CAPABILITY OF HUFF MODEL TO PREDICT MARKET SHARE

Jamal Aimi Jamaludin¹, M. Rafee Majid² and Ahmad Nazri Muhamad Ludin³

¹Tutor, Department of Urban and Regional Planning, Universiti Teknologi Malaysia, jamalaimi@utm.my
²Assoc. Professor, Department of Urban and Regional Planning, Universiti Teknologi Malaysia, rafee@utm.my
³Professor, Department of Urban and Regional Planning, Universiti Teknologi Malaysia, b-anazri@utm.my

ABSTRACT: Gravity model is widely used in the field of transportation planning, trade and allocation of public facilities. Originating from the famous Newton’s Gravitational Law, gravity model has been modified and adapted to various purposes and applications in many fields. However, there are too many varieties of variable being used, especially various types of distance and non-physical variables, putting the viability of the model to explain an ongoing phenomenon and to predict future conditions into question. Application of the gravity model also requires extensive socioeconomic data which, in some countries, are difficult to obtain. Thus, can we use a basic gravity model that only uses the size and distance variables to explain a phenomenon? For that reason, this research compares the accuracy of four (4) modified Huff Gravity Model variants using only i) hypermarket floor area data, and ii) either Euclidean distance or actual distance to predict market share of three selected hypermarkets. The predicted results were then compared to the actual market share of each hypermarket, obtained through customer survey. Results show that the variant that uses Euclidean distance with parameter; power of two (2) function come out with the most accurate market share explanation. The modified Huff gravity model generated by using spatial analyst capability in geographical information system (GIS) further shows the gravitation surface of each hypermarket, and even successfully helped to explain the spatial interaction between each competing hypermarket and their market share - proving the viability of the gravity model to explain existing interaction phenomena and forecast future interaction pattern.

Key words: Huff model accuracy, gravity model evolution, spatial interaction models, geographical information system (GIS), market share prediction.

1.0 INTRODUCTION

Harris in Samat and Masron [10] wrote that, urban planning is an attempt to come up with the best idea or series of actions required in future to create a better urban environment. In most cases, however, these attempts were limited by the complexity of urban-related phenomena in which planners have to deal with huge amount of urban data, complex relationship between urban entities and magnitude of uncertainty of possible future scenarios. These limitations had for long become challenges for planners en route of creating the best urban environment for the future.

In order to predict the future, a planner must first understand the pattern, characteristics and reasons behind a present phenomenon. According to Wheeler [13], models have been used to explain and help us understand a phenomenon for so long. Only by understanding historical and present phenomena would we be able to predict the behaviour of that certain phenomenon in the future and arrive at better planning decisions.

Among of the most used prescriptive and predictive model of urban planning is the gravity model. Ediwan [2] mentioned that the model have been widely used for applications in various fields such social science, business and facilities allocation, the gravity model had undertaken vast evolution and enhancements to fit its various application needs in different fields. This gravity model originated from
The Newton Gravitational Law (NGL) first introduced by Sir Isaac Newton in 1687. He wrote that the interaction between two entities or objects is governed by their mass and the inverse of their distance. This means, bigger entities produce more interaction, while farther distance makes entities interact less. The formula is usually written as:

\[ F_{ij} = \frac{G M_i M_j}{D_{ij}^2} \]

(1)

where \( F_{ij} \) represent degree or value of interaction between the \( i \) and \( j \) entities, \( M_i \) is mass for item or entity \( i \), \( M_j \) represent the mass of entity \( j \), \( D_{ij} \) is the distance between \( i \) and \( j \), while \( G \) is the gravity constant for interaction between \( i \) and \( j \).

1.1. Evolution of Gravity Model in Market Study

Katiman [4] wrote that in 1929, Reilly had revolutionised the use of NGL when he applied the main principles of NGL for economic and business application. For economic purposes, gravity model assumes that the probability for a consumer to choose a facility, in this case a retail outlet, is parallel to the facility attraction force and inverse to distance between the facility and consumer. Later, Reilly’s Retail Gravitation Law was modified and popularised by Converse in 1949, as explained by Chasco & Vincens [1]. Skogster [11] further added, that Reilly introduced the concept of breaking point and identified the inertia factors that visualize unwillingness of consumers to travel beyond certain distances for shopping purpose. Reilly’s Retail Gravitation Law is:

\[ B = \frac{P}{D} \]

(2)

where \( B \) is sale value of a shopping outlet in an area \( a \) or \( b \). \( P \) represent population of the area \( a \) or \( b \). While \( D \) is the distance between a shopping outlet with area \( a \) and \( b \). Converse Gravity Model whereas is as follows:

\[ D_A \]

(3)

\( D_A \) is distance of city \( A \) to breaking point. \( D_{AB} \) is the distance between city \( A \) and city \( B \) while \( P \) is population of the city \( A \) and city \( B \).

However, Eppli and Shilling [3] explained how the Reilly model had been criticized for being too dependent on the distance factor alone, and only considering competition or interaction of two retail outlets. While in real situation, there are multiple ways interaction between several retail outlets with their market within an area. For that reason, Huff upgraded the Reilly and Converse models based on Luce opinion on consumer discrete choice which says that consumers will only choose a facility with most optimal benefit and interest, in comparison to the other alternative facilities within their reach.

Huff in 1963 later theorised that when a consumer had plenty of alternatives to go shopping, they probably chose to stop by several shopping malls rather than just sticking to one nearest shopping mall. That means every shopping mall within consumers’ willingness distance to travel had a certain percentage of probability to be visited by consumers. The resulting gravity model uses travel time distance and shopping malls' retail floor area as variables to determine probability of a consumer to shop at any one retail entity. This model also agreed that the market is complex, aggregate and based on spending probability. This particular gravity model has been the most applied gravity models as it is widely accepted and used in many fields.

Spatial Interaction Model by Huff proposed that a bigger shopping outlet has higher chance or probability to attract more shoppers while a distant shopping mall had a smaller chance to be visited, compared to a nearer shopping mall. This model introduces the use of travel time, rather than physical distance, as its distance variable to improve the Reilly Model and also suggest the use of variable parameter in order to explain the different interaction and different travel purposes. According to Skogster [11], Lakshmanan and Hansen then used retail floor space to determine attractiveness of a
shopping mall compared to its competitors, allowing overlapping of the market catchment area and competition or interaction of every shopping mall in an area.

The Huff Gravity Model is represented as:

\[
P_{ij} = \frac{S_j D_{ij}^{-b}}{\sum_{k} S_k D_{ik}^{-b}}
\]

where \( P_{ij} \) is probability percentage consumer in \( i \) to shop at outlet \( j \), \( S_j \) is the attractiveness level or size of outlet \( j \), while \( D_{ij} \) is distance between area \( i \) and outlet \( j \). \( b \) is parameter of distance base on travel time.

Riza [9] wrote that most empirical findings admit that Gravity Model is able to predict market share of each retail outlet within acceptable accuracy. However, several scholars mention that to make its prediction even more accurate, geo-psychology factors must be taken into account to support the influence of size, population and distance factors. Since most gravity models only concern about the influence of outlets size and distance, without counting in physical and consumer psychology factors related to Consumer Behaviours Theory. Nakkanishi and Cooper [7] later inserted and manipulated several retail store attraction factors such environment, facilities level, image, service level and brand affect inside their gravity model which was later known as Multiplicative Competitive Interaction Model (MCI). Later, Okoruwa, Nourse and Terza in Lee [5] added other factors such outlets age, type, economy and consumer social status obtained through survey. The Multiplicative Competitive Interaction Model (MCI) has therefore become:

\[
\frac{U_{ij}}{\sum_{k} U_{ik}} = \beta_k
\]

IN MCI, \( P_{ij} \) is a probability of consumer from zone \( i \) to shop at outlet \( j \). Where, \( U_{ij} \) represent the attraction value of outlet \( j \) toward the consumer in \( i \), which is defined by a set of the outlet attributes or even external factor that attract the consumer in \( i \) to shop in \( j \). \( \beta \) is parameter of each attribute that represent consumer sensitivity toward that certain outlet attribute \( k \).

1.2 Issues Regarding Viability of Gravity Models’ Predictions

Despite the amount of applications and modifications made, there are still unresting issues regarding the viability of gravity model prediction or even their capability to explain an ongoing phenomenon. Todes in Riza [9] observed that Gravity model used in Chicago was unable to explain travel distribution in Cape Town as there was a constraint of Apartheid Law. While Riza [9] observed the same weakness when he applied the gravity model to explain travel trend of people in Kota Bharu, Malaysia. Pace and Lee [8] on the other hand find out that the uses of the parameter in gravity model most of the time resulted in unsatisfactory prediction.

Kolter in Pace & Lee [8] remarked that size of shopping outlets was significantly related to their total sale value. While, Stanley and Seewal, in Pace and Lee [8] agreed that gravity model’s capability to explain and predict the interaction between a shopping entities with its market is very limited. Eppli and Shilling [3], on the other hand, doubted the importance of the distance factor in determining outlets attraction level. While, Brown in Skogster [11] quoted to agree that the basic gravity model with distance and population variable with parameter would not always perform well whenever applied. However, Suarez-Vega et al. [12] successfully apply the Huff model en route of determining the new location for hypermarket.

Several researchers, such Nakanishi and Cooper [7], include various social economic variables that are too subjective and abstract to be valued with the hope they might simulate or imitate the real world phenomenon. For that purpose, they required many data, complex calculation and evaluation of abstract elements in order to determine the interaction between a shopping outlets and its market. However, the consumer shopping location choice most of the time may not be influenced by that entire factor. According to Clarke in Samat & Masron [10], one of gravity model weaknesses is that it required
numerous numbers of household data within the market catchment area, social economy and travel data. In developing countries, the availability of such complex data is very limited and the data structure are unorganised as wrote by Samat and Masron [10]. This affects the rationality of gravity to be used as planning support system (PSS) and its capability to assist ad-hoc planning decisions.

The question now is whether we can just use a simple or basic gravity model that only uses size and distance variable without including complex social–economy, geo-physical and consumer psychology factors, to predict consumer shopping location? For that purpose, the Huff Gravity Model was selected to be tested, as it is among the simplest variation of gravity model and needs only two variables; namely entities’ size and distance between entities.

This research shows whether the usage of only distance and entities’ size factors is adequate in helping planners to explain the shopping phenomenon, and capability of a gravity model variant to predict and explain consumer behaviour in choosing their shopping location. If it is proven that simple or basic Huff Gravity Model is able to explain and predict this phenomenon well, then the gravity model may be used as a planning support device for planner to produce better development decisions.

2.0 METHOD

In order to evaluate the viability of the Huff model in explaining and predicting a phenomenon regarding urban commercial land uses, this research capitalises on the spatial analysis capability of geographic information system (GIS) to generate the Huff model and visualize gravitational surface of each sampled hypermarket in Johor Bahru. Gravitational surfaces of each hypermarket show patterns and interaction level of each sampled hypermarket with its market and competitors. In order to evaluate the precision of the model, the output of the model was later plotted against the actual origins (distance) of its customers obtained from a customer survey.

This research evaluates prediction precision of four variants of the modified Huff Model. Rather than using the time travel variable as suggested by Huff, this research uses distance variable (either actual distance or Euclidean distance) which is the simplest data to be acquired using existing spatial data. Each of the variant uses either Euclidean distance or actual distance and different power parameter function - the power of two or without any power function. For comparison purposes, three Giant hypermarkets in Johor Bahru District were selected as unit of analysis. Selection of hypermarket from the same franchise aimed to reduce the effect of different prices, variety of products, facilities, environment, consumer perception and effects of hypermarket promotion factors, in influencing the interaction level between a shopping entity with its market and competitors. This ensures an assessment that is only based on distance and hypermarket size.

This study uses a modified Huff Model shown below from which all the four variants of the model originate:

\[
P_{ij} = \frac{S_i D_{ij}^b}{\sum S_i D_{ij}^b}
\]

\[P_{ij}\] is the probability percentage of dwellers in a residential area i opt to visit or shop in hypermarket j compared to j competitors. \(S_i\) is the retail floor space area of hypermarket j in meter square, \(D_{ij}\) is the distance (actual or Euclidean) of hypermarket j from residential i, while b is the power function parameter of distance from i to j.

The difference between surveyed percentage and predicted percentage of customers from origin i to shop at hypermarket j, determines each variant’s accuracy through total absolute error (refer to Eq. 7). Thus, the variant with the lesser amount of total absolute error would be the most accurate model.

\[
\text{Variant Accumulated Error (\%)} = \frac{1}{n} \sum |\text{Surveyed Percentage (\%)} - \text{Predicted Percentage (\%)}| 
\]

(7)
3.0 RESULTS AND DISCUSSIONS

3.1 Most Accurate Huff Model Variant

Based on the gravity model variant from Eq. 6 and the errors determined using Eq. 7, the Huff model variant using Euclidean distance raised to the power of two produced the most accurate prediction with its total average accumulated error of only 2.8% (refer to Table 1). Usage of power function as parameter for both Euclidean and actual distance helped make the predictions for hypermarket and market interaction phenomena to be even more accurate as had been suggested by Huff. This is parallel with the finding by Majid and Yaakup [6] which proved the inverse correlation between distance and number of customers, which resulted in decaying exponential power of two for the distance variable to be used in Huff Model. In that study, they observed that the number of customers decreased in pattern of power of two when plotted in a graph.

<table>
<thead>
<tr>
<th>Gravity Model Variant</th>
<th>Euclidean Distance</th>
<th>Euclidean Distance^2</th>
<th>Actual Distance</th>
<th>Actual Distance^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulate Error (%)</td>
<td>2.894</td>
<td>2.842</td>
<td>2.904</td>
<td>2.874</td>
</tr>
<tr>
<td>Correlation of Surveyed and Predicted based on R^2 value</td>
<td>0.096</td>
<td>0.112</td>
<td>0.091</td>
<td>0.096</td>
</tr>
<tr>
<td>Accuracy Percentage (%)</td>
<td>67.9</td>
<td>76.6</td>
<td>67.1</td>
<td>74.6</td>
</tr>
<tr>
<td>Accuracy Rank</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

It is also observed that by using Euclidean distance the prediction will be even more accurate compared to the use of actual distance. This might be due to the vast and intensive road networks of our study area; Johor Bahru, which made the use of Euclidean distance adequate and even more significant than using actual distance. The small accumulated error differences and R^2 value among all four distance variants also suggested that perhaps a distance factor is not among the most influencing factor considered by consumer in Johor Bahru in deciding their spending location. This condition might result from the small differences or the range between Euclidean distance and the actual distance travelled by each consumer to reach their preferred hypermarket. Due to a high level of accessibility in Johor Bahru, the varied use of distance variable might have less impact towards the market share proportion.

![Figure 1 Prediction accuracy of four modified Huff model variants.](image)

According to Riza [9], at least 60 percent accuracy is needed in order for a model to be viable. Based on result in Table 1, each tested variant have the accuracy of more than 60 percent, with the variant that use Euclidean distance without power function emerged as the most accurate variant with 65.4 percent accuracy and the less average accumulate error value of 2.842 percent. For that reason, the variant was used in assisting us to predict the market share proportion and interaction for three hypermarkets in Johor Bahru.
### 3.2 Interaction of Hypermarket and Market Share

The most accurate variant (Eq. 8) was used to develop and display the gravitation surface of each sampled hypermarket. This allowed researchers to evaluate and compare the modified model capability in explaining market share and market interaction.

\[
P_{ij} = \frac{S_j}{\pi D_{ij}^2}
\]

(8)

\(P_{ij}\) is the probability percentage of dwellers in a residential area \(i\) opt to visit or shop in hypermarket \(j\) compared to \(j\) competitors. \(S_j\) is the retail floor space area of hypermarket \(j\) in meter square, and \(D_{ij}\) is the Euclidean distance of hypermarket \(j\) from residential \(i\).

This analysis compares the model predicted and the surveyed percentage of customers from a specific residential area to two competing hypermarkets. Although there are three competing hypermarkets surveyed in this study, we are only able to provide a simultaneous comparison for two hypermarkets at each time due to data limitation on customer origins.

The generated model successfully shows its capability to visualize the interaction and market share proportion of comparing hypermarkets (refer Table 2). The model indicates preferred shopping location of dwellers in respective residential areas. It is obvious from this result that distance factor is among the most influential factors that facilitate customer shopping preferences. This finding may be evidence of the Central Place Theory introduced by Christaller which states that most customers prefer to shop at the nearest preferable location to their residential area.

Comparison of survey data and the modelled ones obtained using modified Huff Model for Taman Daya and Taman Johor proves that the model is able to predict and explain the interaction between a shopping outlet against its competitor and its market. As for both residential areas, the proportion of market share predicted by the model is almost equal to the surveyed data, with less than 10 percent difference (refer to Table 2).

<table>
<thead>
<tr>
<th>Resident Origin</th>
<th>Market Share Proportion – Based on Customer Origin Survey (%)</th>
<th>Market Share Proportion – Based on Modified Huff Model (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Giant Skudai</td>
<td>Giant Plentong</td>
</tr>
<tr>
<td>Kg. Melayu Majidee</td>
<td>4.55</td>
<td>95.45</td>
</tr>
<tr>
<td>Taman Daya</td>
<td>20.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Pangsapuri Bukit Saujana</td>
<td>14.29</td>
<td>85.71</td>
</tr>
<tr>
<td></td>
<td>Giant Southern City</td>
<td>Giant Plentong</td>
</tr>
<tr>
<td>Bandar Baru Permas Jaya</td>
<td>43.75</td>
<td>56.25</td>
</tr>
<tr>
<td>Taman Desa Harmoni</td>
<td>42.86</td>
<td>57.14</td>
</tr>
<tr>
<td>Taman Johor Jaya</td>
<td>3.70</td>
<td>96.30</td>
</tr>
<tr>
<td>Taman Saujana</td>
<td>27.78</td>
<td>72.22</td>
</tr>
<tr>
<td></td>
<td>Giant Skudai</td>
<td>Giant Southern City</td>
</tr>
<tr>
<td>Kampung Pasir Putih</td>
<td>66.67</td>
<td>33.33</td>
</tr>
</tbody>
</table>

For the other six areas of customer origins, which are Kampung Melayu Majidee, Pangsapuri Bukit Saujana, Bandar Baru Permas Jaya, Taman Desa Harmoni, Taman Saujana and Kampung Pasir Putih, even though the model is unable to precisely predict the percentage of their preferable shopping location, the model is still able to predict and visualize which hypermarket has a more dominant market share compared to its competitors (refer Figure 2).
The lack of accuracy in predicting market share proportion for the six residential areas, might be due to not taking into account the other factors mentioned in Customer Behaviour Theory in choosing shopping location, such traffic congestion, local customer perception toward each hypermarket, familiarity of customer with each hypermarket, dwellers daily activities routine which include their route to the workplace, promotion and even competition from other outlets in their respective residential area.

Figure 2  The modified Huff Model helps predict market share proportion and visualize interaction.

The research also made use of the same model to simulate the gravitation surface of the three hypermarkets in order to visualize market share proportion and the interaction among them (refer to Figure 3). It is obvious from the gravitation surface of each hypermarket, Giant Plentong receives the biggest market share compared to Giant Skudai, while Giant Southern City is the one with smallest market share. This is due to the store size factor. The gravitation surface indicates that Giant Plentong and Giant Skudai might probably attract more population to their store.
This gravitation surface generated can be used for the purpose of evaluating market competition, to project income or profit of each hypermarkets and to determine the shopping location for the future. It can also be used to efficiently plan for the optimum distribution of hypermarkets in an area or a city on the basis of various factors. Transport planners can also use the model to plan for more efficient routes connecting residential areas and commercial areas.

4. CONCLUSION

The study shows that modified Huff Gravity Model using only hypermarket size and distance variables is capable of predicting consumers’ shopping location preferences and explaining market share and interaction of competing hypermarkets. Assessed against surveyed data of eight different customers origins (residential areas), the model quite successfully predict the shopping location preferences of these customers. However, its lack of precision in predicting the percentage of market shares of competing hypermarkets reveals the need to incorporate other variables, such as socioeconomic variables, into the model. Expanding the model to include more than just a few competitors at any one time would also improve the model.

REFERENCES


