Journal of Traffic and Transportation Engineering 3 (2015) 255-265
doi: 10.17265/2328-2142/2015.05.001

Engaging Nash-Sutcliffe Efficiency and Model Efficiency Factor Indicators in Selecting and Validating Effective Light Rail System Operation and Maintenance Cost Models

Xiaohui Zhong[1] and Utpal Dutta[2]

1. Department of Mathematics and Computer Science, University of Detroit Mercy, Detroit MI 48221, USA

2. Department of Civil, Architectural and Environmental Engineering, University of Detroit Mercy, Detroit MI 48221, USA

Abstract: A set of LR (light rail) transit O&M (operation and maintenance) cost models were derived as functions of system attributes, namely Directional-Route-Miles, Train-Revenue-Hours, Train-Revenue-Miles, Peak Passenger Cars and Annual Passenger Trips. These models were developed based on the data of 12 LR facilities in the United States from 2004~2008 reported on the National Transit Database. The models were then validated with newer data of the same systems from 2009~2011. Instead of using the traditional R-squared, NSE (Nash-Sutcliffe efficiency) was employed for model selection. Furthermore, *MEF* (model efficiency factor) indicators were used to validate model's performance. When compared to the actual data, the prediction ability of these new models was superior to those of previously developed PFM (power factor models).

Key words: Light rail, operation and maintenance cost, cost models, model efficiency factor, Zhong-Dutta model.

1. Introduction

The projection of O&M (operating and maintenance) costs is an important component of the planning of any new LR (light rail) transit system. According to FTA (Federal Transit Administration), determination of O&M costs of any planned LR transit system is significant for two reasons [1]:

- The design year O&M costs estimate is a critical factor while determining cost effectiveness of a LR transit system;
- The projections of yearly planned O&M costs are crucial to the development of financial plans that cover the service lifetime process of construction and operation.

According to the NTD (National Transit Database), an LR transit system's total O&M costs has four components: (1) vehicle operation costs; (2) vehicle maintenance costs; (3) non-vehicle maintenance costs; (4) general administrative costs. Vehicle operation costs represent over 40% of the total O&M costs, followed by vehicle maintenance costs at 23% and general administrative and non-vehicle maintenance costs at close to 17% each [2].

At the Transportation Research Board 2013 Annual Meeting, the authors presented a set of LR transit cost models predicting total O&M costs and its four component costs based on some system attributes using 2004~2008 NTD data [3]. These models are called PFM (power factor model) and are presented in Table 1.

2. Definition of Light-Rail Systems, System Attributes and System Costs

The definitions of independent/dependent variables included in the PFM models, as well as models

Corresponding author: Xiaohui Zhong, Ph.D., professor, research fields: applied statistics and mathematics. E-mail: Zhongk@udmercy.edu.

developed as a part of this study, are presented in this section. It is to be noted that no attempts were made to convert units of independent variables into metric system, since model is based on the data reported in NTD, representing system attributes of the transit systems in the US cities.

All variables were reported in annual basis. More detailed explanations can be found in the author's previous paper [3] and NTD glossary [4]:

(1) LR (light rail): a transit mode that typically is an electric railway with a light volume traffic capacity compared to HR (heavy rail);

(2) Peak Passenger Cars in operation/vehicles operated in annual maximum service (A): the number of LRV (LR unit vehicles) required to providing peak headways including spare vehicles;

(3) Annual Train Revenue Hours (B): the hours that trains are scheduled to or actually travel while in revenue service (actual train revenue hours) plus deadhead hours in one year. Actual train hours include layover/recovery time but exclude hours for charter services, operator training and vehicle maintenance testing;

(4) Annual Train Revenue Miles (C): the miles that trains are scheduled to or actually travel while in revenue service in one year. Train revenue miles exclude:

- deadhead;
- training operators prior to revenue service;
- vehicle maintenance tests;
- charter services;

(5) Directional-Route-Miles (D): the mileage in each direction over which LR transit travels while in revenue service;

(6) Annual Passenger Trips (G): the number of passengers who board operational revenue vehicles. Passengers are counted each time they board vehicles, no matter how many vehicles they use to travel from their origin to their destination;

(7) O&M costs: the operation and maintenance costs which can be incurred or categorized to one of the following areas:

- vehicle operations: all activities associated with vehicle operations including but not limited to: (1) revenue vehicle movement control; (2) scheduling of transportation operations;
- vehicle maintenance: all activities associated with revenue and non-revenue (service) vehicle maintenance, including: (1) inspection and maintenance; (2) repair due to accident and vandalism; (3) servicing (cleaning, fueling, etc.) vehicles;
- non-vehicle maintenance: all activities associated with facility maintenance, including but not limited to: (1) operation of electric power facilities; (2) structures, tunnels and subways; (3) passenger stations, operating station buildings, grounds and equipment; (4) communication systems;
- general administration: all activities associated with the general administration of the transit agency, including but not limited to: (1) transit service development, data processing and planning; (2) injuries, damages, safety, insurance and legal service; (3) finance and accounting, real estate management and marketing; (4) office management and services, personnel administration.

In this study, annual costs reported by NTD under vehicle operation, vehicle maintenance, non-vehicle maintenance and general administration are called the vehicle operation costs, vehicle maintenance costs, non-vehicle maintenance costs and general administration costs, respectively. In our models, they were designated as Y_1, Y_2, Y_3 and Y_4. The summary of all four costs of the system is called the total O&M costs, denoted as Y. The relation between all these costs is simply $Y = Y_1 + Y_2 + Y_3 + Y_4$.

The PFM models cited before were developed by using related cost data and system attributes of 12 systems from 2004 to 2008. Those data were collected from the FTA managed NTD database. These 12 LR transit systems coded as LR by the NTD existed since 2004 thus provided maximum data source. The 12 LR systems selected are:

- Baltimore: MTA (Maryland Transit Administration);
- Dallas: DART (Dallas Area Rapid Transit);
- Denver : RTD (Regional Transportation District);
- Houston: METRO (Metropolitan Transit Authority of Harris County);
- Minneapolis: Metro transit (METRO);
- New Orleans: NORTA (New Orleans Regional Transit Authority);
- Portland: TriMet (Tri-County Metropolitan District of Portland);
- Philadelphia: SEPTA (Southeastern Pennsylvania Transportation Authority);
- Sacramento: RT (Regional Transportation District);
- Salt Lake City: UTA (Utah Transit Authority);
- San Jose: Santa Clara VTA (Valley Transportation Authority);
- St. Louis: Bi-State Development Agency (METRO).

As we reviewed our previously-derived PFM models, as well as their output, we observed the following:

- Directional-Route-Mile (*D*) variable is a significant factor driving all costs determined by stepwise regression process;
- Deviation between predicted costs versus actual costs are in millions of dollars;
- R^2 values were computed based on transformed data (such as log of independent variables) and not actual data (more about R^2 can be found in Section 4).

In light of above mentioned fact, the authors attempted to derive a new set of models (Zhong-Dutta models) with an objective to minimize deviation between predicted and actual cost. Since the variable Directional-Route-Mile (*D*) is common to all models and the behaviors of the costs vary significantly by this factor, it is decided that cost and system data will be divided into two sets based on *D* upto 40 miles and *D* greater than 40 miles. Relative uniformity of cost values were observed among systems with *D* greater than 40 miles. Also annual ridership data were included as a part of this study. A close review of NDT data indicated that total O&M costs generally increase with the increase in ridership. Data presented in Table 2 also showed this general trend.

The objective of this study is to develop a set of more efficient models addressing above mentioned issues using 2004~2008 data and then validate the models with 2009~2011 system performance and cost data. In this context, attempts were made to develop:

- better cost models than previously-derived PFM models by including the influence of ridership in cost;
- two distinct sets of models considering Directional-Route-Miles (*D*) greater than 40 miles and upto 40 miles;
- better performance indicator, such as *NSE* (Nash-Sutcliffe efficiency) factor, to evaluate the performance of the newly-derived models;
- the *MEF* (model efficiency factor) for model validation by using new sets of data from 2009 to 2011.

3. Model Development

A set of non-linear models (Zhong-Dutta models) were developed for total O&M costs and its components by using stepwise regression analysis**.** These models are black box systems using stepwise regression techniques to link the inputs: Directional-Route-Miles (*D*), Train-Revenue-Hours (*B*), Train-Revenue-Miles (*C*), Peak Passenger Cars in Operation (*A*) and Annual Passenger Trips (*G*), to model the costs of the LR system in the United States. These models represent some general aspects of a cost's response without going deeply into the real physical processes of operations of LR transit systems, and physical units of the attributes were not considered. The stepwise regression identifies the dominance of system variables on total O&M costs and its components.

The derived models were based on the reported data elements of various LR systems in the NTD database from 2004 to 2008. The NTD data were self-reported by the transit systems on annual basis. Even though the transit agencies carefully follow the specification requirement in reporting, there are still chances for misclassification or omission of information due to interpretation of each specification, in spite of FTA's strict quality assurance review process. The derived models will have the influence of reporting errors, as well as other error. The objective of this effort is to have the predicted costs to be as close as possible to the actual values as a system of whole.

4. Model Selection Criteria

Traditionally, when a set of data (x_{1i}, x_{2i},..., x_{mi}, y_i, i = 1, 2,..., n) is available, a mathematical model $\hat{y} = f(x_1, x_2,..., x_m)$ can be created to predict the values of y based on the values of x_j, j = 1, 2,..., m, where y is respondent variable and values of x_j (j = 1, 2,..., m) are the independent variables. The common method to evaluate such model is to calculate the coefficient of determinant, the R-squared. The detailed definition of R-squared is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ is the arithmetic mean of the observed values of y_i.

When the model is linear, it measures the "goodness" of fit of the data to the line and this value is always between 0 and 1. It also measures relative magnitude of the residual variance compared to the measured data variance [4]. The closer the value is to 1, the better the data "fit" to the line. It is a common practice to seek for an R-squared value of at least 0.5 and as high as possible.

However, if the data are transformed to obtain a pseudo-linear model, the R-squared is calculated based on the transformed data and it measures the fit of predicated value on the transformed data. Because of the distortion of the original data, the sense of direct measurement of deviation between the predicted and actual value is lost. For example, a power model $y = ax^b$ can be obtained by the linear regression on logarithmic transformation of x_i, y_i. The R-squared is calculated by:

$$R^2 = 1 - \frac{\sum_i(\log y_i - \log\hat{y}_i)^2}{\sum_i(\log y_i - \overline{\log y})^2} \tag{2}$$

A high value (close to 1) of this R-squared indicates that the predicated values of $\log \hat{y}_i$ is a good fit to the values of $\log y_i$, but this "goodness" is not necessarily translated back to the "closeness" of the values of $\hat{y}_i$ to the values of y_i. Sometimes, it may result in a very bad prediction. In case like this, the R-squared read directly from the output of software (such as EXCEL, SAS, SPSS, etc.) may give a false impression of an efficient model.

Once a nonlinear transform is made to the data, the influences of the data values will change, as well as the error structure of the model and the interpretation of any inferential results. These may not be desired effects of the application. Therefore, use of a nonlinear transformation requires extra caution. In addition, depending on what the largest source of error is, a nonlinear transformation may distribute the errors in a normal fashion, so the choice to perform a nonlinear transformation must be informed by modeling considerations.

As a result, R-squared is not the best indicators to evaluate the trans-linear models [5, 6]. If the model is trans-linear, there are many ways to determine the efficiency of models. One straightforward way to select the model is to use the *RMSE* (root-mean-square error). *RMSE* represents the differences between values predicted by a model or an estimator and the values actually observed [7, 8]. The formula for *RMSE* is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

One drawback in using *RMSE* is that it can be very large in magnitude depending on the units of the predicted values. It is also difficult to compare models for different systems or in different formats.

To overcome the shortfall of *RMSE* and the authors decided to use the *NSE* factor introduced by Nash and Sutcliffe [9] to determine the efficiency of the models:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

This formula is the same as the *R*-squared of the linear regression but applied directly to the original data on any model. For the non-linear models, *NSE* can be negative. It actually ranges from $-\infty$ to 1. As usual, investigators seek for an *NSE* value close to 1. A negative *NSE* indicates an unacceptable model performance. This is one of the recommended indicators for model efficiency in the field of hydrology, as well as many other applications [6, 10-13]. Similar to *RMSE*, *NSE* alone is not an adequate indicator [14]. Together with *RMSE*, they form a set of model selection criteria which offsets the limitation of each other. Moreover, *MEF* is introduced later in model validation section to give a full interpretation of *NSE* from different prospective.

Using the technique of stepwise regression, a number of non-linear models were developed for each component of the costs and the total O&M costs for systems with *D* upto 40 or greater than 40 miles, respectively. It is to be noted that step-wise regression has identified Directional-Route-Miles (*D*) as the most significant independent system variable among the four variables considered (*P*-values less than 0.05).

Based on the lowest values *RMSE* and highest values of *NSE*, a set of models were finally adapted. These models and previously developed PFM are presented in Table 1. The *RMSE* and *NSE* values of these models were computed using Eqs. (3) and (4), and are presented in Table 3.

5. Characteristics of Zhong-Dutta Models

A review of Tables 1 and 2 reveals the following:

- Directional-Route-Miles (*D*) appears in all models irrespective of the format of the models and the size of the systems. This is consistent with the finding of the previous study [3] that *D* was a dominant factor driving all costs (reflected by its large coefficients). This can be interpreted that the costs of operating and maintaining a LR transit systems are heavily impacted by the infrastructure, that is, the length of the routes;
- Peak Passenger Cars in Operation (*A*) is the next important factor affecting the costs (also has relatively large coefficients). This variable appears in all models except for the general administration costs for the systems with *D* upto 40 miles. The distinct behavior of the general administrative cost will be explained in more detail in next section;
- The newly introduced variable Annual Passenger Trips (*G*) affects the models in similar manner in both systems. First of all, for both systems, the ridership does not significantly affect the vehicle operation costs. Such result is consistent with findings of other research [15]. Second, this variable affects the total O&M costs positively with interaction with *D* in both systems. Finally, even though the effects are mixed, this variable appears in all other models except for the non-vehicle maintenance costs in the systems with *D* greater than 40 miles;
- The variables *B* (Train-Revenue-Hours) and *C* (Train-Revenue-Miles) generally indicate the effectiveness of the transit system [15] their influence on the non-vehicle-maintenance costs is least in either system according to the Zhong-Dutta models. This result is very different from the PFM models where *C* drives up this part of costs and *B* did not show significant effects;
- As stated before, selection of each model was done independently based on *RMSE* and *NSE* values, irrespective of its connection to total cost, since prediction efficiencies of each model also varies. For example, Y_1 (*D* greater than 40 miles) has higher prediction efficiency than the *Y* model. Furthermore, the models are non-linear, mathematically Y_1, Y_2, Y_3 and Y_4 should not add up to *Y*;

Table 1 Zhong-Dutta models (new)and PFM models (previously developed).

	Zhong-Dutta models		PFM models
	Systems with D greater than 40 miles	Systems with D upto 40 miles	
Vehicle operation costs	$Y_1 = 256,939.367D + 1,574.809AD - 0.115BD - 0.014CD$	$Y_1 = 45,067,989.0A^{0.459}C^{0.18}D^{0.417}$	$Y_1 = 63,831.4D^{0.872}B^{0.2}$
Vehicle maintenance costs	$Y_2 = 530,000e^{0.055}A^{0.238}B^{0.177}C^{-0.278}D^{1.674}G^{-0.225} + 41,640.935D + 217.849AD - 0.00470BD + 469.303D^2$	$Y_2 = 126,077.194A - 3,000.690A^2 - 0960AB + 4,489.002AD + 0.005AG$	$Y_2 = 23,155.8D^{0.844}B^{0.211}$
Non-vehicle maintenance costs	$Y_3 = 70,027.710A - 995.965A^2 + 1,663.730AD$	$Y_3 = 226,973.058A - 3,353.878A^2 - 3,532.577AD + 0.012AG$	$Y_3 = 11,579.6A^{-0.530}C^{0.345}D^{0.831}$
General administration costs	$Y_4 = -0.044G + 7,276.135D + 0.008AG + 479.941AD - 0.203BD - 0.004BC - 3.114 \times 10^{-6}BG - 4.251 \times 10^{-8}CG + 0.010\,DG$	$Y_3 = 12.895G - 0.285G\ln(B) + 0.171G\ln(C) + 0.287G\ln(D) - 0.772G\ln(G)$	$Y_4 = 63,767.6D^{0.592}A^{0.556}$
Total O&M costs	$Y = 387,371.9D + 3,916.145AD - 1.209BD - 0.039CD + 3,866.576D^2 + 0.006DG$	$Y = 34,655.637A + 738,171.830D - 4,322.511A^2 - 2,075.244AD + 0.058AG + 0.052CD - 15,766.996\,D^2 + 0.011DG$	$Y = 223,462.7C^{0.144}D^{0.744}$

A: Peak Passenger Cars;
B: Train-Revenue-Hours;
C: Train-Revenue-Miles;
D: Directional-Rout-Miles;
G: Annual Passenger Trip.

Table 2 The 2009~2011 data for systems with D upto 40 miles.

City/institution	A	B	C	D	G	Y_1 (\$)	Y_2 (\$)	Y_3 (\$)	Y_4 (\$)	Y (\$)
Houston (METRO) 2009	17	64,652	903,668	14.8	11,613,720	6,881,434	3,458,215	4,840,587	590,723	15,770,959
Houston (METRO) 2010	17	65,372	900,517	14.8	10,616,292	6,681,180	3,488,835	4,582,020	65,113	14,817,148
Houston (METRO) 2011	17	63,481	901,218	14.8	10,618,061	7,229,515	4,227,349	5,260,492	785,315	17,502,671
New Orleans (NORTA) 2009	23	105,536	816,890	33.5	5,342,126	4,240,617	1,819,293	1,065,617	12,420,393	19,545,920
New Orleans (NORTA) 2010	21	127,435	947,790	25.3	6,784,713	723,731	618,339	27,472	22,487,284	23,856,826
New Orleans (NORTA) 2011	22	127,184	923,681	25.3	8,984,813	7,963,284	4,631,866	534,067	11,874,857	25,004,074
Minneapolis (METRO) 2009	27	68,687	1,955,070	24.7	9,863,042	6,938,709	3,344,475	4,620,750	10,098,490	25,002,424
Minneapolis (METRO) 2010	27	72,216	2,013,961	24.7	10,455,860	7,730,722	4,182,243	5,224,328	8,598,830	25,736,123
Minneapolis (METRO) 2011	27	67,681	2,054,607	24.7	10,400,864	7,877,680	3,891,550	4,305,583	9,641,007	25,715,820

Y: total O&M costs;

Y_1: vehicle operation costs;

Y_2: vehicle maintenance costs;

Y_3: non-vehicle maintenance costs;

Y_4: general administrative costs.

Table 3 ***RMSE*** **(root mean square error) and** ***NSE*** **(Nash-Sutcliffe efficiency) values of Zhong-Dutta models and PFM models.**

Costs	Systems with D greater than 40 miles					Systems with D upto 40 miles				
	Zhong-Dutta models		PFM models		Decrease in	Zhong-Dutta models		PFM models		Decrease in
	RMSE	*NSE*	*RMSE*	*NSE*	*RMSE* (%)	*RMSE*	*NSE*	*RMSE*	*NSE*	*RMSE* (%)
Vehicle operation costs	2,439,333	0.7516	3,469,437	0.4975	29.7	772,518	0.7830	2,221,552	−0.794	65.2
Vehicle maintenance costs	2,462,664	0.5536	2,848,807	0.4027	13.6	694,329	0.8407	979,848	0.6828	29.1
Non-vehicle maintenance costs	2,513,329	0.6276	3,441,896	0.3016	27.0	945,540	0.6661	1,599,661	0.0443	40.9
General administration costs	2,067,007	0.8519	4,486,117	0.3039	53.9	1,982,160	0.3205	2,635,015	−0.2008	24.8
Total O&M costs	6,715,436	0.7916	10,201,002	0.5191	34.2	1,938,491	0.8418	4,354,074	0.2018	55.5

RMSE is in \$.

Table 4 ***MEF*** **factors calculated for new models and old PFM models.**

Models	System with D greater than 40 miles		System with D upto 40 miles	
	Zhong-Dutta	Old PFM	Zhong-Dutta	Old PFM
Y	0.4963	0.7371	1.1258	1.2306
Y_1	0.6014	0.6533	0.8441	1.5523
Y_2	0.7080	0.7360	1.1125	1.5302
Y_3	0.6829	0.8595	0.7883	1.1752
Y_4	0.7057	0.9518	1.2080	1.3178

Table 5 Average distribution of cost components.

Systems	Percent of total O & M cost (%)			
	Y_1	Y_2	Y_3	Y_4
D greater than 40 miles	40	20	20	20
D upto 40 miles	31	16	17	36
Nationwide	42	23	17	18

- For systems with D greater than 40 miles, B has the adverse effect on all costs except the vehicle maintenance costs. Usually in a closed system with fixed Directional-Route-Miles, the slight increase of the level of Train-Revenue-Hours will cause the similar increase of other system variables. The negative terms of B will have the effect of balancing the increase of these variables on the costs such that the costs are kept at the same level. For the vehicle maintenance costs, it can be explained that the effects are offset by positive and negative terms involving with B. Similar effects can be found for C. Such findings indicate that the systems with D greater than 40 miles are cost-effective in terms of productivities, agreeing with the findings of other studies [16, 17];
- For systems with D upto 40 miles, B has no significant effect on the vehicle operation costs, the total O&M costs, as well as the non-vehicle maintenance costs. But this variable does lower the vehicle maintenance costs and general administration costs. The systems with higher B are more cost-effective. But the effect of C is on the opposite direction; It drives up the vehicle operation costs, general administration costs and the total O&M costs, while not affecting the other two costs;
- For the efficiency of the Zhong-Dutta models, they all achieve the better accuracy than the PFM models measured in terms of *RMSE* and *NSE*. For the systems with D greater than 40 miles, the decreases in *RMSE* range from 13.6% to 53.9%, with the greatest improvement in general administration costs. For system with D upto 40 miles, the decreases in *RMSE* are even bigger, ranging from 24.8% to 65.2%, showing greater improvement. These smaller *RMSEs* mean that the overall prediction of the costs will be a lot more closer to the actual ones than those predicted by the PFM models which is one of the objective of this study;
- While comparing the *NSE* values between the Zhong-Dutta models and PFM models of non-vehicle maintenance cost, it is observed that the newer models are able to provide better predication. PFM models did not achieve the desired accuracy.

Overall, the newly-developed models are superior to the originally developed PFM models. Meanwhile, they also inherit the useful and significant information and findings from the previous study. To reaffirm these findings, we will use the newer information from NTD to validate these new models.

6. Model Validation

One approach of validating the models is to evaluate the models with newer information or data that was not used for developing the models. Once models were adopted based on highest value of *NSE*, the validity of models were evaluated by using new set of cost and system data of facilities from 2009~2011.

One can use *RMSE* to measure the accuracy. As mentioned above, the *RMSE* is the average magnitudes of the errors in predictions for overall observations in a single measure of predictive power. However, even *RMSE* is a good measure of accuracy, it is scale-dependent. This indicator itself can be very large in value and it is hard to determine its magnitude without a reference point. For the purpose of this study, we derived a factor named *MEF* by normalizing *RMSE* to reflect the performance indicator of the model:

$$MEF = \frac{RMSE}{STDEV} = \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

where:

$$STDEV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} \tag{6}$$

The validation of models aims for smaller values of *MEF*. While this value ranges from 0 to ∞, the lower the *MEF*, the better the models perform. The ultimate state is that $MEF = 0$ indicating an error-free model which is not realistic.

It is noted that:

$$MEF = \sqrt{1 - NSE} \quad (7)$$

This relationship gives a better interpretation of *NSE* which was addressed by McCuen [18].

The *MEF* factors calculated for new models, as well as old PFM models are presented in Table 4.

7. Discussion on Model Validation

Following summarizes the findings of model validation based on *MEF* values (Table 4) of the models:

- When comparing *MEF* factors of Zhong-Dutta models and PFM models, it is found that the *MEF* values are smaller than those of PFM models for both systems with *D* greater than 40 miles and *D* upto 40 miles, representing robustness of Zhong-Dutta models;
- When comparing *MEF* factors for models between systems with *D* greater than 40 miles and the systems with *D* upto 40 miles, the *MEF* values of systems with *D* greater than 40 miles are always smaller, reflecting models perform better in these systems;
- For systems with *D* greater than 40 miles, *MEF* of all models always showed values smaller than 1, indicating that all models including Zhong-Dutta and PMF are capable of predicting the values at the acceptable levels. But the Zhong-Dutta models consistently have better quality than the old models.

Even though Zhong-Dutta models for systems with *D* upto 40 miles shown improvement over the PFM in terms of MFE values, these values are relatively high with some values slightly greater than 1. These relatively high MFE values may indicate that the predictions with new models for the systems with *D* upto 40 miles did not reach the satisfactory level as desired.

To investigate the causes behind the relatively higher *MEF* values of all models for systems with *D* upto 40 miles, the authors analyzed the 2009~2011 data presented in Table 2 for all three facilities remained in this system. These facilities are Houston (*D* = 14.8 miles), New Orleans (*D* = 33.5 miles in 2009 and *D* = 25.3 miles in 2010~2011), and Minneapolis (*D* = 24.7 miles). Furthermore, average percent distribution of cost components is presented in Table 5.

The analysis of data and statistics in Tables 2 and 5 reveals the following:

- In comparing the systems of Houston and New Orleans, it is found that New Orleans is a bigger system in terms of Peak Passenger Cars (*A*), Train-Revenue-Hours (*B*) and Directional-Route-Miles (*D*). Though the Train-Revenue-Miles (*C*) are at the similar level, but the ridership (Annual-Passenger-Trips (*G*)) of New Orleans is only half of that of the Houston. The values of all these independent system variables are consistent with those reported in 2004~2008. But the average total O&M costs (*Y*) of New Orleans in 2009~2011 is about 7 million more than that of Houston, which is 50% of the average total O&M costs of Houston. Such huge difference in total O&M costs was not shown in the 2004~2009 data. With these data, Zhong-Dutta models gave a very close prediction in total O&M costs (*Y*) for Houston but underestimated the total O&M costs of New Orleans by an average of $6,000,000 because this cost was driven down by the lower ridership in the model;
- In case of Houston and Minneapolis systems, it can be seen that Minneapolis is bigger in terms of Peak Passenger Cars (*A*), Train-Revenue-Miles (*C*) and Directional-Route-Miles (*D*) with similar levels of Train-Revenue-Hours (*B*) and Annual Passenger Trips (*G*). The total O&M costs of the system of Minneapolis were the highest one among all systems in years 2004~2008. In compensating the estimate of costs of other systems, the models did not predict this system well to begin with. So the model inherited the bias estimate costs for this system;
- Comparison of the New Orleans and Minneapolis shows that these two systems are very similar, except

that the proportions of Train-Revenue-Hours (B) and Train-Revenue-Miles (C) are reciprocal. When estimating Y, B is not a factor and C is a significant factor, the difference in values of C making the total costs of these two systems off the target;

- The cost component distributions are totally different from system to system. Houston's general administrative costs were unrealistically low while New Orleans costs were very high.

None of the systems had cost components in line with the general rule: Y_1 = 42%, Y_2 = 23%, Y_3 = 17%, Y_4 = 18% (Table 5);

- Total O&M costs of two of the three systems increased dramatically even though system attributes were unchanged. For example, for New Orleans, the total O&M costs changed from 12~14 million in 2004~2008 to 20~25 million in 2009~2011. That was an increase of 60% (average) over three years. Minneapolis experienced a slightly lower increase and Houston's cost remained at the similar level.

The authors contacted NTD program manager by email regarding the inconsistencies of this set of data. Following responses were provided:

- Allocation of cost distribution is not native to their accounting system for some facilities and is often done on a proportional basis (which can look a bit arbitrary);
- Also, "you happen to have chosen two of our more challenging reporters (transit systems)".

These inconsistencies are reflected by higher *MEF* values in systems with D upto 40 miles.

8. Conclusions

In this study, annual data of total O&M cost and its components of the 12 existing LR transit systems of US cities from 2004~2008 were analyzed to derive a relationship with a number of system attributes namely: Directional-Route-Miles, Train-Revenue-Hours, Train-Revenue-Miles, Peak Passenger Cars, and Annual Passenger Trips. Particularly in this effort, data were divided into two sets based on Directional-Route-Miles greater than and upto 40 miles. In our previous study, Directional-Route-Miles was identified as the most dominant factor. While examining data elements of various facilities, it was observed that uniformity of data exists in the systems with Directional-Route-Miles greater than 40 miles. Based on this observation, it was decided to split the data into two sets in terms of the size of Directional-Route-Miles. Such split resulted in the development of five cost models for each data set respectively. Furthermore, the models were validated by using 2009~2011 system performance and cost data.

Using the technique of stepwise regression, a number of non-linear models were developed for each component of the costs and the total O&M costs for systems with D upto or greater than 40 miles, respectively. Based on the lowest values *RMSE* and highest values of *NSE*, a set of models were finally selected. The selected Zhong-Dutta models were also compared to the previously developed PFM models by using *RMSE* and *NSE* factor. The predictabilities of the new models are superior. Such superiority was also validated by examining the *MEF* factor using new data. The use of *NSE* and *RMSE* in model selection, instead of the traditional R-squared on the transformed data, is recommended for model formation, especially for non-linear models. Inconsistency on data reporting was observed on smaller LR facilities. More reliable models can be produced with consistent data. It is to be noted that no attempts were made to exclude data reporting error while developing models. However, Zhong-Dutta models have the ability to detect data inconsistency (higher *MEF* values). Furthermore, it is recommended that more attention should be placed while reporting operating data for facilities with smaller Directional-Route-Miles.

These newly-developed Zhong-Dutta models should produce better and more accurate projection of total O&M costs of LR transit systems for the US cities. In the absence of a site specific model, these models should be very useful for any agency in the future operation of existing LR systems and the planning phase of any city that is in the development of the system as a guideline. While the models are limited to the cost study for the LR transit systems in the US only, the study method, as well as model selection indicators, can be applied to study the O&M costs of the public transit systems in general. As new systems are being developed and NTD includes more LR transit systems reporting in the database, further researches can be done along this line to include some implicit influential aspects, such as geographical/demographical information, history of the systems and human factors. Also, the developed models can be further refined by including variables, such as PMT (passenger-miles traveled), track-miles, system route-miles and others which are mostly available in NTD. It is our goal to gauge the usefulness of the cost analysis in a wide range of public transit systems under various operating conditions in order to determine the factors which are likely to be more effective or less effective. The ultimate goal is to pinpoint some specific practices that may facilitate the efficacy of the cost management of public transit systems.

References

[1] FTA (Federal Transit Administration). 2010. *Procedures and Technical Methods for Transit Project Planning.* Draft Version 3. Washington, D.C.: FTA.

[2] Parsons Corporation. 2006. *Ann Arbor Downtown Detroit AA/DEIS Transit Study: Draft Operating and Maintenance Cost Methodologies and Costs.* Prepared for Southeast Michigan Council of Governments.

[3] Dutta, U., and Zhong, X. 2013. "A LRT Operation and Maintenance (O&M) Cost Model to Perform Sensitivity Analysis." In *Proceeding of 2013 Annual Meeting of Transportation Research Board,* 1-15.

[4] NTD Program. "NTD Glossary." NTD Program. Accessed March 15, 2014. http://www.ntdprogram.gov/ntdprogram/Glossary.htm.

[5] Devore, J. 2012. *Probability and Statistics for Engineering and the Sciences.* Boston: Brooks/Cole.

[6] Legates, D. R., and McCabe, G. J. 1999. "Evaluating the Use of 'Goodness-of-Fit' Measures in Hydrologic and Hydroclimatic Model Validation." *Water Resources Research* 35 (1): 233-41.

[7] Chu, T. W., and Shirmohannadi, A. 2004. "Evaluation of the SWAT Model's Hydrology Component in the Piedmont Physiographic Region of Maryland." *Transactions of the American Society of Agricultural Engineers* 47 (4): 1057-73.

[8] Vazquez-Amabile, G. G., and Engel, B. A. 2005. "Use of SWAT to Compute Groundwater Table Depth and Streamflow in the Muscatatuck River Watershed." *Transactions of the American Society of Agricultural Engineers* 48 (3): 991-1003.

[9] Nash, J. E., and Sutcliffe, J. V. 1970. "River Flow Forecasting through Conceptual Models: Part I. A Discussion of Principles." *Journal of Hydrology* 10 (3): 282-90.

[10] ASCE (American Society of Civil Engineering). 1993. "Criteria for Evaluation of Watershed Models." *Journal of Irrigation Drainage Engineering* 119 (3): 429-42.

[11] Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D., and Veith, T. L. 2007. "Model Evaluation Guidelines for Systematic Qualification of Accuracy in Watershed Simulations." *American Society of Agricultural and Biological engineers* 50 (3): 885-900.

[12] Sevat, E., and Dexetter, A. 1991. "Selection of Calibration Objective Functions in the Context of Rainfall-Runoff Modeling in a Sudanese Savannah Area." *Hydrological Science Journal* 36 (4): 307-30.

[13] Singh, J., Knapp, H. V., and Demissie, M. 2005. "Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT." *Journal of American Water Resources Association* 41 (2): 361-75.

[14] Jain, S., and Sudheer, K. 2008. "Fitting of Hydrologic Models: A Close Look at the Nash–Sutcliffe Index." *J. Hydrol. Eng.* 13 (10): 981-6.

[15] Poister, T. H., Pasha, O. Q., and Edwrds, L. H. 2013. "Does Performance Management Lead to Better Outcome? Evidence from the US Transit Industry." *Public Administration Review* 73 (4): 625-36.

[16] Fielding, G. J., Babitsky, T. T., and Brenner, M. E. 1985. "Performance Evaluation for Bus Transit." *Transportation Research Part A, General* 19 (1): 73-82.

[17] Ryus, P. 2010. *A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry.* Transit Corporation Research Program Report No. 141, Transportation Research Board.

[18] McCuen, R., Knight, Z., and Cutter, A. 2006. "Evaluation of the Nash-Sutcliffe Efficiency Index." *J. Hydrol. Eng.* 11 (6): 597-602.