TRANSPORT PLANNING AND MODELING

TRIP GENERATION – 1^{ST} MEETING

Trip Generation Introduction

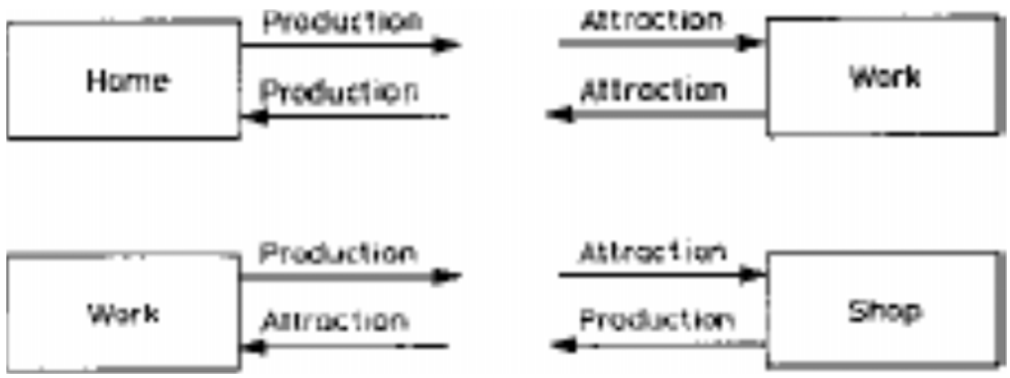
- The trip generation stage of the classical transport model aims at predicting the total number of trips generated by (O_i) and attracted to (D_j) each zone of the study area.
- This can be achieved in a number of ways:
    1. Starting with the trips of the individuals or households who reside in each zone or,
    2. Directly with some of the properties of the zones: population, employment, number of cars, etc.
- The subject has also been viewed as a trip frequency choice problem: how many shopping (or other purpose) trips will be carried out by this person type during a representative week?

Trip Generation Terminology

- Journey (a.k.a. trip): one-way movement from a point of origin to a point of destination to satisfy the need or demand for activity
- Home-based (HB) Trip: Home is the origin or destination
- Non-Home-based (NHB): Neither end of the trip is the home of the traveler
- Trip Production: Home end of a HB trip or origin end of a NHB trip

Trip Generation Terminology (continued)

- Trip Attraction: non-home end of the HB trip and the destination end of the NHB trip
- Trip Generation: total number of trips generated by households in a zone (HB and NHB), where the task remains to allocate NHB to various zones
- Trip chaining: multiple trips are performed in sequence as a matter of efficiency, performing several activities



Classification of Trips—Trip Purpose

- Homebased (HB)
  - Work (HBW)
  - School (HBS)
  - Shopping (HBSH)
  - Social and recreation (HBR)
  - Other (HBO)
- Non-homebased (NHB)→not classified into categories

Classification of Trips—Person Type

- Income level
- Car ownership
- Household size
- Household structure
  - group housing
  - single
  - family-head
  - family-worker

Trip Generation Studies

- Household based

- Zonal based

Factors affecting Trip Generation—Personal Trips (Production)

- income
- car ownership
- household structure
- family size
- value of land
- residential density
- accessibility

Factors affecting Trip Generation—Personal Trips (Attraction)

- office space
- commercial space
- educational space
- number of employees
- type of employment (e.g., government, retail, industrial)

Trip Generation Modeling—Growth Factor Modeling

Growth-factor Modelling

- Since the early 1950s several techniques have been proposed to model trip generation.
- Most methods attempt to predict the number of trips produced (or attracted) by household or zone as a function of (generally linear) relations to be defined from available data.
- It is important to be clear about the following aspects mentioned above:
  1. what trips to be considered (e.g. only vehicle trips and walking trips longer than three blocks);
  2. what is the minimum age to be included in the analysis (i.e. five years or older).

- Consider a zone with 250 households with car and 250 households without car.
- Assuming we know the average trip generation rates of each group:
  1. car-owning households produce: 6.0 trips/day
  2. non-car-owning households produce: 2.5 trips/day
- we can easily deduce that the current number of trips per day is:

$$t_i = 250 \times 2.5 + 250 \times 6.0 = 2125 \text{ trips/day}$$

- Let us also assume that in the future all households will have a car; therefore, assuming that income and population remain constant (a safe hypothesis in the absence of other information), we could estimate a simple multiplicative growth factor as:

$$F_i = C_{di} / C_{ci} = 1/0.5 = 2$$

- We could estimate the number of future trips as:

$$T_i = 2 \times 2125 = 4250 \text{ trips/day}$$

- However, the method is obviously very crude. If we use our information about average trip rates and make the assumption that these will remain constant (which is actually the main assumption behind one of the most popular forecasting methods, as we will see below), we could estimate the future number of trips as:

$$T_i = 500 \times 6 = 3000$$

- which means that the growth factor method would overestimate the total number of trips by approximately 42%.
- This is very serious because trip generation is the first stage of the modelling process errors here are carried through the entire process and may invalidate work on subsequent stages.

- $Ti = a*X_0 + b*X_1$
  - Parameters (a = 2.5 trips/hh; b = 6 trips/hh)
  - Variables (X_0 = no-auto hh's; X_1 = auto hh's)
  - Base year X0 = 500 hh and X1 = 500 hh
    - Ti = 4250 trips generated
  - Forecast year everyone will own a car
    - Ti = 8500 trips → based on growth factor 1000/500 = 2
    - Ti = 6000 → based on changes in explanatory variables

Trip Generation Modeling—Cross-Classification

Cross-Classification or Category Analysis

Classical Model

- The method is based on estimating the response (e.g. the number of trip productions per household for a given purpose) as a function of household attributes.
- Its basic assumption is that trip generation rates are relatively stable over time for certain household stratifications
- Let $t^p(h)$ be the average number of trips with purpose p (and at a certain time period) made by members of households of type h.
- Types are defined by the stratification chosen; for example, a cross-classification based on m household sizes and n car ownership classes will yield mn types h.

- The rate $t^p(h)$ is then the total number of trips in cell h, by purpose, divided by the number of households $H(h)$ in it.

$$t^p(h) = T^p(h) / H(h)$$

	Income	
Family structure	Low	High
1 or less	0.8	1.0
2 or more	1.2	2.3

- There are various ways of defining household categories. The first application in the UK employed 108 categories as follows: six income levels, three car ownership levels (0, 1, and 2 or more cars per household) and six household structure groupings

Table 4.6 Example of household structure grouping

Group	No. employed	Other adults
1	0	1
2	0	2 or more
3	1	1 or less
4	1	2 or more
5	2 or more	1 or less
6	2 or more	2 or more

Cross-classification (category analysis)

- Trip production: $$O_i^p = \sum_h a_i(h) t^p(h)$$
  - p = trip purpose
  - i = zone
  - h = household type grouping
  - $a_i(h)$ = number of households of type h in zone i
  - $t_p(h)$ = trip rate for trip of type p for households of type h

Cross-classification (category analysis): Example

Situation: Zone 23 characteristics are as follows:

Households	Income level		
Household size	<10,000	10,000 to 30,000	>30,000
1	10	100	50
2	10	200	50
3+	30	100	50

Home based work (HBW) trip production data are as follows:

HBW trip rate	Income level		
Household size	<10,000	10,000 to 30,000	>30,000
1	1.5	2.5	2.5
2	2.5	4	5
3+	3	5	7

Cross-classification (category analysis): Steps to create table

- Establish household groupings
- Assign households to the groupings
- Total, for each grouping the observed trips [$T_p(h)$]
  - p is the trip purpose
  - h is the grouping
- Total, for each grouping the observed households [$H(h)$]
  - H is the number of households observed
  - h is the grouping
- Calculate the trip rates by grouping [$t_p(h) = T_p(h) / H(h)$]

Cross-classification (category analysis)

- Advantages of cross-classification
  - Independent of zone system
  - No regression related assumptions necessary
- Disadvantages
  - No extrapolation
  - No trip rate for cells with no observations
  - Difficult to add additional stratifying variables
  - Difficult to choose household groups

Matching Production and Attractions

- trip production models are more reliable than trip attraction
- RESULT: force total trip attractions to equal total trip productions
  - P_i = trips produced by zone *i*
  - A_i = total trips attracted by zone *i*

Matching Generations and Attractions (cont.)

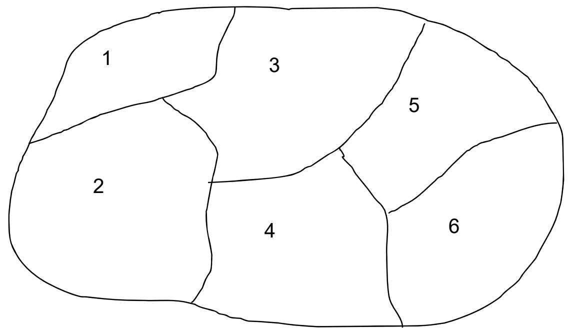
- The adjusting factor to adjust the attractions

$$f = \frac{T}{\sum_i A_i}$$

$$A_i^{adjusted} = A_i * f$$

$$\sum_i P_i = \sum_j A_j^{adjusted}$$

Trip Attraction Adjustment Example



Zone	HBW Productions	HBW Attractions	Adjusted HBW Attractions
1	200	300	
2	100	100	
3	300	300	
4	400	200	
5	200	500	
6	100	400	
Total	1,300	1,800	

Trip Generation Modeling—Regression

Regression Analysis

- The method will not discuss in this class, since other course has discussed it
- Try an example by using data analysis in ms excel
  1. Determining the coefficient of variables
  2. Check t test value
  3. Check multicollinearity
  4. Check R^2

Zonal-based Multiple Regression

- In this case an attempt is made to find a linear relationship between the number of trips produced or attracted by zone and average socioeconomic characteristics of the households in each zone. The following are some interesting considerations:

1. **Zonal models can only explain the variation in trip making behavior between zones**. For this reason they can only be successful if the inter-zonal variations adequately reflect the real reasons behind trip variability. For this to happen it would be necessary that zones not only had a homogeneous socioeconomic composition, but represented as wide as possible a range of conditions
2. **Role of the intercept**. One would expect the estimated regression line to pass through the origin; however, large intercept values (i.e. in comparison to the product of the average value of any variable and its coefficient) have often been obtained. If this happens the equation may be rejected; if on the contrary, the intercept is not significantly different from zero, it might be informative to re-estimate the line, forcing it to pass through the origin.

3. **Null zones**. It is possible that certain zones do not offer information about certain dependent variables (e.g. there can be no HB trips generated in non-residential zones). Null zones must be excluded from analysis;
4. **Zonal totals versus zonal means**. When formulating the model the analyst appears to have a choice between using aggregate or total variables, such as trips per zone and cars per zone, or rates such as trips per household per zone and cars per household per zone. In the first case the regression model would be:

$$Yi = \theta_0 + \theta_1 X_{1i} + \theta_2 X2_i + \ldots + \theta_k X_{ki} + E_i$$

whereas the model using rates would be:

$$yi = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_k x_{ki} + e_i$$

with $y_i = Y_i/H_i$; $x_i = X_i/H_i$; $e_i = E_i/H_i$ and H_i the number of households in zone i.

- To end this theme it is important to remark that even when rates are used, zonal based regression is conditioned by the nature and size of zones

Table 4.2 Inter-zonal variation of personal productions for two different zoning systems

Zoning system	Mean value of trips/household/zone	Inter-zonal variance
75 small zones	8.13	5.85
23 large zones	7.96	1.11

Household-based Regression

- Intra-zonal variation may be reduced by decreasing zone size, especially if zones are homogeneous. However, smaller zones imply a greater number of them and this has two consequences:
  1. more expensive models in terms of data collection, calibration and operation;
  2. larger sampling errors, which are assumed non-existent by the multiple linear regression model.
- Proceed stepwise, testing each potential explanatory variable in turn until the best model (in terms of some summary statistics for a given confidence level) is obtained.
- In actual fact, stepwise methods are not recommended; it is preferable to proceed the other way around, i.e. test a model with all the variables available and take out those which are not essential (on theoretical or policy grounds) and have low significance or an incorrect sign.

Example

- Consider the variables trips per household (*Y*), number of workers (X_1) and number of cars (X_2).
- Table 4.3 presents the results of successive steps of a stepwise model estimation; the last row also shows (in parenthesis) values for the *t*-ratio

Table 4.3 Example of stepwise regression

Step	Equation	R^2
1	$Y = 2.36\,X_1$	0.203
2	$Y = 1.80\,X_1 + 1.31\,X_2$	0.325
3	$Y = 0.91 + 1.44X_1 + 1.07X_2$	
	(3.7) (8.2) (4.2)	0.384

- The third model is a reasonable equation in spite of its low R^2. The intercept 0.91 is not large (compare it with 1.44 times the number of workers, for example) and the regression coefficients are significantly different from zero.
- An indication of how good these models are may be obtained from comparing observed and modelled trips for some groupings of the data (see Table 4.4).
- As can be seen, the majority of cells show a reasonable approximation (i.e. errors of less than 30%). If large bias were spotted it would be necessary to adjust the model parameters; however, this is not easy as there are no clear-cut rules to do it, and it depends heavily on context

Table 4.4 Comparison of trips per household (observed/estimated).

	Number of workers in household			
No. of cars	0	1	2	3 or more
0	0.9/0.9	2.1/2.4	3.4/3.8	5.3/5.6
1	3.2/2.0	3.5/3.4	3.7/4.9	8.5/6.7
2 or more	–	4.1/4.6	4.7/6.0	8.5/7.8

The Problem of Non-Linearity

- the linear regression model assumes that each independent variable exerts a linear influence on the dependent variable.
- It is not easy to detect non-linearity because apparently linear relations may turn out to be non-linear when the presence of other variables is allowed for in the model.
- Multivariate graphs are useful in this sense; the example of Figure 4.9 presents data for households stratified by car ownership and number of workers.
- It can be seen that travel behavior is non-linear with respect to family size.

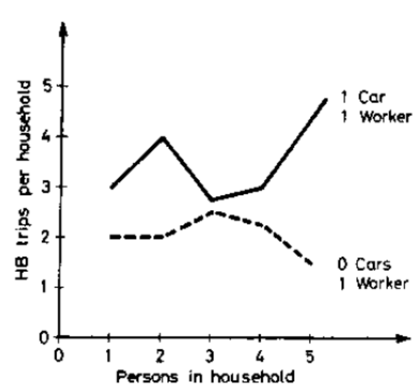


Figure 4.9 An example of non-linearity

- There are two methods to incorporate non-linear variables into the model:
  1. Transform the variables in order to linearize their effect (e.g. take logarithms, raise to a power). However, selecting the most adequate transformation is not an easy or arbitrary exercise,
  2. Use *dummy* variables. In this case the independent variable under consideration is divided into several discrete intervals and each of them is treated separately in the model. In this form it is not necessary to assume that the variable has a linear effect, because each of its portions is considered separately in terms of its effect on travel behavior.

     For example, if car ownership was treated in this way, appropriate intervals could be 0, 1 and 2 or more cars per household. As each sampled household can only belong to one of the intervals, the corresponding dummy variable takes a value of 1 in that class and 0 in the others. It is easy to see that only (n − 1) dummy variables are needed to represent n intervals.

Example 4.4

Consider the model of Example 4.3 and assume that variable X_2 is replaced by the following dummies:

- $Z1$, which takes the value 1 for households with one car and 0 in other cases;
- $Z2$, which takes the value 1 for households with two or more cars and 0 in other cases.

Non-car-owning households correspond to the case where both $Z1$ and $Z2$ are 0. The model of the third step in Table 4.3 would now be:

$$Y = 0.84 + 1.41X_1 + 0.75Z_1 + 3.14Z_2$$
$$(3.6) \quad (8.1) \quad (3.2) \quad (3.5)$$

R^2 = 0.387

$$Y = 0.91 + 1.44X_1 + 1.07X_2$$
$$(3.7) \quad (8.2) \quad (4.2) \qquad 0.384$$

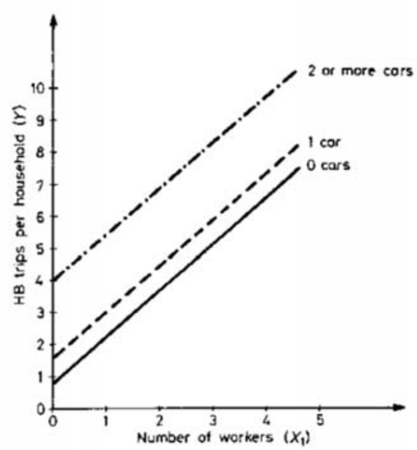


Figure 4.10 Regression model with dummy variables

Obtaining Zonal Totals

- In the case of zonal-based regression models, this is not a problem as the model is estimated precisely at this level.
- In the case of household-based models, though, an aggregation stage is required.
- However, it must be noted that the aggregation stage can be a very complex matter in non-linear models, as we will see in Chapter 9.
- Thus, for the third model of Table 4.3 we would have:

$$T_i = H_i\,(0.91 + 1.44X_{1i} + 1.07X_{2i})$$

- where T_i is the total number of HB trips in zone i, H_i is the total number of households in it and X_{ji} is the *average* value of variable X_j for the zone.
- On the other hand, when dummy variables are used, it is also necessary to know the number of households in each class for each zone; for instance, in the model of Example 4.4 we would require:

$$T_i = H_i\,(0.84 + 1.41X_{1i}) + 0.75H_{1i} + 3.14H_{2i}$$

where H_{ji} is the number of households of class j in zone i.

Bayesian Updating of Trip Generation Parameters

- Assume we want to estimate a trip generation model but lack funds to collect appropriate survey data; a possible (but inadequate) solution is to use a model estimated for another (hopefully similar) area directly.
- However, it would be highly desirable to modify it in order to reflect local conditions more accurately.
- This can be done by means of Bayesian techniques for updating the original model parameters using information from a small sample in the application context.
- Bayesian updating considers a *prior* distribution (i.e. that of the original parameters to be updated), new information (i.e. to be obtained from the small sample) and a *posterior* distribution corresponding to the updated model parameters for the new context.
- Updating techniques are very important in a continuous planning framework

Variable	Prior information	New information
Mean trip rate	t_1	t_s
No. of observations	n_1	n_s
Trip rate variance	S_1^2	S_s^2

Table 4.8 Bayesian updating notation for trip generation

following parameters:

$$t_2 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_s^2} t_1 + \frac{1/\sigma_s^2}{1/\sigma_1^2 + 1/\sigma_s^2} t_s \quad (4.21)$$

$$\sigma_2^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_s^2} \quad (4.22)$$

which, substituting by the known values $\mathbf{S}^2$ and $\mathbf{n}$, yield:

$$t_2 = \frac{n_1 S_s^2 t_1 + n_s S_1^2 t_s}{n_1 S_s^2 + n_s S_1^2} \quad (4.23)$$

$$\sigma_2^2 = \frac{S_1^2 S_s^2}{n_1 S_s^2 + n_s S_1^2} \quad (4.24)$$

Example

- The mean trip rate, its variance and the number of observations for two household categories, obtained in a study undertaken 10 years ago are shown below:
- It is felt that these values might be slightly out of date for direct use today, but there are not enough funds to embark on a full-scale survey. A small stratified sample is finally taken, which yields the values shown below:

	Household categories	
Variable (prior data)	1	2
Trips per day	8	5
No. of observations	65	300
Trip rate variance	64	15
Mean trip variance	0.98	0.05

	Household categories	
Variable (new data)	1	2
Trips per day	12	6
No. of observations	30	30
Trip rate variance	144	36
Mean trip variance	4.80	1.20