Discrete Choice Models

Overview of Choice Theory

Transportation demand can be characterized as the aggregation of the decisions of all individual trip makers within a metropolitan area. A number of techniques for predicting demand directly at the *aggregate level*. An alternative approach, which first emerged in the 1970, is to model directly the decisions process of individual trip makers and then sum over all trip makers in order to obtain the aggregate demand predictions typically required by the evaluation process.

The most common starting point for individual choice models is the notion of utility maximization. That is, decisions makers are assumed to assign at least an ordinal ranking to the trip alternatives available in terms of their relative desirability or utility. Being a rational person, the decisions maker will choose the alternative with the maximum utility – the one that maximizes benefits. Utility maximization is central to microeconomic theory but is not restricted to it in its applications. In particular, given the derived nature of transportation demand, it seems reasonable that travelers will want to minimize travel time and cost, maximize comfort and convenience, and so on, whenever possible.

In this lecture, utility simply represents a convenient generalized function that accounts for the positives and the negatives involved in trip making and that forms the basis for a traveler's decision.

Conventional microscopic theory assumes that traveler is able to use perfectly all of the trip information available and relevant to the decision and to make a completely rational, consistent decision give this information. A major relaxation of these assumptions is possibly by introducing the concept of *random utility*. Primarily originating in the field of psychology, such models represent an attempt to retain the analytical tractability provided by economic assumption of a human being as a rational utility maximizer within a more flexible or realistic world view.

These models recognize that, in practice, people do not always choose the objectively best course of action, nor do they necessarily exhibit consistent choices over time. That is, random utility theory still assumes that an individual will choose the alternative that appears to maximize his or her utility when the choice is being made. This utility is assumed consist of two components:

1. The systematic, observable utility that is identical to the conventional microeconomic utility function.
2. A random term that is intended to capture such effect as variations in perceptions and tastes of individual trip makers, misspecifications of the utility function by the analyst, and measurement errors on the part of the analyst.

If one can assume that this random term enters the utility function additively, then the utility of some course of action I for individual t can be expressed as:

$$U_{it} = V_{it} + \epsilon_{it} \qquad 1$$

Where:
U_{it} : Random utility of alternative I for individual t.
V_{it} : Systematic (observable) portion of utility.
$\in_{it}$: Random portion of utility.

Further, the systematic utility V_{it} is assumed to be a function of the attributes of the alternative X_i and the characteristics of the individual S_t. In particular, it is typically assumed for reasons of analytical tractability that V_{it} is given by:

$$V_{it} = b_1 Z_{it1} + b_2 Z_{it2} + \cdots b_n Z_{itn} \qquad 2$$

Where:
b: row vector of parameters

$$Z_{it} = f(X_i, S_t) \qquad 3$$

If the modeler could observe the value of the random terms for any given decision maker, these values would be incorporated within the systematic or observable portion of the utility function and would no longer be treated as random. However, with randomness incorporated into the decision- making formulation, the modeler cannot say with certainty which alternative will have the maximum utility for a specific decision maker and thus which alternative will be chosen. What can be assessed is the probability that a given alternative i from a set of alternatives available to individual t will be the maximum utility alternative for that individual and hence be chosen. That is, given by Equ. 2 and given a set of alternativesC_t, the probability of individual t choosing alternative I from this set of alternatives (P_{it}) is:

$$P_{it} = P(U_{it} \geq U_{jt}) \qquad \forall j \in C_i \qquad 4$$

Or substituting Equ. 1 in to Equ. 4,

$$P_{it} = P(V_{it} + \in_{it} \geq V_{jt} + \in_{jt}) \qquad \forall j \in C_t \qquad 5$$
$$= P(\in_{jt} - \in_{it} \geq V_{it} - V_{jt}) \qquad \forall j \in C_t$$

Equ. 5 is an expression for the joint cumulative distribution function of the random variable $\in_{jt} - \in_{it}$ evaluated at the points$V_{it} - V_{jt}$. Thus, if the distribution of the ∈ ’s is known or assumed, this equation can be used to compute the probability of an individual making a given choice. Perhaps the most obvious assumption to make is that the ∈ ’ s are distributed multinomial normal. This assumption generates what is known as a *probit* model.
Unfortunately, multinomial probit models cannot be expressed easily in an analytically closed form and hence are computationally cumbersome and expensive to use.

An alternative assumption concerning the distribution of the $\in$ ' s is that they are each independently and identically distributed (iid) with a Gumbel Type I distribution whose cumulative distribution function is given by:

$$F(w) = e^{-e^{-w}} \qquad 6$$

Choosing this particular distribution is motivated entirely by considerations of analytical convenience, since when Equ.6 is integrated in order to evaluate Equ. 5. It can be shown that final expression for P_{it} is the multinomial *logit* model given by:

$$P_{it} = \frac{e^{V_{it}}}{\sum_j e^{V_{jt}}} \qquad 7$$

As an example of a multinomial logit model, consider a three- mode choice situation in which a worker must choose between auto, bus, and walking for the journey to work. The systematic utility functions associated with these alternatives might take the form:

$$V_{auto} = 1.0 - 0.1(TT_{auto}) - 0.05(TC_{auto}) \qquad 8$$
$$V_{bus} = -0.1(TT_{bus}) - 0.05(TC_{bus}) \qquad 9$$
$$V_{walk} = -0.5 - 0.1(TT_{walk}) \qquad 10$$

<u>Where:</u>
TT_i = travel time by model i, minutes
TC_i = travel cost by mode i, dollars

Assume that a given individual is faced with travel times of 5, 15, and 20 minutes for the auto, bus, and the walk modes, respectively. Similarly, out –of-pocket travel costs by auto and bus are:

V_{auto}= 0.42 V_{bus}= -1.575 V_{walk}=-2.5

Substituting these values into Equ. 7, the probability of this worker choosing the auto mode is:

$$P_{auto} = \frac{e^{0.42}}{e^{0.42} + e^{-1.575} + e^{-2.5}} = \frac{1.522}{1.811} = 0.841$$

Similarly, $P_{bus} = \frac{0.207}{1.811} = 0.114$

And $P_{walk} = \frac{0.082}{1.811} = 0.045$

Characteristics of the Logit Model

The logit model has a tractable and convenient functional form. In particular, its parameters can be statistically estimated relatively easily and efficiently using fairly standardized maximum likelihood techniques. Major characteristics and issues associated with the use of this model include:

1. The independence of irrelevant alternatives assumption
2. Representation of the individuals decision-making structure
3. Specification of the utility function
4. Aggregation of predictions
5. Data requirements
6. Model transferability

Independence of Irrelevant Alternatives

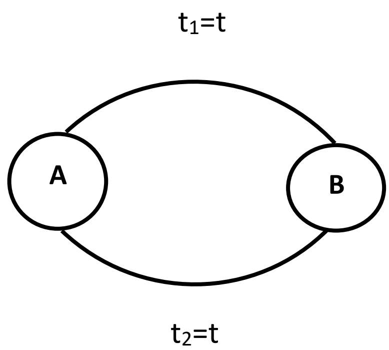
The logit model belongs to a class of models that possesses the so –called independence of irrelevant alternatives (IIA) property. This property can illustrated most easily by observing from Equ. 7 that the relative probability of an individual t choosing alternative I rather than j, another alternative in choice set is, is:

$$\left(\frac{P_{it}}{P_{jt}}\right) = \frac{e^{V_{it}}}{e^{V_{jt}}} \qquad 11$$

The key point to note about Equ. 11 is that the relative probability of choosing i rather than j depends only on the characteristics (utility) of the alternatives I and j. That is. It is independent of any other alternative that might be available. Further, as long as the values of V_{it} and V_{jt}do not change, this relative probability will not change, regardless of weather other alternatives are added or deleted from the choice set.

The IIA property is both one of the strengths of the logit model and its major weakness. The property is advantageous in that it means the model can be estimated based on one choice set and then used to predict choices from a modified choice set. Thus, for example, a mode split model can be estimated based on currently available modes and then used to examine the impact of the introduction of a new mode into the system. The property can be also exploited in cases where the choice set is potentially very large (e.g., shopping destination choice, residential location choice, etc.) to eliminate the need for explicitly including the entire choice set in the calculations. That is, a subset randomly selected from the overall choice set can be used to generally statistically consistent estimation and predictions results.
The problem with the IIA property is that care must be taken to ensure the alternatives included in the choice set are, indeed, independent of each other. Figure 1 provides a case in which the independence assumption is violated, with disastrous results. Figure 1 a presents a simple route choice problem in which two route with equal travel times are available and in which the probability of either route being chosen is clearly 0.5. Figure 1 b presents a modification of the first in which one route has been split into two sub routes that are

identical except for an arbitrary small link at one point. The travel times on all three routes remain equal. Obviously, this arbitrarily small changes in the network should have no practical effect on the system state: there are still essentially two real routes available, and the traffic should spilt evenly between them. As shown by Figure 1 b, however, a simpleminded application of the logit model to the second case results in a prediction of 0.33, 0.33, 0.33 for three routes, or a 1/3 to 2/3 split between the two real routes. This is a direct result of the IIA property (note that the ration P_1/P_2 equals 1.0 in both cases; that is independent of what other alternatives are available) or rather a direct result of applying the logit model to a choice set that clearly violates the IIA assumption. Alternatives 2 and 3 are not independent of each other; rather they are highly dependent, and the probability of the choosing one is highly correlated with the probability of choosing the other.

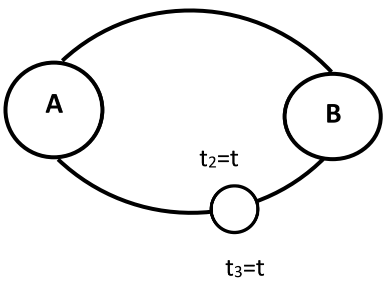


$$V_r = -t_r \; for \; route \; r$$

$$P_r = \exp(V_r) \,/ \sum_r expV_r$$

$$P_1 = \frac{e^{-t}}{e^{-t} + e^{-t}} = 0.5 = P_{2r}$$

a)



$$P_1 = \frac{e^{-t}}{e^{-t} + e^{-t} + e^{-t}} = 0.33 = P_2 = P_3$$

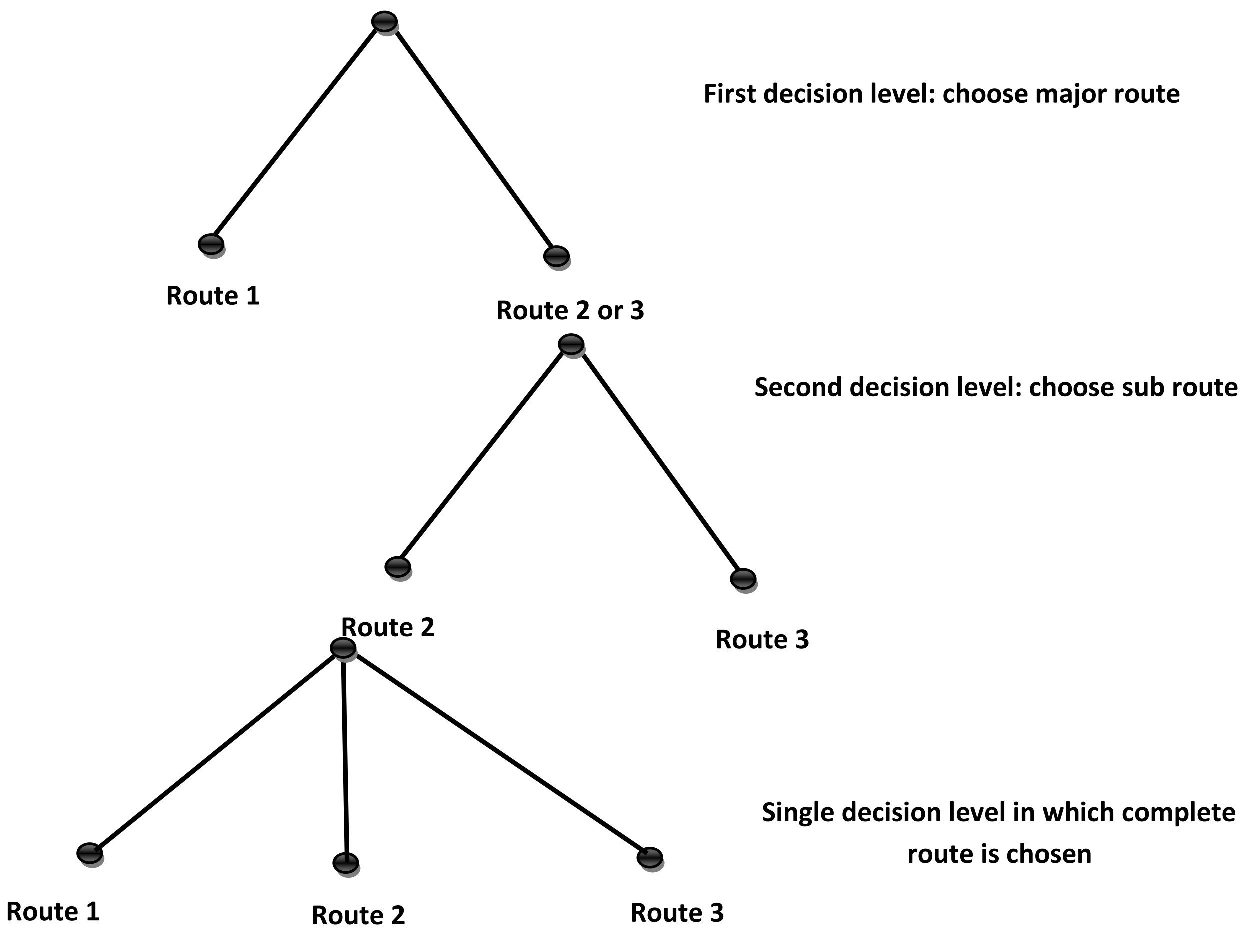
b)

Figure 1: Examples violation of the independence of irrelevant alternatives property: a) sequential decision process b) joint decision process.

Decision Structure

One approach for resolving the IIA violation in the route choice problem previously discussed is to consider the problem as a two stage decision process in which choice is first made between the two major routes, and then a second choice is made, if required, between the two sub routes. Figure 2 presents the decision tree representation of the two stage or sequential, process and contrasts it with the corresponding one stage, or joint, decisions process previously discussed.

In any complex choice situation, a number of decision structures are generally conceivable. The UTMS implicitly assumes a sequential process consisting of decisions concerning whether to make a trip, where to go given that trip is made, what mode to use given the trip destination, and what to go given that trip is made, what mode to use given the trip destination, and what route to use through the chosen modes network to reach the chosen destination. An alternative decision structure is to assume that the decisions of whether to make a trip, where to go, and what model to use to get there are all made simultaneously; that is a joint decision process exists.



a)

b)

Figure 1: Alternative decision structures for a route choice problem: a) sequential decision process b) joint decision process.

The choice hierarchy is shown in Figure 3 for representing travel decisions. Higher level decisions in choice hierarchy are made prior to lower level decisions, which in turn are conditional decisions based on the higher level choices. Thus, non-work travel decisions are assumed to depend on prior work trip decisions that, among other things, determine the number of household autos that will be available for non-work trips. Decisions within each level are generally assumed to be made jointly, although sub hierarchies are conceivable.
The determination of what choice structure to adopt is primarily based on theoretical grounds, although data availability, problem context, and calibration issues can also play a role. The key point, however, is that an assumption of a decision structure is exactly that – a behavioral assumption concerning the trip makers decision-making process. As such, its validity should be tested to the extent that is possible.

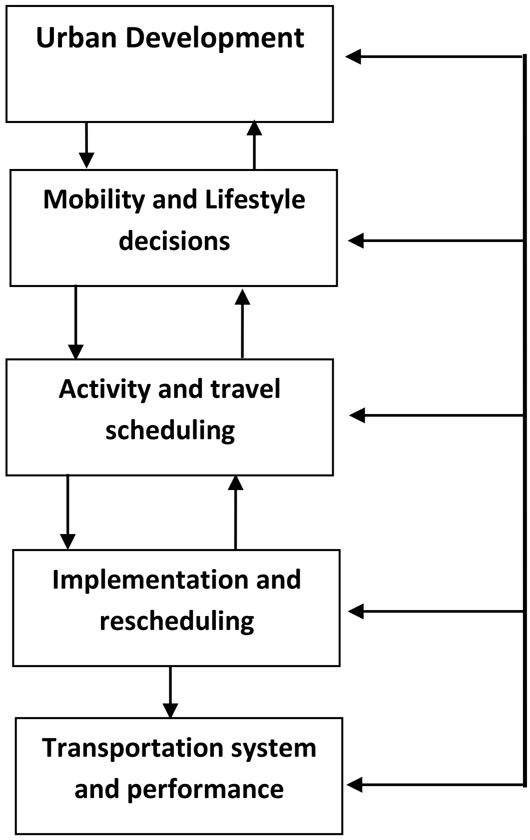


Figure 3: Choice hierarchy for travel decisions.

One approach to empirically testing decision structure hypothesis, as well as to providing an alternative decision structure "in between" the pure joint and pure sequential structure discussed to this point, is to adopt the so called nested decision structure.
In a nested structure, decisions are still made sequentially but a higher level decision (i.e. one made early in the decision process) may include in its calculations expectations concerning subsequent lower level decisions (i.e. ones made later in the decision process). In particular,

the expected maximum utility associated with the next stage in the decision process is included in the current stages utility function.

Mathematically, the nested logit model decomposes into two ordinary logit models. If, for example, one models the choice of mode m and destination d for shopping trips, a typical nested logit model for this process would be:

$$P_{m/d} = \frac{e^{\frac{V_{m/d}}{\mu}}}{\sum e^{\frac{V_{m/d}}{\mu}}} \; \mu \;\acute{\in} M_d \qquad 12$$

$$P_d = \frac{e^{\frac{V_d+\mu I_d}{\mu}}}{\sum e^{(V_d+\mu I_d)}} \; \mu \;\acute{\in} M_d \qquad 13$$

<u>Where:</u>

P_d= Probability of choosing destination d from the choice set D.

$P_{m/d}$= Probability of choosing mode m from the choice set M_d given that d has been chosen as the trips destination.

I_d = Inclusive value for destination d= expected maximum utility associated with choosing a travel model, given the choice of destination d

$$= log_e\{\sum e(V_{m/d}/\mu\} \;\; m \in M_d \qquad 14$$

μ $= scale\ parameter\ (0 \le \mu \le 1)$

Eqs. 12 and 13 can be estimated sequentially as two separate logit models, using ordinary logit model estimation software (with loss of statistical efficiency), or simultaneously, using commercially available software such as ALOGT and LIMDEP. Nested logit models are very commonly used for modelling mode choice, both for implementation within urban UTMS modeling systems and for use in intercity travel demand modeling applications.

Nested logit models are actually special cases of an even more general class of models known as generalized extreme value (GEV) models. Considerable experimentation with various, more complex GEV models underway, since they possess the potential to deal with a variety of more complex choice situations for which even nested logit models are inadequate. The mathematical complexity of these models, however, is quite high, and software to support their application in practical planning contexts does not yet exist.

Utility Function Specification

In the theoretical development of the choice model, it is simply assumed that a systematic utility function V_{it} exists for each individual t and alternative i. In practice, the specification of this utility function constitutes a major task in the model building process. While it is possible to develop utility functions for each individual, conventional practice involves either categorizing individuals into relatively homogenous groups and then developing utility functions for each group or developing generalized utility functions within which an individual's socioeconomic characteristics enter directly.

Variables within a utility function can either genetic or alternative specific in nature. A general variable is one that is included in every alternative utility function with exactly the same weight (i.e. the same parameter value). An alternative specific variable, on the other hand has a different weights for different alternatives, including an a priori specified weight of zero (i.e., it does not enter into a particular alternatives specific constant or bias term that is often employed to capture systematic, "all else being equal" preferences exhibited within a sample.

Some of these concepts can be illustrated with a simple modal choice problem consisting of two modes, auto (a) and transit (t). Transportation variables chosen to characterize the system are in vehicle travel time (IVTT), out-of-vehicle travel time (OVTT), and out-of-pocket travel costs (OPTC). Two socioeconomic variables are used to characterize each traveler:

- Household income (INC)
- Household auto ownership (AO)

A simple modal split model using these variables might be specified by the following functions:

$$V_a = \beta_1 + \beta_2 IVTT_a + \beta_3 OVTT_a + \beta_5\left(\frac{OPTC_a}{INC}\right) + \beta_6 AO \qquad 15$$

$$V_t = \beta_3 IVTT_t + \beta_4 OVTT_t + \beta_5\left(\frac{OPTC_t}{INC}\right) \qquad 16$$

Several points concerning Eqs 15 and 16 should be notes:

1. OVTT and the composite variable OPTC/INC are generic variables because they enter both utility functions with the same parameter value (that is, $\beta_4$and β_5, respectively).
2. IVTT is alternative-specific variable because it enters the two equations with different weights (that is, β_2 andβ_3). This reflects the hypothesis that a minute spent riding a bus is perceived (weighted) differently than a minute spent driving in a car.
3. While the utility functions are "linear in the parameters", nonlinear composite variables (such as the OPTC/INC) can be included.
4. β_1 is an alternative- specific constant for the auto mode. No transit constant is specified (or, more correctly, the transit constant is arbitrarily set equal to zero)

because in a choice set of n alternatives, at most n-1 alternatives are statistically identifiable.

5. Because socioeconomic characteristics for a given individual do not vary across alternatives, socioeconomic variables must enter the utility functions as alternative-specific variables (such OA) or generically in functional combination with a system variable (e.g., the OPTC/INC term). That is. A generic socioeconomic variable will add exactly the same value to every alternatives utility function and will thus have absolutely no impact on the choice probability.

Aggregation

Individual choice models generate predictions of the probabilities associated with given individuals choosing a particular outcome from a set of alternatives. As such, these probabilities are of little direct use for planning process. That, is a planner is rarely interested in the probability that a specific individual will choose transit, but rather the total number of people in a zone or in a study area likely to choose transit. Thus, some procedure must be employed to aggregate the individual choice predictions of model to yield the total demand predictions required for planning purposes.

In principal, the simplest aggregation procedure is to enumerate all individuals within the study area and sum their probabilities of choosing a given alternative. While such a *total enumeration*, as this procedure is called, is being experimented with within emerging microsimulation models, it is not currently represent a practical approach within most operational planning environments. Some other aggregation procedure is generally required. While a range of procedure exist, the three most commonly considered are:

- Native aggregation.
- Classification with native aggregation
- Sample enumeration

Native aggregation involves treating the individual choice model as if it were an aggregate model by using zonal average values for the utility function variables in order to compute an "average" zonal probability. Because logit model probabilities are nonlinear functions of the utility function variables, however, such use of average values will not generate the correct average probability. The errors that can occur through the use of native aggregation can be substantial, and one should avoid the use of this technique whenever possible.

Aggregation errors can generally be reduced if the population is classified into relatively homogenous groups with respect to one or more key variables prior to the use of native aggregation.

While the total enumeration is generally impractical, if not infeasible, very often a reprehensive sample of the population is available to the analyst. In such cases, this sample can be enumerated, and the resulting sample prediction can be used as an estimate of the population prediction (with appropriate "grossing up" if required).

For short-run analysis, the calibration data set or in the absence of a current sample. A sample can often be synthesized from census data, zonal population forecasts, and so on, if reasonable assumptions about the distributions of sample characteristics can be made.

Provided that a reprehensive sample is available or can be reliably generated, sample enumeration is generally the preferred aggregation procedure.

Data Requirements

Individual choices models are more efficient than corresponding aggregate models in their use of data. This is because aggregate models typically employ zonal averages, which require a fair number of observations to construct, whereas individual choice models (and disaggregate models in general) employ every observation directly in their calibration. Thus, individual choice models require fewer observations to construct, for a given level of accuracy. Given the high cost of data collection and the very large samples typically gathered in the past to construct aggregate models, this is a very important strength of individual choice model technique.
The nature and the detail of the data required by individual choice models, however, are often greater than that collected for aggregate models. A wider range of socioeconomic information and more detailed representation of the level of service variables (including service characteristics for “unchosen” alternatives) experienced by the observed travelers are typically required. Further, this more detailed information must be available for the future situations for which predictions are required. As a rule, then, while individual choice models require less quantity of data (in terms of the number of observations in the sample), they often require higher data (in terms of the informations obtained per observation).

Modal Transferability

Because individual choice models are not tied to a specific zone structure within a specific city and because they possess the potential for a relatively rich representation of the factors affecting a traveler’s decision making, it has been argued that such models should be capable of being transferred from one geographical location to another. The benefits of such transferability would be enormous in that it would significantly reduce in a number of cases the data requirements, calibration time and costs, and detailed analytical expertise required by planners to perform demand analysis. Considerable research has gone into investigating the transferability properties of discrete choice models within practical planning contexts. While no universally transferable model exists, logit mode split models can, with care, be transferred from one “context” (e.g. one city or one time period) to another, particularly if some updating of the model parameters is performed, using either available aggregate data that is used to adjust the model constants within the model or a small local sample, in which case all or most of the model parameters can be adjusted.