THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 59, No. 6, July–August 1980
Printed in U.S.A.

Relationships Between the Demand for Local Telephone Calls and Household Characteristics

By W. J. INFOSINO

(Manuscript received November 13, 1979)

The demand for local telephone calls varies among households. This paper develops models, based on limited data from California and Cincinnati, which predict the demand for local calls from household characteristics. The models enable one to stratify a metropolitan area into regions of expected high, low, and medium telephone demand and thus provide a mechanism for efficiently estimating the demand in a metropolitan area from the demand of a sample of telephone customers in the area. Although the data and models are too limited to establish definite causal relationships, the models suggest that the demand for local calls might be related to the number of people in the household and the age and sex of the household head. Furthermore, while there is some ambiguity between the California and Cincinnati results, there is also the suggestion that local call demand might be related to income, the race of the household head, and the telephone density in the wire center.

I. INTRODUCTION

This paper shows that residence telephone calling rates (local calls per day) are related to household characteristics (e.g., number of members, age of the head). The relationships are quantified in the form of models which estimate the number of local calls made from a household telephone as a function of the household's characteristics. These models are then converted into models which estimate the average calling rate in a given neighborhood from census-type population and housing statistics.*

* In this paper, the number of calls made from a telephone refers to the number of local calls that are made from all the telephones (both the primary telephone and its extensions) billed to a particular telephone number. The average calling rate in a given

Telephone companies need precise estimates of average local calling rates in metropolitan areas in order to design local tariffs. These models were originally developed to help obtain these estimates. Metropolitan area calling rates are often estimated by initially selecting a sample of telephone switching systems and then observing the calling rates of a sample of telephones served by those switching systems. A switching system provides a group of telephones in a particular geographic area (neighborhood) with access to the rest of the telephone network. In past studies, large variations in average calling rate among switching systems have been observed.[1] This implies that a large number of switching systems needs to be sampled to get a precise estimate of the average calling rate in a metropolitan area. Unfortunately, sampling many switching systems is expensive and sometimes impossible. With the aid of a calling rate model, however, the precision from a small sample of switching systems can be improved. The model can be used to stratify a metropolitan area into regions of expected high, low, and medium calling rates. Switching systems can then be sampled from each strata. The household characteristics may account for some of the previously "unexplained" variation among switching system average calling rates. Since the amount of "unexplained" variation has been reduced, a more precise estimate of the average calling rate can be obtained.

Furthermore, for reasons of cost and technical feasibility, most metropolitan area calling rate studies are based on samples of electronic switching systems only. In the absence of other information, a telephone company would have to assume that the average calling rate of customers served by the nonelectronic switching systems is the same as in the electronic switching systems. But the calling rate model may enable the telephone company to replace this assumption with the more realistic assumption that areas with similar household characteristics have similar calling rates. The telephone company either can estimate the average calling rate in the nonstudy switching systems directly from the model or can estimate their calling rates by averaging the calling rates of those study switching systems that serve geographic areas similar in the significant household variables.

As noted above, the calling rate model was originally developed as a sampling tool. However, its success suggests that household characteristics may prove useful in demand modeling. The typical demand model relates the demand for a good (e.g., telephone usage) to the price of the good and average income. Household characteristics (other

area is the average with respect to the number of primary telephones (main stations) in the area. And the calling rate of a sampled household is the calling rate (local calls per day) of the telephones billed to the sampled telephone number. This does not necessarily include all the calls the household makes, since some households have additional telephones that are billed to a different (nonsampled) number.

than income) are often ignored. The inclusion of household characteristics may improve these models. Furthermore, typical demand studies analyze *aggregate demand directly*. In this paper, however, we first analyze individual customer behavior and then aggregate. This procedure may lead to better aggregate models, not only because individual customer data have more variability than aggregate data, but also because individual customer models may suggest the correct specification for the aggregate model.

II. DATA

This study is based on a sample of 705 California and 293 Cincinnati residence telephones. First, a sample of telephone switching systems was selected. Ten California and nine Cincinnati switching systems were selected for convenience (i.e., they were easily studied), not at random. Then random samples of telephones served by the switching systems were selected and the corresponding households were mailed socioeconomic questionnaires. These questionnaires included questions on household income, the number of household members, and the age, sex, and marital status of the household head. Approximately 40 percent of the households in both locations returned fully completed questionnaires. Section VI discusses the potential response bias. Each telephone's calling rate (denoted CR) was calculated by dividing the number of local calls made while the telephone was on study by the number of days it was on study. Different telephones were on study at different times and for different lengths of time. However, most of the California and Cincinnati telephones were on study for one and two months, respectively. No adjustment was made for the number of holidays or weekends a telephone was on study. Variations in the number of holidays or weekends were assumed to be distributed evenly over all socioeconomic groups such that no bias was introduced. The California data were collected between May 1972 and September 1973; the Cincinnati data were collected between March 1975 and January 1976.

In addition to the individual household data, some 1970 census population and housing statistics were available for the areas served by the wire centers corresponding to the sampled switching systems. A wire center is a building that houses one or more switching systems and serves a specific geographic area. The census statistics were estimated by weighting census tract statistics according to the percentage of the tract's geographic area that lies within the area served by the wire center. In the rest of this paper, we refer to the "area served by the wire center" as simply "the wire center."

The California telephone subscribers had a choice of three billing options: one flat-rate option and two measured-rate options. Flat-rate customers could make an unlimited number of calls to a "local calling zone" for a fixed monthly fee. Measured-rate customers paid a smaller fixed monthly fee but were charged for each call over a specified monthly allowance. Eighty-four percent of the California residence subscribers chose flat rate. Only flat rate customers were analyzed in this study.

The Cincinnati telephone subscribers had a choice between single- and multiparty service. Only single-party lines were analyzed, and all single-party lines were flat rate. However, the telephone subscribers from one switching system (Hamilton) had a choice between ordinary local area service (LAS) or extended area service (EAS). Both classes were flat rate, but EAS was priced higher and had a larger local (or "free") calling zone than LAS. Both LAS and EAS subscribers were analyzed.

III. ANALYSIS

Models for estimating average calling rate could be built by regressing each study switching system's average calling rate on the household characteristics of the geographic area served by the switching system. The household characteristics could be obtained from the 1970 census. Since there are only ten observations (switching systems) in California and nine in Cincinnati, it would not be reasonable to regress more than two or three variables at the same time in each of these two areas. However, with the aid of the questionnaires, we can develop a model that would estimate the calling rate of an individual telephone as a function of household characteristics. Here we will be able to study many variables simultaneously because we have as many observations as we have sampled telephones (705 in California and 293 in Cincinnati). After determining which variables are statistically significant, we can convert this model for predicting an individual telephone's calling rate into a model for predicting the average calling rate in a specific geographic area.

3.1 Initial model

The models to estimate telephone calling rates are of the form

$$\sqrt{CR_i} = B_0 + B_1X_{1i} + B_2X_{2i} + \cdots + B_pX_{pi} + \epsilon_i.$$

The B_j's are constants estimated by ordinary least squares. The X_{ji}'s are dummy variables (they take on only values of 1 and 0) to indicate the values of the household characteristics, and ϵ_i is an error term. The use of dummy variables for quantitative household characteristics (such as age or income) as well as qualitative characteristics (such as sex or marital status) enables us to automatically estimate nonlinear relationships in the quantitative characteristics. For example, as shown in Table I, the age of the head of the household characteristic is represented by five dummy variables that correspond respectively to the age intervals of 25–34, 35–44, 45–54, 55–64, and over 65. If the head of the household is under age 25, each of the five age dummy variables will have a value of zero. Thus, an age of under 25 is implied unless one of the age dummy variables is set equal to one. If the head of the household is between 25 and 34, the dummy variable corresponding to that age interval is set equal to one and all other age dummy variables will have a value of zero. Thus the model coefficient corresponding to the age 25-to-34 dummy variable indicates how much higher (or lower) the dependent variable of the model (the square root of the calling rate) is expected to be for households whose heads are between the ages 25 and 34 than for households whose heads are under age 25 (assuming all other variables are equal). Similarly, the coefficient for the age 35-to-44 dummy variable shows how much higher (or lower) the square root of the calling rate is expected to be for households whose heads are between the ages 35 and 44 than for households whose heads are under age 25. The relationship between the square root of the calling rate and age of the head of the household is nonlinear. This nonlinear relationship can be seen on Fig. 2. By way of comparison, Fig. 1 shows the more linear relationship between the square root of the calling rate and the number of people in the household.

The initial fit of the models (before eliminating statistically nonsignificant variables) considered the following household characteristics: the number of people in the household, income, education of the head, employment status of the head, type of housing, number of years at the current address, and wire center. In addition, the Cincinnati model contained variables for race and home ownership. The California questionnaires did not include a race question. Later, however, race was added to the California model through the analysis of census data. Home ownership was not included in the California model in order to limit the number of parameters that had to be fit. It was found not to be statistically significant in Cincinnati.

The Cincinnati model does not include as many dummy variables for some characteristics as the California model because the smaller sample size in Cincinnati would not support the additional variables in the sense that some of the dummy variables would be represented by only a few observations. Dummy variables for wire centers were included to allow for the possibility that the environment around a household affects the household's calling rate. Two Cincinnati switching systems were combined to form one dummy variable because they are located in the same wire center. On the other hand, the Cincinnati switching system (Hamilton) which contained both EAS and LAS accounts was divided into two dummy variables such that one dummy variable corresponded to the EAS customers and the other dummy variable corresponded to the LAS customers. This separation allowed for a difference in calling rate between the EAS and LAS customers.

The $\sqrt{CR}$ is the square root of the telephone's average daily calling rate. In fitting the model, this number was calculated from the days the telephone was on study. The square root transformation was used because the residuals of the fitted model were found to be distributed more like a normal distribution and were more homoscedastic with the transformation than without it. Residual normality and homoscedasticity help assure the validity of the F test, which is used to identify significant variables.

3.2 Eliminating nonsignificant variables

Nonsignificant variables were eliminated from each model through a backward elimination procedure. At each iteration of the procedure, the household characteristic with the least significance according to an F test[2] was eliminated from the model. The remaining variables were then refitted. Household characteristic refers to a set of dummy variables. For example, "eighth grade," "high school," and "college graduate" are the dummy variables corresponding to the education characteristic. The single-dummy variable "college graduate" would not be eliminated at a particular iteration, but the set of three dummy variables corresponding to education might be eliminated.

In California, the household characteristics were eliminated in the following order: length of time at the current address, type of housing, employment status, marital status, and education. All the eliminated characteristics were not significant at the 95-percent confidence level.

In Cincinnati, the household characteristics were eliminated in the following order: type of housing, years at address, employment status, own/rent, non-Hamilton wire centers, marital status, education, and income. None of the eliminated characteristics was statistically significant at the 90-percent confidence level. The remaining characteristics (number of people in the household, age of the head, sex of the head, and race) were statistically significant at the 95-percent confidence level.

The Hamilton wire-center effects were not considered for removal from the model because the author believes that the local tariff structure in Hamilton differs from the other study areas in a way that is likely to affect calling rates. It is reasonable to assume that the EAS customers have higher local calling rates than the LAS customers because their local calling zone is larger and because they would probably not purchase EAS service unless their calling rate to the extended area was high.

Tables I and II show the estimated parameters, their standard errors, and t statistics for the California and Cincinnati models after elimination of the variables that were not statistically significant. Both multiple correlation coefficients squared (R^2) are 0.35.

Table I—California regression coefficients for the square root of the calling rate model

Variable	Coefficient	Standard Error	t
Number of People in the Household:			
1 person (implied)	0.0	*	*
2 people	0.23	0.07	3.08
3 people	0.53	0.09	6.02
4 people	0.77	0.09	8.31
5 people	0.93	0.10	9.04
6 people	1.26	0.13	9.56
7 people	1.19	0.19	6.16
8 or more people	1.88	0.28	6.74
Age of Head of Household:			
Age <25 (implied)	0.0	*	*
Age 25–34	−0.25	0.11	−2.29
Age 35–44	−0.10	0.12	−0.87
Age 45–54	−0.08	0.12	−0.71
Age 55–64	−0.08	0.12	−0.67
Age 65+	−0.37	0.13	−2.97
Sex of Head of Household:			
Male (implied)	0.0	*	*
Female	0.21	0.07	3.01
Wire Center:			
Alhambra (implied)	0.0	*	*
Beverly Hills	0.58	0.11	5.46
Bush Pine	−0.03	0.11	−0.27
Franklin	0.09	0.10	0.90
Madison	−0.07	0.10	−0.71
McCoppin	0.27	0.11	2.38
Republic	0.38	0.11	3.51
San Mateo	−0.10	0.11	−0.95
Santa Ana	0.01	0.10	0.06
Palo Alto	−0.15	0.12	−1.27
Income:			
<$3K (implied)	0.0	*	*
$3–5K	−0.30	0.15	−2.00
$5–8K	−0.40	0.13	−3.14
$8–10K	−0.09	0.12	−0.71
$10–15K	−0.26	0.12	−2.13
$15–20K	−0.21	0.13	−1.64
$20–30K	−0.32	0.13	−2.39
$30K+	−0.26	0.15	−1.73

Constant = 1.45
R^2 = 0.349
Mean square error = 0.423

* Each coefficient represents the difference in the average square root of the calling rate between the implied household characteristic and the characteristic corresponding to the coefficient. Therefore, the coefficient associated with the implied household characteristic is by definition zero with no standard error.

Table II—Cincinnati regression coefficients for the square root of the calling rate model

Variable	Coefficient	Standard Error	t
Number of People in the Household:			
1 person (implied)	0.0	*	*
2 people	0.58	0.14	4.14
3 people	0.75	0.16	4.82
4 people	0.93	0.18	5.18
5 people	1.19	0.20	5.91
6 or more people	1.63	0.24	6.86
Age of Head of Household:			
Age <25 (implied)	0.0	*	*
Age 25–34	−0.30	0.21	−1.43
Age 35–44	−0.04	0.23	−0.19
Age 45–54	0.04	0.22	0.19
Age 55–64	−0.13	0.21	−0.61
Age 65+	−0.60	0.22	−2.73
Sex of Head of Household:			
Male (implied)	0.0	*	*
Female	0.45	0.12	3.71
Wire Center × Billing Option:			
Non-Hamilton (implied)	0.0	*	*
Hamilton LAS	−0.28	0.12	−2.23
Hamilton EAS	0.20	0.13	1.51
Race:			
White (implied)	0.0	*	*
Black	0.31	0.15	2.09

Constant = 1.31
R^2 = 0.350
Mean square error = 0.590

* Each coefficient represents the difference in the average square root of the calling rate between the implied household characteristic and the household characteristic corresponding to the coefficient. Therefore, the coefficient associated with the implied household characteristic is by definition zero with no standard error.

3.3 California environmental regressions

The California wire center coefficients in Table I were regressed against wire-center characteristics available from the 1970 census and telephone company records. Statistics on the population of each wire center were considered involving sex, race, age, marital status, education, income, employment, types of housing, housing values, rents, and telephone equipment. The two variables that fit best individually in terms of the R^2 statistic were the residence main telephone density (main telephones per square mile) and the fraction of people in the wire center who are black or Spanish. Beverly Hills, however, appears as an extreme outlier. Rather than to try to develop a model that would improve the fit of Beverly Hills, Beverly Hills was ignored because it is atypical of most areas in the Bell System. According to a 1971 Pacific Telephone Company planning study, Beverly Hills has more telephones per hundred population than any other exchange* in the Bell System, and many residential customers have complex key telephone sets and private switching systems. In addition, Beverly Hills has an unusually large number of foreign exchange lines.†

The least-squares regression of the wire-center coefficients against the residence main telephone density (according to Pacific Telephone Company records) and the fraction of black and Spanish (i.e., the ratio of the number of black and Spanish people to the total population in the wire center according to the 1970 census) indicated that both variables were statistically significant at the 99-percent confidence level. The race and density coefficients were, respectively, 0.51 and 0.000043. The constant term was −0.24, and R^2 was 0.99. The Beverly Hills wire center was not included in the regression for the reasons cited above. The Madison wire center also was not included because there is evidence that the study telephones in Madison are not representative of the entire Madison wire center. Madison is an extreme outlier on a plot of 1970 census income data versus questionnaire average income data. The Madison wire center consists of 10,000 residence customers who are served by more than one switching system. However, the particular switching system from which the study telephones were sampled serves only 178 of these customers.

A model that merges the California household and environmental effects is obtained by replacing the wire-center coefficients in Table I with the above race and density coefficients and adding the constant. This is equivalent to the procedure for fitting nested variables recommended by Daniel and Wood.[3] While race has been developed as an environmental effect in California, it may actually be a household effect. If households headed by both blacks (or Spanish) and whites have higher calling rates in predominantly black (or Spanish) neighborhoods, then race is an environmental effect. If black (or Spanish) households have higher calling rates than white households regardless of the neighborhood, then race is a household effect. Since the California questionnaire did not include a question on race, race had to be treated as an environmental effect. In the Cincinnati model, race is a household effect. The hypothesis that the average calling rate is higher in both black *and* Spanish neighborhoods in California is based on very limited data. The fraction of black and Spanish was used in the model because it gives a better fit ($R^2 = 0.99$) than the fraction of black alone ($R^2 = 0.96$) or the fraction of Spanish alone ($R^2 = 0.72$).

---

* An exchange is the territory within which telephone service is provided without toll charges and covered by a specific rate basis, usually consisting of a single city or environs. A customer's local calling area may include one or more exchanges.

† Foreign exchange lines are lines which are served (at the customer's request) by a wire center other than the one that serves the area in which the telephone is located.

IV. MODEL COMPARISONS

In the next few sections, the statistical entities of the California and Cincinnati models are compared. The more alike the models are, the more confident we may be that the household variables in these models would successfully stratify other metropolitan areas into regions of expected high, low, and medium average calling rate. The reader is cautioned that it is not the intent of these sections to establish the causal relationship between household characteristics and the demand for local calls. To make such an interpretation would be a mistake because, as discussed in Section VI, such an interpretation would require more data and much more analysis and is beyond the scope of this paper.

The California and Cincinnati models have several variables in common: household size, age, race, and sex. On the other hand, the race variables are defined differently, and income and telephone density are statistically significant (at 95-percent confidence) in California but not in Cincinnati. The lack of significance of telephone density in Cincinnati may be due to insufficient data. The range of telephone densities is different (lower) and much smaller in Cincinnati than in California. (In the Cincinnati data, the telephone densities range from 117 to 2917 telephones per square mile; in the California data, the telephone densities range from 1683 to 6794 telephones per square mile.) Furthermore, the wire-center sample sizes are much smaller in Cincinnati than in California.

The qualitative relationships between the demographic variables and calling rate are also the same in California and Cincinnati for those variables that appear in both models. Furthermore, while the point estimates of those relationships (i.e., the point estimates of the coefficients) are different, most differences are not statistically significant at the 95-percent confidence level. We discuss each variable in more detail in the next few sections. Note that regression methodology allows comparisons within a particular demographic variable to be made in the context of all other variables being held constant.

4.1 Number of people in household

Figure 1 is a plot of the California and Cincinnati coefficients for household size. The plot shows that, in both California and Cincinnati, the square root of the calling rate increases with the number of people in the household. Since the standard errors of the coefficients are large, most Cincinnati coefficients are not significantly different from the California coefficients at 95-percent confidence when the coefficients are tested one at a time. Only the coefficients for two people in the household are statistically significantly different.

The California coefficients increase linearly with the number of

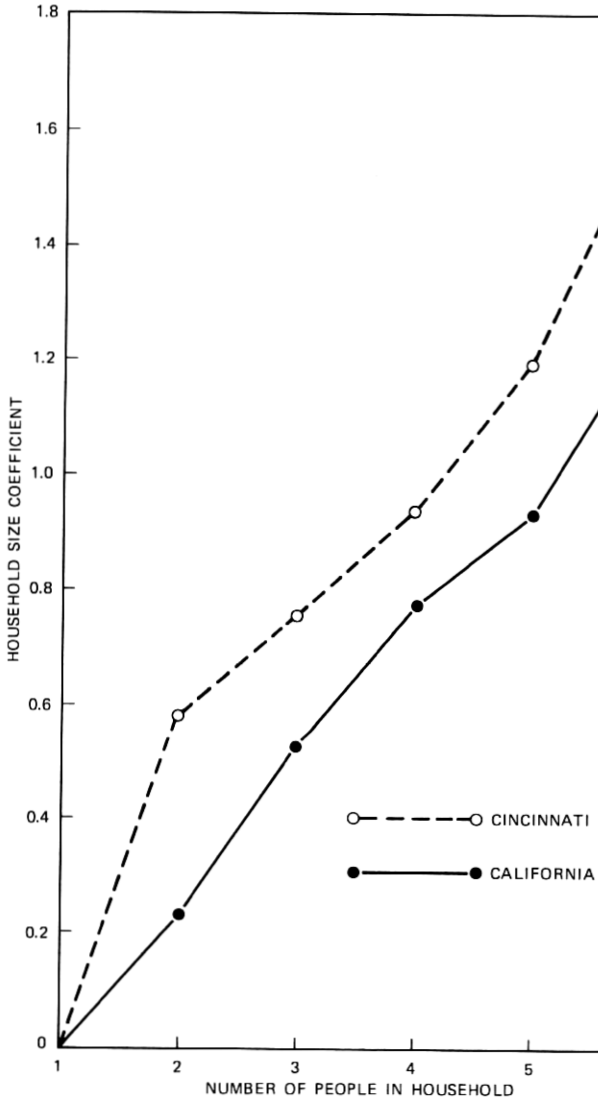

Fig. 1—Household size coefficients vs household size.

people. Each additional person adds about 0.25 to the square root of the calling rate. An F test indicates that the Cincinnati coefficients, however, are statistically significantly nonlinear at the 95-percent confidence level. However, Fig. 1 indicates that the departure from linearity is not too severe. One can argue that a straight line parallel to the straight line which best fits the California data (but with a higher intercept) is reasonably close to the Cincinnati data. Such a line suggests that it is reasonable to consider the relationship between household size and the square root of the calling rate to be the same in California and Cincinnati.

4.2 Age of head

Figure 2 is a plot of the California and Cincinnati coefficients for age of the household head. The plot shows that, in both California and Cincinnati, the square root of the calling rate is lowest when the head of household is over age 65. Calling rates are also low when the head of household is between the ages of 25 and 34. The point estimates of the coefficients indicate that the effect of age > 65 is stronger in

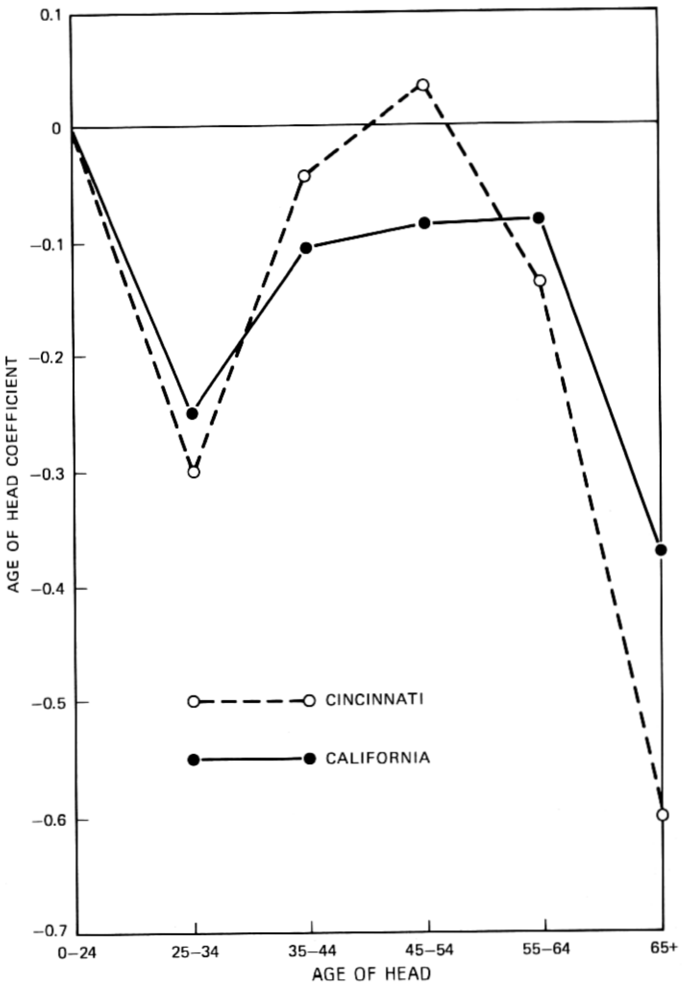

Fig. 2—Age of head coefficients vs age of head.

Cincinnati than in California. However, all the standard errors are large, and none of the differences in age coefficients between Cincinnati and California is statistically significant at 95-percent confidence.

4.3 Sex

In both California and Cincinnati, the square root of the calling rate is higher for households with female heads than for those with male heads. The Cincinnati coefficient for female head is higher (0.45) than the California coefficient (0.21). However, the standard errors of these estimates are large and the difference is not significant at 95-percent confidence.

4.4 Race

The California and Cincinnati race variables are defined differently. The California variable is the fraction of black and Spanish in the *wire center* in which the telephone is located; the Cincinnati variable is whether or not the head of the *household* in which the telephone is located is black. No Spanish households are in the Cincinnati data. From both models, we may conclude that the *average* local calling rate in predominantly black wire centers is higher than in predominantly white wire centers if all other variables are equal.

4.5 Income

The California income coefficients indicate that the square root of the calling rate does not rise or fall monotonically with income. The calling rate is higher if the income is less than \$3000 or between \$8000 and \$10,000 than it is for other income levels. In the Cincinnati data, the income coefficients are not statistically significant. However, a Chicago study[4] indicates that local calling rate increases with income. The difference between the income effects in Chicago and this study may be due to the fact that most Chicago telephone subscribers have measured rate service (where there is an incremental charge for telephone usage) while the telephone subscribers in this study have flat rate service.

V. Wire-center model

In the next few sections, we modify the models developed thus far to make them easier to use to stratify a metropolitan area into regions of expected high, low, and medium average calling rates. We convert the models that estimate the square root of the calling rate of an individual telephone into models that estimate the average calling rate in a wire center. When appropriate, we also reformulate the household characteristics into more readily assessible forms and eliminate those household characteristics that appear to have very little impact on the variability of the average calling rate across wire centers. While some of these modifications are of necessity based on judgment, the test for their validity is the predictive power of the models that result.

The model for estimating an individual telephone's square-root calling rate can easily be converted into a model for estimating the average square-root calling rate in a wire center. According to the model, the square root of the calling rate of the ith telephone is given by

$$\sqrt{\mathrm{CR}_i} = B_0 + B_1X_{1i} + \cdots + B_pX_{pi} + \epsilon_i$$

where the B_j's are constants, the X_{ji}'s are dummy variables that take on the values of 1 or 0 depending upon whether or not the account has a particular household characteristic, and ϵ_i is a random variable with an approximately normal distribution $N(0, \sigma^2)$.

If we sum both sides of the equation over all the telephones in a wire center and divide by N, the number of telephones in the wire center, we get

$$\text{AVG } \sqrt{\mathrm{CR}} = B_0 + B_1P_1 + B_2P_2 + \cdots + B_pP_p + \epsilon'$$

where the P_j's are the fraction of telephones in the wire center that have a particular household characteristic and ϵ' is a random variable with the approximately $N(0, \sigma^2/N)$ distribution. Note that, if we know the exact fraction of telephones with each household characteristic, the variance of the wire center estimate is inversely proportional to the size of the wire center in telephones. This explains why a model that gives relatively imprecise estimates of an individual telephone's calling rate can give precise estimates of a wire-center average calling rate. The models developed thus far estimate wire-center average *square-root* calling rates. In a later section, we convert these models into models that estimate average calling rates.

5.1 Sensitivity analysis and model simplification

The California model can be simplified to

$$\text{AVG } \sqrt{\mathrm{CR}} = 0.59 + \underset{(0.03)}{0.25P} + \underset{(0.06)}{0.52R} + \underset{(0.000008)}{0.000046D}$$

where
- P = average people per household
- R = fraction of black and Spanish people
- D = residence main telephone density in the wire center (i.e., residence main telephones per square mile).

The numbers in parentheses are standard errors. The simplification was obtained (as detailed below) by eliminating the income, age, and sex of head variables and replacing the seven household size variables with the single variable: average people per household.

Income was eliminated despite its relatively high statistical significance because the income coefficients indicate that calling rate does not rise or fall monotonically with income but follows an erratic pattern. This makes the variable very difficult to use for prediction. It is hard to measure the effect of inflation on different income bands. Furthermore, an income distribution is required and is difficult to obtain for specific areas. At best, only average income or an indication of whether or not the area is relatively rich or poor may be available. The model was refit when income was eliminated. That is, the model was fit to the individual household data without an income variable. The estimates of the other coefficients changed only slightly.

The age and sex of head variables were eliminated because they do not usually vary enough from wire center to wire center to cause a large difference in average square-root calling rate among the wire centers. This can be seen in the following:

Maximum People/House Effect	Maximum Age Effect	Maximum Sex Effect	Maximum Race Effect	Maximum Density Effect
0.39	0.06	0.07	0.40	0.24

The table shows the largest difference in average square-root calling rates among the 10 wire centers due to each variable. Note how much smaller the age and sex effects are compared to the other variables. The percent female heads of household in the 10 study wire centers varied from 13 to 41 percent. The percent heads of household over age 65 varied from 2 to 22 percent, and the percent heads of household between the ages 25 and 34 varied from 12 to 51 percent. If one were to consider an area where these variables vary a great deal more than is shown here, it would be appropriate to keep age and sex in the model.

The seven household size variables can be replaced by the average people per household because the household size coefficients are a linear function of the corresponding household size. The model is adjusted by replacing the household size coefficients with the appropriate linear function of household size and algebraically simplifying the result.

In Cincinnati, age and sex cannot be eliminated because their effect on wire-center average square-root calling rates is of the same magnitude as the other variable effects. This can be seen in the following:

Maximum People/House Effect	Maximum Age Effect	Maximum Sex Effect	Maximum Race Effect
0.40	0.15	0.25	0.19

The table shows the largest difference in average square-root calling rate among the Cincinnati study wire centers. The age and sex effects are relatively more important in Cincinnati because the age and sex coefficients are larger in Cincinnati than in California and the race coefficient is smaller.

As pointed out in Section 4.1, the Cincinnati household size coefficients are not linear. Therefore, the individual household size coefficients were not replaced by an average-people-per-household coefficient. Such a replacement would add error to the model. However, judging from the plot of the coefficients on Fig. 1, the error would not be large. Therefore, if the average number of people per household were the only measure of household size available, the Cincinnati model could be simplified without seriously affecting the fit.

5.2 Average calling rate models

The plots in Figs. 3 and 4 suggest that the relationship between the average square-root calling rate and average calling rate is approximately linear in both California and Cincinnati (at least, for the ranges of calling rates under consideration). This means that both the California and Cincinnati average square-root calling-rate models can be converted into average calling-rate models by linear transformations. The equation

$$\text{AVG } \sqrt{\text{CR}} = 0.227 \text{ AVG CR} + 0.893$$

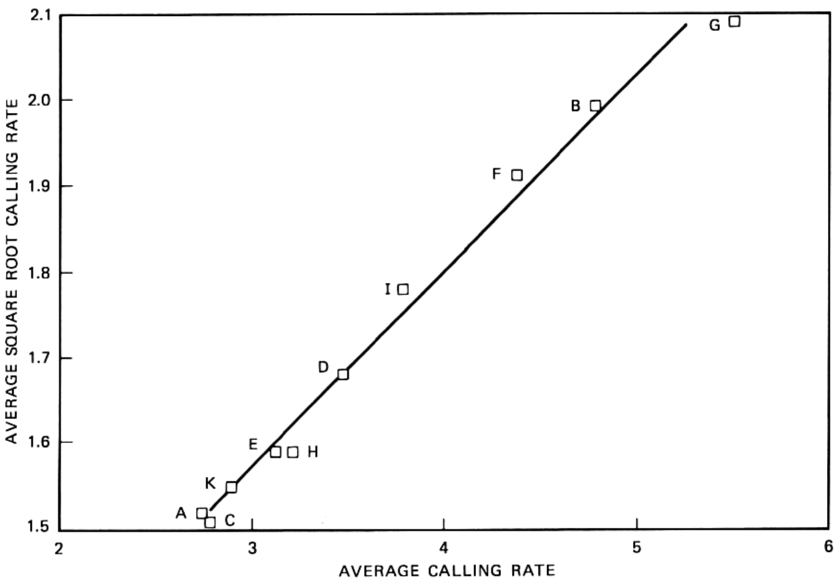

Fig. 3—California square root vs average calling rate.

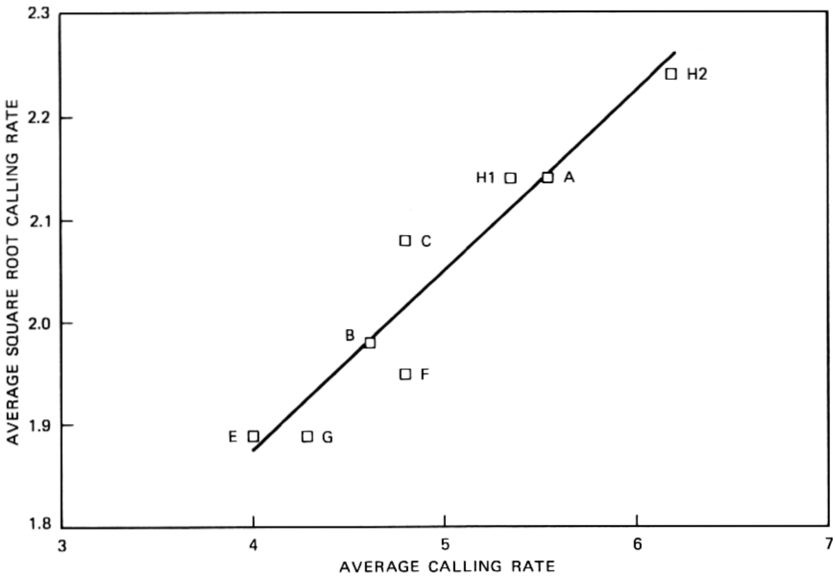

Fig. 4—Cincinnati square root vs average calling rate.

shows the relationship between average calling rate and average square-root calling rate in California. The equation was obtained by fitting a straight line to a plot of the average square-root calling rate versus average calling rate of the sampled telephones in each wire center. The plot is shown in Fig. 3. Substituting this equation into the California average square-root calling-rate model yields the following model for average calling rate.*

$$\text{CR} = -1.34 + \underset{(0.15)}{1.10P} + \underset{(0.28)}{2.30R} + \underset{(0.000036)}{0.000204D}.$$

The parameters P, R, and D are defined in Section 5.1. The numbers in parentheses are standard errors.

The coefficients for the Cincinnati average calling-rate model are shown in Table III. The following equation relates average square-root calling rates to average calling rates in Cincinnati.

$$\text{AVG}\ \sqrt{\text{CR}} = 0.175\ \text{AVG CR} + 1.176.$$

The plot is shown in Fig. 4. The Hamilton wire center is not included

* Since the relationship between average square root calling rate and average calling rate is approximately linear (rather than nonlinear), we could alternatively have estimated the coefficients of the average calling-rate model by directly regressing the individual household calling rates on a linear function of the significant household variables.

in the plot because the EAS/LAS option may cause atypical calling-rate distributions. While Hamilton EAS fits the line reasonably well, Hamilton LAS is an extreme outlier. In this plot, the averages for the two switching systems that serve the same area (West Seventh Street, H1 and H2) were computed and plotted separately.

5.3 Goodness of fit

Figure 5 is a plot of the California model's estimate of the average calling rate versus the actual average calling rate of the sampled telephones in the 10 California wire centers used to develop the model. The estimates for the eight wire centers that were used to develop both the household and environmental components of the model are very good. The Madison (*E*) estimate, on the other hand, is a little high and the Beverly Hills (*B*) estimate is extremely low. When Beverly Hills is excluded, the fraction of the sums of squares of wire-center average calling rate explained by the model (R^2) is 0.93. When both Beverly Hills and Madison are excluded, R^2 is 0.98.

Figure 6 shows how well the Cincinnati model estimates the average calling rate of the sampled telephones in non-Hamilton Cincinnati wire centers. Hamilton EAS and LAS are not included because the model contains dummy variables for Hamilton and because the above linear relationship between average calling rate and average square-root calling rate may not apply to the Hamilton EAS and LAS calling-rate distributions. The fraction of the sums of squares of wire-center average calling rate explained by the model (R^2) is 0.52. This fit is better than the individual telephone model because we are averaging over several telephones in each wire center. On average, we are

Table III—Cincinnati average calling rate model

Variable	Coefficient	Standard Error
2 people	3.33	0.92
3 people	4.27	1.08
4 people	5.32	1.25
5 people	6.82	1.46
6+ people	9.31	1.84
Age 25–34	−1.74	1.23
Age 35–44	−0.26	1.33
Age 45–54	0.23	1.27
Age 55–64	−0.74	1.21
Age 65+	−3.45	1.34
Female	2.56	0.77
Hamilton LAS	−1.58	0.72
Hamilton EAS	1.14	0.76
Black	1.75	0.89

Constant = 0.75

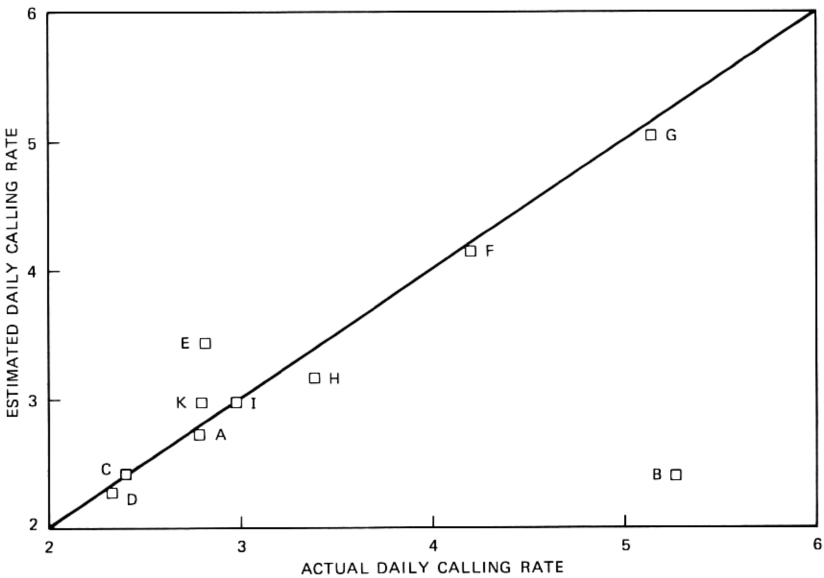

Fig. 5—California estimated vs actual calling rate.

examining 28 telephones per wire center. If we were averaging over more telephones in each wire center, the fit might be better.

5.4 Area to area differences

Figure 7 shows now well the Cincinnati model estimates the average calling rate of the sampled telephones in the eight* California wire centers used to develop both the household and environmental components of the California model. Note that the relative rank of the wire centers according to the model's calling rate estimates is similar to the rank according to the true average calling rates of the sampled telephones. Also note, however, that the Cincinnati model overestimates the level of the California average calling rates. A similar result is obtained when the California model was used to estimate Cincinnati average calling rates. The California model approximately ranks the Cincinnati wire centers but underestimates their average calling rate. The ability to rank the wire centers (according to their average calling rate) is all that is needed for stratifying samples.

The difference between the Cincinnati model's estimate of a wire-center average calling rate and the California model's estimate of the calling rate may be viewed as the "area to area" difference in calling rates between Cincinnati and California. This difference varies with

* The California model was actually developed from 10 wire centers; however, as discussed in Section 3.3, two wire centers were excluded from part of the analysis.

the demographic characteristics of the wire center. The average difference among the 15 wire centers on Figs. 6 and 7 is 1.80 calls per day. The minimum difference is 0.78 and the maximum difference is 2.65 calls per day. In each case, the Cincinnati model's estimate is higher than the California model's estimates. Thus it appears that, given the same household size, age, sex, racial, and telephone density characteristics, Cincinnati wire-center average calling rates are higher than California wire-center average calling rates. The difference varies with the demographic characteristics, but the average is 1.80 calls per day.

VI. CONCLUSION

The results of this study suggest that models similar to those developed here may be a valuable aid in studies of residence average local calling rates in metropolitan areas. The ability of the California and Cincinnati models to predict wire center average calling rates suggests that these models can be used to estimate the average calling rate in the nonsampled wire centers in their respective areas, as discussed in Section I. Furthermore, the fact that the California model preserves the relative rank of the Cincinnati wire centers with respect to the average calling rate (and vice versa) suggests that the household characteristics included in these models may be used to stratify other metropolitan areas into regions of expected high, low, and medium average calling rates. As discussed in Section I, such a stratification

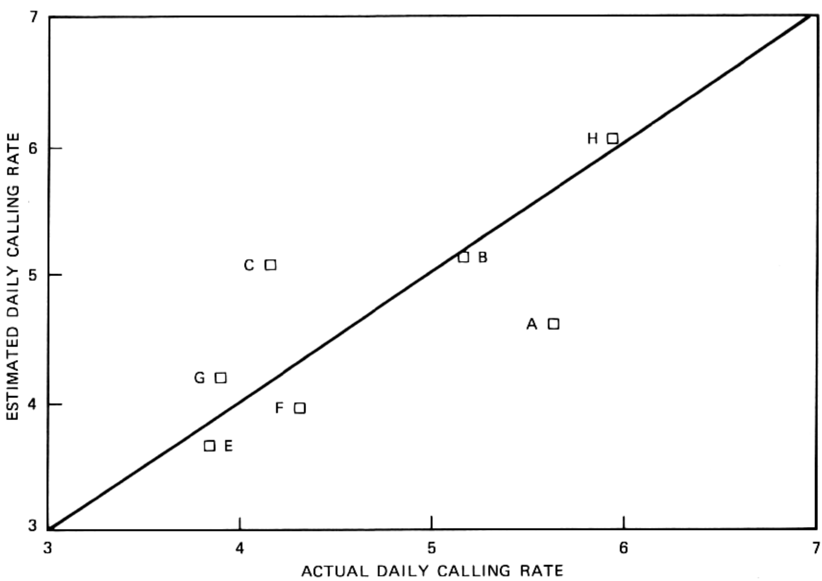

Fig. 6—Cincinnati estimated vs actual calling rate.

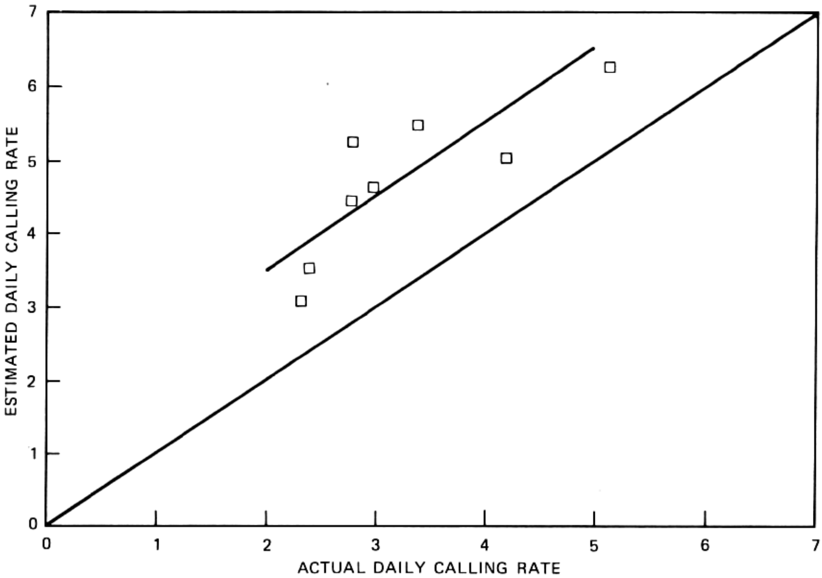

Fig. 7—Cincinnati model vs California calling rate.

could improve the precision of a study to estimate the average calling rate in a metropolitan area.

In addition, the evidence presented here suggests that household characteristics may be a fruitful area of research to obtain a more general understanding of the demand for telephone services. More data and more analysis are needed to study alternative model specifications and to precisely determine the relationships between household characteristics, the demand for telephone services, and the prices of telephone services. In particular, a pooled analysis of data from many different areas is probably needed to determine the cause of the area-to-area difference between California and Cincinnati, correlations between the household variables (multicollinearity) should be investigated, and the interactive effects of household characteristics should be considered. Multicollinearity was not investigated in this paper because our primary purpose was to investigate the feasibility of using household characteristics for prediction rather than to achieve a complete understanding of the relationship between calling rate and household characteristics. The effect of interaction terms was not investigated because it was felt that the values of interaction variables would be difficult (if not impossible) to obtain on a wire-center basis. However, the interaction between race and income was investigated in Cincinnati and found to be not statistically significant.

A potential source of bias in the results presented here lies in the fact that the analysis is limited to the households that returned questionnaires. In both California and Cincinnati, the questionnaire respondents had lower average calling rates than the nonrespondents. The difference was statistically significant at 95-percent confidence in California (3.77 vs 4.40 calls per day) and at 75-percent confidence in Cincinnati (4.75 vs 5.28 calls per day). While this bias is cause for concern, it does not necessarily imply that the model coefficients (or predictions) are biased. For example, suppose calling rates increase linearly with the number of people in the household. If small households are more likely to respond to a questionnaire than large households, then the questionnaire respondents would give an estimate of overall average calling rate that is biased on the low side. Nevertheless, the respondents would give unbiased estimates of the average calling rate among small households and the average calling rate among large households. Thus the respondents' estimate of the coefficient of the linear model that relates calling rate to household size would also be unbiased. In short, a model would not be biased if the respondent calling rates in each socioeconomic subset are representative of that subset.

The following is another example of when a biased (in terms of usage) response does not lead to biased model coefficients. Suppose the calling rate distributions for various household sizes differ only in their means and that, for each household size, the households whose calling rates are in the top 10 percent for that household size do not respond. Under these circumstances, the coefficients of the linear model that relate calling rate to household size would be unbiased, although the constant term would be biased on the low side. Since we do not know whether or not circumstances similar to those illustrated by the above examples apply, we do not know whether or not the model coefficients estimated in this paper are biased.

The reader is cautioned that the differences in household characteristics that are observed to be related to the differences in calling rate may not be the *cause* of the calling-rate differences. Some variables may appear to be significant because they are correlated with other unknown variables which are the real causes of calling rate differences. Furthermore, the wire centers used in this analysis were located in large metropolitan areas. The relationships observed here may not hold in more rural areas.

VII. ACKNOWLEDGMENT

The author wishes to acknowledge the work of Susan Collins of Cincinnati Bell, who participated in the analysis of the Cincinnati data.

REFERENCES

1. C. Pavarini, "Effect of Flat to Measured Rate Conversions on Local Telephone Usage," Conference on Pricing in Regulated Industries: Theory and Regulation, Keystone, Colorado, August 27–30, 1978.
2. K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, 2nd ed., New York: Wiley, 1965, pp. 441–447.
3. C. Daniel and F. S. Wood, *Fitting Equations to Data*, New York: Wiley-Interscience, 1971, Chapter 8.
4. B. B. Brandon, ed., *The Effect of the Demographics of Individual Households on Their Telephone Usage in Chicago*, Cambridge, Mass.: Ballinger, to be published.