Issues on the Design of Stop-Level Transit Demand Models

Summarized by Zhong-Ren Peng

Revised August 20, 1997

Basic Unit of Analysis: Route, Route-Segment or Stop Level Models

Transit ridership varies at the route, route segment and bus stop level, but transit level of service varies at the route or route-segment level, at least from the viewpoint of service planning. So transit models that estimate demand and service supply can be developed at the route, route-segment and stop level.

Previous studies used route or route segment as a basic unit of analysis in transit demand models. These studies represent improvements over the system-wise models by recognizing the variations of ridership and service supply among routes and route segments. The underlying assumption of these route or route-segment level models is that the social-demographic characteristics along the route or route segment are homogenous. But as is well known, this assumption is invalid in most cases. Transit ridership fluctuates among different route segments and stops because of the different land uses or trip generators surrounding transit stops. It is difficult for route and route-segment level model to take into account of stop-level land use characteristics such as pedestrian environment. Furthermore, adjacent transit routes compete with each other for passengers. At the transit route or route-segment level, these competing effects are controlled by the percentage of overlapping service area over the entire route or route segment. This is only a surrogate because transit services compete with riders at the stop.

Bus stop seems to be the better unit of analysis for transit demand model. With stop-level models, the stop-specific variables can be incorporated into the models. These stop-level variables include stop amenities such as shelters, lighting and patrols, and pedestrian environment such as sidewalks and slopes. But the lack of stop-level data hinders the development of stop-level model. The installations of the automatic passenger counter (APC) provider a good source of ridership data at the stop level. The use of GIS provides a technique to allocate and integrate social-demographic data and transit service data at the stop level. The Automatic Vehicle Locator (AVL) provides an accurate measurement to examine bus on-time performance at each time points by comparing the schedule time with the real bus arrival and departure time at the time points. However, the on-time performance cannot be accurately assessed on each bus stop because there is no scheduled time on every stop.

Therefore, on-time performance at non-timepoint stops needs to be assumed as the same as the on-time performance at previous time point. Alternatively, model can be developed at the time point level rather than at the stop level. But the time points are sparsely located on the bus route. They usually represent some major intersections and land use attractions. Ridership at those time points may not represent those in other non-timepoint stops. Thus modeling transit demand at the time-point may introduce bias. One way to reduce this bias is to aggregate ridership of stops that surround a time point to that time point. This will downgrade the stop-level model to the route-segment model, with the segment being smaller than the previous route-segment level model. This seems to be avoided in this stop-level study.

It should be realized, however, that the variations of transit level of service do not occur at the stop level. Transit service usually varies at the route or at most the route segment level. Therefore, the service variable included may not have much variance, and may thus turn to be insignificant in the model output. Even the effective head at stops may not have enough variations for the variable to be statistically significant in the model. On-time performance is mostly a measure of route performance rather than the service performance at the stop level. Furthermore, sometimes the on-time performance is often negatively associated with ridership. More riders usually take the bus longer to load and unload passengers, and therefore negatively affect the bus on-time performance. Therefore, there are confounding effect of on-time performance and ridership. On the one hand, a better on-time performance may increase ridership, while on the other hand, a ridership increase may negatively effect on-time performance. So the coefficient of the on-time performance in the model may be positive, negative or insignificant.

Therefore, the stop-level model is more useful to relate transit demand with demographic and land use characteristics, and it is of more interest to service operation controls. The better understanding of transit demand and land use and demographic characteristics can help transit service planning to better plan services at the route or route-segment level. A good design of stop-level model is essential to serve the needs of ridership estimates, service planning and operation controls.

Models for Service Planning and Operation Controls

Service planning is most concerned with transit ridership and level of services (LOS). The level of service is designed to vary across routes and time in responding to passenger demand variations. The level of service is usually planned at the route or route-segment level. Operation controls, on the other hand, are concerned with the smoothness of operations. That is, the goal of operation controls is to reduce the variations of effective headways between buses. The focus of operation controls is the effective headways at the bus stops.

Therefore, different models are needed to model the behaviors of transit service planners and operation controllers. For service planning, the model is best estimated at the route or route segment level. It is in the following general form:

Ridership = F(LOS, riders social-demographic data, land use data)
LOS = F(Ridership, Land use data)

For operation controls, the model is best estimated at the stop level, and is in the following general form:

Ridership = F(Effective headway, on-time performance, riders social-demographic data, land use data)
Effective headway = F(Load factor, ridership, on-time performance of the previous bus)

A composite model may be constructed to consider the needs of both service planning and operation control as follows. Notice this is a recursive model between Ridership and LOS. The model has to be estimated on at least two stages. That is, the first stage is to estimate the simultaneous model between Ridership and Effective headway at the stop level. The second stage is to aggregate the estimated ridership to the route level, and to estimate the LOS model at the route level.

LOS = F(∑Ridership, Land use data) (-- at the route level)
∑Ridership = F(Effective headway, on-time performance, riders social-demographic data, land use data) (-- at the stop level)
Effective headway = F(Load factor, ridership, on-time performance of the previous bus) (-- at the stop level)

Models Structure

With the abundant ridership data at the stop level, route-specific models can be developed for individual routes. But the question is whether we need to develop models for individual routes. For the purpose of service planning, models for each route typology such as cross-town route, feed routes, radial routes and express routes should be sufficient.

For each route typology, there needs to be different models for different time periods and directions (inbound and outbound) to capture the temporal variations of ridership. These time variations include time of day (peak, midday, evening and night) and day of week (weekday, Saturday and Sunday). Furthermore, models may be needed to estimate boarding and alighting.

Ridership Data Sampling

The model development on route typology and time of day involves spatial and temporal data sampling.

Spatial data sampling

Decision has to be made as to what routes need to be selected as representatives of a route typology. Once the representative routes are chosen, stops need to be selected. One option is to use all stops at every selected route in the model estimation. The advantage of this approach is that there is no sampling error. But the disadvantage is that the service area of adjacent stops overlap, more allocations are needed for social demographic variables. Another option is to use sample stops. The advantage of using sample stops is the reduction of allocation of social demographic variables; only do selected stops need to do social demographic data allocation. But the disadvantage is that there may be some sampling errors. A carefully designed sampling technique such as stratified sampling needs to be developed to make sure that the sample stops represent the stop population.

Temporal data sampling

To capture the temporal variations (hourly, daily and seasonal variations) of ridership, data need to be selected from different time periods. The hourly variation of the same time period such as midday, morning peak and afternoon peak can be captured using the ridership distribution (e.g., mean and standard deviation) of that time period. The daily variations of the same time period can be captured using the ridership distribution of days of the week. And the seasonal variations of the ridership are captured using the ridership distribution of the months in seasons.

Therefore, there needs to be weekly data for at least one month in every season, such as the second week of the month of February, May, August, and November. The weekly variations and seasonal variations (or seasonality) can be indexed using moving average. The purpose of indexing weekly and seasonal variations is to add the variations in ridership forecasting.

For the purpose of data collection in Tri-Met, the full APC data set of every APC-equipped route in the four-week time period is recommended to collect.

Spatial and Temporal Autocorrelation

The data set to be used in the model is a pool of spatial and temporal data. So the model estimation involves both spatial and temporal autocorrelation. Ridership of the bus stops at the same route is autocorrelated. And ridership of the same stop at different time periods (e.g., days and months) is also autocorrelated. There is, however, no known method to deal with the spatial and temporal autocorrelations. A mechanism need to be developed to remedy these two autocorrelations.

However, the temporal autocorrelation can be eliminated if the ridership of a stop to be used is averaged across days and months. Spatial autocorrelation cannot be eliminated unless route-specific model is estimated.

Spatial Data Allocations

Like the route and route-segment level model, the stop-level model requires that social demographic and land use data be allocated to individual bus stops. A fifteen-minute isochron (or a quarter mile walking distance) has to be defined as service area using the allocation function in Arc/Info. Parcel or block level data (or grid data) from RLIS (regional land use information system) are recommended to use. But car ownership and income data need to be derived from census block groups.

Special cautions need to be paid to spatial allocation on overlapping service areas. For overlapping service area on adjacent stops of the same route, the overlapped service area needs to be equally divided among overlapped stops. However, for overlapped service area on adjacent stops or the same stop at different route, the allocation method depends on the relationship of the routes (see below).

Competing And Complementary Effects Among Competing Routes

It is a common phenomenon for one bus stop to serve multiple transit routes. Passengers have a choice of getting to one of many bus routes if these bus routes serve the same destination of the riders. In other words, these bus routes compete with each other for passengers. On the other hand, the same bus stop may serve as a transfer point for passenger to transfer from one route to another if bus routes serve different destinations. In other words, these bus routes are complementary with each other. It is difficult to identify whether two routes are complementary, competing, or independent without knowing passengers origins and destinations.

Since there is no information in passengers origins and destinations, some ad hoc criteria must be used to identify the complementary and competing nature of routes. One criterion may be the sharing of the origin and destination ends. Routes that have the same destination end, such as ending at downtown transit center may be considered as competing routes; while routes that have different origins and destinations may be considered as complementary routes. An another criterion is the amount of overlapped service areas. If the overlapped service area of two routes (not stops) is over 50 percent of that of the entire route, these two routes could be considered as competing routes.

Furthermore, service planners in Tri-Met need also to be consulted since they know the route system well.

Bus stops serving competing routes have two effects on ridership. It may reduce ridership on individual buses at a particular time because of the competing effects. But it may increase the ridership on the route because the increased service frequency tends to attract more transit users (synergistic effect). So in the model, two variables can be used: the number of competing routes the stop served and the number of complementary routes the stop served. But the sign of the coefficient of the first variable may be positive or negative depending on the relative strength of synergistic and competing effects. A bus stop that serves complementary routes should have positive effect on the ridership. The use of these two variables reduces the need to proportionately allocate the served population and employment to each route that goes through the same stop.

Other Issues

Lagged effects of on-time performance

If a bus is earlier than the scheduled time, it may miss some passengers who have to take the next bus. This lagged effect must be considered to estimate ridership on individual bus route at a particular time. But it can be ignored if the average ridership is estimated for a particular route at a time period, especially if a random snapshot of one bus sample is drawn from a time period. Furthermore, this effect can be ignored if the aggregated data are used.

Long term effect of on-time performance on ridership cannot be estimated without a long-term time series data set.

Load Factor

A full-load bus will deter more passengers to get on the bus. A load factor, a ratio of passenger to seat capacity, can be calculated if we have detailed data for boarding and alighting. Unless the loading factor is 100 percent, it will have no impact on ridership. So a dummy variable is only needed to represent the load factor.

A side benefit of load factor is that we can gain knowledge about the passengers travel length by making some assumptions. The assumption may be the passenger who get on the bus first will get off the bus first. The benefit of being able to calculate trip length is to calculate service output in terms of passenger miles. But it will have little impact on stop-level transit demand modeling.

Accessibility to and amenities of transit stops

Accessibility to transit stops is defined by three variables: walking distance, slopes, and the availability and connectivity of sidewalks. Walking distance and slopes can be calculated on the street patterns using Arc/Info. Amenities on transit stops include the availability of lighting, shelters and security patrols.