The Estimation of Combined Model Parameter Based On Traffic Count in Equilibrium Assignment and Study the Factors Affecting these Accuracy

Rahayu SULISTYORINI  Ofyar Z. TAMIN
Lecturer  Professor
Department of Civil Engineering  Department of Civil Engineering
Lampung University  Institute Technology Bandung (ITB)
Jl. Sumantri Brojonegoro No.1  Jl.Ganesha 10 Bandung
Phone&Fax : 062 – 0721 704947  Phone&Fax : 062 - 022 2502350
E-mail: Sulistyorini_smd@yahoo.co.uk  E-mail: ofyar@trans.si.itb.ac.id

Ade SJAFRUDIN
Lecturer
Department of Civil Engineering
Institute Technology Bandung (ITB)
Jl.Ganesha 10 Bandung
Phone&Fax : 062 - 022 2502350
E-mail: ades@trans.si.itb.ac.id

Abstract: The alternative of using traffic counts to estimate O-D matrices is attracted. The stressing of the research is to estimate unknown parameter of transport demand model by combining trip distribution, mode choice and route choice in single process. Furthermore, the factors affecting the accuracy of the estimated O-D matrices will be studied is the effect of incorporating errors in traffic count information and the effect of trip assignment techniques.

Several important findings can be concluded as: (1). Parameter of ($\beta$) and ($\gamma$) value depends on starting value; (2). If there are some error in traffic count, it could be implicate in the estimated O-D matrix, and (3). It can be finally proven that the use of equilibrium assignment in the trip assignment stage also gave significant impact compared to the use of all-or-nothing assignment.

Key Words: traffic counts, combined model, equilibrium assignment

1. INTRODUCTION

Travel has become an integral part of our daily life. This activity generates its good share of problems to any community, including traffic congestion, delay, air pollution and visual intrusion. In order to alleviate these problems, it is necessary to understand the underlying travel pattern. The concept of an “O-D matrix” has been adopted by transport planners to represent the most important features of this travel pattern. An O-D matrix gives a very good indication of travel demand, and therefore, it plays a very important role in various transport studies, transport planning and management tasks.

The conventional methods to estimate O-D matrices requires very large surveys such as: home and roadside interviews; which are very expensive, lengthy, labor intensive, subject to large errors, and moreover, time disruptive to trip makers. All of these have led researchers to investigate alternative, less expensive methods for estimating O-D matrices.
The need for inexpensive methods, which require low-cost data, less time and less manpower generally called as ‘unconventional method’ is therefore obvious due to time and money constraint. Traffic counts, the embodiment and the reflection of the O-D matrix; provide direct information about the sum of all O-D pairs which use those links. Some reasons why traffic counts are attractive as a data base: firstly, they are routinely collected by many authorities due to their multiple uses in many transport planning tasks. All of these make them easily available. Secondly, they can be obtained relatively inexpensive in terms of time and manpower, easier in terms of organization and management and also without disrupting the trip makers.

However, previous research (Tamin 2003) concluded there are some factors that affecting the accuracy of O-D matrices estimated from traffic counts. These factors are as follows:

- The choice of the transport demand model itself to be used in representing the trip behavior within the study area;
- The estimation method used to calibrate the parameters of the transport model from traffic count information;
- The trip assignment techniques in determining the route choice;
- The location and number of traffic count data;
- The level of errors in traffic counts; and finally
- The level of resolution of the zoning system and the network definition.

The problem then becomes one of estimating the unknown O-D demands on the basis of the known link use proportions and the observed link traffic flows. It may be characterized as an integrated two-mode traffic equilibrium method. This method combines a zonal aggregate-demand model with an equilibrium-type road assignment and a transit assignment method. The model examined was the Gravity (GR) model combined with the Multi-Nomial-Logit (MNL) model. Non-Linear-Least- Squares (NLLS) estimation methods were used to calibrate the parameter of the combined model. Three of the stages of this process (four step model), trip distribution, modal split and traffic assignment, combine to estimate expected O-D demands, and as such, are of relevance to this research. The separation distance is referred to as the deterrence or impedance function and is often measured by travel time. The kind of friction factor functions can be selected is negative exponential.

Iterative solution algorithms, that are modifications of the Newton Raphson and Gauss Jourdan’s techniques, are proposed to solve each of the model formulations. The methods proposed in this research are based on the least-squares (LS) estimation technique. In the formulation of the problem in a transportation network, link-flow proportions play a key role. The model that is proposed assumes the sum of the squared deviations between the observed and estimated link flows as the error function that is to be minimized, which are used in the road assignment to calibrate parameters.

The main objective of this research is to continue the development of methods of estimating parameters of combined model tip distribution and mode choice using traffic counts data, in assessing the influence factors of the estimated O-D matrices accuracy from traffic counts data. In this research, in addition to all-or-nothing assignment, equilibrium assignment model is used as types of the trip assignment methods. The other influenced factors of transport demand model estimation from traffic counts were then tested such as trip assignment techniques in determining the route choice and the level of errors in traffic counts.
2. MODEL CALIBRATION AND VALIDATION

A model can be simply represented as a mathematical function of variables and parameters. Calibrating a model requires choosing its parameters, assumed to have a non-null value, in order to optimize one or more goodness-of-fit measures which are a function of the observed data. Estimation involves finding the value of parameters which make the observed data more likely under the model specification; in the case one or more parameters can be judged non-significant and left out of the model. Estimation also considers the possibility of examining empirically certain specification issues; for example, structural and/or functional form parameters may be estimated. There has been a tendency to interpret model validation exclusively in terms of the goodness of fit achieved between observed behavior and base year prediction. Validation requires comparing the model predictions with information not used during the process of model estimation (Willumsen, 1995).

The main use of models in practice is for conditional forecasting; the model will produce estimates of the dependent variables given a set of independent variables. Typical forecast are conditional in two ways:
- In relation to the values assigned to the policy variables in the plan, the impact of which is being tested with the model;
- In relation to the assumed values of other variables

3. CONVENTIONAL FOUR STAGE TRANSPORT MODEL

The general form of the model is depicted in Figure 1. The classic model is presented as a sequence of four sub-models: trip generation, distribution, modal split and assignment. At the end of the traffic assignment stage, new flow level and therefore new travel time will be obtained. The approach in this research is the contrary of conventional four stage models.

![Figure 1. Conventional four stage transportation model](image-url)
The information of traffic flow is used to producing a trip matrix by each mode when the distribution and mode choice were run.

4. ESTIMATION TRANSPORT DEMAND MODEL BASED ON TRAFFIC COUNT

One can interpret link flows (or traffic counts) as resulting from a combination of two elements: an O D matrix and the route choice pattern selected by drivers on the network. These two elements may be linearly related to traffic counts, see equation (1) below, but under normal circumstances there will never be enough traffic counts to identify a single O D matrix as the only possible source of the observed flows. Traffic counts alone are not enough to estimate O D matrices, something else is needed.

The transport demand model estimation approach assumes that the travel pattern behaviour is well represented by a certain transport model, e.g. a gravity model. The main idea is to apply a transport model to represent the travel pattern. It should be noted here that the transport demand models are described as functions of some planning variables like population or employment and some parameters. Whatever the specification and the hypothesis underlying the models, the main tasks is to estimate their parameters on the basis of traffic counts. Once, the parameters of the postulated transport demand models have been calibrated, they may be used not only for the estimation of the current O-D matrix, also for predictive purposes. The latter requires the use of future values for the planning variables.

Consider a study area divided into \( N \) zones inter-connected by a road network which consist of a series of links and nodes, see Figure 4. The trip matrix consists of \( N^2 \) cells, or \((N^2-N)\) cells if intra-zonal trips can be disregarded. The most important stage for the estimation of a transport demand model from traffic counts is to identify the paths followed by the trips from each origin to each destination. In principle, Given all the \( p_{id}' \) and all the observed traffic counts \( \hat{V}_l \), then there will be \( N^2 \) unknown \( T_{id} \)'s to estimated from a set of \( L \) simultaneous linear equations (1) where \( L \) is the total number of traffic (passenger) counts. In principle, \( N^2 \) independent and consistent traffic counts are required in order to determine uniquely the O-D matrix \([ T_{id} ]\). In practice, the number of observed traffic counts is much less than the number of unknown \( T_{id} \)'s.

The total volume of flow \( \hat{V}_l \) in a particular link \( l \) is the summation of the contributions of all trips interchanges between zones within the study area to that link. Thus the flow on each link is a result of:

- trip interchanges from zone \( i \) to zone \( d \) or combination of several types of movement travelling between zones within a study area (=\( T_{id} \)); and
- the proportion of trips travelling from zone \( i \) to zone \( d \) whose trips use link \( l \) which is defined by \( p_{id}' \) \((0 \leq p_{id}' \leq 1)\).

Mathematically, it can be expressed as follows:

\[
V^k_l = \sum_i \sum_d T_{id} P_{id}^{lk}
\]
\[ T_{id} \] = the observed O-D matrix from origin \( i \) to destination \( d \).

\( O_i \) = the total trips generated by origin \( i \).

\( D_d \) = the total trips attracted by destination \( d \).

\( A_i, B_d \) = the balancing factors for origin \( i \) and destination \( d \).

\( C_{id}^m \) = the trip cost of traveling from origin \( i \) to destination \( d \) by mode \( m \).

\( \beta \) = the unknown estimated parameter to be calibrated.

\( V_i \) = the observed traffic counts.

\( p_{id}^l \) = the trip assignment proportion for trips by mode \( m \) from origin \( i \) to destination \( d \) which use link \( l \).

\( L \) = the total number of links observed.

\( M \) = the total number of modes.

\( N \) = the total number of origins or destinations.

The value of \( p_{id}^l \) is determined by using trip assignment technique. The previous research (Tamin et al., 2001) used all-or-nothing assignment to obtain the value of \( p_{id}^l \). By using this method, the value of \( p_{id}^l \) is either 0 or 1. In this research, the uses of equilibrium assignment method which consider the congestion effect of route choice selection was introduced. Hence, the value of \( p_{id}^l \) obtained is between 0 and 1 (0 ≤ \( p_{id}^l \) ≤ 1).

5. **COMBINED GRAVITY-MULTINOMIAL LOGIT MODEL**

The Gravity (GR) model is developed by analogy with Newton's law of gravitation. Newton asserted that the force of attraction, \( F_{id} \), between two bodies is proportional to the product of their masses, \( m_i \) and \( m_d \), divided by the square of the distance between them \( d_{id}^2 \). In geography, ‘force’ is associated with the numbers of movements or trips between two regions; ‘mass’ is replaced by a variable such as population size and measures a region's capacity either to generate or to attract trips; and distance is either measured in physical terms or replaced by a more relevant variable such as travel cost or time. The analogous transport gravity model is:

\[ T_{id} = k \frac{O_i D_d}{d_{id}^2} \quad \text{k is a constant} \]  \hspace{1cm} (2)

This model has some sensible properties. It says that the number of trips from zone \( i \) to zone \( d \) is directly proportional to each of \( O_i \) and \( D_d \) and inversely proportional to the square of the distance between them. Hence, if a particular \( O_i \) and a particular \( D_d \) are each doubled, then the number of trips between these zones would quadrate according to equation (1), when one would be expected that they would only double.

Therefore, the following constraint equations on \( T_{id} \) should always be required, such constraints are not satisfied by equation (1):

\[ \sum_d T_{id} = O_i \quad \text{and} \quad \sum_i T_{id} = D_d \]  \hspace{1cm} (3)
where \( O_i \) and \( D_d \) directly represent the total number of trips originating and terminating at \( i \) and \( d \) respectively. Suppose now there are \( M \) modes traveling between zones, the modified gravity model (Doubly Constrained Gravity Model) can then be expressed as:

In gravity model, trip distribution is described as accessibility, production and attraction from origin zone to destination zone. Description of accessibility to reach the destination zone in this model is expressed in function of traveling cost or impedance function. The Model is inspired by analogy with Newton’s law of gravitational force. Gravity model can be expressed as:

\[
T_{id} = O_i \cdot D_d \cdot A_i \cdot B_d \cdot f(C_{id})
\] (4)

where: \( A_i \) and \( B_d \) = the balancing factors expressed as:

\[
A_i = \frac{1}{\sum_d (B_d \cdot D_d \cdot f_{id})} \quad \text{and} \quad B_d = \frac{1}{\sum_i (A_i \cdot O_i \cdot f_{id})}
\] (5)

With exponential deterrence function as bellow:

Power function: \( f_{id} = C_{id}^{-\alpha} \) (6)

Exponential function: \( f_{id} = e^{-\beta \cdot D_{id}} \) (7)

Tanner function: \( f_{id} = C_{id}^{-\alpha} \cdot e^{-\beta \cdot D_{id}} \) (8)

By substituting equation (4) to equation (1), then the fundamental equation for the estimation of a transport demand model from traffic counts can be expressed as:

\[
V_i = \sum_i \sum_d \left( O_i \cdot D_d \cdot A_i \cdot B_d \cdot f_{id} \cdot p_{id} \right)
\] (9)

This process would be iterated until the values of \( A_i \) and \( B_d \) converge to certain unique values. The fundamental equation (9) has been used by many literatures not only to estimate the O-D matrices but also to calibrate the transport demand models from traffic count information (see Tamin, 2008).

A modal split model will be required if public transport proposals are being considered, or if it is thought likely that proposals for the highway or parking system will lead to a significant transfers of trips between modes, hence altering mode split. The purpose of this stage is the modeling of the choice between modes of conveyance for each trip, usually car and one or more public transport modes in the case of passengers. The most general and simplest mode choice model (Multi-Nomial Logit Model) was used in this study. The Logit model estimates the proportion of trips by a special mode \( m \) according to the relative utility of each mode as a summation of each modal attribute. It can be expressed as:

\[
T_{id}^k = T_{id} \cdot \frac{\exp\left(-\gamma \cdot C_{id}^k\right)}{\sum_m \exp\left(-\gamma \cdot C_{id}^m\right)}
\] (10)
By substituting equations (4)-(9) to equation (1), then 'the fundamental equation' for the estimation of a combined transport demand model from traffic counts is:

$$V_i = \sum_l \sum_d \left( O_i A_l B_s D_d f(C_{ld}) \right) \sum_m \frac{\exp(-\gamma_s C_{ld}^m)}{\exp(-\gamma_m C_{ld}^m)} p_{id}$$  \hspace{1cm} (11)

Theoretically, having known the values of \( \hat{V}_i \) and \( p_{id}^l \), \( T_{id} \) can then be estimated. Equation (9) is a system of \( L \) simultaneous equations with only (2) unknown parameters \( \beta \) and \( \gamma \) need to be estimated. The problem now is how to estimate the unknown parameters so that the model reproduces the estimated traffic flows as close as possible to the observed traffic counts.

Trip assignment used in this study is based on all-or-nothing assignment and equilibrium assignment. For both type assignment, the value of \( p_{id}^l \) is used as trip proportion between origin zone \( i \) and destination zone \( d \) which use link \( l \). In the all-or-nothing assignment, the value of \( p_{id}^l \) is calculated separately and not related in the estimation process. The value of \( p_{id}^l \) in all-or-nothing assignment is 0 or 1. 0 value means that link \( l \) is not used by trip users. Meanwhile, 1 value means that all trip users use link \( l \).

6. EQUILIBRIUM ASSIGNMENT

If one ignores stochastic effects and concentrates on capacity restraint as a spread of trips on a network, one should consider capacity restraint relating flow to the cost on a link. This is usually attempt to equilibrium conditions as formally enunciated by Wardrop 1952 in Willumsen 1995

*Under equilibrium conditions, traffic arranges itself in congested networks in such a way that no individual trip maker or driver can reduce his route costs by switching routes*

If all trip makers perceive costs in the same way (no stochastic effects):

*Under equilibrium conditions, traffic arranges itself in congested networks such that all used routes between an O-D pair have equal and minimum costs while all unused have greater or equal costs*

The algorithm to solve the equilibrium traffic assignment problems is based on non linear optimization technique developed by Frank-Wolf. Application of the Frank-Wolf method is then involves in this following equation.

Minimize \( Z = \int_0^{V_i} C_i(V) dV \)  \hspace{1cm} (12)

\[ V_i = \sum_l \sum_d \left[ T_{ld} \cdot p_{ld}^l \right] \text{ and } T_{id} = \sum_r T_{idr} \]  \hspace{1cm} (13)

\[ T_{idr} > 0 \]  \hspace{1cm} (14)
where:

\[ p'_{idr} = \begin{cases} 
1 & \text{if route } r \text{ using link } l \text{ between } i \text{ and } d \\
0 & \text{otherwise} 
\end{cases} \]

\[ T_{idr} = \text{movement from } i \text{ to } d \text{ using route } r \]

\[ p'_{idr} = \text{proportion of movement from } i \text{ to } d \text{ using route } r \text{ and link } l \]

\[ C_i(V) = \text{mathematic expression between traffic flow and cost} \]

### 7. ESTIMATION METHOD

Tamin (2000) explains one method of estimation which have been developed so far by many researchers is Least-Square estimation method (LLS or NLLS). The main idea of Least-method is to estimate the unknown parameter which minimize the sum of the squared differences between the estimated and observed traffic counts. The problem now is:  

\[
\text{to minimize } S = \sum_i \left[ \frac{1}{V_i} (V_i - \hat{V}_i)^2 \right] 
\]

\[ \hat{V}_i = \text{observed traffic flows} \]

\[ V_i = \text{estimated traffic flows} \]

Having substituted (11) to (12), the following set of equation is required in order to find a set of unknown parameter \( \beta \) which minimize equation (13) and (14):

\[
\frac{\partial S}{\partial \beta} = \sum_i \left[ \frac{1}{V_i} \left( 2 \left( \sum_i \sum_d T_{id} \cdot p'_{id} - \hat{V}_i \right) \left( \sum_i \sum_d \frac{\partial T_{id}}{\partial \beta} \cdot p'_{id} \right) \right) \right] = 0
\]

(16)

\[
\frac{\partial S}{\partial \gamma} = \sum_i \sum_k \left[ \frac{2}{V_i} \left( \sum_i \sum_d T_{id}^k \cdot p_{id}^k - \hat{V}_i \right) \left( \sum_i \sum_d \frac{\partial T_{id}^k}{\partial \gamma} \cdot p_{id}^k \right) \right] = 0
\]

(17)

Those equations have two (2) unknown parameters \( \beta \) and \( \gamma \) need to be estimated. Then it is possible to determine uniquely all the parameters, provided that \( L > 1 \). Newton–Raphson’s method combined with the Gauss–Jordan Matrix Elimination technique can then be used to solve this equation

### 8. METHODOLOGY

Calibration process of parameter estimation based on traffic count can be seen on figure 2. Socio-economic parameter \( (O_i \text{ and } D_d) \) and trip cost each mode \( k \left( C_{id}^k \right) \) using as input.
Figure 2 Calibration’s procedure of parameters’s estimation
9. APPLICATION IN ARTIFICIAL NETWORK

Mathematical trip distribution and mode choice model development is tested in simple artificial network (consist of 6 zones, 8 nodes and 30 links one way and 2 transit links), see Figure 3 and complex network (consist of 12 zones, 405 nodes and 535 links one way and 7 transit links) in Figure 4. Finally to get unknown parameter’s value, the model is applicated in real network system in Bandung (consist of 125 zones, 1088 nodes and 2534 links one way and 7 transit links), see Figure 5.

Figure 3 Simple artificial network

Figure 4 The route of bus in complex artificial network
Two modes, auto (au) and transit (tr), are considered to operate over independent networks, so that the generalized costs associated with each mode are separate. The costs of the transit mode are fixed and given by a timetable and fare schedule.

10. ANALYSIS AND RESULT

Giving value for $\beta$ and $\gamma$ and using Gravity (GR) model with double constraint, we built Original O-D Matrices.

![Figure 5 Real network of Bandung](image)

![Figure 6 Procedure of model verification](image)
This matrix is used to get traffic volume in each link through assignment and to get \( p_{id}' \) value. We give \( \beta \) value and \( \gamma \) value to built matrices and observed traffic count. Traffic Volume is the result of assignment procedure from the matrices using equilibrium assignment. The value of \( p_{id}' \) is determine in the same stage with procedure of parameter estimated.

For artificial data we create traffic observed \( \hat{V}_j \) with giving error factor in traffic volume above. Using Traffic Volume observed and others data (zoning system, network system, Generation and Attraction Data) we run estimation process to get new parameter value.

As shown in these figures we can conclude that the proces of mathematical model development is run well with whatever we put \( \beta \) and \( \gamma \) value as starting point, the value is end in single value solution. If we put value as starting value far from solution, it needs more iteration process to achieve convergence. We also get the result that the unknown parameter of real network system in Bandung is \( \beta = 0.061166 \) and \( \gamma = 0.166011 \). Statistic value if traffic volume estimated compare than traffic volume from survey with \( R^2 = 0.361 \) for private car.
and $R^2 = 0.556$ for bus.

### 10.1 Error in Traffic Counts

We try to see how far some factors that affecting the accuracy of O-D matrices estimated from traffic counts, as follows: trip assignment techniques in determining the route choice; the location and number of traffic count data and the level of errors in traffic counts.

![Flowchart depicting the process of error in traffic count](image)

**Figure 9** Effect of error in traffic count method

![Graph showing the relationship between R² and error percentage](image)

**Figure 10** Error traffic counts and R² relationship in simple artificial network
The effect of error in traffic count is tested using procedure which can be shown in Figure 9. We create error from 10% error, 20%, 30%...until 100% error in traffic count as observed to see how far it affecting the level of accuracy matrices estimation.

We plot the result (the accuracy level and percentage error in traffic count) in graph to see the performance of this research. The result shows that the increase of error in traffic counts is not affecting the level of accuracy ($R^2$). From Figure 10, with 100% error in traffic count data, the level of accuracy ($R^2$) is 0.7 for simple artificial network.

Same with simple network, in complex artificial network we can see that if we use traffic counts data which have 100% error, we can reach $R^2 = 0.77$. With these result we can conclude that error in traffic counts is not affecting the level of accuracy significantly.

![Figure 11 Error traffic counts and $R^2$ relationship in complex artificial network](image)

### 10.2 Trip Assignment Techniques in Determining The Route Choice

In terms of $R^2$, equilibrium assignment gives best result compared to all or nothing assignment related to the level of error in traffic counts, see Figure 12. The result shown that equilibrium assignment give better result than all-or-nothing assignment.
11. CONCLUSION

In this paper, we have developed a general approach to the explanation of transport flows that combines into a consistent format the trip distribution (gravity), mode choice (multinomial-logit) and trip assignment (equilibrium condition). **Non-Linear-Least-Squares (NLLS)** estimation methods were used to calibrate two unknown parameters of the combined model. We developed a two mode (private car and transit) network equilibrium model where the most important features are the distinction between the flow of vehicles and flow of transit passengers and the means of modeling the interaction between both types of vehicles.

It is also shown that the TDMC model with negative exponential deterrence function produced estimation parameter ($\beta$ and $\gamma$) for NLLS estimation methods. This result is very important in terms of time and money for estimating the demand of public transport and also for forecasting purposes.

The paper explains the impact of factors developed to calibrate the parameters of transport demand model from traffic counts information. Some conclusions can be drawn from the result obtained:

- The convergence of ($\beta$) and ($\gamma$) value depends on starting value of ($\beta$) and ($\gamma$). If the starting value of ($\beta$) and ($\gamma$) more away from the solution, it needs iteration more to achieve convergence level
- It is found that the unknown parameter of real network system in Bandung is $\beta = 0.061166$ and $\gamma = 0.166011$. Statistic value if traffic volume estimated compare than traffic volume from survey with $R^2 = 0.361$ for private car and $R^2 = 0.556$ for bus.
- It can be conclude that error in traffic count is not affecting the level of accuracy significantly
- In terms of $R^2$, equilibrium assignment gives better result compared than all-or-nothing assignment
REFERENCES


Tamin, O.Z., (1988), The Estimation of Transport Demand Model From Traffic Counts, PhD Dissertation of the University of London, University College London.


