

- Left-Turn Lanes*. M.S. thesis. Northwestern University, Evanston, Ill., 1970.
8. Institute of Traffic Engineers. The Use and Effectiveness of Double Left-Turn Movements. *Traffic Engineering*, Vol. 45, 1975, pp. 52-57.
  9. W. Kunzman. Another Look at Signalized Intersection Capacity. *ITE Journal*, Vol. 48, 1978, pp. 12-15.
  10. *SAS User's Guide: Statistics*. SAS Institute Inc., Cary, N.C., 1982.
  11. R. J. Freund and R. C. Littell. *SAS for Linear Models: A Guide to the ANOVA and GLM Procedures*. SAS Institute, Inc., Cary, N.C., 1981.

---

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

---

# Use and Effectiveness of Simple Linear Regression To Estimate Saturation Flows at Signalized Intersections

ROBERT W. STOKES, VERGIL G. STOVER, AND CARROLL J. MESSER

In this paper, data from 14 intersection approaches with exclusive double left-turn lanes in three Texas cities are used to illustrate the use of simple linear regression to estimate saturation flows. Some theoretical considerations and potential bias in saturation flows estimated by simple linear regression are also briefly discussed. Average left-turn saturation flows in excess of 1,800 vehicles per hour of green per lane for the majority of the study sites within each city are suggested by the regression estimates. Relative to generally accepted straight-through flow rates, the regression models appear to have substantially underestimated average left-turn departure headways.

When a queue of vehicles is released by a traffic signal, the departure flow rate quickly increases until after the first few vehicles, when a uniform average departure rate is reached. This uniform departure rate is called the saturation flow rate of the intersection approach. Because the flow at signalized intersections is controlled by the amount of green time allotted, saturation flow under these conditions is defined as the flow rate that would result if there were a continuous queue of vehicles and they were given 100 percent green time (1). Saturation flow is generally expressed in vehicles per hour of green time (vphg).

A method commonly used to estimate saturation flows at

signalized intersections is the headway method. In this method, interarrival times (headways) of all saturated vehicles are measured at the intersection stop line, a saturated vehicle being one that has had to stop or almost stop in the queue behind the traffic signal. In the headway method, saturation flow is calculated directly as the reciprocal of the average headway of saturated vehicles.

One of the problems in estimating saturation flows from observed headways is accurately defining the saturation flow region of the queue. This problem is usually addressed by one of two procedures. The first and more straightforward of the procedures is to simply plot the average time headways of a queue of vehicles entering an intersection from a stopped position. These plots typically take the form shown in Figure 1. The saturation flow region of the queue can be identified by examining the plot and making a subjective determination of the vehicle storage positions for which departure headways could be assumed to be equal. Formal statistical procedures such as analysis-of-variance and multiple comparisons can be used to examine the departure headways in a more objective manner (e.g., see paper by Stokes, Messer, and Stover elsewhere in this Record).

A second approach to the problem of determining the saturation flow region of the queue involves the use of a formal optimization process such as simple linear regression. The regression models used in this process are typically of the following form:

$$T_c = B_0 + B_1 n_i + e_i \quad (1)$$

where

$T_c$  = elapsed time since the start of green for the vehicle in queue storage position  $n_i$  to enter the intersection,

$B_0, B_1$  = regression coefficients, and

$e_i$  = random error term.

The parameters of interest in Equation 1— $B_0$  and  $B_1$ —represent starting delay and average headway, respectively.

Presented in this paper is an example of the use of simple linear regression to estimate the saturation flows of exclusive double left-turn lanes. The basic procedures utilized are applicable to straight-through movements as well. Perhaps more important, some theoretical considerations and potential bias in saturation flows estimated by simple linear regression are also discussed.

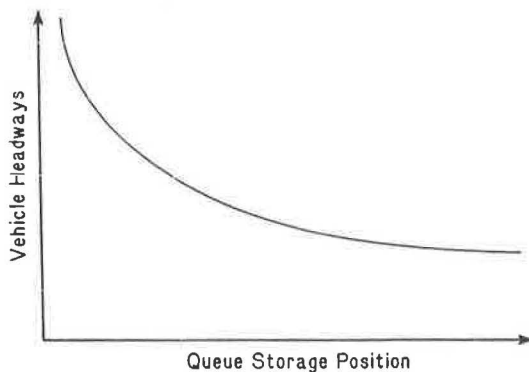


FIGURE 1 Generalization of the relationship between departure headways and queue storage positions.

## DATA BASE

The data used in this paper are from a larger study directed at estimating the saturation flows of exclusive double left-turn lanes (2). The data consist of the time (sec) elapsed since the start of the left-turn green phase for successive left-turning vehicles to cross the intersection stop line. The data were collected using time-lapse photography (9 frames/sec) at 14 intersection approaches with exclusive double left-turn lanes in three Texas cities. Table 1 gives a summary of the sample in terms of green phases, left turns, and average queue lengths observed. As indicated by the data in the table more than 3,400 completed left turns were observed at the 14 study sites. This was considered to be a sufficient sample for estimating left-turn saturation flows.

## DATA ANALYSIS

The regression analysis involved evaluation of the following model:

$$T_{c(i)} = B_0 + B_1 n_i + e_i \quad (2)$$

where  $T_{c(i)}$  is the time (sec) after the start of the left-turn green phase for the rear axle of the vehicle in queue storage position  $n_i$  to cross the intersection stop line from Turn Lane  $i$ , for  $i = 1, 2$  (see Figure 2).

In Equation 2, the parameter of interest is  $B_1$ , where the sample estimate of  $B_1$ , denoted by  $b_1$ , is an estimate of average departure headway (sec/veh). The parameters of the model given by Equation 2 were estimated by the least-squares method using the regression (REG) procedure of the Statistical Analysis System (SAS) Computer Program Package (3). The

TABLE 1 SUMMARY OF THE SAMPLE

Study Site <sup>a</sup>	No. of Phases Observed	No. of Completed Left-Turns Observed (veh)		Average Queue Length <sup>b</sup> (veh)	
		Lane 1 (inside)	Lane 2 (outside)	Lane 1 (inside)	Lane 2 (outside)
WB US 183 at Burnet Rd., Austin, Tex.	22	175	179	6.5	6.7
EB US 183 at Burnet Rd., Austin, Tex.	12	133	121	8.4	8.6
NB Texas at University, College Station, Tex.	22	171	199	7.0	8.9
WB University at Texas, College Station, Tex.	19	124	118	8.6	9.8
EB University at Texas, College Station, Tex.	13	89	116	6.2	7.5
NB Texas at Jersey, College Station, Tex.	23	231	221	8.0	10.4
EB Jersey at Texas, College Station, Tex.	13	82	95	6.3	7.0
SB Texas at Harvey Rd., College Station, Tex.	15	118	136	6.4	8.9
EB Westheimer at Gessner, Houston, Tex.	23	143	132	6.6	6.6
SB Gessner at Westheimer, Houston, Tex.	22	114	110	6.4	6.6
NB Post Oak Blvd. at San Felipe, Houston, Tex.	15	70	80	4.6	6.7
NB Hillcroft at Westheimer, Houston, Tex.	23	124	134	8.7	8.5
SB Gessner at Bellaire, Houston, Tex.	14	64	41	4.9	2.7
NB Gessner at Bellaire, Houston, Tex.	16	69	69	4.4	5.9
Total	252	1,707	1,751		
Average				6.8	7.6

<sup>a</sup>WB = westbound, EB = eastbound, NB = northbound, and SB = southbound.

<sup>b</sup>Queue Length = number of vehicles stopped at the onset of the left-turn green phase.

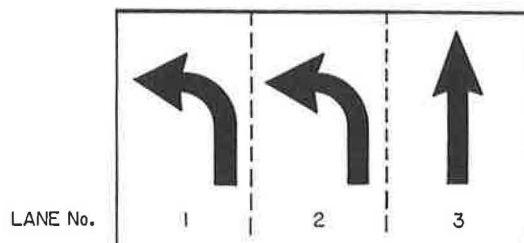


FIGURE 2 Lane numbering scheme

SAS Computer Program Package has been extensively tested and is widely accepted for statistical analyses.

The parameters of the Equation 2 regression model were estimated from the entire sample (i.e., for all vehicles), as well as from subsets of the sample. The subsets of the sample considered included the following regions of the queue:

$$1 \leq n_i \leq n_{k(i)}$$

•  
•  
•

$$5 \leq n_i \leq n_{k(i)}$$

where  $n_i$  is the queue storage position for left-turning vehicles in Lane  $i$ , for  $i = 1, 2$ ; and  $n_{k(i)}$  is the total number of vehicles queued in Lane  $i$  at the start of the left-turn green phase.

## RESULTS

Table 2 gives the 95 percent confidence intervals for the  $b_1$ 's for the fitted regression models. Note in this table that the most precise estimates of minimum departure headways were obtained for the region of the queue defined by  $2 \leq n_i \leq n_{k(i)}$ . Also note that the 95 percent confidence interval for average headways for the region of the queue defined by  $3 \leq n_i \leq n_{k(i)}$  contains the point estimate for the region defined by  $2 \leq n_i \leq n_{k(i)}$ . That is, the estimates obtained from these two regions of the queue differ only in the amount of variation exhibited by the estimates. In addition, the precision in the estimates obtained for the region of the queue defined by  $3 \leq n_i \leq n_{k(i)}$  is within 0.1 sec of the estimates obtained for the region given by  $2 \leq n_i \leq n_{k(i)}$ . Hence, it is suggested by the regression analyses that the saturation flow region of the queue can be defined by the region  $3 \leq n_i \leq n_{k(i)}$ .

Table 3 gives a summary of the fitted Equation 2 regression models for each of the three cities studied for the saturation flow region of the queue. The saturation flow estimates developed from Equation 2 are given in Table 4.

## DISCUSSION OF RESULTS

The concept of saturation flow dictates that when a queue of vehicles is released by a traffic signal, average departure headways gradually decrease until after the first few vehicles, when the headways become uniform. The theory, then, suggests that

TABLE 2 95 PERCENT CONFIDENCE INTERVALS<sup>a</sup> FOR AVERAGE LEFT-TURN DEPARTURE HEADWAYS ESTIMATED FROM EQUATION 2 (SEC/VEH)

Lane	Average Headways ( $\bar{h}_L$ ) for Data Sets Shown					
	All Vehicles	$j \leq n_i \leq n_{k(i)}$ , for $j =$				
		1	2	3	4	5
1	2.2 <sup>b</sup>	2.1 <sup>b</sup>	2.1 <sup>b</sup>	$1.9 \leq \bar{h}_L \leq 2.1$	$1.8 \leq \bar{h}_L \leq 2.0$	$1.8 \leq \bar{h}_L \leq 2.0$
2	2.3 <sup>b</sup>	2.2 <sup>b</sup>	2.1 <sup>b</sup>	$2.0 \leq \bar{h}_L \leq 2.2$	$2.1 \leq \bar{h}_L \leq 2.3$	$2.1 \leq \bar{h}_L \leq 2.3$

<sup>a</sup>Confidence interval =  $b_1 \pm 2 \times$  standard error of  $b_1$ .

<sup>b</sup>Confidence interval < 0.1 sec.

TABLE 3 SUMMARY OF EQUATION 2 REGRESSION MODELS BY CITY FOR  $3 \leq n_i \leq n_{k(i)}$

City and Lane	Sample Size	$b_0$ (Starting Delay)		$b_1$ (Headway)		Mean Square Error	$r^2$
		Estimate (sec)	Standard Error	Estimate (sec/veh)	Standard Error		
Austin							
Lane 1	175	2.2	0.37	1.9	0.06	2.79	0.84
Lane 2	180	1.3	0.31	2.2	0.05	2.09	0.91
College Station							
Lane 1	486	2.5	0.20	1.9	0.04	2.57	0.86
Lane 2	592	2.2	0.26	2.0	0.04	5.07	0.80
Houston							
Lane 1	339	2.4	0.22	1.8	0.05	1.04	0.79
Lane 2	329	2.5	0.24	1.7	0.06	1.33	0.74

TABLE 4 95 PERCENT CONFIDENCE INTERVALS<sup>a</sup> FOR AVERAGE LEFT-TURN SATURATION FLOW BY CITY

Lane	Average Saturation Flow, $\bar{s}_L$ (vphg)		
	Austin	College Station	Houston
1	$1780 \leq \bar{s}_L \leq 2020$	$1820 \leq \bar{s}_L \leq 1980$	$1895 \leq \bar{s}_L \leq 2120$
2	$1565 \leq \bar{s}_L \leq 1715$	$1730 \leq \bar{s}_L \leq 1875$	$1980 \leq \bar{s}_L \leq 2280$

<sup>a</sup>Confidence interval =  $1/(\bar{h}_L \pm 2 \times \text{standard error } \bar{h}_L) \times 3,600$ .

if Equation 2 were to be evaluated by using data sets that successively ignore the first, first and second, . . . , first through fifth vehicles in the queue, the resulting regression lines should tend to become successively more linear. Specifically, it would be expected that the coefficient of determination ( $r^2$ ) would increase, and that the standard error of  $b_1$  [ $S(b_1)$ ] would decrease as data beginning at successively larger queue storage positions were used to estimate the parameters of the model.

Table 5 gives a summary of the analysis-of-variance tables and headway estimates for Equation 2 as estimated from the total sample and subsets of the sample. An examination of the slopes (i.e., headways) given in Table 5 indicates that they appear to behave in a manner consistent with the concept of saturation flow. That is, as vehicles near the head of the queue are successively deleted from the sample, the slopes tend to decrease, indicating a flattening of the regression lines. However, at first glance the  $r^2$ 's, and the  $S(b_1)$ 's to a lesser extent, appear to exhibit somewhat counterintuitive behaviors. For example, there is a tendency for the  $S(b_1)$ 's to increase and a tendency for the  $r^2$ 's to decrease as vehicles near the head of the queue are successively eliminated from the data set.

The apparent anomalies in the behavior of  $r^2$  and  $S(b_1)$  are probably due to the subsetting of the data, not to some inconsistency with the concept of saturation flow. To avoid misinterpreting the regression models, the nature of the mathemati-

cal components of the statistics involved should be examined. In this regard, the following points deserve note.

The value taken by  $r^2$  in a given sample tends to be affected by the range in the observations of the independent variable, the range in the  $n_i$  observations in this case. An examination of the components used to calculate  $r^2$ , as shown in Equations 3–3e, demonstrates the nature of its seemingly anomalous behavior.

$$r^2 = [\Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 - \Sigma (T_{c(i)} - \hat{T}_{c(i)})^2] / \Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 \quad (3)$$

$$= \Sigma (\hat{T}_{c(i)} - \bar{T}_{c(i)})^2 / \Sigma (T_{c(i)} - \bar{T}_{c(i)})^2 \quad (3a)$$

$$= 1 - [\Sigma (T_{c(i)} - \hat{T}_{c(i)})^2 / \Sigma (T_{c(i)} - \bar{T}_{c(i)})^2] \quad (3b)$$

where  $\bar{T}_{c(i)}$  and  $\hat{T}_{c(i)}$  are means and least-squares estimates, respectively. By using a simplified notation, Equations 3–3b can be restated as follows:

$$r^2 = (SST - SSE) / SST \quad (3c)$$

$$= SSR / SST \quad (3d)$$

$$= 1 - (SSE / SST) \quad (3e)$$

TABLE 5 SUMS-OF-SQUARES AND HEADWAY ESTIMATES FOR EQUATION 2

Data Set and Lane	Sample Size	Sums of Squares		Mean Square Error	$b_1$ (Headway)		$r^2$
		Error	Total		Estimate (sec/veh)	Standard Error	
All vehicles							
Lane 1	1,707	7,409	74,542	4.4	2.2	0.02	0.90
Lane 2	1,751	10,697	81,841	6.1	2.3	0.02	0.87
$1 \leq n_i \leq n_{k(i)}$							
Lane 1	1,500	3,028	35,077	2.0	2.1	0.02	0.91
Lane 2	1,598	4,831	49,182	3.0	2.2	0.02	0.90
$2 \leq n_i \leq n_{k(i)}$							
Lane 1	1,251	2,606	23,143	2.1	2.0	0.02	0.89
Lane 2	1,347	4,541	34,615	3.4	2.1	0.02	0.87
$3 \leq n_i \leq n_{k(i)}$							
Lane 1	1,000	2,212	15,431	2.2	2.0	0.03	0.86
Lane 2	1,101	4,193	24,669	3.8	2.1	0.03	0.83
$4 \leq n_i \leq n_{k(i)}$							
Lane 1	752	1,763	10,054	2.4	1.9	0.03	0.82
Lane 2	859	3,792	17,513	4.4	2.2	0.04	0.78
$5 \leq n_i \leq n_{k(i)}$							
Lane 1	509	1,297	6,318	2.6	1.9	0.04	0.79
Lane 2	623	3,356	12,074	5.4	2.2	0.05	0.72

where

SST = total sums of squares,  
 SSE = error sums of squares, and  
 SSR = regression (or model) sums of squares.

In general, SSE is not affected systematically by the spacing of the  $n_i$ 's. However, the wider the spacing of the  $n_i$ 's in the sample, the greater the spread of the observed  $T_{c(i)}$ 's around  $T_{c(i)}$  will tend to be, and hence the greater will be SST. Consequently, the wider the  $n_i$ 's are spaced, the higher  $r^2$  will tend to be (5). Thus, the observed reductions in the  $r^2$ 's are due to the narrowing of the range in the  $n_i$ 's associated with the successive elimination of vehicles near the head of the queue.

A similar explanation applies to the behavior of  $S(b_1)$  across the subsamples.  $S(b_1)$  is given by

$$S(b_1) = [\text{MSE}/\Sigma(n_i - \bar{n}_i)^2]^{1/2} \quad (4)$$

where MSE is the mean square error as given by SSE divided by the degrees of freedom for error.

As the sample is truncated (by successively eliminating vehicles at the head of the queue), the  $\Sigma(n_i - \bar{n}_i)^2$  term in Equation 4 must decrease. Consequently, as long as the MSEs remain fairly stable across the subsamples,  $S(b_1)$  must increase as one moves across the subsamples. An examination of the data in Table 5 suggests that for saturated vehicles the MSEs for each lane are fairly uniform across the subsamples. Consequently, the increases in  $S(b_1)$  would appear to be entirely logical.

The point of the preceding discussion is that a simple comparison of the  $r^2$ 's and  $S(b_1)$ 's for the queue segments considered may not be an efficient procedure for identifying the saturation flow region of the queue.

It also appears that at least one of the basic assumptions of regression analysis was not satisfied. Specifically, the variance of the error term in the regression models does not appear to be constant across the sample. Figures 3 and 4, which show plots of the residuals against the queue storage positions by lane for the region of the queue defined by  $1 \leq n_i \leq n_{k(i)}$ , illustrate the problem. The patterns of the plots suggest that the error vari-

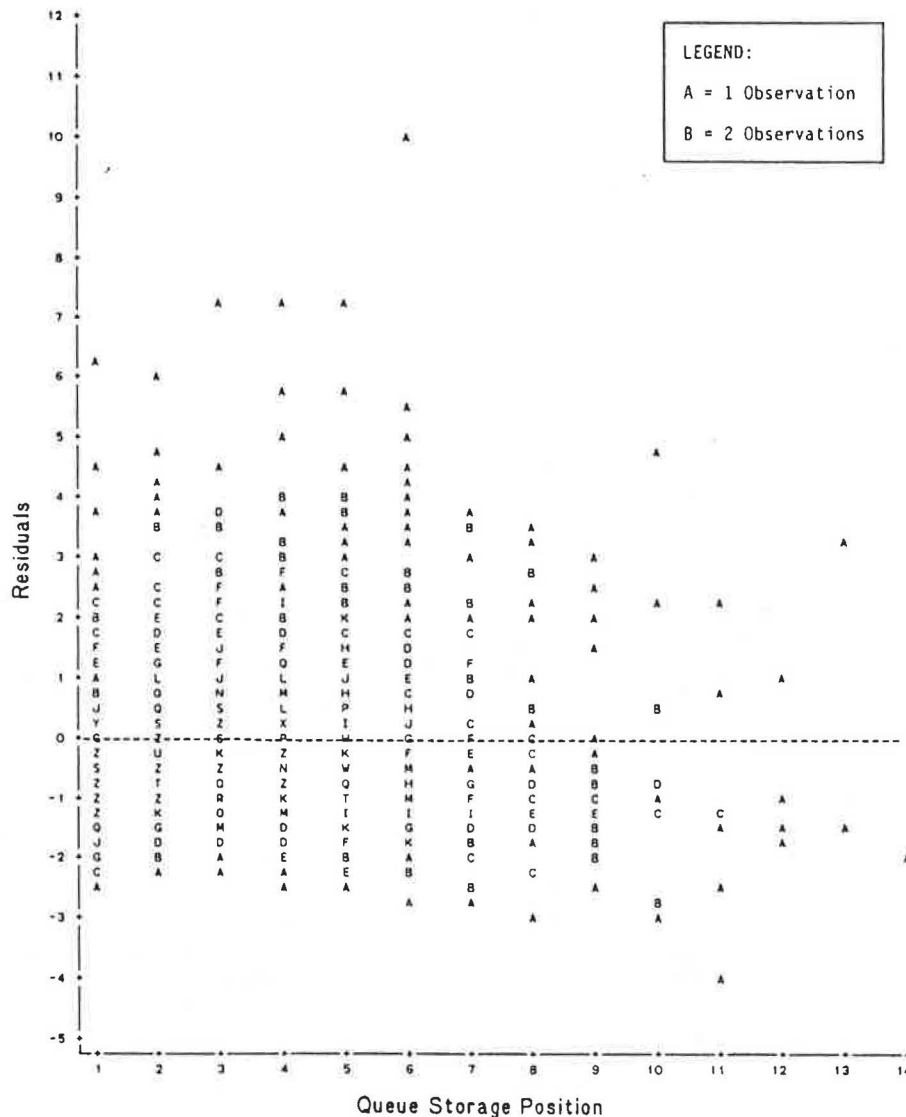


FIGURE 3 Plot of residuals against queue storage positions for Equation 2, for all approaches, for  $1 \leq n_i \leq n_{k(i)}$ : Lane 1.

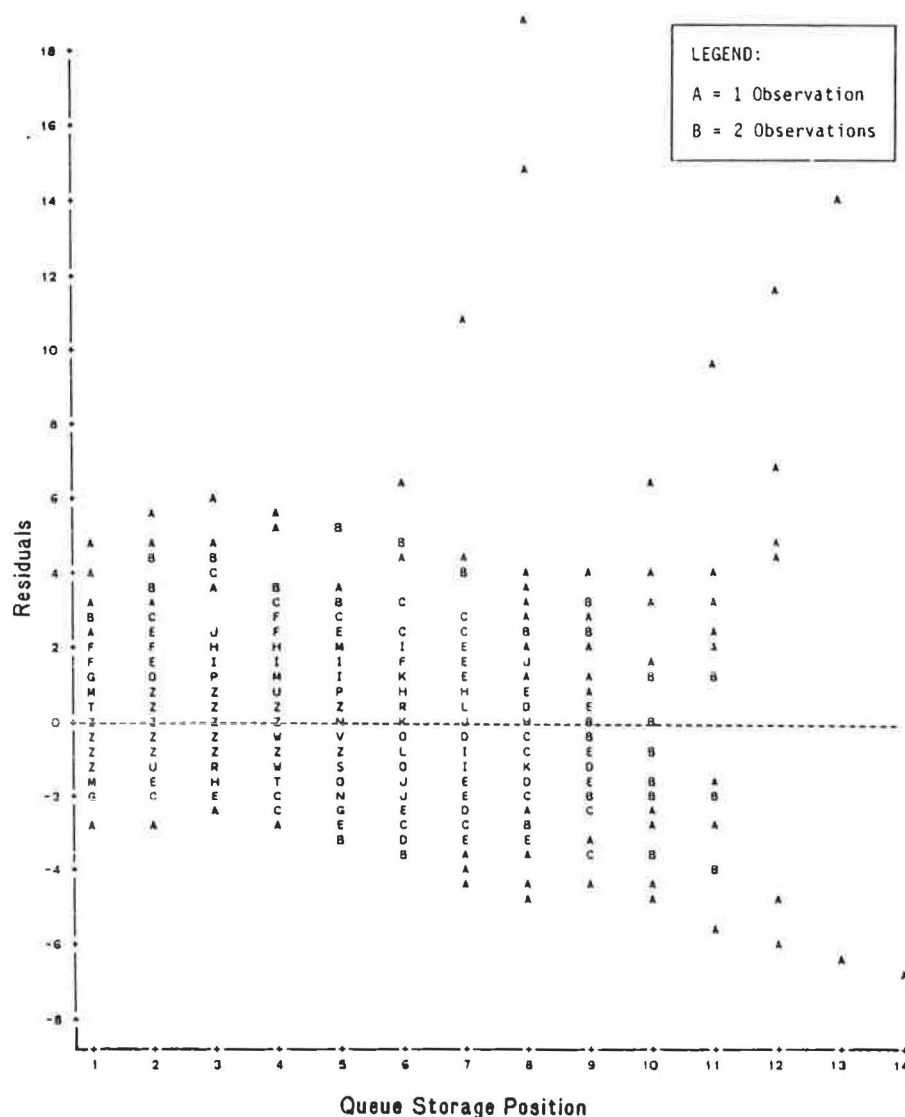


FIGURE 4 Plot of residuals against queue storage positions for Equation 2, for all approaches, for  $1 \leq n_1 \leq n_{k(t)}$ : Lane 2.

ance is a function of queue storage position. The patterns depicted for the two lanes are intriguing. For example, in the case of Lane 1 (Figure 3), the error variance tends to increase with queue storage position up to about the sixth position, where it begins to fall off and then stabilize. Note in Figure 4 that Lane 2 (the outside lane) exhibits the opposite trend; that is, the error variance is fairly constant until about the sixth storage position, where it begins to increase.

In addition to the subjective analysis of residuals just described, the constancy of the error variance was evaluated by using a formal statistical procedure. The procedure used was to

1. Array the observations by queue storage position and lane,
2. Divide the total observations into two equal data sets for each lane,
3. Fit separate regression functions to each half of the total observations,
4. Calculate the MSE for each, and
5. Test for equality of the error variances by the F-test.

The resulting variance ratios were significant at the 5 percent level.

Thus, evaluation of the error variance suggests that the regression estimates of the departure headways may be biased. Table 3 presents data on the nature of the suspected bias. Note that, with the exception of Lane 2 for the Austin sites, the estimated headways (i.e., the slopes) are all less than or equal to 2.0 sec. The regression estimates suggest average left-turn flow rates in excess of 1,800 vehicles per hour of green time per lane (vphgl) for the majority of the study sites within each city. Relative to generally accepted straight-through flow rates of 1,700-1,800 vphgl, the regression models appear to have underestimated average left-turn departure headways.

## SUMMARY

In using least-squares regression techniques in model-building exercises, there are a number of potential problem areas that are frequently overlooked by researchers using this method. For example, it is important that the researcher validate the



distributional assumptions on the model errors. The standard assumptions are that the dependent variable is normally distributed and that the errors are independent and have homogeneous variances.

In this study, the possible consequences of overlooking the distributional assumptions about the error variances have been examined. The results of this limited study suggest that non-constancy of the error variances (heteroscedasticity) may result in regression models that substantially underestimate left-turn departure headways. It is hoped that the discussion relating to model estimation and validation will encourage others to address these basic issues in the literature.

#### ACKNOWLEDGMENTS

The authors wish to thank the following individuals for their contributions to the study: D. L. Pugh, J. deJong, D. A. Max-

well, L. J. Ringer, and A. M. Elmquist, all with Texas A&M University; D. W. Hall, City of Austin; J. R. Black, City of College Station; and W. E. Hensch, City of Houston. The authors remain solely responsible for the contents of this paper.

#### REFERENCES

1. F. V. Webster and B. M. Cobbe. *Traffic Signals*. Her Majesty's Stationery Office, London, England, 1966.
2. R. W. Stokes. *Saturation Flows of Exclusive Double Left-Turn Lanes*. Ph.D. dissertation. Texas A&M University, College Station, 1984.
3. *SAS User's Guide: Statistics*. SAS Institute, Inc., Cary, N.C., 1982.
4. J. Neter, W. Wasserman, and M. H. Kutner. *Applied Linear Regression Models*. Richard D. Irwin, Inc., Homewood, Ill., 1983.

---

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

---

# Freeway Weaving Sections: Comparison and Refinement of Design and Operations Analysis Procedures

JOSEPH FAZIO AND NAGUI M. ROUPHAIL

Weaving sections represent the common right-of-way that occurs when two or more crossing freeway traffic streams are traveling in the same general direction. In conjunction with the development of the 1985 *Highway Capacity Manual* (HCM), several procedures have evolved for the purpose of updating, revising, and replacing the 1965 HCM procedure for design and operations analysis of freeway weaving sections. The objectives of this paper are twofold: to present and review the latest three weaving procedures available to highway and traffic engineers, and to propose specific refinements to a simple weaving section procedure to account for the lane distribution of traffic upstream of the weaving section. These adjustments primarily involve the development of a lane-shift variable, which represents the average amount of peak-period passenger car lane shifts occurring under a given geometric configuration and prevailing traffic volumes. Statistical testing of the refined procedure against the three procedures at more

than 50 sites nationwide indicated that the proposed procedure tends to predict observed average running weaving and nonweaving speeds more closely than do the other procedures in most cases.

---

A weaving section represents the physical space along a freeway where two (simple weaving) or more (multiple weaving) traffic streams traveling in the same general direction cross each other. Four basic movements are serviced in a simple weaving section, two weaving and two nonweaving (outer flows), as indicated in Figure 1a. Weaving traffic originating from the freeway mainline is denoted  $V_2$  and nonweaving traffic is denoted  $V_1$ . Weaving traffic originating from the minor approach or entrance ramp is denoted  $V_3$  and nonweaving traffic is denoted  $V_4$ . The length of a weaving section ( $L$ ) and the number of lanes ( $N$ ) are the two design parameters that dictate the mode of traffic operation to be expected, as illustrated in Figure 1b. (Note in this figure that  $N_b$  is the basic