Evaluating Potential Effectiveness of Headway Control Strategies for Transit Systems

Mark A. Turnquist and Steven W. Blume

Holding strategies for control of headways between transit vehicles are often considered as a means of improving the reliability of transit service. This paper describes simple tests that can be used to identify situations for which control is potentially attractive. These tests depend only on a simple measure of headway variability and the proportion of total passengers who will be delayed as a result of the holding strategy. Thus, this analysis provides transit operators with a simple screening model to evaluate potential effectiveness of controls.

Headway control has been proposed as one way to improve the reliability of transit service. By reliability we mean the ability of transit to adhere to schedule or to maintain regular headways and a consistent travel time. This ability is important to both the transit user and the transit operator. To the user, nonadherence to schedule results in increased wait time, makes transferring more difficult, and creates uncertainty about arrival time at the destination. To the operator, unreliability results in less effective utilization of equipment and personnel and reflects itself in reduced productivity and increased cost in the system's operations.

A study of the potential effectiveness of various strategies for control of unreliability in transit services is thus a vital element in the search for ways to improve transit productivity and efficiency. Such control strategies have important implications for both planning and management of transit systems. Control strategies may be divided into two basic groups: planning and real time. In general, the distinction is that planning strategies involve changes of a persistent nature. Examples include restructuring of routes and schedules, changes in the number and location of stops, or provision of exclusive rights-of-way. On the other hand, real-time control measures are designed to act quickly to remedy specific problems. These actions have immediate effects but seldom exert any influence on the general nature of operations over a longer time period.

Several real-time strategies for correcting service disruptions have been discussed in the literature. A good summary of the state of current knowledge in this area has been provided by Abkowitz and others (1). One commonly considered control strategy is the holding of selected vehicles at control points along a route to regularize headways between successive vehicles. That is, a vehicle that arrives at the control stop too close behind the preceding vehicle would be deliberately delayed to make the headway between these vehicles more nearly equal to the scheduled headway.

The major incentive for making headways more regular is to reduce waiting time of passengers who board at or beyond the control point. If passengers arrive at a stop without regard to the schedule of service (i.e., randomly), a well-known formula [see Welding (2)] gives the average wait time as

$$E(W) = [E(H)/2] + [V(H)/2E(H)] \tag{1}$$

where

E(W) = average wait time,
E(H) = average headway between vehicles, and
V(H) = variance of headway.

Thus, making headways more regular (i.e., reducing the variance) serves to reduce average wait.

On the other hand, the major costs of such a policy are borne by passengers who are already on the vehicle, since they are delayed when the bus is held up. Thus, the implementation of a holding control strategy involves making some passengers better off at the expense of others. At a minimum, if control is to be effective, it must reduce aggregate waiting time by more than it increases aggregate in-vehicle time (possibly allowing for some differential weighting of these two elements of total trip time).

The purpose of this paper is to provide some basic rules of thumb to indicate the conditions under which a holding strategy might be effective. By implication, we also wish to describe those situations in which such a strategy is not likely to be effective. These rules of thumb are based on relatively modest data requirements about the route and, hence, should be useful in making basic planning decisions about whether or not to implement such a control strategy on a given route.

PREVIOUS ANALYSIS

An article by Barnett (3) has provided several important ideas for the work contained here. He formulated a model based on a simple discrete approximation to the probability distribution of vehicle arrival times at bus stops. Based on this simple model, an optimal holding strategy can be derived to minimize the total delay to all passengers who use the route. The resulting strategy depends on (a) the mean and variance (or standard deviation) of the headway distribution, (b) the ratio of average vehicle load at the control point to average number of boarding passengers at subsequent stops, and (c) the correlation between successive vehicle arrival times at the control stop. This last information is a measure of the degree of bunching or pairing of vehicles on the route: A route on which vehicles have bunched in pairs would have a large negative correlation between successive headways because a very short one (between two paired vehicles) will be followed by a very long one (between bunches). Statistical estimation of this correlation is difficult, however, because of the small sample sizes available and the notorious unreliability of the estimators of covariance.

The objectives of this paper are to analyze holding strategies by using a more general probability model of vehicle arrival times at the control stop and to shed some additional light on the question, Under what conditions is control likely to be of value? Specifically, we wish to allow a transit operator to address this question without detailed knowledge of the covariances between successive vehicle arrival times at stops, as this information is seldom available.

Our approach is to use a general model of the probability distribution of headways between successive vehicles and then examine two simple cases that provide approximate upper and lower bounds on the potential benefits of a holding strategy. By doing this, basic conclusions can be reached regarding situations in which control is likely to be beneficial and those in which it is not.

We will examine a holding strategy that holds each early vehicle (i.e., each vehicle preceded by a short headway) until the headway preceding it reaches a minimum allowable value (h_{min}). The structure of the analysis is to find the value of h_{min} that minimizes total delay to passengers (including both wait time and in-vehicle delay). This optimal value of h_{min} will be denoted h^*_{min}. Once h^*_{min} is found, those situations for which control is advantageous can be identified.

UPPER BOUND ON EFFECTIVENESS OF HOLDING

Control of headways will make the greatest reduction in total delay when headways alternate (i.e., short, long, short, long). This happens on routes where vehicles are influenced substantially by the operation of the vehicle in front of them. For example, this would tend to be the case where loading delays are relatively more important than traffic congestion in determining overall vehicle operating speed. Routes in which pairing is prevalent would be of this type. In such a situation, holding a vehicle to lengthen a short headway also serves to reduce the long one that follows. Thus, the variance of headways is reduced by a greater amount for a given delay to the held vehicle than if short headways might be followed by another short headway.

The extreme case is when the observed sequence of headways alternates between two discrete values. In this case, the sum of any two consecutive headways is a constant. That is, if one headway is 2 min too short, the next one must be 2 min too long. By the same argument, if the second headway is 2 min too long, the third must be 2 min too short, and so on. In a statistical sense, successive headways are perfectly correlated, so that knowledge of one headway implies knowledge of the entire set. For this case, headway control will have maximum benefits.

If we denote the scheduled headway by $\overline{H}$ and the magnitude of the deviation by x, the marginal probability density function for headways before control is given by p(H):

$$p(H) = \begin{cases} 0.5 & H = \overline{H} - x \quad (2a) \\ 0.5 & H = \overline{H} + x \quad (2b) \end{cases}$$

For the probability distribution of headways described by Equations 2a and 2b, the expected headway is $\overline{H}$ and the variance is x^2. The control action lengthens the short headways to a value $h_{min} = \overline{H} - \rho x$, where $0 \leq \rho \leq 1$.

We will define an optimal holding strategy to be one that minimizes total delay to passengers. Total delay is expressed as

$$T = \gamma E(D) + (1 - \gamma) E(W) \quad (3)$$

where

T = total delay to all passengers,
E(D) = expected delay to passengers already on board the vehicle,
E(W) = expected wait time for passengers arriving at or beyond the control stop, and
γ = weighting constant to reflect the relative number of passengers already on board to those waiting to board at subsequent stops.

The expected delay to passengers already on the vehicle is simply the average length of time a vehicle will be held. If we assume that passengers arrive at stops at random times, Equation 1 can be used to determine expected wait time.

A holding strategy that minimizes T will be defined by the value h^*_{min}. Because $h_{min} = \overline{H} - \rho x$, we can find h^*_{min} by finding the optimal value of ρ. Note that after control the headway distribution is given by

$$p'(H') = \begin{cases} 0.5 & H' = \overline{H} - \rho x \quad (4a) \\ 0.5 & H' = \overline{H} + \rho x \quad (4b) \end{cases}$$

This distribution has expected value still equal to $\overline{H}$, but has variance $\rho^2 x^2$. This reduces wait time for passengers yet to board to

$$E(W') = (\overline{H}/2) + (\rho^2 x^2 / 2\overline{H}) \quad (5)$$

The delay to passengers already on the vehicle is equal to $(1 - \rho)x$ if the vehicle is held. Since the probability of a short headway is 0.5, the expected in-vehicle delay is

$$E(D) = 0.5(1 - \rho)x \quad (6)$$

By substituting Equations 5 and 6 into Equation 3, we obtain total expected delay as

$$T = 0.5\gamma(1 - \rho)x + (1 - \gamma)[(\overline{H}/2) + (\rho^2 x^2 / 2\overline{H})] \quad (7)$$

To find the optimal value of ρ, we can differentiate the expression for T with respect to ρ, and set the result equal to zero.

$$dT/d\rho = -0.5\gamma x + (1 - \gamma)\rho\,(x^2/\overline{H}) = 0 \quad (8)$$

This implies an optimal value for ρ:

$$\rho = 0.5\gamma x\, \overline{H}/(1 - \gamma)\, x^2 = [0.5\gamma/(1 - \gamma)]\, (\overline{H}/x) \quad (9)$$

The resulting value for h^*_{min} is then

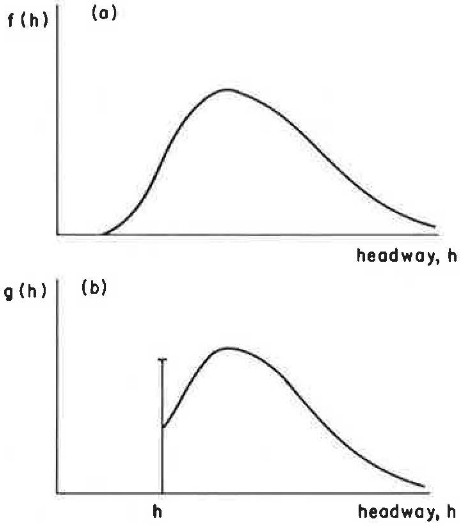
$$h^*_{min} = [(1 - 1.5\gamma)/1 - \gamma]\, \overline{H} \quad (10)$$

For control to be effective, we must have $h^*_{min} > \overline{H} - x$; that is, the optimal minimum headway after control must be greater than the short headways before control, or it does not pay to control at all. This means that we must have $\rho < 1$, which implies that we must satisfy the condition $x/\overline{H} > 0.5\gamma/1 - \gamma$. However, recall that the variance of the headway distribution before control was x^2. Thus, the quantity $x/\overline{H}$ is simply the coefficient of variation (standard deviation divided by mean) of the headway distribution. Thus, for control to be effective, the coefficient of variation of the headway distribution must exceed $0.5\gamma/1 - \gamma$. If it does not, the optimal value of ρ is 1, which implies no control.

This condition, then, provides a simple test for potential effectiveness of a control policy. It is based on two simple pieces of information: (a) the coefficient of variation in the headway distribution and (b) the relative proportion of riders who are already on board the vehicle to those who are yet to board at subsequent stops.

It must be kept in mind that this condition is derived for the best possible case (i.e., when successive headways are perfectly correlated). Thus, if the condition is not met, we can be confident that control will not be effective. However, we must look more closely at situations for which the condition is met because the actual

Figure 1. Headway distribution (a) before control and (b) after control.



situation may be less favorable to control than is reflected in this model.

LOWER BOUND ON THE EFFECTIVENESS OF HOLDING

In order to establish a lower bound on the effectiveness of holding, we will examine the opposite extreme case, which corresponds to the situtation in which headways between successive vehicles are statistically independent. This means that knowledge that a given headway is short gives us no additional information about the probable values for the next headway. Such a situation would arise, for example, when traffic conditions have a much greater effect on vehicle operations than does the loading time at stops. In this case control will be less effective because we have no guarantee that by lengthening a short headway we are also reducing a long headway. We might be simply reducing another, already short, headway. This case of independent headways thus provides a lower bound on the effectiveness of control strategies, which will allow us to further refine our evaluation of situations likely to be favorable for control.

We assume that the distribution of headways before control is applied is described by a cumulative distribution function [F(h)] with a density function [f(h)]. The effect of the control strategy is to make all headways less than some value (h_{min}) equal to that value. The distribution of headways before and after control is shown in Figure 1. There is a nonzero probability that the headway will take on the discrete value h_{min}, and for values of $h > h_{min}$, there is a continuous density function.

The expression for the distribution of headways after control is applied can be derived by considering a sequence of two successive headways after control, which we will denote H'_{i-1} and H'_i. The probability that $H'_i \leq h$ depends on both the headways H_{i-1} and H_i before control of vehicle i (if any), as well as the value of the minimum allowable headway (h_{min}). On one hand, $H'_i \leq h$ if $H_{i-1} > h_{min}$ (and thus not changed by the control strategy) and $H_i \leq h$. If $H_{i-1} \leq h_{min}$, it becomes $H'_{i-1} = h_{min}$ (after control), and the ith headway is shortened. In this case $H'_i \leq h$ if the sum of H_{i-1} and H_i before control of i was less than $h_{min} + h$. Of course, because the control policy enforces a minimum headway, the probability is that $H'_i \leq h$ will be 0 for $h < h_{min}$. These statements can be summarized in the form of a cumulative probability distribution function [G(h)] as shown in Equation 11:

$$G(h) = P(H'_i \leq h) = \begin{cases} 0 & h < h_{min} \\ P(H_{i-1} > h_{min})P(H_i \leq h) + P(H_{i-1} \leq h_{min}) \times P(H_{i-1} + H_i \leq h + h_{min} \mid H_{i-1} \leq h_{min}) & h \geq h_{min} \end{cases} \quad (11)$$

where P(·) denotes the probability of the event described by (·). From the distribution function in Equation 11, we can obtain (at least in theory) the probability density function [g(h)] for $h \geq h_{min}$, shown in Figure 1b, by differentiating with respect to h. The probability that $h = h_{min}$ is given by $G(h_{min})$.

As in the previous case of perfectly correlated headways, our analysis proceeds by solving for the optimal value of h_{min} and then using this to describe the conditions for which control is potentially beneficial. The process of finding the optimal h_{min} involves trading off reductions in wait time (due to reduced headway variance) against in-vehicle delays due to holding of vehicles.

The variance of headways after control can be written as shown in Equation 12:

$$V(H') = [h_{min} - E(H')]^2 G(h_{min}) + \int_{h_{min}}^{\infty} [h - E(H')]^2 g(h)dh \quad (12)$$

The rate of change of this variance with changes in h_{min} is

$$(d/dh_{min}) V(H') = 2G(h_{min})[h_{min} - E(H')] [1 - (d/dh_{min})E(H')] \quad (13)$$

For relatively small values of h_{min}, we can argue (to a first-order approximation) that changes in h_{min} will not affect the mean headway. Thus, $(d/dh_{min})E(H') \simeq 0$. While this approximation is not strictly accurate, a good case can be made that an operator is unlikely to implement a control policy that increases mean headway significantly. This would have negative impacts on vehicle productivity and also on passenger wait and travel time. Thus, the magnitude of control delays applied is likely to be small, and hence the approximation is a reasonable one. For small values of control delay, we can also approximate $G(h_{min})$ by $F(h_{min})$. These two approximations allow us to obtain the result in Equation 14:

$$(d/dh_{min})E(W') = (d/dh_{min}) \{[E(H')/2] + [V(H')/2E(H')]\} \simeq (h_{min} - \bar{H}/\bar{H}) F(h_{min}) \quad (14)$$

This provides the ability to evaluate (approximately) the marginal rate of reduction in waiting time as the minimum allowable headway increases. Total delay (T) will be minimized when the marginal rate of reduction in waiting time is just equal to the marginal rate of increase in in-vehicle delay.

The delay incurred by passengers already on a vehicle that is held is given by:

$$D = \begin{cases} h_{min} - H & H < h_{min} \quad (15a) \\ 0 & H \geq h_{min} \quad (15b) \end{cases}$$

From this, we can derive the expected delay, as shown in Equation 16:

$$E(D) = \int_0^{h_{min}} (h_{min} - h) f(h)dh = h_{min} F(h_{min}) - \int_0^{h_{min}} hf(h)dh \quad (16)$$

Figure 2. Areas of potential usefulness for headway control.

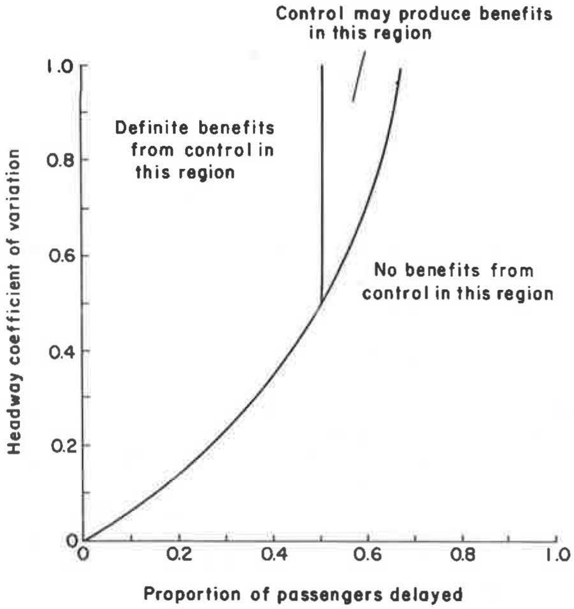
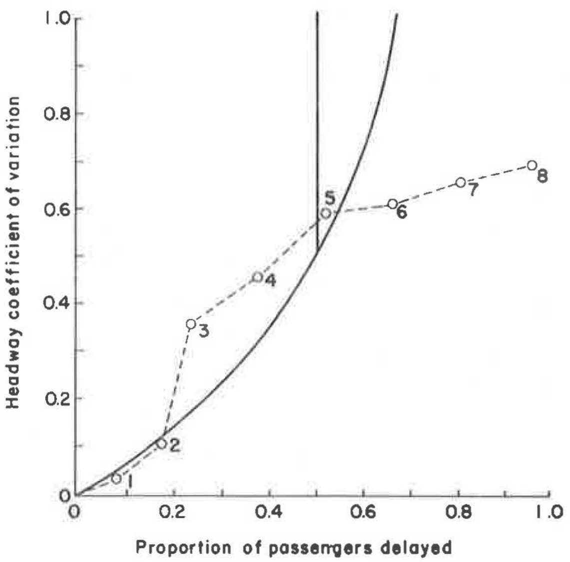


Figure 3. Trajectory of stops along a bus route.



The marginal change in expected delay is then

$$(d/dh_{min})\,E(D) = F(h_{min}) + h_{min}\,f(h_{min}) - h_{min}\,f(h_{min}) = F(h_{min}) \quad (17)$$

By using the expressions in Equations 14 and 17 we can then solve for an optimal value of h_{min} by setting $dT/dh_{min} = 0$.

$$dT/dh_{min} = \gamma(d/dh_{min})\,E(D) + (1-\gamma)\,(d/dh_{min})\,E(W') = 0$$

$$\Rightarrow \quad \gamma F(h_{min}) + (1-\gamma)[(h_{min} - \overline{H})/\overline{H}]\,F(h_{min}) = 0$$

$$\Rightarrow \quad h_{min} = [(1-2\gamma)/(1-\gamma)]\,\overline{H} \quad (18)$$

Since $\gamma > 0$, $h_{min} \le \overline{H}$. Because we must also have $h_{min} > 0$, this solution is only valid if $\gamma < 0.5$. We can summarize this as shown in Equations 19a and 19b, which give the expression for the optimum value of h_{min}, denoted h^{**}_{min}.

$$h^{**}_{min} = \begin{cases} [(1-2\gamma)/(1-\gamma)]\,\overline{H} & 0 \le \gamma < 0.5 \quad (19a) \\ 0 & 0.5 \le \gamma \le 1 \quad (19b) \end{cases}$$

The condition for which nonzero holding is beneficial ($0 \le \gamma < 0.5$) provides important information on the situations in which holding strategies are potentially useful, in the same way that the condition $x/\overline{H} > 0.5\gamma/1 - \gamma$ from Equation 9 did.

IMPLICATIONS OF THE ANALYSIS

These two pieces of information can be combined, as illustrated in Figure 2, to yield a convenient representation of situations for which headway control is likely to produce benefits for passengers—those for which it is unlikely to be worthwhile and those for which more careful analysis is required. By analyzing the two extreme cases of independent headways and perfectly correlated headways in detail, we can bound the regions of effectiveness for a class of headway control strategies, as shown in Figure 2. For situations in which control produces benefits under both extremes, we can be fairly confident that it will be beneficial. On the other hand, there are situations in which control does not appear to be desirable under the best of circumstances; hence, control in these situations is unlikely to be useful. There remains one reasonably small region in which control would probably produce benefits on routes where vehicles are substantially influenced by the vehicles in front of them but not on routes where vehicles move relatively independently of one another. For situations in this region, more detailed and specialized analysis is required.

A major implication of the result shown in Figure 2 is that it is wise to control a route at a point where relatively few people are on the vehicle and relatively many are waiting to board at subsequent stops, in order that the value of γ be small. Generally, this means that the control point should be located as early along the vehicle's route as possible. However, reliability problems worsen as one proceeds along a route. If dispatching at the route origin is effective, the headways will be reasonably regular at the early stops along the route, which implies that the coefficient of variation will be small. At stops further along the route, however, the coefficient of variation in headways will tend to be larger. Thus, the decision of whether or not to implement a control strategy is tied to identification of a logical control point along the route.

Each stop along a route will have a particular headway distribution (with implied coefficient of variation) and value of γ associated with it. Thus, each stop could be plotted as a point in the space defined by these two variables, as shown in Figure 3. Then, by looking at the trajectory of the route relative to the boundary values, the transit operator can make a decision about whether or not to control the route and, if so, where. For example, for the route illustrated by Figure 3, control at stop 3 might be worthwhile, but at stop 8 it is unlikely to be beneficial.

It is also illuminating to examine the form of the optimal holding policy for the two extreme cases analyzed here, as illustrated by Equations 10 and 19. Note first that, in both cases, the magnitude of the optimal minimum headway is dependent on the scheduled average headway, but not the variability of headways. Thus, determining a policy on minimum headways to be enforced for each extreme case is quite simple and requires very little data and only simple analysis.

Second, note that the optimal minimum headway is always smaller if successive headways are independent than if they are negatively correlated. This follows logically from the fact that a given amount of delay is less beneficial when headways are independent. Thus, we would expect the optimal delay to be smaller.

Clearly, as previously demonstrated by Barnett (2), precise setting of an optimal strategy for a given situation requires knowledge of the covariance between successive headways. However, our analysis, based on much more general models of headway distributions than he used, indicates that the range of possible values is not large, at least for small values of γ (for which control is likely to be most beneficial).

The models described here make several simplifying assumptions in order to make the analysis relatively tractable. For this reason they should be viewed primarily as screening models, whose purpose is to identify situations in which decisions are relatively clear-cut and to distinguish those situations for which further analysis is likely to be required. While measures of benefits (reduction in total delay) can be derived from the models presented here, those estimates are likely to be less useful than the identification of regions of potential benefits because the models omit several important factors. More detailed simulation studies of selected situations have been reported by Bly and Jackson (4), Koffman (5), and Turnquist and Bowman (6). Such models incorporate considerably more detail about specific routes and can be used to estimate actual benefits from control much more precisely.

SUMMARY AND CONCLUSIONS

The major point of the analysis in this paper is that basic and important decisions regarding headway control can often be made by using only limited statistics about system operation. The essential data on which fundamental decisions can be based are the coefficient of variation in the headway distribution and the relative proportions of passengers who are on board the vehicle (and will be delayed) and passengers who have yet to board (and will benefit from reduced wait time). By using this rudimentary information, a transit operator can make preliminary decisions regarding whether or not headway control is likely to produce benefits or whether further analysis is required.

The models are based on the assumption that passengers arrive randomly through time at bus stops; therefore, use of the results should be limited to situations for which that is likely to be true. In most cases, this means that average headways should be 10 min or less. For routes on which average headways are longer than 10 min, an analysis that includes a more sophisticated representation of passenger arrivals is necessary. Examples of such passenger arrival models are discussed by Jolliffe and Hutchinson (7), Turnquist (8), and Turnquist and Bowman (6).

This analysis has had nothing to say regarding the costs of implementing the control strategy. Our objective has been simply to illustrate situations in which headway control is likely to produce positive benefits. Of course, the decision to implement such a control system would involve evaluation of the costs as well as the benefits. We have described a rather general concept of a headway control strategy. Details of the implementation of such a strategy are likely to vary greatly from property to property, and hence the cost of implementation is likely to vary greatly as well. The transit operator may be able to generate relatively good cost estimates for a particular system but is likely to be much more uncertain regarding the potential benefits of the controls. The analysis in this paper should provide useful information in that regard.

ACKNOWLEDGMENT

This work was supported in part by a contract from the Office of University Research of the U.S. Department of Transportation.

REFERENCES

1. M. Abkowitz, H. Slavin, R. Waksman, L. Englisher, and N. Wilson. Transit Service Reliability. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Rept. UMTA-MA-06-0049-78-1, Dec. 1978.
2. P. I. Welding. The Instability of Close Interval Service. Operational Research Quarterly, Vol. 8, No. 3, 1957, pp. 133-148.
3. A. Barnett. On Controlling Randomness in Transit Operations. Transportation Science, Vol. 8, No. 2, 1974, pp. 102-116.
4. P. H. Bly and R. L. Jackson. Evaluation of Bus Control Strategies by Simulation. Department of the Environment, Transport and Road Research Laboratory, Crowthorne, Great Britain, Rept. 637, 1974.
5. D. Koffman. A Simulation Study of Alternative Real-Time Bus Headway Control Strategies. TRB, Transportation Research Record 663, 1978, pp. 41-46.
6. M. A. Turnquist and L. A. Bowman. Control of Service Reliability in Transit Networks. Office of University Research, U.S. Department of Transportation, Rept. DOT-RSPA-DPB-50-79-5, March 1979.
7. J. K. Jolliffe and T. P. Hutchinson. A Behavioral Explanation of the Association Between Bus and Passenger Arrivals at a Bus Stop. Transportation Science, Vol. 9, No. 4, 1975, pp. 248-282.
8. M. A. Turnquist. A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Time. TRB, Transportation Research Record 663, 1978, pp. 70-73.

Publication of this paper sponsored by Committee on Transit Service Characteristics.