

Using Log Transformations to Improve AADT Forecasting Models in Small and Medium Sized Communities

Mehrnaz Doustmohammadi^{1,*}, Michael Anderson¹, Ehsan Doustmohammadi²

¹Department of Civil Engineering, University of Alabama in Huntsville, USA
²Department of Civil Engineering, University of Alabama at Birmingham, USA

Abstract Due to cost limitations, Annual Average Daily Traffic (AADT) data is not typically collected for every roadway segment, therefore, it is necessary to have a means to estimate the AADT value when the need arises for an uncounted roadway or roadway segment. Often, the methodology used to develop an estimate of the AADT is through a regression based, linear model. This research examines the use logarithmic transformations to improve the relationship between key socio-economic, roadway variables and the estimated AADT. In the study, traffic count, socio-economic, and roadway data were collected and different regression based models were developed and tested to determine if the use of logarithmic transformations improves the model accuracy and the model transferability to other communities of similar size. The results of the paper indicate that a linear-log model produced the best results of the logarithmic transformations, and was an improvement over a traditional linear regression model.

Keywords Annual Average Daily Traffic, Statistical Analysis, Urban Communities, Logarithmic Transformations

1. Introduction and Background

Annual Average Daily Traffic (AADT) is a critical input to many transportation analyses including but not limited to safety assessments, maintenance schedules, and capacity improvements [1, 2]. The amount of effort, both time and cost, necessary to collect actual AADT data for every roadway in a community limit a community's ability to obtain data for each unique roadway and roadway segment [3]. The demand for quality AADT data on local roads, coupled with the lack of available accurate data, has prompted previous research to examine and develop models that can accurately estimate AADTs within a small or medium sized community based on socio-economic variables [4-15]. The models use a combination of roadway and socio-economic factors within a given distance of the roadway segment and linear equations to convert the factors into AADT estimates. The linear equations provide a basis for estimating the AADT; however, linear models are not the only possible models that can be used. Logarithmic transformations are often used to improve the linear models [16]. The transformations are especially useful when it is unlikely that the data will follow a continuous trend, as with

the case for traffic volumes which tend to limit themselves at capacity of the roadway regardless of the surrounding socio-economic data.

Linear models, with the associated logarithmic transformations, are shown in Table 1. The transformations, as mentioned, are used to adjust the data to improve the model fit. The use of logarithmic transformations in transportation is prevalent in safety analysis [1, 17] and household travel survey and data [18, 19].

Table 1. Logarithmic Transformations

	X	Ln X
Y	Linear $Y = a + b X$	Linear-Log $Y = a + b \text{Ln } X$
Ln Y	Log-Linear $\text{Ln } Y = a + b X$	Log-Log $\text{Ln } Y = a + b \text{Ln } X$

In this study, two medium-sized cities (with metropolitan populations roughly 300,000) and two smaller cities (with metropolitan populations roughly 80,000) were selected to evaluate the models. Four models were developed for one city in each population group, using the different logarithmic transformations. The models were evaluated using statistical parameters to ensure the robustness of the models. To further evaluate the quality of the best model, a transferability test was performed to test the ability of the model to accurately predict traffic volumes for different communities of similar size. The results of the paper indicate that the Linear-Log

* Corresponding author:
md0033@uah.edu (Mehrnaz Doustmohammadi)
Published online at <http://journal.sapub.org/ijtte>
Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

transformation provides the best statistical validation to the case study city and the best transferability to the similar sized communities.

2. Data Collection

The purpose of this paper is to determine if there is an advantage to including a logarithmic transformation versus a using a basic linear model to predict average daily traffic from a collection of socio-economic variables. The data necessary to support the development of the models were collected and managed using ArcGIS. The database contained roadway characteristics (number of lanes and functional classification) as well as population (obtained from the Census Department), retail employment and non-retail employment for specific business locations obtained from an employment database purchased by the individual communities as part of the long range travel model update.

The traffic count values, dependent variables, were available from the Alabama Department of Transportation's roadway count program. The socio-economic data were collected using a 0.25 mile buffer around the count location. To quantify the roadway functional classification for inclusion in the linear models, the following convention was used:

Collector road = 1,
Minor arterial = 2, and
Principal arterial = 3.

This convention was selected as the highest functionally classified roadway would have the greatest value and this should have the desired effect that the parameter for this variable would be non-negative. Additionally, it should be noted that freeways and interstates were explicitly not included in the study or model development because the surrounding socio-economic variable would not be relevant for these roadways because they would not have access to the facilities due to the controlled nature of the roadway system.

The direct demand forecasting models were therefore developed using the following convention:

The response (dependent) variable is:

- Traffic volume; AADT.

The predictors (independent) variables are:

- Function classification of the road; FCLASS.
- Number of lanes; LANE.
- Population within a 0.25 mile buffer around a traffic count station; POPBUFF.
- Retail Employment within a 0.25 mile buffer around a traffic count station; RETAILEMPBUFF.
- Non-Retail Employment within a 0.25 mile buffer around the traffic count station; NONRETAILEMPBUFF.

3. Model Development

Linear regression and the logarithmic transformation were used in this study to produce the direct demand AADT forecasting models for the communities of Gadsden, AL and Montgomery, AL. All the statistical analyses were conducted using Minitab and in accordance with standard statistical methodologies [20]. Regression analysis was selected as this methodology is used to predict the value of one or more responses, AADT in this study, from a set of predictors, roadway and socio-economic variables. It can also be used to estimate the linear association between the predictors and responses. R^2 is the coefficient of multiple determinations and indicates the proportion of the variability in the observed responses that can be attributed to changes in the predictor variables [20]. For the communities studied in the work, the four models developed for Montgomery are shown in Table 2 and the four models developed for Gadsden are shown in Table 3.

Table 2. Montgomery Models and Statistics

Model	R ²
Linear	
ADT = -5625 + 8493 FCLASS + 219 LANE - 1.16 POPBUFF - 2.58 NONRETAILEMPBUFF + 11.55 RETAILEMPBUFF	79%
Log-Linear	
Ln(ADT) = 6.513 + 1.043 FCLASS + 0.142 LANE + 0.000037 POPBUFF - 0.000220 NONRETAILEMPBUFF + 0.000403 RETAILEMPBUFF	73.7%
Linear-Log	
ADT = -4520 + 8508 FCLASS + 419 LANE - 766 Ln(POPBUFF) - 95 Ln(NONRETAILEMPBUFF) + 807 Ln(RETAILEMPBUFF)	77.4%
Log-Log	
Ln(ADT) = 7.320 + 1.637 Ln(FCLASS) + 0.542 Ln(LANE) - 0.0288 Ln(POPBUFF) + 0.0041 Ln(NONRETAILEMPBUFF) + 0.0751 Ln(RETAILEMPBUFF)	74.2%

Table 3. Gadsden Models and Statistics

Model	R ²
Linear	
ADT = -12590 + 4478 FCLASS + 4644 LANE - 1.15 POPBUFF - 0.86 NONRETAILEMPBUFF + 7.91 RETAILEMPBUFF	79%
Log-Linear	
Ln(ADT) = 5.512 + 1.197 FCLASS + 0.189 LANE + 0.000449 POPBUFF + 0.000047 NONRETAILEMPBUFF + 0.000597 RETAILEMPBUFF	78.4%
Linear-Log	
ADT = -8827 + 4107 FCLASS + 4763 LANE - 807 Ln(POPBUFF) - 376 Ln(NONRETAILEMPBUFF) + 788 Ln(RETAILEMPBUFF)	92.1%
Log-Log	
Ln(ADT) = 6.253 + 1.656 Ln(FCLASS) + 0.925 Ln(LANE) - 0.0114 Ln(POPBUFF) + 0.0227 Ln(NONRETAILEMPBUFF) + 0.1225 Ln(RETAILEMPBUFF)	83.5%

It must be mentioned that although the models have statistical validity, the results of the models can result in unrealistic values if they are applied to situations outside the data used to develop the models. Additionally, it is possible

for the results to be negative, and these values should be taken as indication that the model is not valid for these roadways.

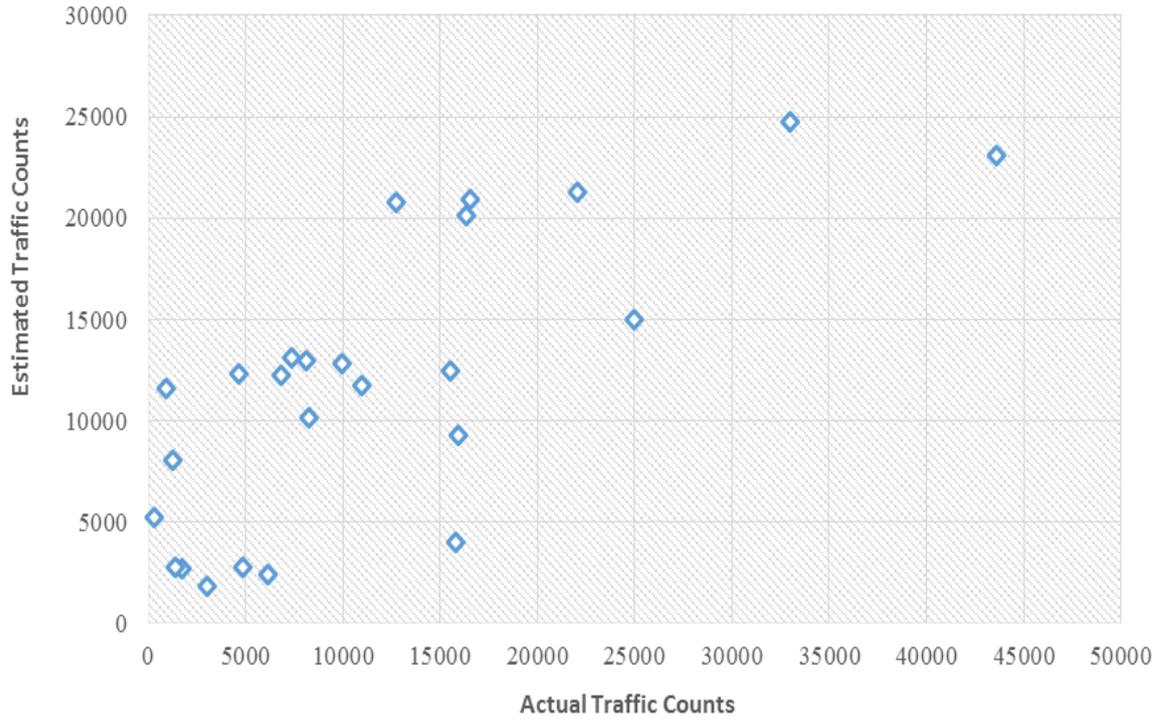


Figure 1. Comparison of Actual and Estimated AADT for Huntsville Using Montgomery’s Linear-Log Model

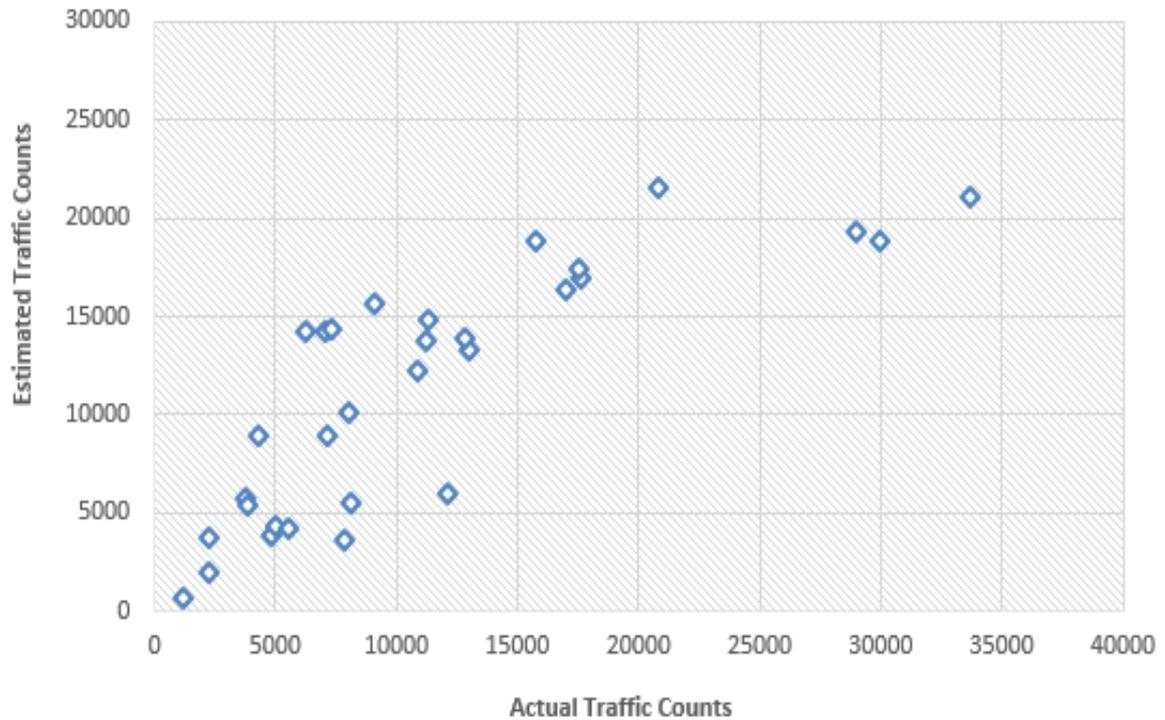


Figure 2. Comparison of Actual and Estimated AADT for The Shoals data Using Gadsden’s Linear-Log Model

4. Model Transferability

The transferability of the models to similar sized communities in the state was also tested to ensure the models would be applicable to other locations and that individual models would not have to be developed for every community. Huntsville, AL was used as the transferability community for the Montgomery models. A paired T-test was performed to test the model result of the forecasted traffic volume in Huntsville versus the actual traffic volumes for roadways in Huntsville. The P-values for the different models are shown in Table 4.

Table 4. Model Parameters for Montgomery Model Applied to Huntsville

Model Type	P-Value for t-test
Linear	0.838
Log-Linear	0.469
Linear-Log	0.965
Log-Log	0.715

The combined cities of Florence, Muscle Shoals, Sheffield and Tuscumbia, known as the Shoals, were used at the transferability test community for the Gadsden model. A paired T-test was performed to test the model result of the forecasted traffic volume in The Shoals versus the actual traffic volumes for roadways in The Shoals. The P-values for the different models are shown in Table 5.

Table 5. Model Parameters for Gadsden Model Applied to the Shoals

Model Type	P-Value for t-test
Linear	0.899
Log-Linear	0.731
Linear-Log	0.8
Log-Log	0.053

The higher P-values indicate the better the model transfers to the other community. Based on the R-square value for the model developed and the P-value for the T-test, for both communities, the Linear-Log model was selected as the best transformation and created the best model. Figure 1 and 2 show the scatterplots of the linear-log models when applied to the transferability city.

5. Conclusions

This paper examined the development and testing of a linear model and three logarithmic transformations techniques as models to estimate traffic volumes from a collection of roadway data and socio-economic factors near a location where an estimate of traffic volume is desired. The data selected for model input included the total number of lanes on the roadway, roadway functional classification, population, retail employment and non-retail employment within a 0.25 mile buffer of the count location. The logarithmic transformations were defined as log-linear, linear-log and log-log. The different models developed in

this work, four for a medium sized area (population roughly 300,000) and four for the smaller area (population roughly 80,000), were tested statistically.

The results of this work indicate that the linear-log model ($\ln Y = a + b X$) had the best combination of statistics for the community developed and the best transferability to the similar community. The use of this model is the best for estimating traffic volumes using the data presented in this work. The transferability scatter plots indicate that the model for the smaller communities had better results. Whether this was from the size of the community or the similarity of the community is unknown. Also, the models results tended to be better at lower volume roadways, which is not necessarily a poor result as the lower the volume the roadway the less likely it is to have a count or to be included in a community count program.

Overall, the ability of the model represents a significant time and cost savings as a means to estimate traffic volume that would be suitable for safety analysis, maintenance scheduling, capacity analysis, as well as other transportation system analyses.

ACKNOWLEDGEMENTS

This paper was made possible by funding provided by the Alabama Department of Transportation, Project Number SPR-0001(058); CPMS Number 1000662945.

REFERENCES

- [1] Brimley, Bradford, Mitsuru Saito, and Grant Schultz. "Calibration of Highway Safety Manual safety performance function: development of new models for rural two-lane two-way highways." *Transportation Research Record: Journal of the Transportation Research Board* 2279 (2012): 82-89.
- [2] Doustmohammadi, Mehrnaz, Michael Anderson, and Ehsan Doustmohammadi. "Examining the Effect of Inaccurate Traffic Impact Analysis on Roadway Infrastructure." *International Journal of Traffic and Transportation Engineering* 4.4 (2015): 103-106.
- [3] Anderson, Michael, Khalid Sharfi, and Sampson Gholston. "Direct demand forecasting model for small urban communities using multiple linear regression." *Transportation Research Record: Journal of the Transportation Research Board* 1981 (2006): 114-117.
- [4] Gecchele, G., Rossi, R., Gastaldi, M., and Caprini, A. (2011). "Data Mining Methods for Traffic Monitoring Data Analysis: A case study." *Procedia - Social and Behavioral Sciences*, 10.1016/j.sbspro.2011.08.052, 455-464.
- [5] Garber, Nicholas J. "A Methodology for Estimating AADT Volumes from Short-Duration Counts." *Virginia Highway & Transportation Research Council*. 1984.
- [6] Zhong, M. and Liu, G. (2007). "Establishing and Managing Jurisdiction-wide Traffic Monitoring Systems: North

- American Experiences." *Journal of Transportation Systems Engineering and Information Technology*, 10.1016/S1570-6672(08)60002-1, 25-38.
- [7] Dadang Mohamad, Kumares C. Sinha, Thomas Kuczek, Charles F. Scholer. "Annual Average Daily Traffic Prediction Model for County Roads, Transportation Research Board 2013 Annual Meeting. 252 1998.
- [8] Sharma, Satish C., Brij M. Gulati, and Samantha N. Rizak. "Statewide traffic volume studies and precision of AADT estimates." *Journal of Transportation Engineering* 122.6 (1996): 430-439. "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [9] Doustmohammadi, Mehrnaz, Michael Anderson and James Swain. "Evaluation of Trip Generation at a Free Standing Discount Superstore" Accepted for Publication in the *International Journal for Traffic and Transportation Engineering*.
- [10] Tao Pan. "Assignment of estimated average annual daily traffic on all roads in Florida" University of South Florida, 2008.
- [11] Wang, Tao; Gan, Albert; Alluri, Priyanka. "Estimating Annual Average Daily Traffic (AADT) for Local Roads for Highway Safety Analysis." Transportation Research Board 2013 Annual Meeting. 267 2012.
- [12] Doustmohammadi, Mehrnaz, and Michael Anderson. "Developing Direct Demand AADT Forecasting Models for Small and Medium Sized Urban Communities." *International Journal of Traffic and Transportation Engineering* 5.2 (2016): 27-31.
- [13] Zhao, Fang and Soon Chung; "Contributing Factors of Annual Average Daily Traffic in a Florida County", Transportation Research Board, 2001.
- [14] Sharma, S., Lingras, P., Xu, F., and Kilburn, P. (2001). "Application of Neural Networks to Estimate AADT on Low-Volume Roads." *Journal of Transportation Engineering*, 10.1061/(ASCE)0733-947X(2001)127:5(426), 426-432.
- [15] Zhao, Fang, and Nokil Park. "Using geographically weighted regression models to estimate annual average daily traffic." *Transportation Research Record: Journal of the Transportation Research Board* 1879 (2004): 99-107.
- [16] <http://www.kenbenoit.net/courses/ME104/logmodels2.pdf>. Access April 12, 2016.
- [17] Wier, Megan, et al. "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning." *Accident Analysis & Prevention* 41.1 (2009): 137-145.
- [18] Koppelman, Frank S. "Non-linear utility functions in models of travel choice behavior." *Transportation* 10.2 (1981): 127-146.
- [19] Vovsha, Peter, Eric Petersen, and Robert Donnelly. "Explicit modeling of joint travel by household members: statistical evidence and applied approach." *Transportation Research Record: Journal of the Transportation Research Board* 1831 (2003): 1-10.
- [20] Montgomery, Douglas, Elizabeth A. Peck, G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. Wiley ISBN-1119180171. 2012.