Sampling and Estimation Techniques for Estimating Bus System Passenger-Miles

PETER G. FURTH

Civil & Environmental Engineering, 400SN
Northeastern University
360 Huntington Avenue
Boston, MA 02115
Email: p.furth@neu.edu

ABSTRACT

Most U.S. bus systems conduct on-off counts on a sample of vehicle-trips to estimate annual passenger-miles, which must be submitted to the National Transit Database. The required sample size depends on the techniques used. This paper reviews alternative methods, including simple random sampling, ratio estimation with a variety of possible auxiliary variables, stratified sampling, cluster sampling, and combinations of these approaches. Most of these alternatives take advantage of electronic registering fareboxes to obtain complete counts of boarding passengers.

Seven alternative estimation techniques are compared in a case study of Santa Cruz Metro. The most efficient approach combined two techniques, stratified sampling and ratio estimation using the combined ratio technique. The latter technique used on-off data from a sample of trips to estimate the ratio of passenger-miles to *potential passenger-miles,* a newly proposed auxiliary variable. This approach reduced the sampling burden by over 80% compared with both simple random sampling and a sampling method published by the Federal Transit Administration.

KEYWORDS: National Transit Database, passenger-miles, sampling.

INTRODUCTION

The Federal Transit Administration (FTA) requires that transit agencies benefiting from federal assistance report annual passenger-miles by mode to the National Transit Database (NTD).[1] Because transit agencies, unlike airlines, do not routinely capture passengers' origin-destination information, measuring passenger-miles is usually done at the level of a trip (i.e., a vehicle-trip), based on on-off counts made at each stop by an onboard surveyor called a checker. Because of the high labor cost involved, on-off counts are generally done on a sample of trips from which an estimate of annual passenger-miles is made. FTA specifies that passenger-miles estimates achieve ±10% precision at the 95% confidence level.

For the bus mode, FTA's precision requirement may be satisfied by following the sampling plan laid out in Circular 2710.1A (USDOT FTA 1990). This sampling plan, based on direct estimation of mean passenger-miles per trip from a random sample of at least 549 trips, is relatively burdensome, requiring roughly one full-time equivalent employee to conduct on-off counts. Alternatively, an agency may use a custom-made sampling and estimation plan, as long as it is applied with a sample size that achieves the specified precision level. By taking advantage of an agency's particular features—its size, route structure, and availability of data on other measures of passenger use that correlate strongly with passenger-miles—custom sampling plans can substantially reduce the sampling burden.

One particular development in the transit industry creates the possibility for more efficient sampling plans. It is the widespread adoption of electronic fareboxes, with which transit agencies can count boardings on every trip. Because boardings are correlated with passenger-miles, an alternative to directly estimating mean passenger-miles per trip from a sample of trips is to estimate the ratio between passenger-miles and boardings, and then expand this ratio by the annual boardings count.

Besides the ratio estimation techniques, a number of other sampling and estimation techniques can improve the precision of an annual passenger-miles estimate. This paper describes several approaches for estimating annual passenger-miles for bus systems, based on experiences developing sampling plans for more than 20 U.S. transit agencies. Numerical results for seven alternatives are compared in a case study of Santa Cruz Metro (California), a system with 47 routes that vary considerably in length and ridership. Alternative estimation techniques are shown to reduce the sampling burden by over 80% compared with Circular 2710.1A. The most efficient sampling approach is found to be one that combines stratified sampling with ratio estimation, estimating the ratio of passenger-miles to a newly proposed auxiliary variable called *potential passenger-miles,* defined for a trip as the product of (passenger boardings) * (route length).

This paper has four main sections. The first describes in more detail how passenger-miles are measured and why transit agencies are looking for more efficient sampling techniques. The second introduces the case study agency. The third describes alternative approaches to estimating passenger-miles, along with results from the case study. The final section compares the alternatives and offers conclusions.

SAMPLING ELEMENT AND THE COST OF MEASUREMENT

As mentioned previously, the measurement unit for passenger-miles is a trip, normally defined as the one-way movement of a vehicle from one terminal to another. To measure passenger-miles for a selected trip, one counts ons and offs, also called boardings and alightings, by stop. From the on-off count, passenger load on every interstop segment may be determined. Multiplying the load on each segment by segment length (a known quantity) yields the number of passenger-miles occurring on each segment, and summing over all segments yields passenger-miles for the trip. Note that with this measurement technique it is neither possible nor necessary to know the trip length of individual passengers. Also note that on-off counts also yield a measurement of trip-level boardings, so that paired measurement of boardings and passenger-miles is no more burdensome than measuring passenger-miles alone.

[1] See the NTD website at www.ntdprogram.com.

The industry norm is for on-off counts to be made by transit agency employees known as checkers. In all but the smallest transit agencies, it is too disruptive of operations for bus operators to make on-off counts. On-off counts can also be obtained using automatic passenger counters (APCs), but only a small number of U.S. transit agencies have APCs due to their high cost.

Anecdotal evidence indicates that the majority of U.S. transit agencies follow the sampling plan of Circular 2710.1A, checking 549 trips per year. Transit agencies find this sampling requirement a rather onerous burden. While sampled trips may last only 30 minutes on average, the checker time involved can run upwards of 2 hours per trip because of travel time to the start of the trip, slack time to ensure catching the right bus, and return time. Coordination and supervision are also difficult. When multiple trips per day are sampled, one may be early in the morning and another late at night; they may be separated by a large geographic distance; and two selected trips may be close enough in time that it is impossible for the same person to check both.

For the vast majority of U.S. transit agencies, then, sampling trips to estimate passenger-miles involves considerable labor cost, a cost that agencies are interested in reducing by employing more efficient estimation and sampling techniques. A recent National Academy of Sciences review of NTD legislation, while acknowledging that the federal requirement for reporting passenger-miles estimates is reasonable (because passenger-miles is part of the legislated formula for allocating federal funding), also acknowledges the burden of this type of sampling (Furth and McCollom 1987). It recommends the development of more efficient sampling plans, particularly plans that take advantage of boarding counts made using electronic fareboxes.

SANTA CRUZ METRO CASE STUDY

Santa Cruz Metro (SCM) affords an interesting case study of sampling methods to estimate annual bus passenger-miles. Its 47 bus routes include short and long local routes, long commuter routes, and one very long and heavily used express route along Highway 17 that crosses the mountains to San Jose. It is common for a route to follow several different routing patterns in the daily schedule, and many of its routes are loops.

Using electronic fareboxes, SCM counts all passenger boardings. Furthermore, its boardings counts are always associated with the route being served, making it possible to use estimation techniques based on route-level boardings.

Because sampling requirements are driven by weekday service, which typically accounts for 85% to 90% of passenger-miles, the case study analysis was confined to weekday service. Historic data available for analyzing estimation techniques was a single day's observation of every trip in the weekday schedule in fiscal year 2001. Each trip record indicates the trip's boardings and passenger-miles, as well as identifiers (route, date, and so forth).

In order to make the case study more representative of other U.S. transit systems, the Highway 17 route was omitted. Unlike SCM's other routes, that route operates mostly along a freeway using over-the-road coaches. Because nearly all its passengers travel express between Santa Cruz and San Jose, average passenger trip length is very high (about 30 miles) and has little variability, making it easy to estimate passenger-miles for this route. Because it accounts for 15% of SCM's passenger-miles, including the Highway 17 route would substantially distort the case study results from the perspective of representing a "typical" transit agency.

Table 1 provides a comparison of necessary sample sizes and other relevant statistics for SCM using alternative estimation approaches. Entries in the table are explained in the following sections of the paper as each estimation alternative is presented. The sample sizes used in these comparisons are not "finished." Their application would require accounting for weekends and the Highway 17 route, and would probably involve some rounding. In addition, all calculated sample sizes for SCM are inflated by 50% relative to the sample size formula given. This degree of oversampling is a reasonable precaution, because sample size calculations are based on historic data, and transit agencies typically use a recommended sample size for several years before recalibrating the sample size requirement using a more recent dataset. Including oversampling

TABLE 1 Comparison of Sampling Strategies

	Sampling and estimation alternative	Unstratified ratio estimation statistics: *cv* of pass-miles	*cv* of auxiliary variable	Correlation coeff between pass-miles and auxiliary variable	Overall performance: Unit *cv*	Necessary sample size (SS)	SS reduction vs. Circular
—	Circular 2710.1A				n.a.	549	
A	Direct sampling and estimation				0.95	522	5%
B	Ratio estimation; auxiliary variable = boardings	0.95	0.77	0.67	0.72	296	46%
C	Stratified ratio estimation (4 strata); auxiliary variable = boardings				0.43	117	79%
D	Combined ratio estimation; auxiliary variable = boardings				0.45	117	79%
E	Ratio estimation; auxiliary variable = (boardings * route length)	0.95	0.91	0.89	0.44	112	80%
F	Ratio estimation; auxiliary variable = (boardings * adjusted route length)	0.95	1.09	0.89	0.51	147	73%
G	Combined ratio estimation; auxiliary variable = (boardings * route length)				0.39	86	84%

KEY: *cv* = coefficient of variation.

also makes comparisons with Circular 2710.1A more "fair," because the sample sizes called for in that circular, being intended for application nationwide, include a certain degree of oversampling.

SAMPLING AND ESTIMATION APPROACHES

Simple Random Sampling (Alternative A)

The population is all of the trips operated by a transit agency in a year; let N be the population size (number of trips operated in a year). Let Y = passenger-miles at the trip level, so that y_i = passenger-miles on trip i. Let $\bar{Y}$ and S_y be the population mean and standard deviation of Y; then $cv_y = S_y/\bar{Y}$ is the coefficient of variation (cv) of Y. Let n be the sample size, that is, the number of trips observed by means of on-off counts. Finally, let $\hat{Y}_{total}$ be the estimate of annual systemwide passenger-miles, the ultimate quantity being estimated.

If the sample of observed trips is drawn at random from the population of trips operated over the year, $\bar{Y}$ can be estimated by the sample mean

$$\bar{y} = \frac{1}{n}\sum y_i \tag{1}$$

The relative standard error (*r.s.e.*) of $\bar{y}$ is

$$r.s.e. = \frac{cv_y}{\sqrt{n}} \tag{2}$$

For unbiased estimators, *r.s.e.* is the *cv* of the estimator, and $(r.s.e.)^2$ is the relative variance of the estimator. Because the population size N is generally large compared with the sample size, the finite population correction is ignored.

The estimate of total annual passenger-miles is found by direct expansion of the sample mean

$$\hat{Y}_{total} = N\bar{y} \tag{3}$$

Because N is a known constant, the relative standard error and precision of $\hat{Y}_{total}$ are the same as those of $\bar{y}$.

Precision at the 95% confidence level (*prec*) is given by

$$prec = 1.96(r.s.e.) \tag{4}$$

The necessary sample size to achieve a specified precision at the 95% confidence level is therefore given by

$$n = \left(\frac{1.96cv_y}{prec}\right)^2 \tag{5}$$

In our experience analyzing data from about 20 transit agencies, we have found the passenger-miles *cv* to almost always lie in the range 0.8 to 1.2, corresponding to a sampling requirement (without oversampling) of 250 to 550 trips. Two passenger-miles *cv* values already reported in the literature are 0.82 for greater Pittsburgh (Furth 1998) and 1.08 for greater Buffalo (Townes 2001). Only once have we encountered an agency with passenger-miles *cv* exceeding 1.2; the value of its *cv*, 1.3, would have required a sample size of 650 trips if the simple random sampling approach had been chosen.

SCM results for simple random sampling are shown in table 1 as alternative A. SCM's *cv* of passenger-miles was found to be 0.95. The corresponding necessary sample size, with 50% oversampling, was 522 trips.

Circular 2710.1A

The sampling plan in FTA Circular 2710.1A also uses the sample mean as an estimator. It varies slightly from random sampling because it uses a two stage sample, selecting n_1 days within the year in stage 1 and n_2 trips for each selected day in stage 2, with a resulting sample size of $n = n_1n_2$. The circular offers a family of combinations of n_1 and n_2. Choices at stage 1 are sampling every day (n_1 = 365), every other day (n_1 = 183), every third day (n_1 = 122), and so forth. The combination with the smallest sample size, which is preferred by most agencies that follow the circular, is to sample 3 trips every other day, for a sample size of 549 trips per year. This sampling requirement is based on analysis done in the late 1970s of data from two transit agencies and was first published in 1978 as Circular 2710.1.

The standard error of a sample mean obtained using two-stage sampling involves variances at the two stages (Cochran 1977). However, it turns out that for passenger-miles, between-day variance of mean passenger-miles per trip is negligible in comparison with between-trip variance within a day, and the latter is essentially the same as between-trip variance over the entire population of a year's trips. Therefore, compared with simple random sampling, no advantage is gained by deliberately using a two-stage sample of the type used by Circular 2710.1A.

For the same reasons, the precision obtained using the two-stage approach of Circular 2710.1A is essentially the same as what would be obtained using simple random sampling with the same sample size. The range of passenger-miles *cv's* reported earlier, therefore, confirms the reasonableness of the Circular 2710.1A sample size, in the sense that most transit agencies following its sampling plan will achieve the specified precision.

Ratio Estimation

Ratio estimation (Furth and McCollom 1987; Cochran 1977) is a sampling and estimation technique that takes advantage of available data on an auxiliary variable that is closely correlated to the variable of interest. In order to use ratio estimation, two conditions must be met: the annual total of the *auxiliary variable* must be known, and sampled trips must provide paired measurements of the variable of interest (passenger-miles) and the auxiliary variable. Auxiliary variables that have been used for passenger-miles estimation include boardings and revenue.

Let X be the name of the auxiliary variable at the trip level; for the sake of definiteness, let X be trip-level boardings. Its population mean and total, $\overline{X}$ and X_{total}, are assumed to be known. Sampling yields a set of n-paired observations (x_i, y_i). Let $\bar{x}$ be the mean of X from this sample. Also of interest are the statistics

S_x^2 variance of X, usually estimated using the sample variance s_x^2,

$cv_x = s_x / \bar{x}$ = estimated coefficient of variation of X,

r_{xy} = estimated correlation coefficient of X and Y.

Here, we are interested in estimating the ratio $R_{population} = \bar{Y} / \bar{X}$, which often has an intuitive meaning. When X represents boardings, this ratio is the average length of an unlinked passenger-trip, usually called average passenger trip length.

$R_{population}$ is estimated from the paired sample by statistic R, the ratio of sample means

$$R = \frac{\bar{y}}{\bar{x}} \tag{6}$$

The estimate of annual system total passenger-miles is then the product

$$\hat{Y}_{total} = X_{total} R \tag{7}$$

Because X_{total} is a known constant, R and $\hat{Y}_{total}$ have the same relative standard error and the same precision. The relative standard error of a ratio estimate is given by

$$r.s.e. = \frac{1}{\sqrt{n}} \sqrt{cv_x^2 + cv_y^2 - 2r_{xy} cv_x cv_y} \tag{8}$$

"Unit cv" as a Measure of Statistical Efficiency

Equation (8) can also be expressed in the form

$$r.s.e. = \frac{ucv}{\sqrt{n}} \tag{9}$$

where the ratio estimator's *ucv,* standing for unit *cv,* is given by

$$ucv = \sqrt{cv_x^2 + cv_y^2 - 2r_{xy} cv_x cv_y} \tag{10}$$

The concept of unit *cv* can also be applied to simple random sampling. Comparing equations (2) and (9), it is clear that, for simple random sampling, the unit *cv* is

$$ucv = cv_y \tag{11}$$

Unit *cv* is a convenient term, first proposed by Furth and McCollom (1987), for comparing the efficiency of estimation techniques. It summarizes the inherent variability in an estimation technique, because the relative variance of an estimate depends only on the unit *cv* and the sample size. By comparing unit *cv's* of various estimation techniques, we can readily see which one requires a greater sample size or yields the more precise estimate for a given sample size.

Using the concept of unit *cv,* a sample size formula that applies to all the estimation techniques presented in this paper is

$$n = \left(\frac{1.96ucv}{prec} \right)^2 \tag{12}$$

and the precision (at the 95% confidence level) obtained for a given sample size is

$$prec = \frac{1.96ucv}{\sqrt{n}} \tag{13}$$

With ratio estimation, bias can become a problem at low sample sizes (Cochran 1977). Equations (12) and (13) are only valid as long as the sample size is neither so small that bias becomes significant, nor so large that the finite population correction applies.

The efficiency of a ratio estimator depends strongly on the correlation coefficient, at the trip level, between the auxiliary variable and passenger-miles. Squaring equation (10) and rearranging, the square of the unit *cv* can be expressed as the sum of two terms

$$ucv^2 = (cv_x - cv_y)^2 + 2cv_x cv_y (1 - r_{xy}) \tag{14}$$

For the kinds of auxiliary variables normally considered when estimating passenger-miles, the second term dominates. Therefore, as a general tendency, the stronger the correlation between the auxiliary variable and passenger-miles, that is, the closer r_{xy} is to 1, the smaller the unit *cv* of the ratio and the more efficient the estimation technique.

Boardings as the Auxiliary Variable (Alternative B)

Since the sampling plan in Circular 2710.1A was first published, nearly all buses in the U.S. transit fleet have been equipped with electronic fareboxes. Besides counting revenue, electronic fareboxes can also be used to count passenger boardings, making it possible to acquire a complete, systemwide count of boardings. Because boardings are correlated with passenger-miles—trips with more boardings tend to have more passenger-miles—boardings can serve as a useful auxiliary variable for ratio estimation. As mentioned before, the ratio of passenger-miles to boardings is average passenger trip length. A study of Buffalo area data (Furth 1998) found the correlation of boardings to passenger-miles to be 0.59—not optimal, but enough to reduce the sample size requirement by 33% compared with direct estimation of the sample mean.

The approach of using boardings to help estimate passenger-miles can only be adopted by agencies that, like SCM, count all passenger boardings. While nearly every U.S. transit agency uses electronic fareboxes, they do not all get reliable boardings counts. Boarding counts using electronic fareboxes are partly automated and partly manual. Essentially, passengers who interact with the farebox by entering a standard fare or swiping a card through an attached magnetic card reader are registered automatically. To register passengers who do not have a standard farebox interaction (e.g., passengers using a nonmagnetic transfer or pass or those paying a reduced fare because they are seniors or pupils), bus operators have to push a button corresponding to the appropriate fare category. In many large cities, where bus operator duties are particularly demanding, the farebox is not always operated in a way that yields reliable counts of passenger boardings. Where this is the case, boardings cannot be used as an auxiliary variable to estimate passenger-miles. (The Chicago Transit Authority is a good example of a large city transit agency that gets reliable boardings counts using fareboxes. They use advanced fare-collection technology to maximize the fraction of passengers registered automatically and follow management practices that emphasize the need for operators to register remaining passenger boardings.)

Results for SCM are shown in table 1 as alternative B. The correlation coefficient between boardings and passenger-miles is 0.67. The resulting unit *cv* is 0.72; comparing it with the unit *cv* for simple random sampling (0.95), we can see how using boardings as an auxiliary variable reduces the variability inherent in the estimation technique. The necessary sample size, 296 vehicle-trips, represents a reduction of 43% compared with simple random sampling, and 46% compared with Circular 2710.1A.

Revenue as the Auxiliary Variable

The earliest applications of the ratio technique for passenger-miles estimation used farebox revenue as an auxiliary variable. Farebox revenue is correlated with passenger-miles (more revenue on a trip usually means more passengers and, therefore, more passenger-miles). Annual total revenue is certainly known. And, even before the invention of the electronic registering farebox, most transit agencies had mechanical registering fareboxes that allowed checkers making on-off counts to measure trip revenue by reading the revenue register at the start and end of the trip. With this approach, the ratio of passenger-miles per dollar of revenue can be estimated from a sample of trip observations and then expanded by annual revenue to yield an estimate of annual passenger-miles.

Furth and McCollom (1987) found a relatively strong correlation between revenue and passenger-miles using Pittsburgh area data from the early 1980s. Based on that analysis and a similar analysis of data from San Antonio, FTA published a revenue-based sampling and estimation method with a sample size requirement of only 208 trips (USDOT UMTA 1985). However, FTA later withdrew default approval for this sampling plan, because widespread adoption of monthly passes weakened the correlation of passenger-miles to farebox revenue. Agencies may still use this technique, but must justify the sample size they use by analyzing local data.

This technique was not tested as part of the SCM case study, because trip revenue data were not part of the available dataset. However, given the widespread use of passes at SCM, it is likely using revenue, because an auxiliary variable would be less efficient than using boardings.

Stratified Sampling

Stratification is another approach that can improve sampling and estimation efficiency. For passenger-miles estimation, stratification has been mostly used together with the ratio technique for estimating average passenger trip length. The goal of stratification is to divide the population of vehicle-trips in a way that passenger trip length varies as much as possible between strata, rather than within strata. Stratification is usually done by route (Huang and Smith 1993), because routes can differ widely in their average passenger trip length (typically, average passenger trip length is small on short routes and large on long routes). Three variations of stratification by route have been followed, as described below.

Each Route a Stratum

In the first variation of stratified sampling, each route is a stratum. A sample of trips is observed in each stratum, measuring both boardings and passenger-miles on each observed trip. From each sample, the average passenger trip length ratio for the stratum is estimated and then expanded to annual passenger-miles by multiplying by the stratum's annual boardings (assumed to be known). Those annual passenger-miles figures are then aggregated over all the strata to yield the systemwide, annual estimate of annual passenger-miles.

In order to apply this approach, an agency needs not only counts of all passenger boardings during the year, but the ability to break out those counts by route. Among those agencies that get a reliable count of boardings using electronic fareboxes, some are still unable to stratify by route because bus operators do not register (by pushing a sequence of buttons) every time the bus changes route, and so recorded counts cannot be associated with a particular route.

Stratum-level parameters and statistics are defined as follows. Let N_h and n_h be population size and sample size for stratum h, respectively, both measured in trips. The unsubscripted variables N and n retain their meaning as overall population size and sample size; that is,

$$\sum N_h = N$$

$$\sum n_h = n$$

The relative size of stratum h, in terms of population size, is given by

$$w_h = N_h / N \tag{15}$$

Relative size serves as a weighting factor, since

$$\sum w_h = 1 \tag{16}$$

Let $\overline{X}_h$, assumed to be known, be the mean boardings per trip within stratum h, and let $\bar{y}_h$ and $\bar{x}_h$ be the sample means of Y and X within stratum h. Finally, let s^2_{yh}, s^2_{xh}, and r_{xyh} be the sample variance of passenger-miles, the sample variance of boardings, and the sample correlation coefficient between passenger-miles and boardings, respectively, within stratum h.

The ratio estimated within each stratum is

$$R_h = \bar{y}_h / \bar{x}_h \tag{17}$$

The estimate of total annual systemwide passenger-miles involves expansion by stratum, followed by aggregation over strata:

$$\hat{Y}_{total} = \sum N_h \overline{X}_h R_h = N \sum w_h \overline{X}_h R_h \tag{18}$$

The estimate of average passenger-miles per trip is

$$\bar{y}_{strat} = \frac{\hat{Y}_{total}}{N} \tag{19}$$

Because these final two estimates differ by only the known factor N, they have the same relative standard error, and consequently the same unit cv and the same precision.

The variance of an estimate made using stratified sampling depends in part on how the sample is allocated among the strata. In this paper, allocation is assumed to be proportional to stratum size, that is, for a given n,

$$n_h = w_h n \tag{20}$$

Proportional allocation is not, in general, the optimal (i.e., variance-minimizing) way of allocating a sample among strata. However, for the range of parameters typically encountered in passenger-miles estimation, proportional allocation is not much inferior to optimal allocation in terms of variance, and it has other desirable properties including ease in determining sample size and making certain types of estimators self-weighting.

With proportional allocation, the relative standard error of the annual systemwide passenger-miles estimate is

$$r.s.e. = \sqrt{\frac{1}{n}} \left[\frac{\sqrt{\sum w_h \left(s^2_{yh} + R^2_h s^2_{xh} - 2 R_h r_{xyh} s_{xh} s_{yh} \right)}}{\bar{y}_{strat}} \right] \tag{21}$$

The term in brackets is the unit *cv* of the estimator. Precision (at the 95% confidence level) for a given sample size, the sample size necessary to achieve a given precision, can be determined using equations (12) and (13).

While route-level stratification is a compelling concept, it has one serious drawback. Ratio estimators are biased when sample size is small (Cochran 1977). An analysis of transit trip-level ridership data found that in order to effectively limit bias, at least 10 trips should be observed per stratum (Furth and McCollom 1987). For even a mid-sized transit agency, this limitation makes stratification by route of no practical value, limiting the approach to bus systems with a small number of routes. Therefore, stratification by route was rejected as a sampling and estimation approach for SCM.

Stratification by Route Length (Alternative C)

One way to overcome the limitation of a minimum stratum sample size is to use a coarser stratification scheme, grouping trips into strata by route length. Correlation of boardings to passenger-miles can still be expected to be a good deal stronger within a stratum than systemwide, albeit not as strong as if each route were a stratum.

In the SCM case study, routes were grouped into four strata by length. Table 2 presents relevant statistics. Stratum 4 contained SCM's long express routes (but not the excluded Highway 17 route) and accounted for about 2% of the daily vehicle-trips; the other three strata, roughly equal in size, corresponded to short, medium, and longer routes. Within each stratum, correlation of boardings to passenger-miles (at the vehicle-trip level) was rather strong, with correlation coefficients ranging from 0.79 to 0.89. Of particular interest are the average passenger trip length ratios for the four strata: 2.8, 3.2, 7.8, and 10.5 miles, respectively. The large differences of the last two strata from the first two show the benefit of separating them into different strata.

Overall results are shown in table 1 as alternative C. The unit *cv* was 0.43, a large improvement over the previously described methods. The corresponding necessary sample size was calculated to be 109; constraining stratum 4 sample to at least 10 observations results in a required sample size of 117 vehicle-trips.

Combined Ratio Estimation (Alternative D)

A third approach to stratified sampling is to use the so-called combined ratio estimation technique (Furth 1998; Cochran 1977). It uses stratified sampling to select the trips that are observed, but then uses that data to estimate a single, systemwide ratio using the equation

$$R = \frac{\sum w_h \bar{y}_h}{\sum w_h \bar{x}_h} \tag{22}$$

Using a systemwide ratio is a disadvantage relative to conventional stratified ratio estimation (i.e., a ratio estimated for each stratum), weakening the correlation of boardings to passenger-miles. However, the method also offers two advantages. First, it

TABLE 2 Stratified Ratio Sampling Statistics

	Stratum				
	1	2	3	4	All
Stratum weight	37%	31%	31%	2%	100%
Mean passenger-miles per trip	41	123	237	197	129
cv of passenger-miles	0.68	0.63	0.62	0.45	
Mean of boardings per trip	15	39	30	19	27
cv of boardings	0.56	0.71	0.54	0.46	
Ratio of passenger-miles to boardings (average passenger trip length)	2.81	3.21	7.78	10.54	
Boardings to passenger-miles correlation coefficient	0.89	0.88	0.79	0.79	
Unit *cv* of the ratio	0.31	0.34	0.38	0.30	
Contribution to the sum in equation (21) (relative contribution to variance of the systemwide estimate)	60	549	2, 479	57	3, 144

is unbiased regardless of stratum sample size, and, therefore, permits every route to be a stratum. Second, it requires only knowledge of systemwide, not route-level, boardings, and, therefore, can be applied by transit agencies that routinely count all passenger-boardings, even if they cannot break out the counts by route.

In this technique, on-off counts are made for one or more trips on each route, providing paired observations of y (passenger-miles) and x (boardings), from which the combined ratio is calculated using equation (22). Allocation of the sample between strata (i.e., between routes) is again proportional to size. Relative standard error is estimated by

$$r.s.e. = \sqrt{\frac{1}{n}}\left[\frac{\sqrt{\sum w_h\left(s_{yh}^2 + Rs_{xh}^2 - 2Rr_{xyh}s_{xh}s_{yh}\right)}}{\bar{y}_{combined}}\right] \quad (23)$$

where

$$\bar{y}_{combined} = \bar{X}R \quad (24)$$

is the estimated mean passenger-miles per trip. Again, the quantity in brackets in equation (23) is the unit *cv* of the estimator.

The only previously published report that uses this technique for passenger-miles estimation found it to be very efficient. When applied to the eight-route transit system of Kenosha, Wisconsin, it called for a sample size of fewer that 50 vehicle-trips (Furth 1998). However, when SCM used this technique it was not as efficient. As shown in table 1, under alternative D, the unit *cv* (0.45) and the necessary sample size (117) are virtually the same as obtained for alternative C, conventional stratified ratio estimation.

Closer examination of the differences between conventional stratified ratio estimation and combined ratio estimation helps explain why the combined method did not perform as well at SCM as in Kenosha. Equation (23) is the same as equation (21), except that the former uses the combined ratio in place of stratum-specific ratios. In both formulas, the sum in the numerator represents the expected squared difference between observed and predicted passenger-miles. For conventional stratified ratio estimation, this difference for a paired observation (y_{ih}, x_{ih}) is $(y_{ih} - R_h x_{ih})$, while with combined ratio estimation the difference is $(y_{ih} - Rx_{ih})$. Naturally, differences tend to be smaller when using a stratum-specific ratio; the degree to which this factor hurts the performance of the combined ratio technique depends on how much average passenger trip length varies between routes. Because average passenger trip length is closely related to route length, one would expect the technique to be more effective when route length varies little within the network.

Not surprisingly, Kenosha's transit system, like those of many small cities, uses pulse scheduling based around a transit center. In this kind of network, routes are all designed to have roughly the same length. At SCM, in contrast, routes vary considerably in length, and so average passenger trip length varies widely between routes. This explains why for SCM the combined ratio technique holds no advantage over conventional stratified ratio estimation for estimating average passenger trip length. This is a significant finding that most likely extends to other transit systems whose route lengths vary considerably from one another.

Using Potential Passenger-Miles as the Auxiliary Variable (Alternative E)

In an effort to improve sampling efficiency further, a new auxiliary variable is proposed: the product of boardings and route length, which can be called *potential passenger-miles*. This formulation is motivated by the observation that trip-level passenger-miles tend to be proportional to not only the number of passengers on the trip but also to the overall length of the route. The ratio of passenger-miles to potential passenger-miles has an intuitive interpretation: it is the average fraction of a route's length that passengers travel. For example, a ratio of 0.6 would indicate that on average, passengers travel 60% of the length of their chosen route.

This estimation approach requires the usual sample of on-off counts and knowledge of annual boardings by route. Because route length is a known constant, potential passenger-miles can be calculated for both the sample data and the annual totals by simply multiplying every boarding count by the length of the route on which the count was made. In the SCM case study, on routes with multiple routing patterns, "route length" was defined to be the length of the most often used pattern.

Mathematically, alternative E is simply ratio estimation, like alternative B, except that the auxiliary variable X is redefined to be potential passenger-miles. As indicated in table 1, alternative E, the correlation of passenger-miles with potential passenger-miles ($r_{xy} = 0.89$) turns out to be considerably stronger that the correlation with boardings alone ($r_{xy} = 0.67$ in alternative B); as a result, there is an impressive reduction in necessary sample size (from 296 to 112) when the auxiliary variable is changed from boardings to potential passenger-miles.

This result shows the value of the compound auxiliary variable (boardings * route length). However, as an overall approach, unstratified ratio estimation using this auxiliary variable still offers no substantial improvement to stratified ratio estimation using boardings as an auxiliary variable.

Using Adjusted Route Length to Calculate Potential Passenger-Miles (Alternative F)

Alternative F is the same as alternative E, except that in calculating potential passenger-miles, an adjusted measure of route length is used on loop routes. On loop routes—those that return to a main terminal by a substantially different path than the that taken when leaving that terminal—SCM defines route length as the length of the full loop rather than as the one-way distance between terminals. In alternative F, potential passenger-miles on loop routes were calculated using half the length of a loop as the route length.

It turns out that adjusting route length in this manner did not improve the correlation of potential passenger-miles with passenger-miles, as shown in table 1. Compared with alternative E, the correlation coefficient remained essentially unchanged while the *cv* of the auxiliary variable increased, resulting in an increased necessary sample size.

Combined Ratio Estimation Using Potential Passenger-Miles as the Auxiliary Variable (Alternative G)

The final alternative, alternative G, marries the two most efficient techniques found previously: ratio estimation using potential passenger-miles as the auxiliary variable, and stratification by route using the combined ratio estimation method.

Comparisons to this approach can be drawn against two other approaches: unstratified ratio estimation using potential passenger-miles as the auxiliary variable (alternative E), and combined ratio estimation using boardings as the auxiliary variable (alternative D). In alternative E, while both unstratified ratio estimation and combined ratio estimation involved a single, systemwide ratio, the stratification involved in the combined ratio method reduced inherent variability. In alternative D, the weakness was the large degree of variation in average passenger trip length between routes. When the auxiliary variable is potential passenger-miles, the ratio of interest becomes the fraction of route length covered by the average passenger trip, a ratio that does not vary nearly as much between routes.

Mathematically, alternative G is the same as alternative D, except that the auxiliary variable X represents potential passenger-miles. Again, we used proportional allocation between strata.

As indicated by table 1, alternative G turned out to be the most efficient, requiring a sample size of only 86. This represents a reduction of about 25% compared with alternatives D and E, confirming both the advantages of stratified sampling and of using combined ratio estimation with an auxiliary variable that varies little between routes.

Other Sampling Techniques

This overview of sampling techniques for estimating annual passenger-miles would not be complete without mentioning two other techniques that have been found to offer advantages.

Sampling Round Trips

The cost structure of on-off checks is such that it is almost always more efficient to sample round trips rather than independently selected trips: once a checker has surveyed a trip, the return trip can be sampled at nearly no additional cost because the checker usually has to be paid anyway to return to his or her starting point. Sampling round trips is an instance of cluster sampling, that is, selecting predefined clusters of, in this case, two trips for observation.

At transit agencies with labor agreements requiring eight-hour assignments for checkers, clusters lasting three to four hours are preferred, so that a checker can be assigned to one cluster in the morning and another in the afternoon. A cluster of this length is typically a chain of four, six, or eight trips performed by a single vehicle. The larger the cluster, the smaller the per-trip overhead related to getting to the start of the trip, supervision, and returning from the sampled trip.

However, when clusters tend to be homogeneous (which is certainly the case in this application, since the trips performed in a chain by a single vehicle are usually on the same route and take place during the same general time of day), variance per observed trip will be greater with cluster sampling than if trips are sampled independently (Cochran 1977). Therefore, the number of trips that would have to be observed to achieve a given precision using cluster sampling is greater than if trips are sampled independently. The *cluster effect* is defined as the ratio between these necessary sample sizes:

$$cluster\ effect = \frac{n_{cluster} * cluster\ size}{n_{SRS}} \qquad (25)$$

where n_{SRS} = necessary sample size in elementary units (e.g., one-way trips) using simple random sampling,

cluster size = number of elementary units per cluster, and

$n_{cluster}$ = necessary sample size (number of clusters) with cluster sampling.

In the literature (Cochran 1977), the cluster effect has been called Kish's d_{eff}, where d_{eff} stands for design effect.

The cluster effect can be used to convert a necessary sample size, obtained using a formula for simple random sampling, into a necessary sample size in units of clusters:

$$n_{cluster} = n_{SRS}\left(\frac{cluster\ effect}{cluster\ size}\right) \qquad (26)$$

A study of Los Angeles data (Furth et al. 1988) found that when sampling clusters of four trips to estimate the ratio of boardings to farebox revenue, the cluster effect was 2.2. Therefore, the number of four-trip clusters that would have to be observed is (2.2/4) = 55% as great as the number of one-way trips that would have to be observed if one-way trips were selected independently. Whether cluster sampling is cost-effective depends on if it is less expensive to do on-off checks on n trips selected independently or 0.55 n clusters of four trips.

A study of Madison, Wisconsin, data found that while sampling in round trip clusters was effective because of the small marginal cost of checking a return trip, sampling in larger clusters was not (Huang and Smith 1993). Our experience in analyzing cluster data for passenger-miles estimation from Dayton, Ohio, and Pittsburgh, Pennsylvania, confirms this finding. The larger the cluster, the greater the cluster effect, making clusters larger than a round trip rather ineffective as sampling units. Once a checker has measured passenger activity on a single round trip, little further information can be gained by sampling the next round trip operated by the same vehicle, since it will normally be operating on the same route and at the same period of the day. Large clusters are, therefore, recommended only when labor rules make it such that it costs little more to check multiple round trips than to check a single round trip.

To determine a sample size requirement using round trip clusters, it is often necessary to guess the magnitude of the cluster effect, because cluster data are rarely available for direct analysis. Experience suggests that a conservative estimate of the cluster effect is 1.5 for round trips. Using that value, equation (26) indicates that the number of round trips that would have to be sampled is 75% as great as the number of one-way trips that would have to be sampled using independent sampling. Therefore, if the cost of checking a round trip is no more than the cost of checking a one-way trip, sampling by round trip can reduce cost by 25% compared with sampling trips independently.

Two-Stage Sampling

In very small transit systems, the number of trips sampled over the year can approach the number of trips in the daily schedule. Sometimes, even in larger systems, the transit agency has a policy of checking every trip in the daily schedule once per year. If the finite population correction is accounted for, a two-stage design in which all (or most) of the trips in the daily schedule are observed eliminates all (or most) of the between-scheduled-trip variation (Cochran

1977). Because most of the variation in passenger-miles tends to be between scheduled trips (e.g., peak period versus offpeak) rather than between days for a given scheduled trip, such a two-stage approach can be quite efficient. This technique was demonstrated in a study done for the Los Angeles Blue Line light rail (Furth 1993) and has been applied in numerous bus systems as well.

COMPARISON AND CONCLUSION

Table 1 presents summary statistics for SCM comparing seven alternative sampling and estimation approaches using the Circular 2710.1A sampling plan. The key measure used to compare alternatives was the necessary sample size to meet the FTA precision criterion.

Circular 2710.1A, requiring a sample of 549 vehicle-trips, was the benchmark. This analysis found that it was a reasonable sample size to require, in the sense that passenger-miles variability at the trip level are such that, for most transit agencies, following it will achieve the FTA precision criterion. Alternative A used simple random sampling, where the only significant difference from the circular was that its sample size was based on a local *cv* of passenger-miles; for SCM, this alternative required almost as large a sample size (522) as Circular 2710.1A.

The remaining estimation techniques tested involved estimating a ratio between passenger-miles and an auxiliary variable where the annual total value is known. When boardings was the auxiliary variable, the ratio of interest was average passenger trip length. Compared with simple random sampling, this approach (alternative B) substantially improved efficiency, as the sampling requirement fell to 296.

When boardings were known by route, stratifying the population of trips by route length improved sampling efficiency, since average passenger trip length tended to vary systematically with route length, being greater on long routes and smaller on short routes. In alternative C, estimating the average passenger trip length ratio separately in four strata dropped the sampling need to only 117 trips. Making every route a stratum could further improve sampling efficiency; however, the constraint that ratio estimation be based on samples of at least 10 observations per stratum (in order to limit bias) makes it an impractical approach for an agency with a large number of routes.

The combined ratio method permits stratification by route in sample selection without concerns about bias, but it involves estimating a single, systemwide ratio. This technique, used to estimate the average passenger trip length ratio (alternative D), had the same sampling requirement as alternative C. It had the advantage that, unlike alternative C, it required only that an agency know annual system boardings, without requiring that annual boardings be known by route. The effectiveness of this technique was considerably greater for the transit agency in Kenosha, Wisconsin. Analysis of the technique suggests that its effectiveness will be greatest when the routes in a transit system vary little in length.

A new auxiliary variable was introduced, called potential passenger-miles, which is the product of boardings and route length. It performed better than boardings as an auxiliary variable. Without stratification (alternative E), it dropped the sampling requirement from 296 to 112; with stratification using combined ratio estimation (alternative G), it dropped the sampling requirement from 117 to 86. Attempts to use a modified definition of potential passenger-miles (alternative F) failed to improve efficiency. Alternative G, the most efficient approach, reduced the sampling burden by 84% compared with Circular 2710.1A. Conveniently, alternative G required only knowledge of system-level boardings, not route-level boardings.

While SCM's available dataset did not permit analysis of sampling trips in clusters, evidence from the literature and from unpublished studies indicates that sampling using round trip clusters improves cost-effectiveness, because a round trip cluster generally carries more statistical information than a single trip, while costing no more to observe due to the need of the checker to return to his or her starting point.

Evidence from only a few transit agencies is not sufficient to make a broad conclusion about the most efficient estimation and sampling method or sample size needed. Analysis of data from other transit agencies is needed to determine which results are transferable. Nevertheless, the results of this study show a promising direction for any transit agency considering ways to reduce its passenger-miles sampling burden.

ACKNOWLEDGMENT

The author would like to acknowledge the careful and helpful reviews of three anonymous referees, whose comments guided the paper's revision.

REFERENCES

Cochran, W.G. 1977. *Sampling Techniques,* 3rd ed. New York, NY: John Wiley and Sons, Inc.

Furth, P.G. 1993. Ridership Sampling for Barrier-Free Light Rail. *Transportation Research Record* 1402:90–97.

____. 1998. Innovative Sampling Plans for Estimating Transit Passenger-Miles. *Transportation Research Record* 1618:87–95.

Furth, P.G., K.L. Killough, and G.F. Ruprecht. 1988. Cluster Sampling Techniques for Estimating Transit System Patronage. *Transportation Research Record* 1165:105–114.

Furth, P.G. and B. McCollom. 1987. Using Conversion Factors to Lower Transit Data Collection Costs. *Transportation Research Record* 1144:1–6.

Huang, W.J. and R.L. Smith. 1993. Development of Cost-Effective Sampling Plans for Section 15 and Operational Planning Ride Checks: Case Study for Madison, Wisconsin. *Transportation Research Record* 1402:82–89.

Townes, M., Chair of National Academy of Sciences Committee for the National Transit Database Study. 2001. Letter report to H. Walker, Acting Administrator of the Federal Transit Administration. June 1.

U.S. Department of Transportation (USDOT), Federal Transit Administration (FTA). 1990. *Sampling Procedures for Obtaining Fixed-Route Bus Operation Data Required Under the Section 15 Reporting System,* FTA Circular 2710.1A. Washington, DC. Also available at www.ntdprogram.com.

U.S. Department of Transportation (USDOT), Urban Mass Transit Administration (UMTA). 1985. *Revenue Based Sampling Procedures for Obtaining Fixed Route Bus Operation Data Required Under the Section 15 Reporting System,* UMTA Circular 2710.4. Washington, DC.