AN ALTERNATIVE APPLICATION OF LOGIT MODELLING IN MANAGEMENT OF METROPOLITAN AREAS: THE CASE STUDY OF POZNAŃ URBAN AREA

Zbyszko Pawlak
Poznan School of Logistics / The Poznan University of Economics, Poznan, Poland

ABSTRACT. The article proposes a new modelling framework for simulating flows of people between suburban areas and the metropolis. The model is based on a logit relationship used in researches of transport mode choice problems. After explaining the properties of the model, it is than applied to analyze Metropolitan Area - Poznań Urban Area, Poland. As results it has been found that the main factors influencing the number of trips between each suburban area and Poznań as a proportion of all trips to and from that area are distance from Poznań and population of the area. It was also found that some unknown factors also affect the phenomena. It was found that model can be used in such a context, although it is of a limited use in making inferences about areas for which data are not available and this area remains a subject for further research.

Key words: metropolitan area, sustainable management, logit model, binary trip split model, isotrip lines.

INTRODUCTION

The year 2006 marked an important moment in the history of humankind. For the first time in history more than half of global population were urban dwellers. While in 1900 only 13% of world’s population were urbanites, it is expected that by the year 2030 this proportion will have reached 60% [UN Habitat 2006]. Definitions of urban area may differ from country to country [Latham, McKormack, Macnamara, McNeil 2008], but worldwide rapid urbanization is a fact. A remarkable feature of that process is the emergence of mega- (population above 10 million people) and metacities (population above 20 million people). While Tokyo was the first to reach the threshold of 20 million in mid-60s, today at least 9 other agglomerations can be named metacities (at the same time population of Greater Tokyo reached 35 million people i.e. more than the whole population of Canada).

On one hand one could claim that such concentrated population is easier to manage than population dispersed over a large area e.g. it is easier and economically more viable to provide urbanites with basic amenities, education services or public transport. In fact, this would be true if and only if the newly emerging cities had the characteristics of so-called compact cities (Jenks, Burton, Wiliams 1996). Unfortunately, this is rarely the case. Omnipresent noise, low quality of air, confined spaces along with growing affordability of private cars make people move away from the city centres in a process of suburbanization leading to urban sprawl [Rodrigue, Comtois, Slack 2006].

The result is emergence of cities with their adjacent spheres of socioeconomic influence i.e. their metropolitan areas, extending on nearby suburban towns. This results in increased population flows
between these towns and central metropolis (e.g. commuting). The cost of these flows is increasing congeston on transport links (roads, public transport). But such problems are not a domain of solely mega- or metacities, but in fact a domain of every big and medium size city possessing its sphere of socioeconomic influence. The problem thus arising is how to manage these metropolitan flows sustainably i.e. in a way taking into consideration possible economic, environmental and social costs.

METROPOLITAN AREAS MANAGEMENT

The subject matter of the metropolitan areas management has been present in the scientific literature since as early as 1940s. This was the time when first comprehensive studies, surveys and land-use plans concerning American cities became available [Turton 1992]. This enabled scientists to extensively investigate and describe the dynamics of such areas. On the other hand, in case of London the literary boom in was driven by the concern about uncontrolled urban sprawl. This resulted in a massive amount of literature on metropolitan planning from Sir Patrick Abercrombie's famous idea of 'the metropolitan green belt' in 1940s [Munton 1986] to contemporary strategies such as The London Plan [Mayor of London 2004]. In general the idea behind remained the same: to accommodate different phenomena such as migrations, goods and information flows or trade in a way minimizing their associated costs: pollution, congestion or social costs such as mental discomfort. The attention with which policymakers and scientists approach the metropolitan areas comes from the fact that cities and their surrounding spheres of influence can become, as [Camagni, Capello, Nijkamp 1998] claim, a core of sustainable development for the whole region and even a country. The condition is that they need proper management which, as they point out, must also include its surrounding metropolitan area. In other words a city cannot develop sustainably if its adjacent areas do not follow that path (and vice versa). Thus it is widely recognized in the literature of the subject [e.g. Nijkamp, Perrels 1994, Hadaś 2004] that promoting 'spatial order and sustainable development are basic functions [...] of planning and spatial development' [Rzeczyński 2007].

In order to research and administer properly these vast metropolitan areas populated by thousands of people, the domain of metropolitan management is largely based on mapping and modelling. These techniques allow decision-makers and scientists to evaluate potential impacts of each proposed policy scenario. This allows them to conduct a cost-benefit analysis and seek for the optimal solution. As a result numerous models have been proposed for the management of urban areas. These range from traditional gravitational relationships [Yeates 1968] or Christaller's central place theory [Christaller 1966], through more complex mathematical models e.g. urban transport models [Ortuzar, Willumsen 1994, Quinet, Vickerman 2004], scenario analysis [Semenov 2004] towards GIS-based computer simulations [Longley, Goodchild, Maguire, Rhind 2005]. It turns out that such frameworks can prove to be quite successful e.g. in traffic congestion modelling [Robitaille, Nguyen 2003, Pawlak 2007].

Metropolitan areas management may also become increasingly challenging task in light of emergence of so-called polycentric metropolises [Pain, Hall 2006]. In such a case several compact megacities, closely interlinked to each other, will need to be managed in such a way as to remain sustainable and at the same time allow neighbouring cities to do so. As different forms of flows (people, goods, information) are inherent to this new urban form, there will be a need for tools facilitating the process of governing. The model presented below is a one of the potential responses to that need.

PROPOSED MODEL

In order to model the split of flows between settlements within the metropolitan area and the metropolis the logit model is applied. This class of models have been widely applied in transportation management, especially in simulating the modal split. The general form of this model applied as suggested by Ortuzar and Willumsen [1994] is given as:
where:

\( P^1_{ij} \) - proportion of trips between destination \( i \) and destination \( j \) via mode 1.

\( e \) - base of a natural logarithm

\( \mu \) - parameter controlling dispersion in mode choice and destinations

\( C^k_{ij} \) - generalized cost of travelling between destination \( i \) and destination \( j \) via mode \( k \).

\( k \) - mode number

The relationship expresses the proportion (probability of use) of certain mode of transport as a function of generalized costs of using that particular as well as all other available transport modes. However, the model can be brought into alternative use, helpful in describing and managing flows in the metropolitan area. Instead of calculating the proportion of trips undertaken via certain mode, model can be transformed to reflect number of trips between the suburban town and its metropolis as a proportion of total trips to and from that suburban town. The choice of such a model has been dictated by two reasons. The first one was the need to get a relationship whose resulting values are proportions (values between 0 and 1 rather than absolute numbers). This is because proportions reflect more conveniently the relative importance of contact of a suburban town with its metropolis as compared to town's contact with the rest of the world. Secondly, the mathematical properties (particularly the shape and limits) of logit relationship reflect well the fact that neither will all the trips concerning certain suburban town be made exclusively between the metropolis nor will the reverse be true i.e. no such trips will take place. Since in this case the only interest is in the distinction between two types of trips, the binary version similar to one described by Quinet and Vickerman [2004] is sufficient, although the model could easily be extended to handle multiple destinations.

Thus the binary trip split model is given as:

\[
P^1_{ij} = \frac{e^{-\mu C^1_{ij}}}{\sum_k e^{-\mu C^k_{ij}}}\]

where:

\( P^1_{ij} \) - proportion of trips between destination \( i \) and destination \( j \) via mode 1.

\( e \) - base of a natural logarithm

\( \mu \) - parameter controlling dispersion in mode choice and destinations

\( C^k_{ij} \) - generalized cost of travelling between destination \( i \) and destination \( j \) via mode \( k \).

\( k \) - mode number

\[
P_{ij} = \frac{e^{\alpha N_i + \beta D_{ij} + \epsilon}}{1 + e^{\alpha N_i + \beta D_{ij} + \epsilon}} = \frac{e^{\theta}}{1 + e^{\theta}}
\]

where:

\( P_{ij} \) - proportion of trips between town \( i \) and metropolis \( J \),

\( e \) - base of a natural logarithm,

\( N_i \) - population of suburban town \( i \) (in 10s of thousands),

\( D_{ij} \) - distance between suburban town \( i \) and metropolis \( J \) (in kilometres),

\( \theta \) - logit parameter used to simplify the expression \( \theta = \alpha N_i + \beta D_{ij} + \epsilon \),

\( \alpha, \beta, \epsilon \) - parameters

The relationship states that within a given metropolitan area, the proportion of trips made from a certain suburban town \( i \) to the metropolis \( J \) is dependent on the population of that town and distance from the metropolis. Moreover, also parameters \( \alpha, \beta, \epsilon \) play significant roles. Let:
\[ P_{i-J} = 1 - P_{iJ} \]

i.e. the number of trips between the suburban town \( i \) and all areas except the metropolis as a proportion of all trips to and from that town is simply 1 minus proportion of trips between the town and the metropolis. As a result:

\[
\frac{\partial P_{iJ}}{\partial D_{iJ}} = \beta P_{iJ} P_{i-J} \\
\frac{\partial P_{iJ}}{\partial N_i} = \alpha P_{iJ} P_{i-J} \\
\varepsilon = \ln\left(\frac{P_{iJ}}{P_{i-J}}\right) - \alpha N - \beta D
\]

Since \( P_{iJ} \) and \( P_{i-J} \) are positive, value (and sign) of \( \alpha \) will indicate whether bigger towns will have higher or lower proportions of their overall number of trips to and from the metropolis than smaller towns. On the other hand value of \( \beta \) will determine whether increased distance from the metropolis will cause a fall or rise in that proportion. Finally, \( \varepsilon \) is an all-catching variable indicating the strength of impact of other factors, not quantified in this paper.

It is expected that the value of \( \alpha \) is negative. This would mean that more populated towns have their proportion of trips to and from the metropolis smaller than towns of a lower population. The reason for such a hypothesis is the fact that more populated town means bigger local market and hence more opportunities for running own business reducing the need for commuting and thus generation of trips to the metropolis. Moreover, such a town itself creates its zone of influence covering adjacent villages or perhaps even smaller towns. Finally, more populated towns can maintain certain public services (high schools, hospitals) which otherwise would only be available in the metropolis thus requiring people to travel. Moreover, it is expected that the value of \( \beta \) parameter is negative i.e. proportion of trips will be lower for the towns located further from the metropolis. This hypothesis has dual explanation. Firstly, greater distance means higher cost of travelling (regardless of the mode) which is a disincentive for trip-making. Secondly, the further the town from the metropolis, the higher the chance that certain demand for a trip (e.g. need to go to a shopping centre) will be satisfied in an alternative way to visiting the metropolis e.g. by going to the shopping mall located in another metropolitan area.

The parameter \( \varepsilon \) is in fact a residual (unexplained factors). By its nature it is hard to estimate in advance, but it is more probable that its value is positive. This would mean that variations in populations and distance are not the only factors influencing changes in the proportion of trips between the metropolis and the suburban town e.g. it may catch the quality of transport links between the areas. In order to verify the hypotheses, model will be tested with the use of data from one of the biggest metropolitan areas in Poland - Poznań Urban Area.

**APPLICATION OF THE MODEL**

The best way to evaluate each model is to confront it with the real world data. To test the binary trip split model, statistical data for metropolitan area (Poznań Urban Area, Poland) will be used (for
the purpose of the paper defined as areas belonging to powiat ziemski). The data include Regional Census Data [GUS 2007] and Extensive Transport Survey 2000 [Kompleksowe Badanie Ruchu 2000, BIT 2000]. Distances are the road distances along main routes to Poznań city centre (assumed to be located at the Plac Wolności). Since data on number of trips between certain areas are from 2000, they needed updating especially in the light of the dynamic suburbanization taking place in the region in recent years [Parysek, Mierzejewska 2006]. This was done in accordance with the fact that number of trips between two places is proportional to the product of populations of these places [Quinet, Vickerman 2004].

Consequently, the growth rate of number of trips between Poznań and each suburban town is, ceteris paribus, the sum of growth rates of populations of Poznań and the suburban town. More formally:

\[ T_{iPOZ} = r P_i P_{POZ} \quad r = \text{const.} \]

\[ \frac{\partial T_{iPOZ}}{\partial t} = \frac{\partial P_i}{\partial t} + \frac{\partial P_{POZ}}{\partial t} \]

where:

- \( T_{iPOZ} \) - number of trips between town \( i \) and Poznań
- \( P_i \) - population of town \( i \)
- \( P_{POZ} \) - population of Poznań (metropolis)
- \( r \) - all-catching parameter, assumed to be constant
- \( t \) - time

Let also:

\[ \theta^* = \ln\left(\frac{T_{iPOZ}^*}{T_{iPOZ}^* + T_{i-POZ}^*}\right) - \ln\left(1 - \frac{T_{iPOZ}^*}{T_{iPOZ}^* + T_{i-POZ}^*}\right) \]

where * denotes empirical (statistical) data and \( \theta^* \) is logit parameter derived from that data.

Thus the relationship simply shows the method of calculating empirical values of logit parameter. The results obtained from regressing \( \theta^* \) on population and road distance are given in Table 1. Trip data for 17 surrounding communes (gmina) were used (where possible only data for towns instead of whole communes were included) and calculations were done with the use of Minitab 15 software.
Table 1. Results of the multiple regression analysis, $\theta^*$ vs. population and distance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>$\sigma$</th>
<th>$t$</th>
<th>$p$</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-0.1455</td>
<td>0.2045</td>
<td>-0.71</td>
<td>0.488</td>
<td>14</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.0776</td>
<td>0.0215</td>
<td>-3.62</td>
<td>0.003</td>
<td>14</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>1.1715</td>
<td>0.6450</td>
<td>1.82</td>
<td>0.091</td>
<td>14</td>
</tr>
</tbody>
</table>

Coefficient of determination: $R^2=0.593=59.3\%$

Source: Own elaboration based on data from BIT 2000 and GUS 2007.

The table presents also the values of estimated standard deviations of the coefficients ($\sigma$), values of t for the Student's t-test and p-values resulting from the test (probabilities of null hypothesis being true). The last column gives the number of degrees of freedom used in conducting the t-test. Thus for the Poznań Metropolitan Area the binary trip split model is given as:

$$P_{iPOZ} = \frac{e^{-0.1455 \times N_i - 0.0776 \times D_{ij} + 1.1715}}{1 + e^{-0.1455 \times N_i - 0.0776 \times D_{ij} + 1.1715}}$$

Clearly, the signs of the parameters conform to the assumed hypotheses and, by explaining almost 60% of variation in $\theta^*$, the model fits the empirical data quite well (although Pobiedziska and Suchy Las deviate significantly which will be discussed later). Yet in case of $\alpha$ and $\epsilon$ estimated standard deviations are large as compared to the parameters. Moreover, the resulting p-values are high as well i.e. statistical significance of the parameters is low. On one hand this means that were the data used a sample data, the resulting values of $\alpha$ and $\epsilon$ could be a product of coincidence (choice of sample) rather than reflection of underlying relationship. This would limit their use, for instance, in making inferences about suburban towns for which data are not available. On the other hand, in the analysis all suburban towns and communes in the metropolitan area were taken into account i.e. population rather than sample was used. As a result it can be claimed with certainty that within the metropolitan area - Poznań Urban Area, there exists a relationship between each town's population and the proportion of trips between this town and Poznań as a fraction of all trips to and from that town. Similar reasoning can be applied to the $\epsilon$ parameter meaning that there are also other factors influencing the investigated proportion of trips.

On the other hand parameter $\beta$ describing the character of impact of distance on $P_{iPOZ}$ was found to be statistically highly significant (probability that the relationship exists by a chance is less than 1%) and the estimated standard deviation is relatively smaller if compared to standard deviation of the remaining parameters. As a result it can be deduced that the relationship between the distance from the metropolis strongly affects the proportion $P_{iPOZ}$. Moreover, were the towns treated as a sample of suburban towns (e.g. if even larger area was investigated), the parameter could be than used to make inferences about the impact of distance on trip split in other towns or chosen area e.g. northern part of certain commune.
Fig. 1. Actual and modelled values of \( \theta \) parameter for different towns and communes in the Poznań Metropolitan Areas

**Legend:**
- Metropolis
- Suburban Town

**Value of \( \theta \):**
- \( > 0,5 \)
- \( 0,0 – 0,5 \)
- \( -0,5 – 0,0 \)
- \( -1,0 – -0,5 \)
- \( -1,5 – -1,0 \)
- \( < -1,5 \)

Source: Own research and elaboration

Rys. 1. Aktualne i modelowe wartości parametru \( \theta \) dla różnych miast i gmin w obszarze metropolitalnym Poznania
Since the research problem of this paper involves investigating spatially dynamic phenomena, the most convenient way to visualize the obtained results is using a map. Figure 1 presents the contour maps of researched metropolitan area - Poznań Urban Area, with contour lines, in this case named isotrip lines, connecting areas where trips between that area and the metropolis consist the same proportion of all trips involving that area. The upper map is the actual state of affairs i.e. the isotrip lines were derived basing on statistical data for the suburban communes. On the other hand, lower map has been created using values of $\theta$ generated with the binary trip split model. As a result it is possible to compare the real world to the model-generated world. The overall pattern is kept - highest values near metropolis and decline towards the borders of metropolitan area, although discrepancies are visible. The [-0.5, 0] contour includes more communes in the reality (importance of trips to and from the metropolis is greater in these areas) than the model would suggest. Furthermore, modelled map does not display the highest nor the lowest contour i.e. distribution of $\theta$ seems to be more even than in reality. Moreover, some communes fall into different categories (contours) and such a deviation is acceptable as long as these contours are neighbouring i.e. category which is just above or just below. If this criterion is used, only one town (commune) does not match it - Pobiedziska (PBD). In this case model predicts far lower proportion of the trips being conducted to and from Poznań than there is in reality. This feature means that there are more trips between Pobiedziska and Poznań than there should be and the reason for that is a matter for further research as is the reason for discrepancies between the model and the reality. But such a model generated picture could enable policymakers or scientists to assess the transportation needs and in fact could be one of the methods of delineating where the metropolitan area ends. For instance it could be assumed that the city belongs to Poznań Metropolitan Area if and only if its $\theta$ parameter is greater or equal to -1 which is equivalent to saying that 26% or more of all trips concerning the area needs to be made between that area and the metropolis). In this case the -1 isotrip line would mark the borderline of the metropolitan area.

In order to look at the problem from a different perspective, the data were plotted in the 3 dimensional space (Figure 2) allowing for better visualization of existing relationships (x-axis is distance from Poznań in kilometres, y-axis is population of suburban town and z-axis is the empirically found value of $\theta^*$). In the diagram the negative relationship between $\theta^*$ and distance is clearly visible (points aligned along a plane sloping down as distance increases). The same cannot be said about the relationship between $\theta^*$ and towns' populations - no clear pattern is visible in this case.
This confirms the results obtained in the course of regression analysis i.e. this link is not very clear nor strong. Last but not least, as already mentioned Suchy Las deviates very strongly from the model. In other words, unusually high proportion of all trips to and from the commune is conducted to and from Poznań. The reason for such a condition may lie in a very dynamic economic development and indeed population growth, along the arterial road 11 (Droga Krajowa 11) leading to smooth transition of built-up area between Poznań and Suchy Las. In fact any strict border between the areas is hardly observable. Such integration with the metropolis (becoming practically part of it) may manifest itself in this exceptionally high importance of trips to and from Poznań. Yet this explanation is only a suggestion and the possibility of such a transition from being an independent suburban commune to being de facto a part of the metropolis is a subject of ongoing research.

CONCLUSIONS

In the course of the research it was found that the proposed binary trip split logit model is a useful tool for the metropolitan management. Having applied the model in a real world situation, it was found that the distance from the metropolis and the population of the commune (or town) have an impact on the distribution (proportions) of trips between the communes (towns) and metropolis. Yet only the estimated relationship between the trip proportion and the distance from the metropolis was statistically strong enough to make a potential use of it in making sound inferences for areas where data was not readily available. Moreover, the model was able to reflect the general pattern of trip distribution in the Metropolitan Area of Poznań, although some discrepancies were found. For some communes and towns it underestimates the significance of trips between those areas and the metropolis and, what is more, predicts less sudden changes in that proportion throughout the metropolitan area. Finally, two areas of exceptional deviation from the model were found. In case of Pobiedziska the proportion of trips between that town and Poznań is far higher than it should be for a comparable city placed in such a distance. Similarly, Suchy Las reveals exceptionally high significance of trips between the commune and the metropolis. Factors lying behind both these discrepancies which could include for instance the quality of transport links, are subjects of ongoing research. Also the meaning of the residual $\varepsilon$ is being investigated. As for now the application of the model in urban and metropolitan management could come through its use in transport infrastructure planning, urban development plans and wider policymaking aimed at more sustainable existence of those areas.

REFERENCES


ALTERNATYWNE ZASTOSOWANIE MODELOWANIA LOGITOWEGO W ZARZĄDZANIU OBSZARAMI METROPOLITARNYMI: STUDIUM PRZYPADKU DLA POZNAŃSKIEGO OBSZARU ZURBANIZOWANEGO

STRESZCZENIE. W artykule została przedstawiona nowa metoda modelowania przepływów ludnościowych pomiędzy obszarem zurbanizowanym (aglomeracji) a jej ośrodkiem centralnym. Prezentowana technika modelowania oparta jest na kształtowaniu logitowym, stosowanym w badaniach nad problematyką doboru optymalnego środka transportu. Oprócz wyjaśnienia założeń modelu, przedstawione jest również jego zastosowanie w opisaniu i zbadaniu obszaru aglomeracji poznańskiej. W toku badań wykazano, iż głównymi czynnikami determinującymi ilość podróży pomiędzy poszczególnymi obszarami aglomeracji a Poznaniem wyrażonymi jako proporcja wszystkich podróży do i z danego obszaru są odległość od...
centrum aglomeracji oraz populacja badanego obszaru. Ponadto wyniki badań wskazują na obecność innych czynników. Model dowiódł zatem swojej przydatności w kontekście praktycznym, jakkolwiek jego potencjalne użycie we wnioskowaniu dla obszarów, dla których dokładne dane statystyczne są niedostępne, jest w obecnej chwili ograniczone i pozostaje tematem do dalszych badań.

Słowa kluczowe: obszar metropolitalny, rozwój zrównoważony, model logitowy, binarny model podziału podróży, izotropy.