Subway Reliability and the Odds of Getting There on Time

GARY HENDERSON, HEBA ADKINS, AND PHILIP KWONG

The most common service reliability measure—on-time performance—has drawbacks transit managers and consumers should be aware of. By converting on-time performance to odds ratios, a fresh look at what is meant by transit reliability is obtained. Although differences in on-time performance may seem slight, measurement with odds ratios can indicate quite different service levels. Transit managers, if they rely exclusively on on-time performance and do not heed the lesson to be learned from odds ratios, may be deceived when they estimate how much improvement can be achieved with a particular policy or action. The magnitude of effects will vary with level of effort, depending on the initial level of reliability.

Herein, reliability indicates timeliness (especially travel time), not the reliability of the rolling stock, infrastructure, or other separate components. In general, regularity measures are most useful in defining the reliability of high-frequency service, but the reliability of travel time is also a vital dimension of service. Timeliness of the passenger's trip is the end result of all the material and organizational aspects of transit service delivery; it is the consumer's experience at the point-of-service. Perhaps the most common performance measure in the transit industry for estimating timeliness is on-time performance. On-time performance is a useful and seemingly straightforward mathematical concept, but in some ways it distorts the perception of transit reliability. Reliability measurement can also be done by transforming on-time performance into odds ratios, the odds of being on time. Adopting this concept of the gambling industry can reveal aspects of transit reliability relevant both to transit operators and to passengers.

ODDS OF BEING ON TIME

A statistician thinks of on-time performance as on-time probability. The percent of trains arriving on time is another way of saying the probability that a given train will arrive on time. Although trains can be late by varying amounts, there are no different shades of being on time. Once a standard is chosen, e.g., 5 min, every arrival, on the basis of this cutoff, can be classified as being late or being on time. On-time probability is a yes or no proposition, just as flipping a coin is heads or tails. A value of such a binary variable can be calculated as a ratio of successes to failures (of on-time to late trains). This value is known as an odds ratio.

For example, if 220 of 250 trains arrive on time, the on-time probability is (220/250 = 0.88, or 88 percent). The on-time odds are the ratio of on-time trains to late trains, 220/30 or 7.3 to 1. This ratio is the same as that of the on-time probability to the late probability (0.88/0.12 = 7.3). In mathematical terms, if P is the probability of being on time and $1 - P$ the probability of being late, the odds of being on time are $P/(1 - P)$.

On-time performance increases linearly, whereas the odds of being on time increase geometrically. If there are 250 trains, every on-time train contributes 0.4 percentage point to on-time performance. So, if instead of 220 on-time trains, an operating adjustment makes it possible for 240 trains to arrive on time, on-time performance increases 8 percentage points (20×0.4), from 88 percent to a new level of 96 percent. On-time performance improved by 8/88, about 9 percent. The odds, however, increased from 7.3:1 to 24:1 (240/10). Reliability as measured by the odds improved by about 230 percent [(24 − 7.3)/7.3].

Figure 1 shows the relationship between on-time probabilities and odds. The line starts at 50 percent, at which the odds are 1:1. Below 50 percent on time, it is more appropriate to speak of the odds of being late, which is an identical curve, but inverted in the range from negative infinity to negative one. The rate of change (as measured by the slope of a line drawn tangent to the curve) around 60 percent on time is smaller than the rate of change at 85 percent. The odds are increasing slowly at 60 percent on time; but as on-time performance steadily improves to over 85 percent, the rate of change increases sharply.

In fact, probabilities and odds say the same thing in different ways; for the statistician, probabilities are usually the expression of choice. However, in describing transit reliability as passengers experience it, the odds ratio may be the measure that is in best accord with rider psychology.

Furthermore, when transit managers determine what actions or investment of effort or resources will improve on-time performance and wish to measure the size of the effects these actions will achieve, they prefer to consider odds ratios. This process also requires statisticians to determine the best technique for measuring the strength of causal factors in improving reliability.

PASSENGERS' VIEW

The average passenger is not likely to conceive of the likelihood of arriving on time in the mathematically abstract terms of probability. The odds of being on time are more in harmony with the concrete experience of riders. Recently, the New York State Office of the Inspector General for the Metro-

New York State Office of the Inspector General, Metropolitan Transportation Authority, 100 Park Avenue, 14th Floor, New York, N.Y. 10017.

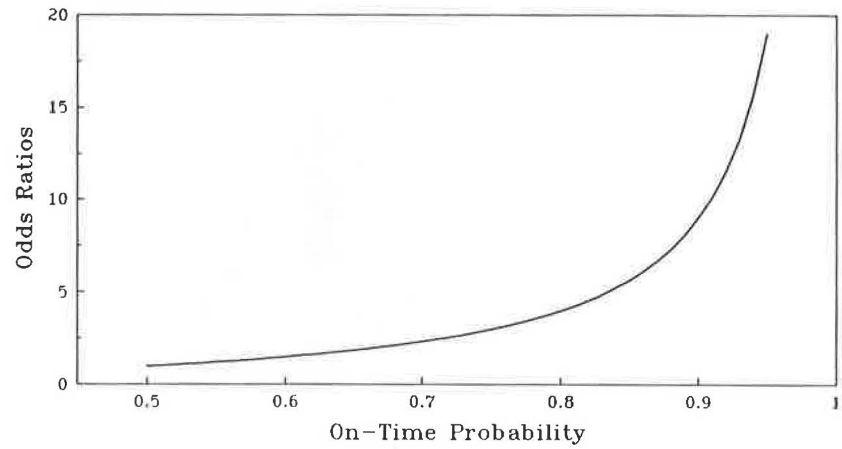

FIGURE 1 Odds ratios compared with probabilities. [Source: N.Y. State Office of the Inspector General for the MTA (Service Review Unit).]

politan Transportation Authority (MTA) monitored waiting times for the Brooklyn B35 and B46 bus routes (*1*). *The New York Times* interviewed passengers and reported that riders thought of catching the B46 bus as the "lotto wait." The "chances of a (B46) bus . . . running on time are about the same as winning the lottery" (*2*).

Odds ratios are represented in the most basic terms: how many days is the rider on time compared with how many days the rider is late. For example, a rider may recognize that he or she is late, say, once a week, which translates to being on time 4 days for every 1 day late, an odds ratio of 4:1. If on-time performance is 60 percent, passengers are on time 3 days and late 2 days in a 5-day week, a 3-to-2 chance of being on time. At 80 percent on time, riders are late 1 day a week. At 90 percent on time, 1 day in 2 weeks (more than double the reliability of the 80 percent level). Above 90 percent on time, a change in on-time performance has an enormous effect on the odds. For example, at 95 percent on time, riders are late once in 4 weeks. At 96 percent on time, riders are late once in 5 weeks; at 97 percent, once in 6½ weeks; and at 99 percent on time, only once in 100 work days, a little more often than twice per year. At such high levels of on-time performance, passengers are late so infrequently that they may not even remember the last time they were late.

Some examples from a report by the inspector general for New York's MTA will illustrate the situation (*3*). Evaluating 1988 morning rush hour subway service, the inspector general's report found that the D/Q Manhattan-bound route was 58 percent on time, 9 percentage points higher than in 1987, whereas the L Manhattan-bound route was 80 percent on time in 1988, 9 percentage points lower than in 1987. Although the magnitude of change of the on-time performance of the two lines was the same in terms of percentage points, the difference in the change in reliability was great. Figures 2 and 3 show that the change in the odds of being on time for the D/Q line was minimal—from 1 to 1.4—whereas the reliability of the L line declined by more than 50 percent—from 8.3 to 4.1.

A number of interesting conclusions can be made from the foregoing discussion. First, the difference in the quality of service between a transit service that is 70 percent on time (six times late in 4 weeks) and one that is 95 percent on time (once late in 4 weeks) is extremely large. Reliability for the latter is eight times better, because the odds at 70 percent on time are 7:3 whereas the odds at 95 percent are 19:1, or 57:3. Another implication is that when on-time performance is at a mediocre level, a large percentage point improvement is needed to make an impression on passengers. When on-time

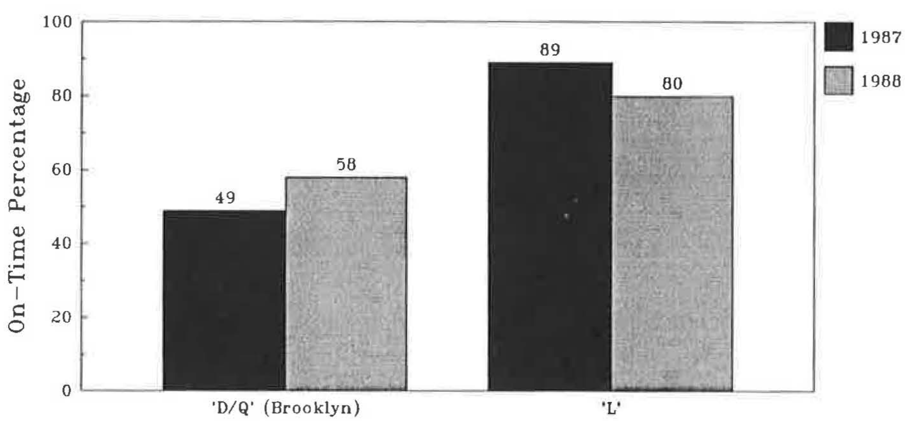

FIGURE 2 On-time percentage for the D/Q and L lines. [Source: N.Y. State Office of the Inspector General for the MTA (Service Review Unit).]

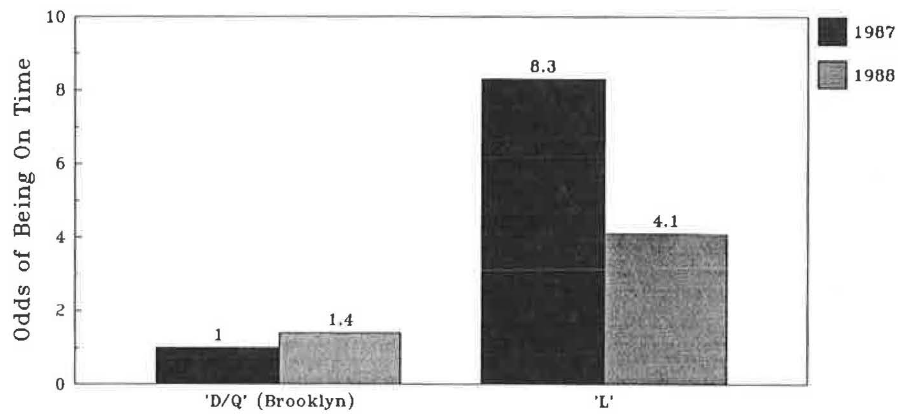

FIGURE 3 Odds ratios for the D/Q and L lines. [Source: N.Y. State Office of the Inspector General for the MTA (Service Review Unit).]

performance is high, small percentage point improvements in on-time performance have a great impact. However, as will be seen, the higher the level of reliability, the more resources are needed to improve on-time percentages.

TRANSIT MANAGER'S VIEW

It has been noted that small changes in on-time performance are more noticeable to passengers at higher levels of performance. The improvement from 95 percent to 98 percent—for which the odds of being on time increase 2½ times from 19:1 to 49:1—requires a level of effort (in terms of investment, operating costs, or degree of change in service configuration) similar to that needed to improve on-time performance from 60 to 79 percent—for which the odds are 1.5:1 and 3.8:1, respectively. Just as a change in on-time performance from 95 to 98 percent is more significant for passengers than a change from 60 to 63 percent, this improvement requires significantly more resources to achieve, all other things being equal. Lines operating at the 95 percent level of on-time performance have few major equipment, infrastructure, or crew problems; lateness of trains is caused by much more subtle problems or by random occurrences. However, at the lower levels of reliability, an improvement in a major area would affect a large number of trains and significantly increase reliability.

This fact has implications for any operational program for improving reliability. Clearly, the greatest payoff can be realized by working on the worst lines (as long as they are not too bad). If reliability is considered the result of a variety of policies or operating inputs, the same effect may not result from the same level of effort. Using the jargon of the statistician, if reliability is the dependent variable in a causal model with a variety of other independent variables (e.g., mechanical reliability, track condition, service configurations, headways, and passenger loads), the relationship of reliability to each of the many causal factors is not a linear one.

The statistical reason for this is that on-time probabilities are not continuous variables. They can increase or decrease only within a restricted range—from 0 to 1. In order to use on-time performance as a continuous variable, it must first be transformed so it will have an infinite range, as with the technique of logistic regression (*4*). To ensure that there is no upper bound to restrict improvement in performance, on-time probabilities are converted to odds ratios. Odds ratios stretch from zero to positive infinity. To ensure there is no lower bound, the odds ratios are transformed into logits by taking their natural logarithm.

$$\text{logit} = \log_e \frac{P}{1 - P}$$

Logits extend from negative infinity to positive infinity. When the odds are exactly even—at 50 percent on time—the logit of on-time performance is 0. As shown in Figure 4, below 50 percent on time the logit is negative; above 50 percent on time, it is positive. The negative sign is appropriate, because below 50 percent on time, the likelihood is greater that riders will be late than on time. Figure 4 shows the nonlinear relationship between on-time performance and logits of on-time performance. The changing slope of that curve indicates the changing degree of difficulty to improve reliability at different levels of reliability.

In order to demonstrate how such a model would work, the results from a preliminary causal model of subway performance can be examined using logistic regression. This model uses a few key variables like headway, mechanical reliability, merges, number of stops, etc., not by any means a comprehensive set. Because the coefficients from the techniques of logistic regression are different, depending on the starting condition, it was estimated that an increase in headway of 1 min would improve on-time performance by almost 5 percentage points for a route at 50 percent on time, but only 1 percentage point for a route at 95 percent on time. Similarly, the addition of one merging route would decrease on-time performance 14 percentage points for a route at 50 percent on time but only 2.5 percentage points for a route at 95 percent on time. These coefficients are given not as valid estimates of the effects of these actions but rather to demonstrate the variability of the effect. A study to estimate the strength of such factors is in progress, and this effort will include a more comprehensive set of variables and a larger sample of data.

Figure 4 also shows that when performance is below 10 percent on time, a huge investment would be required to

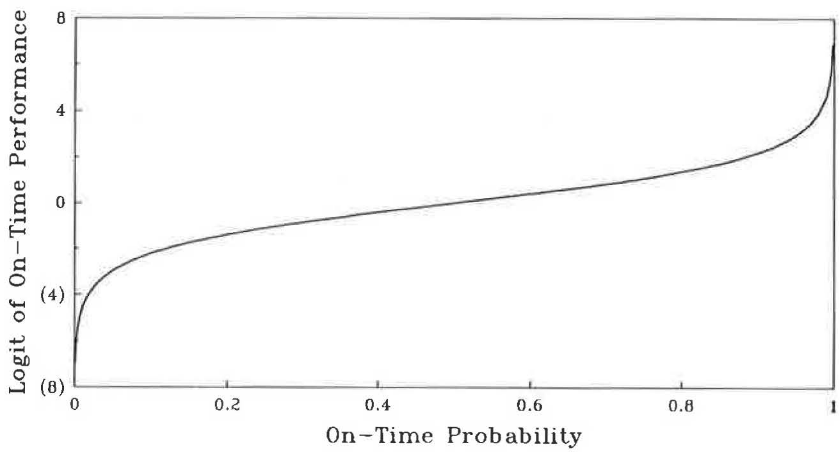

FIGURE 4 Logit curve for on-time performance. [Source: N.Y. State Office of the Inspector General for the MTA (Service Review Unit).]

improve performance. How could performance be so poor unless the infrastructure is in total disrepair, the rolling stock completely antiquated, or the workers utterly demoralized? When it is 50 percent on time, the slope is about level, suggesting that the greatest impact on reliability for a given level of effort can be achieved. Above 90 percent on time, the reliability logit rises steeply, indicating that additional investment will have diminishing returns.

The shape of this curve also has implications for performance goals for transit services. When performance is in the middle range (20 to 80 percent on time), large improvement in the on-time percentage is possible with the least amount of change in effort. When performance is above 90 percent, small percentage increments in on-time performance are significant and only small improvement may be possible. Therefore, operators need to set different targets for on-time performance for different lines depending on the operating conditions and resources available for each line. Performance above 95 percent may be an unrealistic goal for most routes, especially where merges and crowding are factors.

CONCLUSION

Service reliability is most commonly defined by the on-time performance of an operation. However, on-time probabilities hide the 1,611 effects of service performance on passengers. On-time performance percentages are less concrete and expressed less in daily real-life terms than ratios of how often passengers are late versus how often they are on time. By transforming on-time probabilities into odds ratios, transit operators can more effectively measure the reliability of the service delivered from the passengers' point of view. Odds ratios are more intuitively meaningful to passengers than on-time probabilities, as well as being operationally relevant to operating agencies. Using odds ratios, operators can determine how to allocate their resources to achieve the greatest payoffs.

REFERENCES

1. *A Review of Midday Performance for the B35 and B46 Brooklyn Bus Routes*. Report 90–4, New York State Inspector General for the Metropolitan Transit Authority, June 19, 1990.
2. Playing the B46 Lotto: Will The Bus Show Up? *New York Times*. July 25, 1990.
3. *The Inspector General's Evaluation of Morning Rush Hour Subway Performance, 1987–1988*. Report 90–11, New York State Inspector General for the Metropolitan Transit Authority, Aug. 14, 1990.
4. J. H. Aldrich and F. D. Nelson. *Linear Probability, Logit, and Probit Models*. Sage Publications, Beverly Hills, Calif., 1984.

Publication of this paper sponsored by Committee on Transit Management and Performance.