Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 26 (2015) 66 – 69

Procedia Environmental Sciences

Spatial Statistics 2015: Emerging Patterns

Ordinary Least Squares Regression Method Approach for Site Selection of Automated Teller Machines (ATMs)

Kübra Bilginol[a], Hayri Hakan Denli[a], Dursun Zafer Şeker[a]

[a] *Istanbul Technical Universiy, Maslak, Istanbul 34564, Turkey*

Abstract

The importance of alternative distribution channels on finance sector is increasing day by day. Most of the customers consider about ATM accessibility. It is an important criterion for customers' bank choices. Banks are making investments for increasing the number of ATMs to serve more customers and generate ease-of-use. Here, it is important to make the investments to appropriate locations for new ATM establishments. The ATM location distribution must be balanced while ensuring efficiency and answering demand of the costumers. On the site selection problems there are many spatial statistical methods such as Factor analysis, cluster analysis, neural networks, regression analysis and correlation analysis. In this study, canonical correlation analysis and ordinary least squares regression analysis are used. First of all, to find which criteria affect ATM site selection decision, all inputs are analysed with correlation method. Then, the result criteria which affect ATM locations are analysed on ordinary least squares regression model. Finally the optimum ATM locations and the predictive efficiencies of those ATMs are found. This study aims to make predictive analysis and to find optimum locations for ATMs by using ordinary least squares regression method.

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee

Keywords: Ordinary least squares regression analysis; predictive analysis; correlation analysis; decision support

1. Introduction

At many developed countries,access to financial services rate is about %90 and the people who do not have a bank account is very limited.Besides the average distance to financial service points like an ATM or bank branch is an indicator ,at access.Atm or bank branch per kilometer is evaluated as an indicator of accessibility [1]. Each bank makes studies to improve their financial accessibility rate within their direction of strategy.The most striking one of these strategies is the dissemination of distrubition points.

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns Committee
doi:10.1016/j.proenv.2015.05.026

In Turkey, there are 44.443 ATM's [2] and 46 different banks by date 2014 September [3].There are approximately 1000 ATM's for each bank.While this count was about 900 in 2013 ,number of ATM's shows increase about %10.

To increase rivalry and access more client,banks increase the number of ATM's day by day.Increasing ATM's also have brought the risks.A wrong positioned ATM can not perform the expected count and it leads bank to make loss from the aspects of renting,transportation and holding cash.To prevent the loss and provide productivity,ATM's must be positioned correctly for number of operation and efficiency.At this point,selection of location problem shows up.

There are many factor that affecting the desicion of selecting location.The primary factors are,the definition of client mass,the location of the client mass,rival analysis,demographical composition,intensity of social media, point of sales locations,point of interest maps and alternative service locations.To reach optimal and rational solutions,the factors that affect selection of point should be analyzed as spatial and weights which affect decision should be found and used as a referance of selection.

For all factors optimal spatial analysis,inverse distance weighted,least squares method,analytical hierarchy process etc. statistical methods are used.Scope of study,with detection of the datas which affect the selection of ATM's location,with simple linear regression analysis method,datas put into operation with spatial influance area and as a result ; for chosen sample bank,optimum field selection areas had formed and predictions had made for feasibility areas.

2. Method

In this study due to multiple parameters affecting the ATM location selection and number of transactions, to minimize the errors, regression analysis with least squares method is used.

Least squares method is the method that minimized the error due to the parameter modeling of the selection model. Using least squares method, the summation of the squared differences between actual and estimated output values are minimized by gradient descent [8].

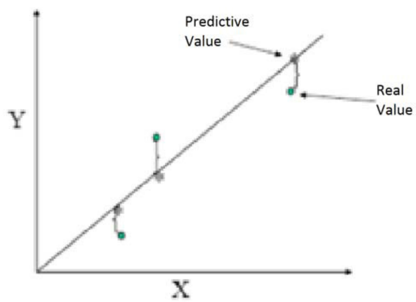

Fig.1. OLS

The implementation of the least squares algorithm on the prepared data is achieved using the OLS (Ordinary Least Squares)-Least Squares (LS) tool of the ArcGIS Spatial Statistics toolbox.

The analysis suggests that the POI and population are negatively correlated with the ATM Transactions while the social media and the rival distribution in there are positively correlated.

Multiple R-Squared is a correlation coefficient specific to the regression modelling. The closer the R-Squared to the 1, the more accurate the data is. In conclusion the R-Squared value is calculated as 0.95. The feasibility of the ATM location is decided with 5 percent error.

```
                                Summary of OLS Results
  Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE  Robust_t Robust_Pr [b]  VIF [c]
 Intercept        0.387119 0.194460    1.990737        0.046510*  0.153825  2.516616     0.011843* --------
COUNTSOCIA       -0.702198 0.065444  -10.729833        0.000000*  0.644033 -1.090314     0.275576  1.308669
 JOINCOUNT       -2.574678 0.499328   -5.156289        0.000001*  5.978828 -0.430633     0.666754  1.689273
TOTAL_AMOU        0.002185 0.000003  708.286540        0.000000*  0.000076 28.823418     0.000000* 1.491757
     DEGER        0.306839 0.024621   12.462442        0.000000*  0.295877  1.037050     0.299710  1.542543


                                       OLS Diagnostics
Input Features:                       all_join2   Dependent Variable:                           TOTAL_TRX
Number of Observations:                   32767   Akaike's Information Criterion (AICc) [d]:   326063.615727
Multiple R-Squared [d]:                0.957974   Adjusted R-Squared [d]:                           0.957969
Joint F-Statistic [e]:            186699.213248   Prob(>F), (4,32762) degrees of freedom:           0.000000*
Joint Wald Statistic [e]:           1298.449820   Prob(>chi-squared), (4) degrees of freedom:       0.000000*
Koenker (BP) Statistic [f]:        14459.257389   Prob(>chi-squared), (4) degrees of freedom:       0.000000*
Jarque-Bera Statistic [g]:   2484079654.150368   Prob(>chi-squared), (2) degrees of freedom:       0.000000*


Notes on Interpretation
 *  An asterisk next to a number indicates a statistically significant p-value (p < 0.01).
[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.
[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.01); if the Koenker (BP)
Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.
[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.01); if the Koenker (BP) Statistic [f] is statistically
significant, use the Wald Statistic to determine overall model significance.
[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.01), the relationships modeled are not consistent (either due to non-
stationarity or heteroskedasticity).  You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic
to determine overall model significance.
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.01) model predictions are biased (the residuals are not normally distributed).
```

Fig.2. OLS results

3. Conclusion and Discussion

In conclusion, an infrastructure which is capable of predicting the ATM transactions using linear regression analysis based on the population, social media, POI and rival information is developed using the geographical information.

In this study, a sample bank which is located in Istanbul is chosen. The feasible locations other than the position where ATMs are located and the positions which are in 500m radius of the ATM are found. The numbers of transactions are estimated for those feasible locations and the risk is minimized to enable the calculation of the efficiency.

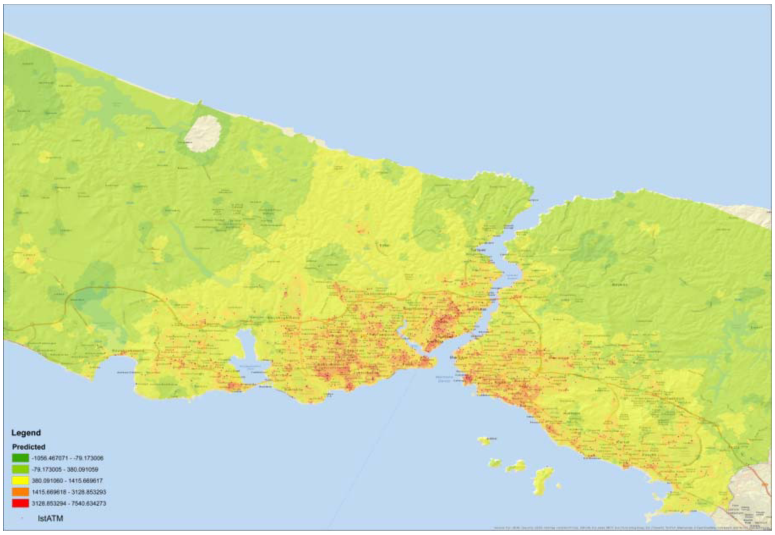

Fig.3. ATM Feasibilty map

References

1. Dünyada ve Türkiye’de Finansal Hizmetlere Erişim ve Finansal Eğitim, Türkiye Merkez Bankası, Mart 2011
2. BKM, Eylül 2014
3. Türkiye Bankalar Birliği, Türkiye'de Bankacılık Sistemi Banka, Şube ve Personel Sayıları, Eylül 2014
4. http://wiki.gis.com/wiki/index.php/Regression_analysis
5. Practical Regression and Anova using R , Julian J. Faraway, July 2002
6. Regression Analysis Using GIS, Jennie Murack, MIT, 2013
7. https://geodacenter.asu.edu/system/files/geodaworkbook.pdf
8. Korelasyon Ve Tekli Regresyon Analizi-En Küçük Kareler Yöntemi, Prof.Dr.A.KARACABEY Doç.Dr.F.GÖKGÖZ, Ankara Üniversitesi, Temmuz 2012
9. http://resources.arcgis.com/en/help/main/10.1/index.html#/An_overview_of_the_Modeling_Spatial_Relationships_toolset/005p0000001w000000/
10. Allen, R. C. Socioeconomic Conditions and Property Crime: A Comprehensive Review and Test of the Professional Literature. American Journal of Economics and Sociology, 55(3), 293-305, 1996.
11. Zar, J.H. Biostatistical Analysis, pp. 328-330 & 380. In J.H. Zar, Simple Linear Regression & Simple Linear Correlation. Prentice Hall, New Jersey, 2010.
12. Bates, D. M. & D. G. Watts. Nonlinear Regression Analysis and Its Applications. New York: Wiley. 1988.
13. Gallant, A. R. “Nonlinear Regression.” The American Statistician 29:73—81. 1975.