by crowding levels. The full model estimation results for different versions of the model specifications were impossible to present within the current paper format. In the subsequent sections, we present the main findings that relate to the crowding and reliability coefficients.

Statistical Analysis and Model Estimation Results: Crowding Effects

Overall, we present six model specifications that correspond to six crowding-related effects reported previously or hypothesized as possible ways to improve the model. In order to facilitate the further analysis we present the crowding weights in a graphical form in Figure 6.
Figure 6: Main Crowding Effects

- **Generic Crowding Effect**
- **Crowding Effect by Trip Purpose**
  - Commute
  - Non-commute
  - Non-commute normalized
- **Crowding Effect by Age**
  - Age under 46 years
  - Age 46+
  - Age 46+ normalized
- **Crowding Effect by Mode**
  - Bus
  - LRT
  - Commuter rail
  - LRT normalized
  - Commuter rail normalized
- **Crowding Effect by Income**
  - Income $0-60K
  - Income $60K+
  - Income $60K+ normalized
- **Crowding Effect by Trip Length**
  - Less than 30 min
  - 30+ min
  - 30+ min normalized

Non-normalized curved represent crowding weights relative to the minimum crowding level for the same segment (for example, commuting to work and non-work trip weights calculated separately). Normalized curves represent crowding weights relative to the minimum crowding level (for example, commuting to work and non-work trip weight calculated relative to the minimum crowding level for commuting trips).
The first specification used a generic formulation where crowding level used as the only segmentation dimension for in-vehicle time coefficient. Overall, the estimation results confirmed the main hypothesis that crowding is perceived by transit users as an extra weight on in-vehicle time that becomes quite significant (1.62) at the extreme crowding level. This number is somewhat lower than the numbers adopted for extreme crowding levels in some previous studies as discussed above (2.0 and higher). However, it is still very significant and affects transit assignment and mode choice results strongly.

The second specification included a segmentation by trip purpose – commuters to work and college where separated from trip for other (non-work) purposes. The original hypothesis was that crowding would be perceived as somewhat more onerous for commuters due to the frequency of the trip while for less frequent trips the users will be more tolerant to (occasional) crowding. This hypothesis was confirmed although the difference between travel purposes was not striking. The most significant difference corresponds to the highest crowding level that the users are more willing to tolerate on an occasional non-work trip but become very negative when it comes to a daily commuting trip.

The third specification included segmentation by person age. The original hypothesis was that younger users might be relatively tolerant to crowding while older users would be more sensitive and crowding-averse. In particular, having a seat should be essential for older users. The estimation results confirmed certain age-related effects. The most statistically significant results were obtained when the transit users were broken into two broad categories – younger users of age 45 or younger and older users of age 46 or older. There are two particular effects intertwined. One of them can be seen when the relative weights for in-vehicle time are normalized versus the in-vehicle time at the lowest crowding level for the same age group. In this case, older users proved to be more sensitive to higher crowding levels than younger users which is expressed in a greater weight (1.65 versus 1.51). The second interesting effect is that the base in-vehicle time coefficient for the lowest crowding level proved to be significantly lower for the older users. This means that while seating, the older users perceived travel time as a less onerous factor compared to younger users. When the ride becomes less convenient and probability of having a seat decreases the perception of time of older users is approaching the perception of younger users.

The forth specification included segmentation by mode. The original hypothesis was that the vehicle design would have a significant impact on the crowding function. Bus curves were hypothesized as relatively flat but starting to “climb up” at relatively low crowding levels since these vehicles are built to make standing convenient. The commuter rail curve was hypothesized to have a longer initial flat part and then a spike at the end when crowding approaches the maximum level since commuter rail cannot accommodate many standing passengers and standing in general is inconvenient. LRT curve was expected to be somewhere between bus and rail. While this initial hypothesis was somewhat confirmed the model estimation results gave some additional insights into the in-vehicle time perception by different transit modes. It proved that there is an overall difference between modes across all crowding levels where more convenient modes like LRT and Commuter Rail are characterized by lower in-vehicle time coefficients than bus. This is an interesting finding that suggests a differentiation of in-vehicle time coefficients by mode that is somewhat contrary to the prevailing practices. Overall, a discounting coefficient of 0.8 for most crowding levels except for very high crowding levels seems reasonable for rail
modes compared to bus. When a relative crowding effect is added on top of this, it manifests itself stronger for rail modes although the curve proved to be not as steep as was expected.

The fifth specification included segmentation by household income groups. The original hypothesis was that transit users with higher income would exhibit more sensitivity to comfort and convenience, hence, a more crowding-averse behavior expressed in a steeper crowding function. This can be supported by the fact that higher-income travelers generally have more alternative options available because of the higher car ownership and higher willingness to pay while many low-income transit riders are captives because the auto for them is either not available or prohibitively expensive (for example, because of the parking cost). Multiple attempts were made with different income brackets to capture a systematic effect but neither of them brought a conclusive and statistically significant difference. The specification reported in this paper included two groups: 1=with a yearly income under $60K, and 2=$60K and more. As can be seen, no significant variation of in-vehicle time coefficient by income was found across the entire range of crowding levels. This finding might look counter-intuitive but it can be explained if the entire combination of behavioral parameters is compared across incomes. This is true, that value of time is strongly correlated with income. However, it does not automatically mean that the in-vehicle time coefficient should be lower for low-income users. It is more behaviorally appealing to assume that the lower value of time for low-income users would rather be a consequence of a higher cost coefficient (i.e. higher sensitivity to cost). In the same vein, there is no particular reason why while having an option to choose low-income users would be more tolerant to crowding and standing. When the willingness-to-pay factor is controlled, the user preferences with respect to convenience (traded against in-vehicle time) proved to be similar across all income groups. When the cost coefficient is taken into account, higher-income users are willing to pay more for convenience but this effect is proportional to their willingness to pay more for travel time savings. The expectation that high-income users would have a somewhat special sensitivity to crowding beyond their overall higher willingness to pay was not confirmed by the data.

The sixth specification included segmentation by trip length. The original hypothesis was that trip length would have a strong effect of the steepness of the crowding function. It is logical to expect that transit users would be tolerant to crowding (and standing) when the trip is short while for longer trips that would try to avoid crowded vehicles (and standing). Multiple statistical trials with different functional forms were implemented to capture this effect. However, the results proved to be either statistically insignificant or inconclusive. An example with trip segmentation by in-vehicle time into two categories: 1=under 20 min, and 2=20 min or longer, is shown in the table to illustrate the typical outcome. The crowding weight proved to be roughly equal and independent of the trip length. While, this result looked originally counter-intuitive and disappointing, it can be explained by the underlying choice model structure. In the choice model context, even if the weight is constant, the resulted crowding effect does grow with trip length because the choice probability is defined by the difference in the utilities, and not by their ratio. Consider an in-vehicle time weight of 1.5 for a certain crowding level. With this weight, 10 min in a crowded vehicle would be equivalent to 5 extra min of travel time while 60 min in a crowded vehicle would be equivalent to 30 extra min of travel time. Thus, trip length would manifest itself in stronger crowding-averse behavior for longer trips even if the weight per min is constant.
Statistical Analysis and Model Estimation Results: Reliability Effects

The SP survey included two variables on reliability – amount of delay and probability (relative frequency in percent) of delay. The first and most straightforward formulation of the corresponding choice utility term is expected average delay calculated as a product of amount and probability. When included in the base model specification along with all other time and cost attributes it proved to be statistically very significant statistically with a coefficient equal to 1.76 of the coefficient on in-vehicle time. This confirms the initial hypothesis that transit users have a strong negative perception of unreliability beyond just the fact that it results in a longer trip time on the average. This can be well explained by the associated (perceived) penalty of being late vs. the planned arrival and also by the “uncertainty” factor.

However, a model with a single variable like expected delay masks some important details of the actual distribution of delays. For example, a delay of 5 min that happens with probability 50% would be equalized with a delay of 25 min that happens with probability 10% since both situations result in an expected average delay of 2.5 min. In reality frequent small delays are perceived differently from infrequent large delays and some important non-linear effects can be lost with a simplified formulation of expected average delay as a single term.

For this reason, we explored multiple different non-linear specifications for both variables – amount of delay and probability of delay. The best function form statistically proved to have three terms with the following coefficients:

\[0.142 \times \text{Delay} \times \text{Probability} + 0.091 \times \text{Delay} \times \text{Probability}^2 + 0.161 \times \text{Delay}^{0.5} \times \text{Probability}\]

This functional form is presented in Figure 7 where the upper part (a) plots this function along the amount-of-delay axes with probability fixed at certain levels while the bottom side (b) plots this function along the probability-of-delay axes with the amount of delay fixed at a certain levels beyond which we did not have enough observations. This functional form suggests that there is a certain convexity with respect to amount of delay and concavity with respect to probability of delay. Convexity with respect to amount of delay means discarding small delays and assigning a greater disutility to bigger delays. Form plots (a), it can be said that delays under 3 min are largely discarded across a wide range of probabilities. Concavity with respect to probability of delay means adaptation by taking a buffer time based on the worst experience. From plots (b) it can be said that adaptation occurs somewhere around probability of 0.5-0.6. This means that if the delay occurs with such a high probability further worsening it (say to a level of 0.8 or 0.9) does not really matter because from the user perspective it is already high enough to make a buffer based on the worst case. Stated otherwise, the probability of delay should be significantly improved to the level of 0.1-0.2 in order to make at least some of the users take a risk of reducing the buffer.
Conclusions

Overall, the web-based SP survey technology proved to be a viable and inexpensive way to collect a unique set of observed transit itineraries as well as evaluate user preferences with respect to in-vehicle crowding and service reliability against a rich set of other LOS and cost variables. We believe that a survey of this type provides a unique opportunity to understand and model transit user choices and preferences.
Based on the collected survey sample that included 2,609 valid complete responses and over 30,000 generated choice observations, a discrete choice model was estimated. This model had in-vehicle time coefficients segmented by crowding levels and reliability-related variables included along with the average travel time components. The estimation results proved to be quite robust and all coefficients of interest came out with a high statistical significance.

With respect to in-vehicle crowding effects, the estimation results confirmed several hypotheses that were stated from the beginning of the study based on the previous research:

- Crowding perceived by transit users as a negative factor expressed as a perceptional weight on in-vehicle time that can reach a value of 1.6 or higher for extreme crowding levels.
- Crowding is more onerous for commuters to work when they experience crowding daily than for occasional trips for non-work purposes.
- Crowding is more onerous for older riders compared to younger riders that is expressed in a higher perceptional weight for older riders (and higher appreciation of having a seat).
- Crowding perceived differentially by mode where bus has a greater (negative) in-vehicle time coefficient while commuter rail has a relative spike only at the extreme crowding levels.

Some originally stated hypotheses on crowding impacts were not confirmed with the following explanation:

- Crowding more onerous for higher income users compared to lower income users. The data suggested that both income groups are equally sensitive to crowding in terms of in-vehicle time coefficient differentiation by crowding levels. The difference between them is in the baseline value of time that affects the cost coefficient rather than time coefficient.
- Crowding weight grows with trip length. While this hypothesis may look the most intuitive it was not actually confirmed. Crowding indeed affects long trips more than short trips but it is fully accounted by the length of the trip with the fixed crowding weight per minute.

With respect to transit reliability effects, different model specifications were explored with such available variables as amount of delay and probability of delay. The simplest specification that was adopted for the model implementation is based on a product of these two variables that represents an expected average delay. This variable proved to be very significant statistically with a relative weight of 1.76 compared to the base in-vehicle time. This confirms that unreliability is perceived by transit users as a strong negative factor beyond just extra trip time involved. Additionally, some finer non-linear effects were explored that showed that transit users tend to discard very small delays and respond more to bigger delays. They also tend to adapt to frequent delays due to taking a buffer time in their scheduling decisions.

The developed crowding functions and unreliability estimates were incorporated in the operation travel model used by LACMTA. Application of the developed methods would be very beneficial in practice and will help evaluate many transit projects in a more comprehensive way. In particular, transit capacity
relief projects in major metropolitan regions where transit services are already overloaded would be portrayed in a very different way since a significant additional portion of User Benefits would be added.

References