A POISSON MODEL OF RURAL TRANSIT RIDERSHIP

Edward S. Neumann, and Bernard F. Byrne, West Virginia University

A Poisson Model for ridership on rural public transportation routes is developed. Models are tested on data collected previously in the research, and some modifications made. Illustrated is a technique of using analysis of variance on ridership rates to determine those which are significantly different, so as to form categories for cross-classifications which are not arbitrary.

One of the technical problems facing Rural Public Transportation Planning is the lack of suitable techniques for estimating route ridership. The U.S. Department of Transportation, Office of University Research has sponsored a project at West Virginia University to develop techniques of demand estimation which are compatible with the resource constraints faced by planners in rural areas (1). The more important requirements were that the models should be simple to understand, easy to apply, and be low cost in nature; that is, capable of utilizing existing sources of data, such as the census. At the same time, the methods were to be accurate enough for reasonable estimates of needed vehicle sizes, and potential revenues. Another desirable characteristic was that the models should offer the possibility of transferability from the areas where the model was calibrated to other areas. Yet another concern was that the models be capable of identifying the needs generated by specific target populations along routes, such as the elderly, carless, or households with low income.

The most comprehensive previous work in this area has emphasized the use of regression anlaysis to forecast route ridership as a linear function of aggregate route characteristics such as total population along routes and route length and destination population (4). The same approach was attempted in the first phase of the reported research, but without good results. Generally, these models have met the requirements of being easy to apply, and low cost in nature. The models have often taken an econometric form with a moderately complex underlying structure in which parameters similar to coefficients of elasticity govern the estimates, for example,

$$\frac{R}{P_o \cdot P_d} = A \cdot F^{B1} \cdot D_R^{B2} \qquad (1)$$

where

R - Route passengers/day
P_o - Origin Population
P_d - Destination Population
F - Frequency - vehicle trips/day
D_R - Round Trip distance
A,B1,B2 - coefficients determined by regression.

The sensitivity of the predictions to errors in parameter estimates with this type of model can be high, which means that the application of the model may be limited to the range of values used in calibration and to the area represented by the values. The route ridership approach based on linear regression models has been found to produce large forecasting errors when applied to other areas of the nation (5). Due to the structure of the models, however, the reasons for the error cannot be easily identified and corrected, in as much as the models involve a set of simultaneously determined exponential coefficients. A second drawback of the linear regression models has been a lack of sensitivity to the causal socioeconomic variables which are usually hypothesized to create a need for transit. Though aggregate route variables such as number of households with incomes below $3,000 have been tested as independent variables in regression analyses, they have seldom shown consistent, significant correlations with ridership and have not appeared in the final models. The authors concluded that a different modelling approach was needed.

First it intuitively appeared that probabilistic models were more appropriate than linear causal models. It was observed that one of the major problems with the route regression approach was that the model linear mathematical structure was better suited to the kind of relationship in which transit ridership along a route could be represented as a total magnitude that every person along the route contributed to in an extremely small but consistent amount. In reality rural transit ridership represented a small number of discrete daily occurences from among a large sample space of people. Thus, a probability model structure which could predict the probable range of a small number of riders, was conceptually more appropriate than a linear model structure which failed to recognize the probabilistic aspects of demand. Given this, a second observation followed: it was realistic to depict ridership generation as a phenomenon that occured at the level of the household, where the decision to use transit was influenced by the socioeconomic characteristics of the household. Thus, some form of cross-classification approach was desirable, in which types of households were grouped on the basis of similar ridership rates. Combining the two observations resulted in the concept of a Poisson Model of Rural Transit Ridership which utilized a household level cross-classification of trip rates to establish a probable range of ridership along a route.

The Model

A poisson Model of Rural Transit Ridership may be developed in the following way: let us assume that we have a typical rural transit route that leaves a central city, goes out through the countryside in a loose loop and returns to the city, picking up people with no drop-offs. Let us suppose that we can dimension the loop in some sense, with N being a linear dimension describing the loop, e.g., distance, population, number of autoless, etc. Let the probability of having n people on the bus be $P_n(N)$. Let the probability of picking up one person in a short length of the route, ΔN, be $p\Delta N$. If ΔN is sufficiently short, then the probability of picking up two riders is negligible, so that the probability of picking up no riders is $1 - p\Delta N$. At the point $N + \Delta N$, the probability of having n people on the bus is the sum of two probabilities:

1. The probability of n-1 people on the bus at N times the probability of picking up someone in ΔN.
2. The probability of n people on the bus at N times the probability of picking up nobody in ΔN.

In other words,

$$P_n(N + \Delta N) = P_{n-1}(N)p\Delta N + P_n(N)(1 - p\Delta N) \quad (2)$$

This can be rearranged to read

$$\frac{P_n(N + \Delta N)-P_n(N)}{\Delta N} = p[P_{n-1}(N) - P_n(N)]. \quad (3)$$

letting $\Delta N \rightarrow o$ in the limit, we obtain

$$\frac{dP_n(N)}{dN} = p[P_{n-1}(N) - P_n(N)]. \quad (4)$$

To solve this equation, let us start by solving the situation in which the bus is empty and remains empty. In this case 4 simplifies to

$$\frac{dP_o(N)}{dN} = -pP_o(N) \quad (5)$$

Since $P_o(0) = 1$, the solution to 5 is

$$P_o(N) = e^{-pN} \quad (6)$$

For n = 1, 6 can be substituted into 4 to yield

$$\frac{dP_1(N)}{dN} = p[P_o(N) - P_1(N)]. \quad (7)$$

This can be solved to yield

$$p_1(N) = (pN)e^{-pN} \quad (8)$$

And, in general

$$P_n(N) = \frac{(pN)^n}{n!} e^{-pN} \quad (9)$$

which is the Poisson distribution with parameter pN.

To apply the Poisson distribution to the problem of estimating ridership, let us suppose that a route has been divided into sections, designated N_i. If, in each section, we find the probable ridership, we can then sum the probable ridership over all sections to find the total probable route ridership. The probable (or expected) ridership in each section will be the number of riders, X, multiplied by probability of obtaining that given number of riders, $P_x(N)$, and summed over all numbers of riders. In other words

$$S_i = \sum_{x=1}^{\infty} xP_x(N_i) \quad (10)$$

$$= \sum_{x=1}^{\infty} x \frac{(pN_i)^x}{x!} e^{-pN_i} \quad (11)$$

$$= pN_i \quad (12)$$

where S_i is the section ridership.
Probable Route ridership, Q, will be

$$Q = \Sigma S_i = p\Sigma_i N_i \quad (13)$$

This also shows that the parameter p is independent of section size and that the route can be considered as a whole, merely by summing properties of sections.

To provide more generality, let us suppose that we can divide riders into G groups with each group having its own probability, P_g, of belonging to the subgroup which uses public transportation and probability, r_g, of riding on a given day. The section measure for each group would be N_{ig}, the population in group g for section i. Total probable ridership on a route would be found by summing over all groups. If S_g is ridership by group, then

$$S_g = \Sigma S_{ig} = \Sigma\Sigma p_g r_g N_{ig} = \Sigma(p_g r_g \Sigma N_{ig}) \quad (14)$$

Again, this implies that each group can be treated as a whole for each route. The parameter $\Sigma_g S_g$ can be used in a Poisson distribution

$$P_n(\Sigma_g S_g) = e^{-\Sigma S_g} (\Sigma_g S_g)^n \quad (15)$$

to estimate the probability associated with different values of total ridership, n, over the route (6).

This approach has a number of advantages which make it theoretically superior to linear regression route models.

1. It is a disaggregate model in the sense that transit users are disaggregated into socioeconomic groups and usage relationships are developed for each group. In this sense, the model

forms the theoretical basis for placing a confidence interval around values obtained with a cross-classification approach (it is relevant to note that cross-classification approaches are rapidly replacing zonal linear regression approaches in urban transportation studies). It is sensitive to the causal factors of transit usage.

2. The model produces the probabilities of attaining given numbers of riders and thus can be used to determine the likelihood that demand would exceed capacity of various sizes of buses. Regression models can give similar information by use of confidence intervals, but the confidence interval has severe drawbacks for forecasting rural transit usage. First, it is possible (and very likely) to give negative values of ridership. Second, there are two confidence intervals that must be considered: one about the regression line itself, the other about a fixed point on a line. The latter is much wider than the former and is more likely to provide negative values. The width of the confidence intervals depends upon the error remaining after the regression line has been fitted; in previous regression models for rural ridership the error has been large, producing very large confidence intervals. In contrast, the probabilities generated by the Poisson model are easier to comprehend, may span a smaller range of values, and there cannot be negative ridership.

3. The model can be applied very simply. Once a calculation of the term $p_g r_g \sum_i N_{ig}$ has been made the planner can go to a table of Poisson probabilities and immediately determine the probabilities associated with different numbers of riders. The planner does not need to understand logarithms. The probabilities p_g and r_g can be explained in a very simple heuristic manner.

Parameter Estimation

Thus far, the means for obtaining the values of p_g have not been discussed. To provide a way to estimate the value of p_g, a maximum likelihood estimator will be derived. Such an estimator would be used on a present route to provide estimates for p_g for use in planning other routes. On a present route, let us suppose that in each section i and for each group g we have n_{ig} people boarding the bus.

For a Poisson distribution, the maximum likelihood is described as the product of all the probabilities for all the observations. One will be done for each class.

$$L_g = \prod_{i=1}^{m} \frac{(p_g N_{ig})^{n_{ig}} e^{-p_g N_{ig}}}{n_{ig}!} \qquad (16)$$

To find the maximum likelihood estimator, we will take the natural log of both sides

$$\ln L_g = \sum_{i=1}^{m} [n_{ig} \ln p_g N_{ig} + \ln e^{-p_g N_{ig}} - \ln n_{ig}!] \qquad (17)$$

Taking the derivative of both sides with respect to p_g and setting it equal to zero yields, for p_g

$$p_g = \frac{\sum_{i=1}^{m} n_{ig}}{\sum_{i=1}^{m} N_{ig}} \qquad (18)$$

This is the maximum likelihood estimator for p_g and is the same estimator one would use in cross-classification trip generation.

Study Area

A portion of northern West Virginia on the boarder of Pennsylvania was chosen for study due to an abundance of existing transit services and nearness to the research institution. The counties studied included Harrison, Monongalia, and Marion. The data base consisted of on-off counts, a survey of rider characteristics, and census data. Maps prepared by the State Department of Highways were used to determine locations of buildings along roads in the study area to estimate the portions of the census population living within 1.21 km of each route. Each of the counties had a county seat of between 26,000 and 35,000 population which served as the focal point for the rural routes. Average density for the rural portions (exclusive of the county seat) of the three county area in which the routes operate is 12 dwelling units per square km (31 per square mile).

Expansion of Survey

The probabilities p_g and r_g were estimated from the rider survey and on-off counts which were conducted as part of the research (1,2). First, the rider survey had to be factored to equal total ridership on the day on-off counts were taken to compensate for people who did not respond to the survey. Among the daily routes, a total of 117 questionnaires were obtained, and there were 277.1 trip ends recorded from the rural areas. Assuming two trip ends were associated with each unique individual in the survey, the factor by which the survey had to be expanded was (277.1)÷(2 x 117)= 1.18. Thus, all responses to the survey questionnaire had to be expanded by this factor. Among weekly routes, a total of 116 questionnaires was returned, and the on-off counts revealed a total of 229.5 trip ends. The expansion factor for this group of riders was estimated as (229.5)÷(2 x 116)= .99, which was close enough to 1 to eliminate the need for any expansion. Next, factors had to be developed to compensate for a sampling bias which caused under-representation of infrequent riders. It was obvious that people who used transit daily or every week on weekly service had a high probability of being included in the survey sample. But the distinct individuals who rode less frequently tended to be under-reported since the odds of surveying them on any given day decreased as their frequency of use decreased.

Mathematically, the problem was expressed as follows: let the expected number of riders who ride x days per month that would appear on a given day of service be S_x. Let the total number of unique individuals who ride x days per month be N_x. If a person rode x days per month and service was offered D days, it was assumed that the odds of surveying him on any random day of the month would be x/D. Thus, the expected number of riders who ride x days per month who would be surveyed on a random day is $S_x = N_x \cdot x/D$. If it is assumed that a sample of riders is taken on one day and the number who indicate they ride x days is S_x, then the total number of unique riders who ride x days can be estimated by

$$N_x = \frac{S_x \cdot D}{x} \qquad (19)$$

The number of weekdays of service offered per month, D, was estimated to be 21.7 days based on an average calculated over a 12-month period (this was approximately 4.34 weeks per month). For daily routes, the number of survey respondents responding to the question, "How often do you ride the bus?" was multiplied by the expansion factor (1.14) for each category of response, to provide S_x. If a person said he or she rode daily, it was assumed to mean 21.7 days per month, the value of x. For responses of "2-4 times a week," "once a week," "2-4 times a month," "once a month," and "less frequently" the number of days per month the individual rode, x, was assumed to be 13.0, 4.3, 2.5, 1.0 and 0.5, respectively. It may be questioned whether people could actually respond to this question in a correct manner. It is possible that they did not provide accurate information but no independent data existed to verify their answers. The answers were assumed correct due to lack of knowledge otherwise. The total number of riders, N_x, who ride x days per month was then estimated by the previous relationship, $N_x = S_x \cdot D/x$. The probability of riding on any given day of service, r_g, was estimated as $\sum_x S_x/\Sigma N$. This approach produced a large number of unique riders and a low probability of riding on a given day. This approach assumed that users were randomly distributed among the days service was offered, and had equal probability of riding on any given day.

Daily Routes

Among the routes operated on a daily basis, total population served (within 1.21 km of route) was estimated as 18,693. The probability that an individual belonging to socioeconomic group g utilized public transportation, P_g, was calculated as the ratio of ΣN_x to the total population living along the route as follows from the development of the maximum liklihood estimator. Previous research done in conjunction with the project showed that there were no significant differences in ridership rates for daily routes among different sub-populations. Age, income, and auto ownership did not appear to influence the rate of ridership. Thus, P_g and r_g were the same for all age, income, and auto ownership groups.

The predictions of the model for each of the daily routes are shown in Table 1. The table lists the minimum number of individuals J which have a cumulative probability greater than .05 of riding on a given day, and the maximum number J which have a cumulative probability less than .95 of riding on a given day. Thus, the J values associated with .05 and .95 cumulative probabilities define a 90 percent confidence interval for the number of riders. Also shown are the expected number of riders $N_g \cdot P_g \cdot r_g$, and the actual number of riders. The cumulative probabilities associated with the actual number of riders are presented, making it possible to determine the likelihood of that many riders actually materializing under the assumptions of the Poisson model. As can be seen in the table, the Poisson model did not fit the data well. Two routes had actual ridership values less than the values associated with the .05 level, and one had actual ridership greater than the .95 level, with the result that only two routes, Crown and Wolf Summit, had values falling within the .90 probability interval of the Poisson model. These results indicated that the model tended to over-predict and under-predict. To further investigate the reasons for this, the relationship between actual ridership rates and accessibility measures were analyzed to determine if accessibility might have had an influence on ridership. The ratio of actual riders to service area population was calculated for each route. This ratio represented the value that the

TABLE 1
POISSON MODEL RESULTS
DAILY ROUTES
Rider estimates based on service area population,
$p \times r = .02600 \times .291 = .00757^a$

Route	90% prob. interval	Expected Riders	Actual Number	Cum Prob. Actual no.
Enterprise	27-45	31.4	18.7	.0120
Crown	8-19	13.8	18.7	.9301
Cheat	24-42	33.2	49.6	.9976
Wolf Summit	31-52	41.7	44.4	.6772
Worthington	14-28	20.9	7.7	.0004

[a]Rider estimates = total trip ends ÷ 2.0

TABLE 2
POISSON MODEL RESULTS
DAILY ROUTES

Rider estimates based on service area population with p x r estimated as a function of route length[a]
pxr = .000304 x length (km)

Route	90% prob. interval	Expected J	Actual Number	Cum Prob, Actual no.
Enterprise	15-30	22.6	18.7	.2650
Crown	11-24	18.1	18.7	.6463
Cheat	39-61	50.4	49.6	.5126
Wolf Summit	28-47	37.8	44.4	.8628
Worthington	5-16	10.9	7.7	.2416
Doddridge Co.	0-4	2.2	6.6	.9981

[a]Rider estimates - total trip ends ÷ 2.0

Poisson parameter $S_g = P_g \cdot r_g \cdot N_g$ would have to take to provide a good Poisson model estimate of ridership for each route. A plot of the ratio S_g versus route length for each route produced an extremely good straight line relationship with a correlation of .971. A least squares model of S_g versus length with suppression of the g intercept was obtained. The model had an R-square value of .94 and the intercept was not significantly different from zero, making it possible to refit the model with suppression of the intercept. The resulting relationship obtained was

$$\hat{S}_g = .000304 \times \text{route length (km)} \qquad (20)$$

This relationship was used to estimate $\hat{S}_g$ for each route. Results of utilizing $\hat{S}_g$ in place of $S_g = N_g \cdot P_g \cdot r_g$ are shown in Table 2. All actual ridership figures fell within the .90 probability interval except for Doddridge County, which was a new daily route initiated after the model had been developed.

Ridership Estimates for Weekly Routes

Because of previous research carried out (2,3) in conjunction with the project, wherein it was shown that there are significant differences in ridership rates among people of various groups for weekly routes, whereas there were none for daily routes, it was felt that it would be necessary to divide the population into groups and form a cross-classification model for weekly routes.

The independent socioeconomic variables tried in the analysis were age, sex, income and auto ownership. Unfortunately, census data at the enumeration district (ED) level was cross-classified only by age and sex. No cross-classification existed for any other sets of variables. Therefore, only one-variable classifications were available for any variables except age and sex.

The selection of categories in cross-classification is usually somewhat arbitrary. To reduce the arbitrariness and to strengthen the estimates, an analysis of variance was applied. This was done by using the category variables as classification variables for analysis of variance. Even though it was believed that the statistical distribution underlying the cross-classification analysis was Poisson, nevertheless, the cross-classification rates derived were estimates of the means of the Poisson distribution and were therefore normally distributed, and thus could be analyzed using analysis of variance. Furthermore, multiple range tests, such as Duncan's test or Tukey's test, could be applied to the means(rates) of the categories to determine those which were not significantly different from one another. Those which were not significantly different were combined to form new categories which, having a greater number of observations, reduced the variance in the new categories and thereby produced better estimates.

The first step in deriving the model was to estimate the age-sex distribution of the ridership. Individual units of observation were considered to be routes. The questionnaire data (as described previously, (2,3)) was used to estimate the age-sex distribution of riders by route. The number of people in each age-sex category within the service area of the route for each ED was then estimated. The numbers in each category were summed along the route. The ratios of ridership to total population in each category were estimated. The breakdown by age followed those used in the Census. The categories were as follows:

5-14	35-44	65-up
15-24	45-54	
25-34	55-65	

An analysis of variance of the rates was then undertaken for both weekly and daily routes, the results of which are illustrated in Table 3. As illustrated by the table, in the daily routes there was a significant difference between sexes, but not a significant difference by age and no interaction. This tends to follow all previous results and to indicate that a total population-based model may be applicable to daily routes. For weekly routes it was found that there was a significant difference among age groups and by sex and there was significant interaction. To find where the significant difference occurred, Duncan's multiple range test was employed. The results are shown in Table 4. Each letter in the grouping indicates groups in which the means were not significantly different. Three groups were evident, of which two were overlapping. One group stands out by itself, women over 65. Another group is women between the ages of 45 and 64 and men over 65. The third group is all men and women below the age of 45. There is an overlap, since the categories "men over 65" and "women between 45 and 54" belong to two groups. For the purpose of forming cross-classification categories, it appeared that these two old categories should be grouped with the women 55-64. Therefore, the final cross-classification categories derived were:

TABLE 3
ANALYSIS OF VARIANCE OF RIDERSHIP RATES

Source of Variation	Degrees of Freedom	Square	F-Ratio	Prob of F> Calculated
		DAILY		
Sex	1	0.00636	13.56	0.0005
Age	6	0.00358	1.27	0.2847
Interaction	6	0.00307	1.09	0.3791
Residual	56	0.00047		
		WEEKLY		
Sex	1	0.02214	19.65	0.0001
Age	6	0.11144	16.48	0.0001
Interaction	6	0.05458	8.07	
Residual	182	0.00113		

TABLE 4
Grouping of Means for Cross-Classification

CLASSIFICATION Sex	Age	Mean	Grouping
F	65 & over	.1210	A
F	55-64	.0398	B
M	65 & over	.0232	BC
F	45-54	.0203	BC
M	45-54	.0081	C
M	35-44	.0065	C
F	35-44	.0053	C
M	55-64	.0049	C
F	15-24	.0038	C
F	5-14	.0013	C
F	25-34	.0000	C
M	15-24	.0000	C
M	25-34	.0000	C
M	5-14	.0000	C

1. Women over 65
2. Men over 65 and women 45-64
3. All others

With these new categories formed, new rates were calculated. These were:

Group		Rate·trips/person/ day of service
I.	Women over 65	.121
II.	Men over 65 and women 45-64	.0398
III.	All others	.0027

These rates were applied to existing routes and the results are shown in Table 5. Also shown in the table are the Poisson probabilities associated with each route. There were significant over- and underestimates. Only seven out of fourteen estimates were within the 90% interval. With respect to underestimates, these occurred on routes which went through hamlets (population 500-1500) as well as the countryside. To correct this problem, the ridership and population were divided into two groups, those in the countryside and those in hamlets. Ridership rates were then calculated for these two groups separately, and a new estimate was made as shown in Table 6. Rates utilized were as follows:

COUNTRYSIDE RATES

Group		Rate-Trips/Person/Day
I.	Women over 65	.1338
II.	Men over 65 and Women 45-64	.0336
III.	All others	.0029

HAMLET RATES

Group		Rate-Trips/Person/Day
I.	Women over 65	.0424
II.	All others	.0018

A comparison of the two estimates is shown in Table 7. Nine out of fourteen estimates were within the 90% interval. A improvement in ridership estimates was immediately evident. In all cases but two the percentage error decreased. In particular for the three gross over-estimates, much better estimates were obtained.

The same analysis was also performed for auto ownership, household size, and income, using a one-way classification. The error of prediction was greater for each of these than for the age-sex cross-classification. Also, the error was reduced when a division between countryside segments and hamlet segments was used. Nevertheless, the error was still greater than with the age-sex cross-classification as segmented.

Conclusions

Presented in this paper has been a Poisson model for estimating rural transit ridership which utilizes cross-classification data. This model has been derived theoretically, and theoretical estimators for parameters have been derived. These derivations provide a theoretical basis for placing confidence intervals around the cross-classification approach which, heretofore, has been a purely empirical approach to trip generation and modal split. Having a theoretical basis for the approach now permits refinements in the approach, such as the setting of confidence intervals around route estimates and development of statistical tests. A further benefit of the approach is its simplicity which permits its use in all sorts of unsophisticated planning operations.

The application to daily routes shows that there is a significant contribution of route length, which strengthens the model greatly. What the relationship suggests for daily routes is not immediately clear. It could reflect the influence of level of service, or the care taken to locate a route within a service area. Although the relationship to route length is strong, further research is necessary to discover if it is universal. For weekly routes, differing rates by population group were found. The rates can be tested by using analysis of variance and multiple range tests, to provide the best grouping, i.e., the one union which has the minimum variance, yet retains statistically significant differences. A further discovery made was that ridership rates vary according to whether riders are in the countryside or in hamlets, and thus much better estimates were obtained by treating these two areas separately.

Use of cross-classification trip rates for specific groups of people with similar socioeconomic characteristics increased the sensitivity of the prediction of the socioeconomic factors hypothesized to create the need for public transportation. Though the models were not tested outside the portion of West Virginia where they were calibrated, the trip rate approach would make it easier to compare results from one area of the

TABLE 5
POISSON MODEL RESULTS
WEEKLY ROUTES
Rider estimates based on age-sex classification[a]

Route	90% prob. interval	Expected no. J	Actual Number	Cum Prob, Actual no
Grafton	13-28	21.2	12.3	.0227
Mt. Heights	27-48	38.2	20.7	.0018
Blacksville	5-14	9.6	13.2	.8938
Carolina	0-7	4.5	9.8	.9929
Fairview	24-44	34.7	12.0	.0
Kingmont	7-20	14.3	11.8	.3324
Mannington	84-116	100.3	19.3	.0
Colfax	1-7	4.6	10.5	.9972
McWhorter	18-33	25.7	28.0	.7183
Kincheloe	18-33	26.3	22.0	.2349
Johnstown	10-23	16.7	14.0	.3029
Sardis	9-21	15.7	18.0	.7668
Wyatt	55-81	68.6	24.0	.0
Wallace	21-38	29.5	28.0	.4402

[a]Rider estimates = total trip ends.

TABLE 6
POISSON MODEL RESULTS
WEEKLY ROUTES - HAMLET SEPARATED
Rider estimates based on age-sex classification[a]

Route	90% Prob, Interval	Expected J	Actual Number	Cum.Prob, Actual No.
Grafton	16-32	24.3	12.3	.0046
Mt.Heights	13-26	19.5	20.7	.5151
Blacksville	6-16	13.3	10.9	.7905
Carolina	2-8	5.2	9.8	.9823
Fairview	11-24	18.0	12.0	.0917
Kingmont	10-22	16.3	11.8	.5362
Colfox	2-8	5.2	10.5	.9823
Monnington	16-32	24.2	19.3	.1701
Mt.Wharter	11-25	19.0	28.0	.9805
Kincheloe	14-28	21.5	22.0	.5987
Johnstown	9-21	14.9	14.0	.4759
Sardis	11-24	18.0	18.0	.5622
Wyatt	27-46	36.4	24.0	.0193
Wallace	19-34	26.7	28.0	.5739

[a]Rider estimates = total trip ends

TABLE 7
COMPARISON OF RIDERSHIP ESTIMATES

Route	Actual Ridership	Entire Route % Error	Hamlet Separated % Error
Grafton	12.3	72.1	97.7
Mt.Heights	20.7	84.2	-6.0
Blacksville	13.3	-27.8	-18.0
Carolina	9.8	-54.1	-46.7
Fairview	12.0	189.2	50.0
Kingmont	11.8	21.2	38.1
Colfax	10.5	-55.2	-50.5
Mannington	19.3	419.7	25.4
McWhorter	28.0	-8.2	-32.1
Kincheloe	22.0	36.4	-2.3
Johnstown	14.0	19.3	6.4
Sardis	18.0	-12.8	0.0
Wyatt	24.0	185.8	51.7
Wallace	28.0	5.4	-4.6

country to another. Rather than comparing the simultaneously determined coefficients of multiple linear regression models, with their often obscurely intercorrelated values, a comparison could be made of simple age and sex specific rates. Level of service variables such as route length could be used to modify the rates.

In most cases, the models have reasonably accurate results. Over seventy percent of the routes were estimated within a 90 percent confidence interval. Though some predictions were not as accurate as the planner would like to obtain, they were realistic and could easily be used to determine a probable range of demand for planning purposes. For example, by examining the minimum and maximum values of the confidence intervals, the planner possibly could begin to make a preliminary assessment of the need for vehicles of different capacities and the probable range of revenues which could be anticipated.

ACKNOWLEDGEMENTS

The authors would like to thank Teh-Foo Tsai and Kamrooz Molavi, who performed significant work in relation to this article. The work was sponsored by the U.S. Department of Transportation Office of University Research. The statements and views expressed represent the opinions of the authors and not those of the Department of Transportation.

REFERENCES

1. Byrne, B.F. and Neumann, E.S., "Feasibility of Developing Low-Cost Measures of Demand for Rural Public Transportation," Final Reprot, Phase I., U.S. Department of Transportation, Office of University Research, August 1976.
2. Neumann, E.S., and Byrne, B.F., "Characteristics of Rural Transit Users," presented at the 57th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1977.
3. Byrne, B.F., and Neumann, E.S., "Low-Cost Models of Demand for Rural Public Transportation," Proceedings, First Annual Conference on Rural Public Transportation June 1976, p 77-89, Greensboro, N.C.
4. Burkhardt, J.E., and Lago, A.M., Methods of Predicting Rural Transit Demand, prepared for the Pennsylvania Department of Transportation, Harrisburg, 1976.
5. Smith, R.L., Jr., "Evaluation of Rural Public Transportation Demand Models Which Include Level of Service Measures," presented at the 57th Annual Meeting of the Transportation Research Board, Washington, D.C. January 1977.
6. Feller, William, An Introduction to Probability Theory and Its Application, Volume I, Wiley, New York, 1950, Chapter 6, 12.