Charging for Local Telephone Calls

How Household Characteristics Affect the Distribution of Calls in the GTE Illinois Experiment

Rolla Edward Park, Bridger M. Mitchell, Bruce M. Wetzel, James H. Alleman
This material is based upon work supported by the National Science Foundation under Grant No. DAR-80-12668. Any opinions, findings, and conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Foundation.

Library of Congress Cataloging in Publication Data
Main entry under title:

"R-2535-NEF."
HE8777.045 334.6'3 81-5179
ISBN 0-8330-0313-5 AAC82

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation
R-2535-NSF

Charging for Local Telephone Calls

How Household Characteristics Affect the Distribution of Calls in the GTE Illinois Experiment

Rolla Edward Park, Bridger M. Mitchell, Bruce M. Wetzel, James H. Alleman

March 1981

Prepared under a grant from the National Science Foundation
This report describes research undertaken as a part of a larger study of the allocative and distributive effects of usage-sensitive pricing of local residential telephone service. The research is funded by grants to The Rand Corporation from the National Science Foundation. Earlier versions of the report were presented at the Seventh Annual Telecommunications Policy Research Conference, Skytop, Pennsylvania, April 1979; the Western Economic Association meetings, Las Vegas, Nevada, June 1979; the Pacific Telecommunications Conference, Honolulu, Hawaii, January 1980; and the General Telephone and Electronics Measured Service Demand Workshop, Milford, Connecticut, April 1980.

When the billing of local telephone service is changed from flat rate to measured service, the distribution of monthly calling rates is altered. This report models the distribution of flat-rate telephone usage in terms of demographic variables and stochastic components; the shift to measured service affects both the systematic and stochastic parameters. The model is fitted by maximum likelihood to data for interviewed households participating in General Telephone's local measured service experiment in Illinois. Households tend to make more calls if they are larger (more people), older, or include teenagers. They tend to reduce calling proportionately more in response to usage charges if they average many calls under flat rate for any of the above reasons or for other, unexplained reasons. There is substantial variation in telephone usage by households with similar demographic characteristics. Consequently, the benefits and costs of local measured service will tend to be diffused across demographic groups.
ACKNOWLEDGMENTS

Hans Kraepelien and Willard Manning reviewed drafts of this report and gave us many useful comments and suggestions. Lee Lillard suggested the LISREL formulation of our model. The data we analyze were supplied by Gerald Cohen’s group at GTE Service Corporation. None of them is responsible for the final form of this report. We greatly appreciate all of the help we have received.


## CONTENTS

**PREFACE** ................................................................. iii

**ABSTRACT** ............................................................... v

**ACKNOWLEDGMENTS** ................................................... vii

**PUBLICATIONS OF RAND'S TELEPHONE PRICING PROJECT** ......... ix

**Section**

I. **INTRODUCTION AND SUMMARY** ...................................... 1

II. **AN OVERVIEW OF THE EXPERIMENTAL DATA** ................. 3
   - The Experiment .................................................. 3
   - The Data .......................................................... 3

III. **THE MODEL** ....................................................... 9

IV. **THE ESTIMATES** .................................................... 14
   - Alternative Specifications .................................... 14
   - The Systematic Part of the Equation ........................ 16
   - The Stochastic Part of the Equation ......................... 19

V. **APPLICATIONS OF THE ESTIMATED DISTRIBUTIONS** ........... 23
   - Comparisons of Fitted Distributions with Sample Data .... 23
   - Regression Toward the Mean ................................... 29
   - Large Users Reduce More than Small Users ................ 30
   - Distributions that Incorporate the Effect of Demographics 31
   - Summary ......................................................... 33

**Appendix**

A. **SAMPLING ISSUES** .................................................. 37

B. **TRANSFORMATION OF CALLING RATE DATA** ................. 45

C. **TREATMENT OF MISSING VALUES FOR EXPLANATORY VARIABLES** 54

D. **DEMOGRAPHIC PORTION OF SURVEY INSTRUMENT** ............ 59

E. **APPLICATION OF LISREL TO ESTIMATE THE TELEPHONE CALLING RATE MODEL** 66

F. **ESTIMATES OF FIRST AND SECOND MOMENTS OF FITTED DISTRIBUTIONS OF MONTHLY CALLING RATES** ................. 82

**REFERENCES** ............................................................ 85
I. INTRODUCTION AND SUMMARY

INTRODUCTION

Most residential telephone service in the United States is provided for a flat monthly fee, with no extra charge for calls to telephones within the local area. An alternative, common in the rest of the world and increasingly discussed in the United States, is to levy a charge for each local call or each minute of local calling. This alternative—referred to as "usage sensitive pricing" or "local measured service"—holds out the possibility of being both more efficient and more equitable than flat rate charges. However, not much is known about the effects of switching from flat rate to local measured service. Will the efficiency gains offset the additional costs of measurement and billing? Who gains and who loses from conversion to measured service?

Some beginnings of answers to these broad questions may be found in Alleman (1977), Kraepelien (1976), and Mitchell (1978). These publications are, however, largely pre-empirical. The present work, like that of Pavarini (1979), Brandon (1981), Infosino (1980), and Park and Wetzel (1981), is empirical and more narrowly focused. The questions we attempt to answer here are: What effect does a change from flat rate to measured service have on the level and distribution of residential telephone usage? How is usage, and the change in usage, related to household characteristics?

We use data collected in a particular experimental setting—the GTE local measured service experiment in Illinois. Section II is a brief overview of the experiment and the data. Section III specifies the model that we fitted to the data. The estimation results are in Sec. IV. In Sec. V the results are applied with several purposes in mind, including a demonstration that the fitted distributions describe the data reasonably well and an illustration of the effects of household demographic characteristics on flat rate and measured rate call distributions.
SUMMARY

We fit a model that describes the number of telephone calls a household makes each month in terms of that household's demographic characteristics, whether it pays a usage sensitive charge for telephone calls, and random errors including household-specific components. For 641 interviewed households participating in GTE's local measured service experiment in Illinois, the data include the number of calls made during each of three flat rate months and during the same three months a year later under measured rates.

Under a flat rate the number of calls tends to be larger in households that include more people, are headed by an older person, include teenagers, or have many local acquaintances. These and other less significant demographic characteristics account for 32 percent of the variance in the number of calls.\footnote{As transformed to induce normality.} Persistent differences among households not associated with any of the measured demographic characteristics account for an additional 55 percent of the variance.

When measured rates are imposed, most households tend to reduce the number of calls they make. The reductions tend to be proportionately larger in households that include more people, have lower incomes, are headed by an older person, and include no small children. However, these and other measured characteristics account for only 11 percent of the total variance in households' response to measured rates. Households that consistently make an unusually large number of calls under flat rates (that is, more than one would predict based on their measured characteristics) tend to reduce their calling proportionately more under measured rates, and this accounts for an additional 10 percent of the response variance.

Mandatory measured rates for residential customers will tend to benefit households that do not use their telephones very much and harm those that use them a lot. There is a wide range of telephone use within any demographic group, so the benefits and harms of measured service will tend to be diffused across groups.
II. AN OVERVIEW OF THE EXPERIMENTAL DATA

THE EXPERIMENT

General Telephone and Electronics (GTE), recognizing the need for better information on subscriber demand for telephone calls and related matters, is conducting a local measured service experiment in three small cities in central Illinois—Jacksonville, Clinton, and Tuscola. Starting in May 1975, GTE began to record information on individual customers' pre-experimental telephone use under the flat rate tariffs in effect in those exchanges. Then, on September 1, 1977, GTE switched to the measured service tariffs shown in Table 1 and continued to record usage information.² In contrast to some measured service plans that are available elsewhere, the experiment tariffs are nonoptional and include no allowance of free calls. Residential subscribers can avoid paying for each outgoing call only by downgrading to multi-party service, which is still on flat rate. Our sample includes only households that subscribed to single-party service throughout the period covered by our data.³

THE DATA

We work with a unique set of data from the GTE Illinois experiment. These data combine for the first time demographic information on several hundred households with information on each household's telephone usage under both flat rate and measured rate tariffs.

The demographic information comes from a telephone survey conducted during April 1978. The basic survey sample was drawn as a


²The experimental tariff is not a cost-based tariff. Information on the appropriate marginal costs of telephone use were not available when the tariff was designed.

³For a discussion of the effect of downgrading on aggregate telephone use and the effect of the $19 ceiling on usage charges, see Park and Netzel (1981).
Table 1

GTE EXPERIMENTAL MEASURED SERVICE TARIFFS\textsuperscript{a}
(Residential single-party telephone service)

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Per Month</th>
<th>Per Call</th>
<th>Per Minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacksonville</td>
<td>$3.15</td>
<td>$.02</td>
<td>$.01</td>
</tr>
<tr>
<td>Clinton and Tuscola</td>
<td>2.50</td>
<td>--</td>
<td>.015</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Discounts of 20 percent apply evenings (5-11 p.m.) and Sunday (8 a.m.-11 p.m.) and 50 percent nights (11 p.m.-8 a.m.). There is a $19 ceiling on usage charges per month.

\textsuperscript{b}These are the monthly charges in urban areas; suburban rates are higher. The experimental monthly charge is approximately 40 percent of the pre-experimental flat rate monthly charge.

Systematic sample from a list of single-party telephone lines ranked by their local minutes of use during June, July, and August 1977. This sampling method assured that all usage levels are equally represented in the basic sample, which was then augmented by drawing additional telephone lines from the highest and lowest 10 percent of users. Thus high and low users are overrepresented in the stratified sample. A total of 728 completed interviews resulted.\textsuperscript{1}

Data were incomplete for a portion of the sample. For 102 households missing interview data values were estimated from regression equations for households with complete data.\textsuperscript{2} Eighty-seven households were omitted because of inconsistent or unusable survey data, change in class of service, more than one telephone line, no service during

\textsuperscript{1}Appendix A contains a more complete description of the sampling procedure and a comparison of sample and population data.

\textsuperscript{2}See Appendix C for details. In principle, the observations with missing variables should be downweighted to account for the lesser amount of information they contain; see Dagenais (1973). In the present instance, however, the variables with missing values are of secondary importance in the estimated equation, so the effect of the adjustment would be slight.
part of the sampled months, or zero use in one or more months.\(^1\) After these exclusions, a total of 641 households (3,846 monthly observations) remained for analysis.

We linked the interview data to telephone usage data for six separate months: the last three months under the flat rate tariff (June, July, and August 1977), and the same three months a year later under the measured service tariff (June, July, and August 1978). All of the analysis here concerns use measured as the number of calls per month, not the number of minutes.\(^2\) Although there are large seasonal variations in telephone use, those three summer months appear to be homogeneous.

For our regression results to represent population characteristics, we must reduce the weight given to the observations drawn from the lowest and highest 10 percent of users to undo the effects of oversampling. For each stratum, the appropriate weight is proportional to the ratio of the number of households in that stratum of the population to the number in the sample. We calculate the weights separately for each exchange.\(^3\)

Table 2 shows the weighted average number of calls for our sample households by exchange and by usage category. Overall, calls per household decreased by 8.1 percent from the three flat rate months to the three measured rate months.\(^4\) There is a decided "tilt" to the response, with the largest users showing the largest decreases, and the smallest users actually showing increases. (The increases are small when measured as change in the number of calls, but large

---

\(^1\) Zero users are excluded to make the distribution of the dependent variable used in our analysis approximately normal; see Appendix B for details.

\(^2\) Accurate data on minutes of calling by individual households were not available when this analysis was conducted. Park and Wetzel (1981) analyze aggregate data on both calls and minutes.

\(^3\) See Appendix A for details of the calculation and use of the weights.

\(^4\) This is somewhat less than the reduction estimated in other studies. See Park and Wetzel (1981) for a summary of the other studies and a partial explanation of the differences in results.
Table 2
WEIGHTED AVERAGE NUMBER OF CALLS PER MONTH
FOR THE STRATIFIED SAMPLE

<table>
<thead>
<tr>
<th>Exchange or Stratum</th>
<th>Number of Households</th>
<th>Flat Rate Calls</th>
<th>Measured Rate Calls</th>
<th>Absolute Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacksonville</td>
<td>322</td>
<td>94.4</td>
<td>86.0</td>
<td>-8.4</td>
<td>-8.9</td>
</tr>
<tr>
<td>Clinton</td>
<td>208</td>
<td>86.8</td>
<td>80.0</td>
<td>-6.9</td>
<td>-7.9</td>
</tr>
<tr>
<td>Tuscola</td>
<td>111</td>
<td>67.6</td>
<td>65.1</td>
<td>-2.5</td>
<td>-3.7</td>
</tr>
<tr>
<td>Lowest 10%</td>
<td>75</td>
<td>17.2</td>
<td>21.4</td>
<td>+4.2</td>
<td>+24.7</td>
</tr>
<tr>
<td>Middle 80%</td>
<td>380</td>
<td>82.5</td>
<td>78.7</td>
<td>-3.8</td>
<td>-4.6</td>
</tr>
<tr>
<td>Highest 10%</td>
<td>186</td>
<td>207.5</td>
<td>162.6</td>
<td>-44.9</td>
<td>-21.6</td>
</tr>
<tr>
<td>All</td>
<td>641</td>
<td>88.5</td>
<td>81.4</td>
<td>-7.1</td>
<td>-8.1</td>
</tr>
</tbody>
</table>

SOURCE: Appendix Table A.5.

in percentage terms.) To some extent the tilted response is a "regression toward the mean" phenomenon.1 However, the results reported in Sec. IV and applied in Sec. V establish that there is a real tendency for larger users to make larger percentage reductions in their calling than smaller users.

We are also interested in the distribution of telephone use by different households. Figure 1 shows the distribution of calls by the sample households in Jacksonville during July 1977. The histogram shows the weighted fraction of sample households making between one and ten calls, 11 and 20 calls, and so on. This is a typical skewed distribution, with most households making few calls and a smaller number of households making many calls.

A common way to summarize distributional information is with such statements as, "The highest 10 percent of telephone users make 30 percent of all calls, while the lowest 20 percent of users make

1That is, some households are classified as low or high users because they had unusually low or high minutes of telephone use during the three flat rate months in which the sample was drawn. When these households later return to normal usage levels, they raise the average number of calls by low users and lower the average by high users.
only 4 percent of the calls." A more complete statement of such information is the so-called Lorenz curve, which displays the cumulative fraction of calls made by any fraction of users.\(^1\) A Lorenz curve for calls under flat rate in Jacksonville\(^2\) is plotted in Fig. 2. It shows, for example, that:

- The illustrative statement at the beginning of this paragraph is correct.
- The smallest 10 percent of users make only slightly over 1 percent of calls.
- The largest 1 percent of users make about 5 percent of calls, and the largest 2 percent make about 9 percent.

\(^1\)For a general discussion of the properties of Lorenz curves, see Kakwani (1977).
\(^2\)The curves for measured rate and other exchanges are nearly the same.
Fig. 1 — Weighted fractions of sample households in Jacksonville making different numbers of calls during July 1977

Fig. 2 — Lorenz curve for flat rate calling in Jacksonville (sample distribution)
III. THE MODEL

To motivate the model that we actually estimate we start with the following simple relationship:

\[ C_{it} = \beta_0 + Z_i \beta + \alpha_0 T_t + \epsilon_{it} \]  

(1)

where

- \( C_{it} \) is the number of calls by household \( i \) during month \( t \);
- \( Z_i \) is a vector of variables characterizing household \( i \);
- \( T_t \) is a dummy variable equal to \( 1 \) in the three months during which the measured service tariff is in effect, and zero otherwise;
- \( \epsilon_{it} \) is an independent, identically distributed error term;

and \( \beta_0 \), a vector \( \beta \), and \( \alpha_0 \) are coefficients to be estimated. This simple relationship explains household telephone calling as a function of demographic characteristics and a (presumably downward) shift when the measured service tariff is in effect.

The model that we actually estimate is more complicated than that for the following reasons:

1. We do not expect the error term \( \epsilon_{it} \) to be independently and identically distributed; rather we expect that there are components of error (\( \mu_i \)) specific to each household that persist from month to month. Because of unobserved influences, or because of taste differences, or for whatever reason, a household that makes an inexplicably large number of calls this month will probably do so next month and next year as well.

2. We do not expect all households to react in the same way to the measured service tariff. Their reactions may vary
according to the same demographic variables that determine telephone usage under the flat rate tariff. Thus a second function of \( Z_i \) should replace the constant shift coefficient \( \alpha_0 \).

3. We expect reactions to the measured service tariff to be stochastic. Demographic variables will not explain the reactions perfectly. Thus the shift function should include a household-specific response error \( (v_i) \).

4. The statistical technique we use to estimate our model yields maximum likelihood estimates if the errors are normally distributed. We achieve normally distributed errors by a transformation that makes the distribution of the dependent variable normal. Analysis of these distributions for the Illinois exchanges by methods introduced by Box and Cox (1964) establishes that a power function of the monthly calling rate \( (C_{it}) \) is approximately normally distributed.¹

5. Some or all of the coefficients may differ among the three exchanges in the sample. For notational simplicity we suppress the exchange subscript.

The more complicated equation that we actually estimate, then, is

\[
(C_{it}) = \beta_0 + Z_i \beta + (\alpha_0 + Z_i \alpha) T_t + u_i + v_i T_t + \epsilon_{it}.
\]

(\( i = 1, \ldots 641; \ t = 1, \ldots 6 \))

The first row in (2) is the systematic portion of our model and the second is the stochastic part; we will estimate both. The systematic part describes central tendencies, and the stochastic part, consisting of three error components, describes dispersions.

¹See Appendix B for details. This finding agrees with results reported by Pavarini (1979) for 73 Bell System switching offices.
The demographic variables ultimately used in the model were chosen from those measured in the household survey. The selected variables, based on preliminary estimates of Eq. (2), are

\[
\begin{align*}
\text{HHLDSIZE} & = \text{logarithm of number of persons in the household} \\
\text{SINGLE} & = 1 \text{ for single-person households, 0 otherwise} \\
\text{INCOME} & = \text{logarithm of household income ($000)} \\
\text{AGE} & = \text{age of principal wage earner, measured in scores (20 years)} \\
\text{CHILDREN} & = 1 \text{ if household includes any children 12 or under, 0 otherwise} \\
\text{TEENS} & = 1 \text{ if household includes any teenagers (13-18), 0 otherwise} \\
\text{YOUNGADULTS} & = 1 \text{ if household includes any young adults (19-25), 0 otherwise} \\
\text{FRIENDS} & = 1 \text{ if household reported having many local acquaintances, 0 otherwise} \\
\text{ESTINC} & = 1 \text{ if income value estimated, 0 if reported} \\
\text{ESTAGE} & = 1 \text{ if age value estimated, 0 if reported.}
\end{align*}
\]

The functional forms in which household size, income, and age enter the equation were determined on the basis of residual plots from exploratory regressions.

Equation (2) can be written more compactly as

\[
y = X\gamma + u, \quad E(u'u') = V.
\]

The model is seen to be a linear regression model with error covariance matrix V. The structure of V depends on the correlations among the error components. We assume the following:

---

1. See Appendix D for the relevant portion of the survey instrument. Demographic variables tried and rejected were education, length of residence, and number of automobiles owned. Brandon (1981) finds race to be significant, but race was not included in the GTE questionnaire.
the error components all have zero means;
the error components have constant variance for all households:

\[ E(\mu_i^2) = \sigma_{\mu}^2; \quad E(\nu_i^2) = \sigma_{\nu}^2; \quad E(\varepsilon_{it}^2) = \sigma_{\varepsilon}^2; \]

the error components are uncorrelated across households;
within a household, \( \varepsilon_{it} \) is uncorrelated with \( \mu_i \) and \( \nu_i \);
within a household, \( \varepsilon_{it} \) follows a first order autocorrelated pattern, so that \( E(\varepsilon_{it}\varepsilon_{is}) = \rho |t-s|\sigma_{\varepsilon}^2 \)
within households, \( \mu_i \) and \( \nu_i \) may be correlated:

\[ E(\mu_i, \nu_i) = \sigma_{\mu\nu} = \delta \sigma_{\mu} \sigma_{\nu}. \]

These assumptions imply that the error covariance matrix is block diagonal, \( V = \text{diag}[\Omega, \Omega, \ldots, \Omega] \), where the six by six (month) blocks \( \Omega \) for each household have the structure shown in Table 3. We estimate Eq. (2) by a maximum-likelihood procedure starting from ordinary least squares estimates of the coefficients.\(^1\)

\(^1\)See Appendix E for details.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 1977</td>
<td>$\frac{\sigma^2 + \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \rho \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \rho \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
</tr>
<tr>
<td>Jul 1977</td>
<td>$\frac{\sigma^2 + \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \rho \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
</tr>
<tr>
<td>Aug 1977</td>
<td>$\frac{\sigma^2 + \sigma^2}{\mu \cdot \epsilon}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
<td>$\frac{\sigma^2 + \delta \sigma \cdot \sigma}{\mu \cdot \mu \cdot \nu}$</td>
</tr>
<tr>
<td>Jun 1978</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
</tr>
<tr>
<td>Jul 1978</td>
<td>symmetric</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
</tr>
<tr>
<td>Aug 1978</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
<td>$\frac{\sigma^2 + 2 \delta \sigma \cdot \sigma + \sigma^2}{\mu \cdot \nu + \sigma^2}$</td>
</tr>
</tbody>
</table>

**NOTE:** The autocorrelation of within-household errors $\epsilon$ is assumed to have a negligible effect between August 1977 and June 1978 (that is, $\rho^{10} \sigma^2_\epsilon$ is very small relative to $\sigma^2_\mu + \delta \sigma \cdot \sigma_\mu$), and the assumption is confirmed by the estimate of $\rho = .23$ below.
IV. THE ESTIMATES

ALTERNATIVE SPECIFICATIONS

The calling rate model (2) may be applied to the three experimental exchanges in several ways. It is commonly observed that local telephone use tends to be higher in larger exchanges, presumably because there are more people to call. We test for this and other differences among exchanges by fitting five different specifications:

$H_1$ assumes that the exchanges are identical in all respects and estimates a single set of coefficients and stochastic parameters that applies to all three.

$H_2$ allows the intercept $\beta_0$ to differ to account for possibly different levels of use in the three exchanges, and also allows the tariff reaction coefficient $\alpha_0$ to differ in exchanges that face different tariffs—that is, one coefficient for Jacksonville and another for Clinton and Tuscola.

$H_3$ additionally allows the other coefficients $\beta$ and $\alpha$ to differ among exchanges. Each of the first three specifications constrains the parameters of the error structure to be the same for all three exchanges.

$H_4$ allows all coefficients and stochastic parameters to differ from one exchange to another.

We also estimate a model of calling rate distributions without conditioning on the demographic variables $Z$:

$H_5$ is the same as $H_2$ except the coefficients of all demographic variables are constrained to equal zero. Comparison of specification $H_5$ with $H_2$ allows us to assess the importance of the demographic variables as a group in explaining calling rates.

The following schematic shows the relationships among the various specifications. An arrow is to be read as "is nested in" or "is a more restrictive version of." Thus $H_2$, for example, is constructed by imposing restrictions on $H_3$ or $H_4$. 

A statistic for testing these partially nested hypotheses can be constructed using the value of the likelihood function under each specification. Under the hypothesis $H_i$, minus twice the logarithm of the likelihood ratio of specification $H_i$ to a less restrictive specification $H_j$ is distributed approximately as $\chi^2$ with degrees of freedom equal to the difference in the number of independent parameters estimated under $H_i$ and $H_j$.

Table 4 shows the results of testing the various restrictions. Because we have a fairly large sample, we choose to test at the .01 level.

Table 4

<table>
<thead>
<tr>
<th>Test of Hypothesis $^a$</th>
<th>Against Hypothesis $^a$</th>
<th>$\chi^2$ Statistic</th>
<th>Degrees of Freedom</th>
<th>Critical Value (.01 Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$H_2$</td>
<td>20.3</td>
<td>3</td>
<td>11.3</td>
</tr>
<tr>
<td>$H_2$</td>
<td>$H_3$</td>
<td>54.3</td>
<td>41</td>
<td>64.9</td>
</tr>
<tr>
<td>$H_2$</td>
<td>$H_4$</td>
<td>70.7</td>
<td>51</td>
<td>76.2</td>
</tr>
<tr>
<td>$H_5$</td>
<td>$H_2$</td>
<td>303.5</td>
<td>20</td>
<td>37.6</td>
</tr>
</tbody>
</table>

$^a$All hypotheses constrain the coefficients and stochastic parameters to be the same across all exchanges except:

$H_1$: No exceptions;
$H_2$: Separate intercepts; separate reaction to different tariffs;
$H_3$: Separate intercepts, reactions, and slope coefficients;
$H_4$: All coefficients and parameters separate;
$H_5$: Separate intercepts; separate reactions to different tariffs; coefficients of all demographic variables constrained to equal zero.

$^1$See Jöreskog (1979).
level in an attempt to balance the expected loss from Type I and
Type II errors. We must reject the hypothesis $H_1$ that all coeffi-
cients and stochastic parameters are the same in all exchanges in
favor of $H_2$, which allows the intercepts and reactions to differ.
Removal of any further restrictions, however, is not compelled by
the data at the .01 level. The test of $H_5$ against $H_2$ strongly con-
irms the importance of the demographic characteristics of households
as determinants of calling rates.

**THE SYSTEMATIC PART OF THE EQUATION**

The maximum likelihood estimates of the systematic part of
Eq. (2) for specification $H_2$ are shown in Table 5. The first set
of coefficients characterizes calling under the flat rate tariff.
For the same level of demographic variables, households in Jackson-
ville make more calls than do those in Clinton and Tuscola. The
coefficients of the demographic variables measure the partial
effect of each variable on monthly calling. They show:

1. The number of calls that a household makes each month is
   strongly dependent on its size (number of people in the
   household);\(^1\)

2. Calling increases with the age of the household head.

3. Households with teenagers make significantly more local
calls.\(^2\)

\(^1\)The effect of number of people in the household is measured by
a combination of our variables HHLDSIZE and SINGLE as \( .667 \times \ln(\text{number of persons}) + .304 \times (1\ \text{if single person household,}
\ 0\ \text{otherwise}) \). The estimated effects are as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Calculation</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( .668 \times \ln(1) + .304 )</td>
<td>.304</td>
</tr>
<tr>
<td>2</td>
<td>( .668 \times \ln(2) )</td>
<td>.463</td>
</tr>
<tr>
<td>3</td>
<td>( .668 \times \ln(3) )</td>
<td>.734</td>
</tr>
</tbody>
</table>

etc.

\(^2\)Readers familiar with our Pacific Telecommunications Conference
paper (1980) will note that this result differs from the surprising
negative effect of teenagers on calling rates reported there. The
earlier result was due to a computational error. We mistakenly pro-
gressed the computer to read the wrong data element; what we labeled
TEENS in the earlier paper was in fact CHILDREN.
Table 5
REGRESSION RESULTS

<table>
<thead>
<tr>
<th>Flat Rate</th>
<th>Measured Rate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-statistic</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td></td>
</tr>
<tr>
<td>Jacksonville</td>
<td>2.116</td>
<td>11.5</td>
</tr>
<tr>
<td>Clinton</td>
<td>1.964</td>
<td>10.3</td>
</tr>
<tr>
<td>Tuscola</td>
<td>1.841</td>
<td>9.5</td>
</tr>
<tr>
<td>HHLDSIZE</td>
<td>.668</td>
<td>5.3</td>
</tr>
<tr>
<td>SINGLE</td>
<td>.304</td>
<td>2.6</td>
</tr>
<tr>
<td>INCOME</td>
<td>-.004</td>
<td>1.1</td>
</tr>
<tr>
<td>AGE</td>
<td>.068</td>
<td>1.9</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>.027</td>
<td>.3</td>
</tr>
<tr>
<td>TEENS</td>
<td>.500</td>
<td>6.1</td>
</tr>
<tr>
<td>YOUNGADULTS</td>
<td>-.004</td>
<td>.1</td>
</tr>
<tr>
<td>FRIENDS</td>
<td>.196</td>
<td>2.5</td>
</tr>
<tr>
<td>ESTINC</td>
<td>.125</td>
<td>1.8</td>
</tr>
<tr>
<td>ESTAGE</td>
<td>-.127</td>
<td>.9</td>
</tr>
</tbody>
</table>

H$_2$: With Demographics

H$_5$: Without Demographics

---

*a Dependent variable is (calls per month)$^{27}$.\)
4. Households reporting many acquaintances in the local community make significantly more local calls.

The second set of coefficients characterizes the systematic reaction to the measured service tariff. They indicate that

1. Large households reduce their calling significantly more than do smaller households;
2. Higher income households may reduce calling less than those with lower incomes, but the difference is not significant;
3. Households with older heads reduce calling more than younger households;
4. Households with children make significantly smaller reductions in calling than households with no children present.

There are several points to make about the tariff reaction coefficients $\alpha_0 = .109$ for Jacksonville and $\alpha_0 = .171$ for Clinton and Tuscola. First, their relative magnitudes are consistent with those in other empirical work (Park and Wetzel, 1981) indicating that the measured service tariff in Jacksonville, which has a 2-cent charge for the initiation of each call, should reduce the number of calls more than does the tariff for Clinton and Tuscola, which charges only for the number of minutes of conversation. Second, the fact that we estimate these coefficients to be positive is not necessarily inconsistent with our expectation that the effect of the measured service tariff should be a reduction in calling. The $\alpha_0$ coefficients are only one part of the calculation that determines the expected reaction of any particular household.\(^1\) Third, the measured rate response coefficients pick up the effect of the measured rate plus all other systematic influences on calling rates that changed between summer 1977.

\(^{1}\) For example, a four-person household in Jacksonville, with 40-year-old head, two teenagers, and $20,000$ income, would have an expected reaction to measured service given by $.109 - .139 \ln(4) + .025 \ln(20) - .057 \ln(2) - .049 = -.110.$
and summer 1978. \(^1\) Thus it is not surprising to find that the net effect is estimated to be positive for some types of households.

Table 5 also shows an estimate of \(H_5\), which is not conditional on the demographic variables. Although demographics are clearly important, it is convenient for some purposes to know the unconditional calling rate distribution. We shall, for example, make extensive use of the unconditional estimates in the concluding section of this report.

**THE STOCHASTIC PART OF THE EQUATION**

Table 6 shows our estimates of the five underlying stochastic parameters of the covariance matrix \(\Omega\). The between-household variance \(\sigma^2_\mu\) is about four times as large as the within-household variance \(\sigma^2_\varepsilon\), confirming our expectation that high or low calling rates for particular households that are not explained by measured household characteristics tend to persist over time. The variance of the household response error \(\sigma^2_\nu\) is somewhat smaller than the within-household variance. The first order autocorrelation coefficient for the within-household errors \(\varepsilon\) is .230. The household-specific error component \(\mu\) and the response error \(\nu\) are negatively correlated (\(\delta = -.334\)), indicating that households with unexplained high calling rates tend to reduce calling more in response to measured service than do households with unexplained low calling rates.

Under \(H_5\), the effects of household demographic characteristics are thrown into the household-specific error \(\mu\) and (to a lesser extent) the household-specific reaction \(\nu\). The household-specific error variance \(\sigma^2_\mu\), which is .307 under \(H_2\), increases to .476 under \(H_5\).

The total variance in transformed calls per month (\(c^{.27}\)) can be decomposed as shown in Table 7. \(^2\) Under both flat rates and measured rates, most of the variance in calling (at least 31 + 55 = 86 percent) is due to persistent differences between households. A little

---

\(^1\)These may include weather, economic conditions, or secular trends in telephone use.

\(^2\)This illuminating way to present variance results is due to Lillard and Acton (1980).


Table 6
ESTIMATES OF THE STOCHASTIC PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>$H_2$: With Demographics</th>
<th>$H_5$: Without Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estimate</td>
<td>t-statistic</td>
</tr>
<tr>
<td>$\sigma^2_{\varepsilon}$</td>
<td>within-household error</td>
<td>.070</td>
<td>27.5</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon \varepsilon-1}$</td>
<td>lagged covariance of $\varepsilon$</td>
<td>.016</td>
<td>7.0</td>
</tr>
<tr>
<td>$\rho$</td>
<td>implies first order autocorrelation of $\varepsilon$</td>
<td>.230</td>
<td>(a)</td>
</tr>
<tr>
<td>$\sigma^2_{\mu}$</td>
<td>household-specific error</td>
<td>.307</td>
<td>16.1</td>
</tr>
<tr>
<td>$\sigma^2_{\nu}$</td>
<td>household-specific reaction to measured service</td>
<td>.055</td>
<td>7.5</td>
</tr>
<tr>
<td>$\sigma_{\mu \nu}$</td>
<td>covariance of $\mu$ and $\nu$</td>
<td>-.044</td>
<td>5.1</td>
</tr>
<tr>
<td>$\delta$</td>
<td>implied correlation of $\mu$ and $\nu$</td>
<td>-.334</td>
<td>(a)</td>
</tr>
</tbody>
</table>

*The parameters $\rho$ and $\delta$ were not estimated directly but were calculated as $\rho = \sigma_{\varepsilon \varepsilon-1}/\sigma^2_{\varepsilon}$ and $\delta = \sigma_{\mu \nu}/(\sigma^2_{\mu} \sigma^2_{\nu})^{1/2}$. Thus t-statistics are not available. When we estimated $\rho$ directly (under $H_2$) using the alternative computational procedure described in Appendix E, we obtained $\rho = -.235$ with a t-statistic of 4.9.*
Table 7

VARIANCE DECOMPOSITIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Magnitude</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Total Variance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta' \Sigma \beta$</td>
<td>Due to demographics</td>
<td>.179</td>
<td>32</td>
</tr>
<tr>
<td>$\sigma^2_{\mu}$</td>
<td>Persistent household-specific error</td>
<td>.307</td>
<td>55</td>
</tr>
<tr>
<td>$\rho^2 \sigma^2_{\epsilon}$</td>
<td>Due to autocorrelation</td>
<td>.004</td>
<td>1</td>
</tr>
<tr>
<td>$(1-\rho^2) \sigma^2_{\epsilon}$</td>
<td>Purely random</td>
<td>.066</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>.556</td>
<td>100</td>
</tr>
<tr>
<td>Measured rates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\beta+\alpha)' \Sigma (\beta+\alpha)$</td>
<td>Due to demographics</td>
<td>.151</td>
<td>31</td>
</tr>
<tr>
<td>$\sigma^2_{\mu} + 2\sigma^2_{\mu \nu} + \sigma^2_{\nu}$</td>
<td>Persistent household-specific error</td>
<td>.274</td>
<td>55</td>
</tr>
<tr>
<td>$\rho^2 \sigma^2_{\epsilon}$</td>
<td>Due to autocorrelation</td>
<td>.004</td>
<td>1</td>
</tr>
<tr>
<td>$(1-\rho^2) \sigma^2_{\epsilon}$</td>
<td>Purely random</td>
<td>.066</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>.495</td>
<td>100</td>
</tr>
<tr>
<td>b. Response Variance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha' \Sigma \alpha$</td>
<td>Due to demographics</td>
<td>.007</td>
<td>11</td>
</tr>
<tr>
<td>$\delta^2 \sigma^2_{\nu}$</td>
<td>Due to correlation with household-specific error $\mu$</td>
<td>.006</td>
<td>10</td>
</tr>
<tr>
<td>$(1-\delta^2) \sigma^2_{\nu}$</td>
<td>Remainder</td>
<td>.049</td>
<td>79</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>.062</td>
<td>100</td>
</tr>
</tbody>
</table>

$^a$S is the weighted sample covariance matrix of explanatory variables, including exchange dummies; $\beta$ and $\alpha$ include the individual exchange intercept and reaction coefficients $\beta_0$ and $\alpha_0$. 
over one-third of the persistent variance (31/86 = 36 percent) is associated with measured demographic characteristics. The response to measured rates in our model is given by the terms $a_0 + Z_1^\alpha + v_1$ in Eq. (2). Only 11 percent of the variance in this quantity (the "response variance") is associated with measured demographic characteristics, and another 10 percent is associated with the level of the household-specific flat rate error $\mu$. 
V. APPLICATIONS OF THE ESTIMATED DISTRIBUTIONS

Tables 5 and 6 contain estimates of telephone calling rate distributions (flat rate and measured rate, with and without demographic variables) that we shall apply in this section. The applications here are meant both to clarify the nature and properties of the distributions and to answer some specific questions. Specifically, we shall

- Demonstrate that the fitted distributions do a reasonably good job of reproducing the data;
- Illustrate the effect of regression toward the mean;
- Demonstrate that most of the aggregate reduction in use is due to large users;
- Illustrate the effects of various demographic characteristics on flat rate and measured rate call distributions.

Except in the final subsection, we work with the distributions without demographic variables because they let us summarize results for an entire exchange, not just households with a particular set of measured demographic characteristics.

Throughout this section, we shall repeatedly make the shift from some normal distribution of transformed calls $C^{27}$ to the corresponding skewed distribution of untransformed calls $C$. For plotting distributions, we must multiply the normal probability density function (pdf) for transformed calls by the Jacobian $3C^{27}/3C$ to get the pdf for untransformed calls. For calculating summary statistics (mean and standard deviation) for untransformed calls, we use a straightforward extension of approximation formulas found in Huang and Grawe (1980); see Appendix F for details.

COMPARISONS OF FITTED DISTRIBUTIONS WITH SAMPLE DATA

Here we shall use the fitted distributions to "reproduce" the various summary descriptions of sample data presented in Sec. II
above. The purpose is to develop some confidence that the fitted distributions describe the sample data reasonably well, and incidentally to illustrate some of the ways in which the fitted distributions can be manipulated.

Fitted Distribution for Flat Rate Calling in Jacksonville

The distribution of transformed calling \( C' \) is given by our estimate of Eq. (2) under \( H_5 \) (without demographics). The flat rate distribution for Jacksonville is normal with mean \( \beta_0 = 3.191 \) (from Table 5) and variance \( \sigma_F^2 = \sigma_\mu^2 + \sigma_\varepsilon^2 = .546 \) (from Table 6). The corresponding untransformed distribution is plotted as the smooth curve in Fig. 3. This fitted distribution shows the relative frequency with which households in Jacksonville would be expected to make different numbers of calls under flat rate. For example, we would expect to find about twice as many households making between 50 and 60 local calls a month as make between 110 and 120 calls per month. The fitted distribution smoothes out the irregularities of the sample histogram but reproduces the general shape of the sample distribution very well.

Fitted Lorenz Curve for Flat Rate Calling in Jacksonville

The fitted distribution plotted in Fig. 3 can be used to calculate a Lorenz curve comparable to that for sample data shown before in Fig. 2. The Lorenz curves for the sample and the fitted distribution, which are both plotted in Fig. 4, are nearly indistinguishable.

Fitted Mean Calling Rates by Exchange

The mean calling rate calculated from the fitted distribution for flat rate calls in Jacksonville (Fig. 3) is 93.6. We make similar calculations for the other exchanges and for measured rates as well as flat rates. (The measured rate means are calculated for distributions of transformed calls that are normal with mean \( \beta_0 + \alpha_0 = 3.114 \) for Jacksonville, and variance \( \sigma^2_M = \sigma^2_\mu + 2\sigma_\mu\nu + \sigma^2_\nu + \sigma^2_\varepsilon = .490 \).) The results are shown in Table 8, along with weighted sample
Fig. 3 — Fitted (solid curve) and sample (bar graph) distributions of households in Jacksonville by number of calls during July 1977

Fig. 4 — Lorenz curves for first rate calling in Jacksonville (fitted versus sample distribution)
means and weighted standard errors of the sample means for comparison. In no case does the fitted mean differ from the sample mean by as much as one standard error.

**Fitted Mean Calling Rates by Usage Category**

So far we have compared fitted with sample values for all except one of the data descriptions in Sec. II, and the correspondence has been very close. The one remaining item is mean calls by usage stratum, sample values for which are shown in the bottom part of Table 2. The calculation of fitted means by usage category is more complicated than it was by exchange, because we have not estimated separate distributions for the different usage strata.

Our approach is to predict measured rate usage separately for each household, based on that household's observed flat rate usage. We then average over all households assigned to a particular usage stratum ¹ to get a fitted number of measured rate calls to compare with the sample value. Transformed flat rate and measured rate calls are jointly distributed as bivariate normal with correlation coefficient

\[
\rho^* = \frac{\sigma^2 + \sigma_{uv}}{\sigma_F \sigma_M},
\]

or about .804 using estimates under H₅ from Table 6. The high positive correlation means that a household making a large number of calls during a particular flat rate month will probably make many calls in a measured rate month as well. In fact, we can calculate the expected number of measured rate calls for a household that made any specified number of flat rate calls, using the properties of the bivariate normal distribution.² The distribution of transformed calls under

¹ Remember that households are assigned to usage strata based on average minutes of use in June, July, and August 1977, so that a given household is always counted in the same stratum regardless of the number of calls it makes during a particular month.

² For the properties of the bivariate normal distribution, see, for example, Bennett and Franklin (1954, pp. 128-132).
Table 8
COMPARISON OF FITTED AND SAMPLE\textsuperscript{a} MEAN CALLS PER MONTH BY EXCHANGE

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Flat Rate Calls</th>
<th>Measured Rate Calls</th>
<th>Absolute Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacksonville</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>93.6</td>
<td>84.4</td>
<td>-9.2</td>
<td>-9.8</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>94.4</td>
<td>86.0</td>
<td>-8.4</td>
<td>-8.9</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(4.1)</td>
<td>(3.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinton</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>84.8</td>
<td>80.7</td>
<td>-4.1</td>
<td>-4.8</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>86.8</td>
<td>80.0</td>
<td>-6.9</td>
<td>-7.9</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(5.7)</td>
<td>(4.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuscola</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted means</td>
<td>71.3</td>
<td>67.6</td>
<td>-3.7</td>
<td>-5.2</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>67.6</td>
<td>65.1</td>
<td>-2.5</td>
<td>-3.7</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(5.5)</td>
<td>(5.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted means</td>
<td>88.0</td>
<td>81.0</td>
<td>-7.0</td>
<td>-8.0</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>88.5</td>
<td>81.4</td>
<td>-7.1</td>
<td>-8.1</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(3.0)</td>
<td>(2.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Sample standard errors are calculated for 3-month averages of individual household calling rates.
measured rate, conditional on a particular number of transformed flat rate calls, $x$, is normal with mean

$$\beta_0 + \alpha_0 + \rho x \left( \frac{\sigma_M}{\sigma_F} \right) (x - \beta_0)$$

and variance $(1 - \rho^2) \sigma_M^2$.

We take the observed flat rate use for each household each month as given and calculate expected measured rate use for each, based on the conditional distribution described in the previous paragraph. Weighted averages by usage stratum are then calculated in exactly the same way as they were for actual measured rate calls for Table 2. The resulting values ("fitted means" for measured rate calls) are compared with sample values (from Table 2) in Table 9. The fitted and sample values are very close for the middle and high usage categories,

Table 9

<table>
<thead>
<tr>
<th>Usage Category</th>
<th>Flat Rate Calls</th>
<th>Measured Rate Calls</th>
<th>Absolute Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest 10 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>--</td>
<td>25.5</td>
<td>+8.3</td>
<td>+48.3</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>17.2</td>
<td>21.4</td>
<td>+4.2</td>
<td>+24.7</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(1.4)</td>
<td>(1.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle 80 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted means</td>
<td>--</td>
<td>78.1</td>
<td>-4.4</td>
<td>-5.3</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>82.5</td>
<td>78.7</td>
<td>-3.8</td>
<td>-4.6</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(2.4)</td>
<td>(2.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest 10 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted means</td>
<td>--</td>
<td>161.9</td>
<td>-45.6</td>
<td>-22.0</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>207.5</td>
<td>162.6</td>
<td>-44.9</td>
<td>-21.6</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(14.6)</td>
<td>(12.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted means</td>
<td>--</td>
<td>81.2</td>
<td>-7.3</td>
<td>-8.2</td>
</tr>
<tr>
<td>Weighted sample means</td>
<td>88.5</td>
<td>81.4</td>
<td>-7.1</td>
<td>-8.1</td>
</tr>
<tr>
<td>Weighted standard errors</td>
<td>(3.0)</td>
<td>(2.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
but the fitted mean for the lowest usage category is significantly larger than the sample mean. This is probably a consequence of our model assumption that the household-specific error $\mu$ and the household-specific reaction $\nu$ are linearly correlated.

REGRESSION TOWARD THE MEAN

Another use for the bivariate distribution with which we constructed Table 9 is to illustrate the effect of regression toward the mean. Assume that the introduction of measured rates had no effect on calling rate distributions. That is, $a_0$, $\delta$, and $\sigma^2_\nu$ all equal zero, so that the measured rate distribution is identical to the flat rate distribution. But we shall see that there is still a regression toward the mean effect. Under these conditions, the distribution of transformed measured rate calls conditional on $x$ transformed flat rate calls is normal with mean $\beta_0 + \rho^*(x - \beta_0)$ and variance $(1 - \rho^2_\mu) \sigma^2_M$. The correlation $\rho^*$ is now $\sigma^2_\mu/(\sigma^2_\mu + \sigma^2_\epsilon)$, or about .894 for our estimates. The correlation $\rho^*$ acts as a shrinkage factor that tends to result in conditional measured rate use that is closer to the mean than is the flat rate use on which it is conditioned. For any flat rate use $x$ greater than the mean $\beta_0$, the mean of the conditional measured rate usage distribution is less than $x$, and conversely for $x$ less than the mean.

We can illustrate the effect of regression toward the mean in this model where nothing else is happening by applying the model in the same way we did with the full model to construct Table 9 in the previous subsection. The results are tabulated, along with sample values for comparison, in Table 10.

The results show no change in overall average use, consistent with the "nothing happening" model, but a decided regression toward the mean effect. The calculated expected increase by the lowest 10 percent of users (6.0 calls per month) more than accounts for the increase observed in the sample usage by that stratum (4.2 calls per month). For the highest 10 percent, regression toward the mean
Table 10

ILLUSTRATION OF REGRESSION TOWARD THE MEAN
IN A "NOTHING-HAPPENING" MODEL
(Calls per month)

<table>
<thead>
<tr>
<th>Usage Category</th>
<th>Flat Rate Calls</th>
<th>Measured Rate Calls</th>
<th>Absolute Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest 10 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>--</td>
<td>23.2</td>
<td>+6.0</td>
<td>+34.9</td>
</tr>
<tr>
<td>Weighted sample mean</td>
<td>17.2</td>
<td>21.4</td>
<td>+4.2</td>
<td>+24.7</td>
</tr>
<tr>
<td>Middle 80 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>--</td>
<td>84.0</td>
<td>+1.5</td>
<td>+1.8</td>
</tr>
<tr>
<td>Weighted sample mean</td>
<td>82.5</td>
<td>78.7</td>
<td>-3.8</td>
<td>-4.6</td>
</tr>
<tr>
<td>Highest 10 percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted mean</td>
<td>--</td>
<td>188.3</td>
<td>-19.2</td>
<td>-9.3</td>
</tr>
<tr>
<td>Weighted sample mean</td>
<td>207.5</td>
<td>162.6</td>
<td>-44.9</td>
<td>-21.6</td>
</tr>
</tbody>
</table>

Weighted average

| Fitted mean             | --              | 88.3                | -0.2            | -0.2           |
| Weighted sample mean    | 88.5            | 81.4                | -7.1            | -8.1           |

with nothing else happening would account for slightly more than 40 percent of the observed reduction.¹

LARGE USERS REDUCE MORE THAN SMALL USERS

We have noted before that high usage households tend to reduce their calling more in response to measured rates than do low usage households. In the sample data (Table 2), however, this effect is confounded with the regression toward the mean effect. Our fitted distributions provide a way to quantify the tendency of large users to reduce more, apart from the regression toward the mean effect.

Regression toward the mean arises because of month-to-month variation in calls by particular households (σ²ε). The procedure that we use here integrates out month-to-month variation and focuses on long-run average flat rate and measured rate use. Consider a set of

¹19.2/44.9 = 43 percent.
households with a particular value of the household-specific error $\mu_i$. (A high value of $\mu_i$ will correspond to high usage households.) Their transformed calls in a particular flat rate month are normally distributed with mean $\beta_0 + \mu_i$ and variance $\sigma^2_{\varepsilon}$; the mean of the corresponding untransformed distribution is their long-run average flat rate calling rate.

We can find the long-run average measured rate calling by these same households by using the distribution of transformed measured rate calls conditional on $\mu_i$. This distribution is normal with mean $\beta_0 + \alpha_0 + \mu_i + \delta(\sigma_v/\mu_i)\mu_i$ and variance $(1 - \delta^2)\sigma^2_v + \sigma^2_{\varepsilon}$. The mean of the untransformed distribution is long-run average measured rate calling.

If we do the calculations for many different values of $\mu_i$, and plot the average change in calling against long-run flat rate calling, we get the curve plotted in Fig. 5 for Jacksonville. Not only do large flat rate users make larger absolute reductions under measured rates than do smaller users in Fig. 5, they make larger proportionate reductions.\footnote{We used the relationship plotted in Fig. 5 in an analysis of the effects of optional measured service; see Mitchell and Park (1981).} Thus, for example, a household with a long-run average usage of 300 flat rate calls per month would be expected to cut back by almost 18 percent under measured rates, while one with 100 flat rate average calls would only cut back by less than 9 percent. Even after elimination of the regression toward the mean effect, we find that the fitted model predicts a small increase in usage by small users, because of the linear correlation between the household-specific errors $\mu$ and $\nu$.

**Distributions that Incorporate the Effect of Demographics**

Here we shift our attention to the distributions estimated under $H_2$: with demographics. This distribution of transformed calls is normal, with flat rate mean equal to $\beta_0 + Z\beta$ and variance $\sigma^2_{\varepsilon}$, and measured rate mean of $\beta_0 + \alpha_0 + Z\beta + Z\alpha$ and variance $\sigma^2_{M}$. We plot the untransformed distributions in Fig. 6 for households with "base case"
Fig. 5 — Change in monthly calls due to measured rate as a function of average calls per month under flat rate for Jacksonville

Fig. 6 — Distribution of households by their calling rates, for households with base-case characteristics
characteristics in Jacksonville. Base case characteristics are:

- three persons
- $20,000 income
- 40-year-old head
- no children, teenagers, or young adults
- many friends.

(This is not meant to be a typical household--one wonders, for example, how old the third person is--but only to form a point of departure for the investigation of the effects on the distribution of each variable separately.) The nature of the response to measured rates is typical of a general pattern--a shift to the left and a reduction in variance, both primarily due to the decrease in relative frequency of large numbers of calls.

Perhaps the most important thing to observe in Fig. 6 is the considerable variation in calling rates by households with identical demographic characteristics. If measured service benefits small users and harms large users, the benefits and harms will tend to be diffused across demographic groups, because both small and large users will be found in all groups.

One could plot similar distributions for households with any chosen set of characteristics. However, the distributions all have similar shapes, so the differences among them can be conveniently and compactly summarized by tabulating their means and standard deviations. We do so in Table 11 for a variety of cases chosen to illustrate the separate effects of each of the variables.

The number of people in the household and whether any of them are teenagers are clearly the most influential of the demographic characteristics in the model. The general effect of the introduction of measured service is to reduce both the mean and the standard deviation of the calling rate distributions.

SUMMARY

The major points made in this section may be summarized as follows:
Table 11
EFFECTS OF DEPARTURES FROM BASE CASE CHARACTERISTICS\textsuperscript{a}
ON CALLING RATE DISTRIBUTIONS

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Flat Rate</th>
<th>Measured Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>(H_3): With Demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case</td>
<td>85.2</td>
<td>59.3</td>
</tr>
<tr>
<td>Clinton</td>
<td>72.2</td>
<td>52.6</td>
</tr>
<tr>
<td>Tuscola</td>
<td>62.6</td>
<td>47.4</td>
</tr>
<tr>
<td>Single person</td>
<td>52.3</td>
<td>41.7</td>
</tr>
<tr>
<td>Six people</td>
<td>135.7</td>
<td>83.0</td>
</tr>
<tr>
<td>$10,000</td>
<td>85.6</td>
<td>59.4</td>
</tr>
<tr>
<td>$40,000</td>
<td>84.9</td>
<td>59.1</td>
</tr>
<tr>
<td>20-year-old head</td>
<td>79.2</td>
<td>56.2</td>
</tr>
<tr>
<td>60-year-old head</td>
<td>91.6</td>
<td>62.4</td>
</tr>
<tr>
<td>Children present</td>
<td>87.7</td>
<td>60.5</td>
</tr>
<tr>
<td>Teenagers present</td>
<td>140.7</td>
<td>85.2</td>
</tr>
<tr>
<td>Young adults present</td>
<td>85.2</td>
<td>59.3</td>
</tr>
<tr>
<td>Friends absent</td>
<td>68.7</td>
<td>50.7</td>
</tr>
<tr>
<td>(H_7): Without Demographics \textsuperscript{b}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacksonville</td>
<td>93.6</td>
<td>76.8</td>
</tr>
</tbody>
</table>

\textsuperscript{a}See the text for base case assumptions. The other cases depart from base case assumptions in only one dimension each, as indicated by their labels. Cases are not meant to be representative of typical households but only to illustrate the effects of each variable separately.

\textsuperscript{b}The coefficient of variation of approximately .8 under both flat and measured rate agrees with Pavarini's (1979) result.
1. The model fits the data quite well.
2. Customers whose long-run average telephone usage under flat rates is high reduce their calling under measured rates by a larger proportion than do smaller users.
3. There is considerable variation in telephone use by households with similar demographic characteristics. Consequently, the gains and losses from local measured service will tend to be diffused across demographic groups.
Appendix A
SAMPLING ISSUES

Opinion Research Corporation of Princeton has conducted four telephone surveys for GTE in the three Illinois test exchanges:

Wave I: September 27-October 8, 1975
Wave II: December 3-13, 1975
Wave III: April 12-22, 1978

The third sample, which is the one used here, included 728 residential customers.

The sample frame constructed by GTE consisted of 9585 telephone lines: all those (9792) that had been in use by the same single party residential telephone subscriber for the 10 months from May 1977 through February 1978, except for 207 lines judged to have erratic usage. The 10-month criterion assured that usage data under both flat rate and measured service would be available for all surveyed households. The exclusion of erratic users was intended to focus attention on households with "typical" usage patterns. Table A.1 shows how erratic usage was defined.

Households in the sample frame were sorted by exchange and, within each exchange, rank ordered by their average monthly minutes of telephone use during June, July, and August 1977. For each exchange i, an interval size $n_i$ was chosen so that $1/n_i$ times the number

1Because some households subscribe to more than one telephone line (e.g., a separate line for teenage children), a sample of lines is not exactly equivalent to a sample of households that subscribe to telephone service. However, only approximately 4 percent of all residential subscribers in the three exchanges have more than one line, so the distinction is unimportant for most purposes; when it does not matter, we frequently use the word "households," when to be precise (but cumbersome) we would have to use "telephone lines" or some more elaborate circumlocution.
of sample frame households in the exchange would yield the target number of completed interviews (taking into account the expected completion rate). A random starting point was picked within the first interval and every nth household was picked after that to create the basic sample.

In addition to the basic sample, a sample was drawn from households in the lowest and highest usage deciles. Low and high users were defined separately for each exchange in terms of their average monthly minutes of use during June, July, and August 1977, as shown in Table A.2.

Completed interviews total 728 households. Of these 87 were dropped for a variety of reasons:

- There were two inconsistent interviews for each of four telephone numbers (eight households dropped).
- Three telephone numbers changed from regular residential tariff to a business or concession tariff (three households dropped).
- Some telephone numbers were not in service during the entire
Table A.2
DEFINITION OF HIGH AND LOW USAGE
(Average monthly minutes of use during
June, July, and August 1977)

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Lowest 10 Percent</th>
<th>Highest 10 Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacksonville</td>
<td>less than 45 minutes</td>
<td>more than 775 minutes</td>
</tr>
<tr>
<td>Clinton</td>
<td>less than 38 minutes</td>
<td>more than 685 minutes</td>
</tr>
<tr>
<td>Tuscola</td>
<td>less than 23 minutes</td>
<td>more than 545 minutes</td>
</tr>
</tbody>
</table>

six months June, July, August 1977 and June, July, August 1978 (28 households dropped).

- One interview was judged to be unusable because the respondent refused to supply data on six of the variables included in the initial set used in our analysis (one household dropped). See Tables C.1 and C.2.
- Households with zero use during one or more months were dropped to make the power function transformation of monthly calls approximately normally distributed (20 households dropped). See Appendix B.
- Households with more than one telephone line were dropped because our data file does not include usage data for the additional line(s) (27 households dropped).

A total of 641 households remained for analysis.

1 As determined by a "yes" answer to question 24 in the following sequence:

23.  How many telephones (instruments) are there in your house?
    1 one  skip to Q. 25
    2 two
    3 three or more

24.  Do you have more than one telephone line in your house?
    (More than one telephone number.)
    1 yes
    2 no

25.  ...
SAMPLE WEIGHTING

The 641 households whose telephone usage we analyze are distributed by exchange and usage level\(^1\) as shown in Table A.3. The corresponding numbers of sample frame households are also shown. High and low users constitute a larger proportion of the sample than they do of the sample frame, because of the oversampling mentioned above. In addition, the two smaller exchanges, Clinton and Tuscola, are overrepresented in the sample relative to Jacksonville. Because we want our analysis to represent the full sample frame, not the distorted sample, we apply weights to correct for the differential sampling.\(^2\) The weights for each of the nine usage/exchange cells are shown in Table A.4. They are calculated as the ratio of the proportion of the population to the proportion of the sample that falls in each cell. For example, the lowest 10 percent category in Jacksonville accounts for 559/9585 = 5.83 percent of the population but 45/641 = 7.02 percent of the sample. The weight for this cell is therefore 5.83/7.02 = .831.

These weights are used to calculate the weighted average calling rates for our sample that are shown in Table A.5. They are also used to construct the moment matrix analyzed in the body of this report. The weight for each observation (depending on which exchange/usage category the observation is in) is applied to that observation's contribution to the moment matrix. Thus the \(j^k\)th element of the matrix is\(^3\)

\[
\sum_{i=1}^{641} \frac{w_i x_{ij} x_{ik}}{.}
\]

\(^1\)Recall that usage level is defined in terms of average monthly minutes of telephone calling during June, July, and August 1977.

\(^2\)Alternatively, one might want the analysis to represent some population even more broadly defined than the sample frame. However, we are not able to calculate the appropriate weights for a broader population because of lack of data, and more fundamentally because of conceptual difficulties.

\(^3\)Here we use \(j\) and \(k\) to index variables (for example, HHILDSIZE and INCOME).
Table A.3
NUMBER OF HOUSEHOLDS BY USAGE CATEGORY AND EXCHANGE
(Sample frame/sample)

<table>
<thead>
<tr>
<th>Usage Category</th>
<th>Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jacksonville</td>
</tr>
<tr>
<td>Lowest 10 percent</td>
<td>559/ 45</td>
</tr>
<tr>
<td>Middle 80 percent</td>
<td>4468/188</td>
</tr>
<tr>
<td>Highest 10 percent</td>
<td>559/ 89</td>
</tr>
<tr>
<td>Total</td>
<td>5586/322</td>
</tr>
</tbody>
</table>

Usage category is based on rankings within each exchange of telephone lines by average monthly minutes of telephone use during June, July, and August 1977.

Table A.4
SAMPLE WEIGHTS

<table>
<thead>
<tr>
<th>Usage Category</th>
<th>Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jacksonville</td>
</tr>
<tr>
<td>Lowest 10 percent</td>
<td>.831</td>
</tr>
<tr>
<td>Middle 80 percent</td>
<td>1.589</td>
</tr>
<tr>
<td>Highest 10 percent</td>
<td>.420</td>
</tr>
</tbody>
</table>

COMPARISON OF SAMPLE AND POPULATION

Because of the way in which the sample was drawn, it is almost necessarily representative of the sample frame. Here we check to make sure that the sample also represents a more broadly defined population. We shall compare sample and population mean calling rates for each of two months, July 1977 and July 1978. We define the population in each month to be all single-party residential telephone lines that were in service for the full month. The relationship of the population to the less-inclusive sample frame is shown in Table A.6.
Table A.5
WEIGHTED AVERAGE\textsuperscript{a} NUMBER OF CALLS PER MONTH FOR THE STRATIFIED SAMPLE

<table>
<thead>
<tr>
<th>Usage Category</th>
<th>Exchange</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jacksonville</td>
<td>Clinton</td>
<td>Tuscola</td>
<td>All Three</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
</tr>
<tr>
<td>Lowest 10 percent</td>
<td>20.04</td>
<td>22.79</td>
<td>13.53</td>
<td>20.53</td>
<td>12.50</td>
<td>17.50</td>
<td>17.17</td>
</tr>
<tr>
<td></td>
<td>(1.91)</td>
<td>(2.67)</td>
<td>(2.08)</td>
<td>(3.23)</td>
<td>(2.13)</td>
<td>(3.04)</td>
<td>(1.36)</td>
</tr>
<tr>
<td>Middle 80 percent</td>
<td>87.63</td>
<td>83.25</td>
<td>82.44</td>
<td>78.31</td>
<td>62.01</td>
<td>61.05</td>
<td>82.53</td>
</tr>
<tr>
<td></td>
<td>(3.19)</td>
<td>(3.36)</td>
<td>(5.09)</td>
<td>(4.36)</td>
<td>(3.99)</td>
<td>(4.24)</td>
<td>(2.43)</td>
</tr>
<tr>
<td>Highest 10 percent</td>
<td>223.20</td>
<td>171.57</td>
<td>195.23</td>
<td>152.68</td>
<td>167.38</td>
<td>145.32</td>
<td>207.53</td>
</tr>
<tr>
<td></td>
<td>(21.29)</td>
<td>(18.10)</td>
<td>(23.34)</td>
<td>(18.06)</td>
<td>(23.70)</td>
<td>(30.80)</td>
<td>(14.59)</td>
</tr>
<tr>
<td>All three</td>
<td>94.43</td>
<td>86.04</td>
<td>86.82</td>
<td>79.97</td>
<td>67.61</td>
<td>65.14</td>
<td>88.50</td>
</tr>
<tr>
<td></td>
<td>(4.14)</td>
<td>(3.69)</td>
<td>(5.65)</td>
<td>(4.53)</td>
<td>(5.51)</td>
<td>(5.54)</td>
<td>(2.99)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The flat rate (FR) averages are for June, July, and August 1977; the measured rate (MR) averages are for June, July, and August 1978. Weighted standard errors for the means of the three-month averages for individual households are in parentheses.
Table A.6

RELATIONSHIP OF POPULATION AND SAMPLE FRAME FOR TWO MONTHS COMPARED IN TABLE A.7

<table>
<thead>
<tr>
<th>Population: All single-party residential lines in service for the full month</th>
<th>July 1977</th>
<th>July 1978</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11,738</td>
<td>12,229</td>
</tr>
<tr>
<td>Less those not in service for full 10 months May 1977 through February 1978</td>
<td>1,946</td>
<td>2,709</td>
</tr>
<tr>
<td>Less others defined to be erratic users</td>
<td>207</td>
<td>194</td>
</tr>
<tr>
<td>In the sample frame</td>
<td>9,585</td>
<td>9,326</td>
</tr>
</tbody>
</table>

Sample and population means are compared in Table A.7; in no case is the difference significant at the .05 level.¹ We conclude that our sample is representative of the populations defined.

¹Population statistics were calculated for us by GTE Service Corporation. Willard Manning pointed out to us that a more powerful test would be based on (approximately normal) distributions of transformed calling rates C²⁷. Because the necessary data are not readily available to us, we chose not to do the more powerful test.
### Table A.7

**COMPARISON OF SAMPLE AND POPULATION MEANS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Exchange</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jacksonville</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
<td>MR</td>
<td>FR</td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted mean, ( \bar{x} )</td>
<td>95.36</td>
<td>85.42</td>
<td>88.06</td>
<td>79.41</td>
<td>62.69</td>
<td>62.85</td>
<td>88.68</td>
</tr>
<tr>
<td>Weighted standard deviation, ( s_x )</td>
<td>84.49</td>
<td>71.55</td>
<td>78.30</td>
<td>60.86</td>
<td>50.76</td>
<td>52.14</td>
<td>79.57</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean, ( \bar{y} )</td>
<td>97.40</td>
<td>84.02</td>
<td>91.81</td>
<td>78.58</td>
<td>68.71</td>
<td>65.56</td>
<td>91.79</td>
</tr>
<tr>
<td>Standard deviation, ( s_y )</td>
<td>89.68</td>
<td>76.80</td>
<td>83.90</td>
<td>69.91</td>
<td>63.03</td>
<td>59.12</td>
<td>85.36</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference in means, ( d = \bar{y} - \bar{x} )</td>
<td>2.04</td>
<td>-1.40</td>
<td>3.75</td>
<td>-.83</td>
<td>6.02</td>
<td>2.71</td>
<td>3.11</td>
</tr>
<tr>
<td>Standard error of difference, ( s_d^a )</td>
<td>4.83</td>
<td>4.09</td>
<td>5.63</td>
<td>4.39</td>
<td>5.06</td>
<td>5.15</td>
<td>3.24</td>
</tr>
</tbody>
</table>

**NOTES:** Flat rate (FR) values are for July 1977; measured rate (MR) values are for July 1978.

\(^a\) Calculated as \( s_d^a = \left( \frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^{\frac{1}{2}} \), where \( n_x \) and \( n_y \) are the number of households in the sample and in the population, respectively.
Appendix B

TRANSFORMATION OF CALLING RATE DATA

In developing the model that we estimate, we analyzed the statistical distributional properties of the monthly calling rates for each telephone exchange by month. Our primary interest was whether the distribution of calling rate data (or some transformation of these data) would reasonably approximate a normal distribution. With this knowledge we could more confidently apply statistical estimation techniques that make this assumption.

Initially we used normal probability plots for both calling rate and log of calling rate by month for each telephone exchange. The general shapes of these plots is illustrated in Figs. B.1 and B.2 for data from Jacksonville during July 1977. These figures plot the standard normal quantiles (vertical axis) against either calling rate (Fig. B.1) or log of calling rate (Fig. B.2) on the horizontal axis. If the random variable along the horizontal axis followed a normal distribution, then the graph would be approximately linear with a positive slope. The curvature in these plots provides evidence of non-normality in both the distribution of calling rate and its logarithm.

The concave shape of the graph in Fig. B.1 and the convex shape in Fig. B.2 suggest that a more nearly normal distribution might result from using a transformation of calling rate that is in some sense "between" these two cases. A useful parametric family of transformations that includes these two as special cases is the simple one-parameter power transformation introduced and analyzed by Box and Cox (1964).

Let $c_{it}$ be calls by the $i$th household in month $t$. Then following Box and Cox, we let

$$c_{it}^{(\lambda)} = \begin{cases} 
\frac{c_{it}^{\lambda} - 1}{\lambda}, & \lambda > 0 \\
\log c_{it}, & \lambda = 0 
\end{cases}$$
Fig. B.1--Normal probability plot for calls in Jacksonville during July 1977
Fig. B.2--Normal probability plot for logarithm of calls in Jacksonville during July 1977
be the simple power transformation of calling rate. Note that \( \lambda = 1 \) gives the untransformed case while \( \lambda = 0 \) yields the log of calling rate. Now, from the above heuristic argument about curvature we expect that a value of \( \lambda \) between 0 and 1 will provide transformed data that better approximate a normal distribution.

Box and Cox suggest a maximum likelihood approach to estimating an optimal \( \lambda \) from the data with the assumptions that \( \mathbb{E}(c_{it}^{(\lambda)}) \) is a nonstochastic linear model and that the \( c_{it}^{(\lambda)} \) are approximately independently normally distributed with constant variance.\(^1\) For our purposes we chose to assume a simple model: \( \mathbb{E}(c_{it}^{(\lambda)}) = \mu_t \), for a given month \( t \) and a given telephone exchange. Clearly, this is a much simpler model than the one we estimate in the report. However, we feel justified in using this simple form to choose a \( \lambda \), as it should yield a reasonable improvement in the distributional properties of the \( (c_{it}^{\lambda}) \) as the dependent variable for our more elaborate model.

For a fixed \( \lambda \) Box and Cox demonstrated that the maximized log likelihood is (except for a constant term):

\[
L_{\text{max}}(\lambda; c_t) = -(1/2)n \log \sigma^2(\lambda; c_t) + \log J(\lambda; c_t^{(\lambda)}),
\]

where, for our assumptions,

\(^1\)Because the power function \( (c_{it}^{\lambda}) \) for \( 0 < \lambda < 1 \) makes sense only for \( c_{it} \geq 0 \), the range of \( c_{it}^{(\lambda)} \) is necessarily the interval \([-1/\lambda, \infty)\], and only a proper subset of the domain \((\infty, \sigma)\) of a usual normal distribution is realizable through this transformation. Thus, the mathematically correct distributional assumption for \( c_{it}^{(\lambda)} \) follows some truncated normal distribution. Also, the form of the expectation \( \mathbb{E}[c_{it}^{(\lambda)}] \) includes a term in addition to the usual "nonstochastic linear model" referred to above (see Poisier and Melino, 1978). As a result we should properly view the maximum likelihood approach referred to above as an approximate maximum likelihood approach. The approximation should improve as less and less of the normal distribution probability is concentrated in the interval \((\infty, -1/\lambda)\).
\[ \hat{\sigma}^2 (\lambda; c_t) = \frac{1}{n} \sum_{i=1}^{n} (c_{it} (\lambda) - \hat{\mu}_t)^2 , \]

\[ \hat{\mu}_t = \frac{1}{n} \sum_{i=1}^{n} c_{it} (\lambda) , \]

\[ J (\lambda; c_t (\lambda)) = \prod_{i=1}^{n} |d c_{it} (\lambda)/d c_{it}| , \]

and \( n \) is the number of households for the particular telephone exchange. To simplify the log likelihood expression we worked with the normalized simple power transformation:

\[ z_{it} (\lambda) = \frac{c_{it} (\lambda)}{[gm(c_{it})]^{\lambda-1}} , \]

where

\[ gm(c_{it}) = \left( \prod_{i=1}^{n} c_{it} \right)^{1/n} , \]

is the geometric mean of the \( c_{it} \). Also \( E[z_{it} (\lambda)] = \nu_t = \mu_t / [gm(c_{it})]^{\lambda-1} \), so that the maximized log likelihood is (except for a constant):

\[ L_{\text{max}} (\lambda; z_t) = -(1/2)n \log \hat{\sigma} (\lambda; z_t) , \]

where

\[ \hat{\sigma}^2 (\lambda; z_t) = \frac{1}{n} \sum_{i=1}^{n} (z_{it} (\lambda) - \hat{\nu}_t)^2 , \]
\[
\hat{\nu}_t = \frac{1}{n} \sum_{i=1}^{n} z_{it}^{(\lambda)}.
\]

Thus, the value of \( \lambda = \lambda^* \) that maximizes \( L_{\text{max}}(\lambda; z_t) \) is the same value that minimizes \( \hat{\sigma}^2(\lambda; z_t) \). Table B.1 shows the optimal \( \lambda \) and the associated maximized log likelihood we computed for each of the six months and for the three telephone exchanges.\(^1\)

We chose \( \lambda = 0.27 \) as our power transformation parameter for the following reason: Pavarini (1979) found that (calling rate\(^{0.27} \) was normally distributed. Our data do not compel us to reject \( \lambda = 0.27 \) in favor of \( \lambda = \lambda^* \) at the 0.01 significance level for any of the \( \lambda \) we estimated. Furthermore, at the 0.05 level of significance only

\(^1\)The formulas used for computing \( \hat{\sigma}^2(\lambda; z_t) \), \( \hat{\nu}_t \) and \( \text{gm}(c_{it}) \) incorporated the weights for various households in each exchange:

\[
\text{gm}(c_{it}) = \exp \left[ \frac{\sum_{i=1}^{n} w_i \log c_{it}}{n} \right],
\]

\[
\hat{\nu}_t = \frac{\sum_{i=1}^{n} w_i z_{it}^{(\lambda)}}{n}, \quad \text{and}
\]

\[
\hat{\sigma}^2(\lambda, z_t) = \frac{\sum_{i=1}^{n} w_i (z_{it}^{(\lambda)})^2}{n} - (\hat{\nu}_t)^2,
\]

where \( w_i \) is the weight for the \( i \)th household in the given telephone exchange.
Table B.1

OPTIMAL POWER TRANSFORMATION PARAMETER (λ) AND MAXIMIZED LOG LIKELIHOOD (L_{max})

<table>
<thead>
<tr>
<th>Month</th>
<th>Jacksonville (n=337)</th>
<th>Clinton (n=215)</th>
<th>Tuscola (n=116)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ*</td>
<td>L_{max}(λ*;z_t)</td>
<td>λ*</td>
</tr>
<tr>
<td>June 1977</td>
<td>0.16</td>
<td>-1370.61</td>
<td>0.19</td>
</tr>
<tr>
<td>July 1977</td>
<td>0.21</td>
<td>-1376.40</td>
<td>0.22</td>
</tr>
<tr>
<td>August 1977</td>
<td>0.25</td>
<td>-1370.15</td>
<td>0.22</td>
</tr>
<tr>
<td>June 1978</td>
<td>0.16</td>
<td>-1335.36</td>
<td>0.20</td>
</tr>
<tr>
<td>July 1978</td>
<td>0.24</td>
<td>-1335.79</td>
<td>0.22</td>
</tr>
<tr>
<td>August 1978</td>
<td>0.19</td>
<td>-1342.62</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The number of observations for each exchange exceeds that in our final sample. This analysis was done before we dropped 27 multi-line households from the sample.

three of the 18 values of λ* we estimated would lead us to reject λ = 0.27. Table B.2 gives the values of the log likelihood ratio statistics, -2(L_{max}(λ*) - L_{max}(0.27)), by month and telephone exchange. Because this statistic is approximately distributed as χ^2 with one degree of freedom (Box and Cox, 1964, p. 216), critical values are χ^2_{.05}(1) = 3.84 and χ^2_{.99}(1) = 6.63.

In Fig. B.3 we show a normal probability plot for the random variable (calling rate)^{0.27} for July 1977 data from Jacksonville. This graph illustrates the general effect of the power transformation of calling rates: The transformed data are more consistent with the normality assumptions than either the untransformed data or their logarithms.
Table B.2
VALUES OF \(-2[L_{\text{max}}(\lambda^a) - L_{\text{max}}(0.27)]\)

<table>
<thead>
<tr>
<th>Month</th>
<th>Jacksonville</th>
<th>Clinton</th>
<th>Tuscola</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 1977</td>
<td>5.25\textsuperscript{a}</td>
<td>1.87</td>
<td>0.17</td>
</tr>
<tr>
<td>July 1977</td>
<td>1.57</td>
<td>0.80</td>
<td>0.45</td>
</tr>
<tr>
<td>August 1977</td>
<td>0.21</td>
<td>0.63</td>
<td>0.02</td>
</tr>
<tr>
<td>June 1978</td>
<td>4.82\textsuperscript{a}</td>
<td>1.21</td>
<td>6.18\textsuperscript{a}</td>
</tr>
<tr>
<td>July 1978</td>
<td>0.42</td>
<td>0.49</td>
<td>0.15</td>
</tr>
<tr>
<td>August 1978</td>
<td>2.58</td>
<td>0.07</td>
<td>1.41</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Significant at the 0.05 level for a $\chi^2$ distribution with one degree of freedom.
Fig. B.3—Normal probability plot for (calls) in Jacksonville during July 1977
Appendix C
TREATMENT OF MISSING VALUES FOR EXPLANATORY VARIABLES

Some of the survey demographic data are missing for some of our sample households. At an early stage in our analysis we elected to estimate values for missing data so we could include these households in our analysis. At that time, we were working with a sample of 668 households (before exclusion of 27 households with more than one telephone line) and with the initial set of variables shown in Table C.1. The initial set includes three variables that do not appear in our final model (EDUC, LRES, and NCAR) and excludes two that do (TEENS and YOUNGADULTS).

Of the variables listed in Table C.1, each of INCOME, AGE, EDUC, and NCAR had at least one value missing.\(^1\) Of the 668 households, the number with INCOME, AGE, EDUC, and NCAR missing were 96, 20, 34, and 10, respectively. In Table C.2 we show the distribution of missing data over all 668 households. Thus, 119 households (18 percent) had at least one variable whose value was missing.

We estimated a set of auxiliary linear regression equations based on the 549 households with complete data and used these to fill in missing values for the other 119 households. The ordinary least squares estimates for the regression coefficients for each equation are shown in Table C.3 together with t-statistics (in parentheses) and \(R^2\) values.

To estimate a household's missing value we simply multiplied the vector of regressor values for that household by the appropriate dependent variable regression coefficients. When a household had more than one variable to estimate, we used the mean or median value (computed from the households with complete data) for a missing regressor variable in the auxiliary regression. The mean values used for INCOME and AGE were 2.79 and 2.35 respectively. The median

\(^1\)The other variables (including TEENS and YOUNGADULTS) had no missing values.
values used for EDUC and NCAR were 3 and 2, respectively.

If households that refuse to answer survey questions differ systematically from those that do answer, the auxiliary regressions may yield biased predictions of missing values. To reduce any consequent problems in our final equations, we include dummy variables ESTINC and ESTAGE, equal to 1 for households with estimated INCOME and AGE, respectively, and 0 otherwise. The final sample of 641 households includes 94 with an estimated value for INCOME and 19 with an estimated value for AGE, distributed as shown in Table C.2.
Table C.1
INITIAL SET OF DEMOGRAPHIC VARIABLES FOR EXPLAINING HOUSEHOLD CALLING RATES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Included in Final Set?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIZE</td>
<td>logarithm of number of household members including the head of household</td>
<td>yes</td>
</tr>
<tr>
<td>INCOME</td>
<td>logarithm of household income ($000)(^a)</td>
<td>yes</td>
</tr>
<tr>
<td>AGE</td>
<td>Age of head of household in scores of years</td>
<td>yes</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>1 if children 12 and under in household; 0 otherwise</td>
<td>yes</td>
</tr>
<tr>
<td>FRIENDS</td>
<td>1 if household reported many local acquaintances; 0 otherwise</td>
<td>yes</td>
</tr>
<tr>
<td>EDUC</td>
<td>Education of head of household = 1 (8th grade or less), 2 (some high school), 3 (high school completed), 4 (some college), 5 (completed Junior college), 6 (completed college), 7 (post graduate study)</td>
<td>no</td>
</tr>
<tr>
<td>LRES</td>
<td>Length of time at this address: = 1 (&lt; 6 months), 2 (6 months to 1 year), 3 (1 to 2 years), 4 (2 to 3 years), 5 (3 to 5 years), 6 (5 to 10 years), 7 (&gt; 10 years)</td>
<td>no</td>
</tr>
<tr>
<td>NCAR</td>
<td>Number of passenger autos owned or regularly used by household members = 0, 1, 2, 3 (if 3 or more)</td>
<td>no</td>
</tr>
</tbody>
</table>

\(^a\)The questionnaire asks for income intervals. We used the midpoint of the closed intervals; for the open interval "$30,000 or more" we arbitrarily used $40,000.
Table C.2

DISTRIBUTION OF MISSING DATA

<table>
<thead>
<tr>
<th>Number of Variables K</th>
<th>Households with K Variables from Initial Set&lt;sup&gt;a&lt;/sup&gt; with Missing Values</th>
<th>Households with K Variables from Final Set&lt;sup&gt;b&lt;/sup&gt; with Missing Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before Exclusion&lt;sup&gt;c&lt;/sup&gt;</td>
<td>After Exclusion&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>0</td>
<td>549</td>
<td>525</td>
</tr>
<tr>
<td>1</td>
<td>85</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>668</td>
<td>641</td>
</tr>
</tbody>
</table>

<sup>a</sup>The initial set of variables is listed in Table C.1.

<sup>b</sup>The final set of variables is listed in Table 5.

<sup>c</sup>The columns totaling 668 apply to the sample before exclusion of households with more than one telephone line; the subset of these observations with no missing values among the initial set of variables (549 households) was used to estimate the auxiliary equations in Table C.3. The columns totaling 641 apply to the sample after exclusion of multi-line households; all of these observations were used to estimate the calling rate distributions in Tables 5 and 6.
Table C.3

LINEAR REGRESSION EQUATION COEFFICIENTS FOR ESTIMATING MISSING VALUES FOR INCOME, AGE, EDUC, AND NCAR
(t-statistics in parentheses)

<table>
<thead>
<tr>
<th>Regressor</th>
<th>INCOME</th>
<th>AGE</th>
<th>EDUC</th>
<th>NCAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.174</td>
<td>2.652</td>
<td>3.159</td>
<td>.526</td>
</tr>
<tr>
<td></td>
<td>(13.7)</td>
<td>(18.3)</td>
<td>(6.7)</td>
<td>(2.5)</td>
</tr>
<tr>
<td>SIZE</td>
<td>.426</td>
<td>-.182</td>
<td>-.348</td>
<td>.654</td>
</tr>
<tr>
<td></td>
<td>(6.4)</td>
<td>(2.7)</td>
<td>(1.9)</td>
<td>(9.0)</td>
</tr>
<tr>
<td>INCOME</td>
<td>--</td>
<td>-.158</td>
<td>.656</td>
<td>.232</td>
</tr>
<tr>
<td></td>
<td>(3.7)</td>
<td>(5.9)</td>
<td>(4.9)</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-.158</td>
<td>--</td>
<td>-.424</td>
<td>-.170</td>
</tr>
<tr>
<td></td>
<td>(3.7)</td>
<td>(3.8)</td>
<td>(3.6)</td>
<td></td>
</tr>
<tr>
<td>CHILDREN</td>
<td>-.188</td>
<td>-.532</td>
<td>.060</td>
<td>-.414</td>
</tr>
<tr>
<td></td>
<td>(2.6)</td>
<td>(7.6)</td>
<td>(.3)</td>
<td>(5.1)</td>
</tr>
<tr>
<td>FRIENDS</td>
<td>.249</td>
<td>.105</td>
<td>.182</td>
<td>.163</td>
</tr>
<tr>
<td></td>
<td>(2.9)</td>
<td>(1.2)</td>
<td>(.8)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>EDUC</td>
<td>.093</td>
<td>-.060</td>
<td>--</td>
<td>-.006</td>
</tr>
<tr>
<td></td>
<td>(5.9)</td>
<td>(3.6)</td>
<td>(.3)</td>
<td></td>
</tr>
<tr>
<td>LRES</td>
<td>.012</td>
<td>.169</td>
<td>-.029</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>(.7)</td>
<td>(10.4)</td>
<td>(.6)</td>
<td>(3.3)</td>
</tr>
<tr>
<td>NCAR</td>
<td>.182</td>
<td>-.134</td>
<td>-.035</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(4.9)</td>
<td>(3.5)</td>
<td>(.3)</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.35</td>
<td>.49</td>
<td>.14</td>
<td>.33</td>
</tr>
</tbody>
</table>
Appendix D

DEMOGRAPHIC PORTION OF SURVEY INSTRUMENT

The following pages reproduce the demographic portion of the survey instrument that was designed and administered by Opinion Research, Inc., for GTE's third customer survey, April 12-22, 1978. The first part of the questionnaire, questions 1 through 55, dealt with attitudes and perceptions of telephone use. See Opinion Research Corporation (1980) for details.
That pretty well takes care of our questions on telephones. In order to classify our results, we would like to know something about your home and you.

56. Now, we'd like to know a little about the building you live in. How many living quarters, occupied and vacant, are at this address?

1 ONE 5 FIVE TO NINE
2 TWO 6 TEN OR MORE
3 THREE 7 MOBILE HOME
4 FOUR OR TRAILER

57. Are your living quarters . . . (READ)

1 owned or being bought by you (or someone else in household)?  → SKIP TO Q. 62
2 rented for cash rent?  → SKIP TO Q. 63
3 occupied without payment of cash rent?

IF "OWNED" OR "BOUGHT," ASK:

58. Is this a cooperative or condominium?

1 NO
2 COOPERATIVE
3 CONDOMINIUM
4 DON'T KNOW

59. Is this a one-family house?

1 YES  → SKIP TO Q. 63
2 NO

(IF "YES," ASK):

60. How would you describe this house? Would you say; it is a house on a place of ten acres or more; part of this property is used as a commercial establishment or medical office; it is a farm?

1 ON TEN ACRES OR MORE
2 USED AS A COMMERCIAL ESTABLISHMENT OR MEDICAL OFFICE
3 FARM (Describe): 

4 NONE OF THESE

(WRITE IN ACTUAL NUMBER IN DOLLARS)

(IF "RENTED" ON Q. 57 ASK):

62. What is the monthly rent for your living quarters?

(WRITE IN ACTUAL NUMBER IN DOLLARS - NO CENTS)
63. How many passenger automobiles are owned or regularly used by members of your household?

   1. NONE
   2. ONE
   3. TWO
   4. THREE OR MORE

64. About how long have you lived at this address?

   1. LESS THAN SIX MONTHS
   2. SIX MONTHS TO ONE YEAR
   3. MORE THAN ONE YEAR TO TWO YEARS
   4. MORE THAN TWO TO THREE
   5. MORE THAN THREE TO FIVE
   6. MORE THAN FIVE TO TEN
   7. MORE THAN TEN

(IF FIVE YEARS OR LESS, ASK):

65. How many times have you moved in the last five years?

   1. ONCE
   2. TWICE
   3. THREE TIMES
   4. FOUR OR MORE
   5. DON'T KNOW

66. On your last move, did you move from within the same community, within the same county, from an adjoining county, within the same state, or out of state?

   1. SAME COMMUNITY
   2. SAME COUNTY
   3. ADJOINING COUNTY
   4. SAME STATE
   5. ANOTHER STATE

67. Which of the following statements apply to the members of your household? As I read each, tell me whether or not it applies.

<table>
<thead>
<tr>
<th>Statement</th>
<th>APPLIES</th>
<th>DOES NOT APPLY</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many relatives live in this community with us</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Have many close friends in this community</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>We know a lot of people in this community</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>We really don't know very many people in this community</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
These last questions are for statistical purposes:

68. What is your current marital status?
   Are you . . . [READ]
   1 Married
   2 Single and have never been married
   3 Widowed
   4 Divorced
   5 Separated

69. Altogether, how many people live in this household?
   [WRITE IN ACTUAL NUMBER]
   SINGLE PERSON HOUSEHOLDS, SKIP TO Q. 73

IN MULTI-PERSON HOUSEHOLDS, ASK:

70. Are there any children twelve and under?
    teenagers?
    young adults (19-25)?
    senior citizens (60 and older)?
    handicapped person or persons?
    (How handicapped?)
    None of these

71. Were any of these people especially affected by the change to measured local rates? (Which ones?)
    IF PEOPLE WERE AFFECTED, ASK:
    72. How were they affected?

73. Which of these best describes you:
    Are employed full-time, that is, 30 hours a week or more for pay;
    employed part-time for pay; or are you not employed at this time?
    1 EMPLOYED FULL-TIME (30 HOURS A WEEK OR MORE) FOR PAY
    2 EMPLOYED PART-TIME (LESS THAN 30 HOURS A WEEK) FOR PAY
    3 NOT EMPLOYED FOR PAY AT THIS TIME
    4 HOUSEWIFE
    5 RETIRED
    6 OTHER (Explain): 
    7 DON'T KNOW

74. What is the last grade you completed in school?
    1 EIGHTH GRADE OR LESS
    2 SOME HIGH SCHOOL
    3 HIGH SCHOOL COMPLETE
    4 SOME COLLEGE
    5 JUNIOR COLLEGE COMPLETE
    6 COLLEGE COMPLETE
    7 POST GRADUATE STUDY
75. What is your age?

(RECORD ACTUAL AGE)

76. How many members of this household are employed?

(WRITE IN ACTUAL NUMBER)  

NONE  

RETIRED

77. Are you the chief wage earner in this household?

1 YES

2 NO — SKIP TO Q. 80

INTERVIEWER: WE ARE PARTICULARLY INTERESTED IN ANSWERS FROM OLDER RESPONDENTS (60 YEARS OF AGE AND OVER.)

PLEASE PROBE FOR FULL INFORMATION ON HOUSEHOLD EMPLOYMENT/INCOME SOURCES AND DESCRIBE BELOW.

CHECK: PENSION, SOCIAL SECURITY, WELFARE, ETC.
(IF RESPONDENT IS CHIEF WAGE EARNER, ASK):
78. What kind of work do you do? (If unemployed, what is your usual occupation?)

79. What kind of business is that? What do they make or do?

(PLEASE CIRCLE):
1. OWNS BUSINESS, HIRES OTHERS
2. SELF-EMPLOYED, HIRES NOBODY
3. WORKS FOR SOMEONE ELSE
4. RETIRED
5. OTHER:

→ SKIP TO QUESTION 84

(IF RESPONDENT IS NOT CHIEF WAGE EARNER):
80. What is the age of the chief wage earner?

(RECORD ACTUAL AGE)

81. What kind of work does the chief wage earner in this household do? (If unemployed, what is usual occupation?)

82. What kind of business is that? What do they make or do?

(PLEASE CIRCLE):
1. OWN BUSINESS, HIRES OTHERS
2. SELF-EMPLOYED, HIRES NOBODY
3. WORKS FOR SOMEONE ELSE
4. RETIRED
5. OTHER:
83. What is the last grade completed in school by the chief wage earner?

1 EIGHTH GRADE OR LESS
2 SOME HIGH SCHOOL
3 HIGH SCHOOL COMPLETE
4 SOME COLLEGE
5 JUNIOR COLLEGE COMPLETE
6 COLLEGE COMPLETE
7 POSTGRADUATE STUDY

84. One final question. For statistical purposes only, we need to know which of these groups describes your total family income (before taxes) for 1977. Include your own income and that of any members of your immediate family who are living with you. As I read the categories, please stop me when I reach the right one.

1 $30,000 OR MORE
2 $25,000 BUT LESS THAN $30,000
3 $15,000 BUT LESS THAN $25,000
4 $10,000 BUT LESS THAN $15,000
5 $8,000 BUT LESS THAN $10,000
6 $5,000 BUT LESS THAN $8,000
7 UNDER $5,000
Y REFUSED

THANK YOU FOR YOUR HELP.

FOR RESPONDENTS 60 YEARS OF AGE AND OLDER:

This finishes the formal part of our survey. We are, however, very interested in your reactions to measured local service. Is there anything you would like to add? Did we leave out anything that we should have asked?
Appendix E

APPLICATION OF LISREL TO ESTIMATE THE TELEPHONE CALLING RATE MODEL

THE GENERAL LISREL MODEL

LISREL is a model and computer program for the analysis of linear structural relationships by the method of maximum likelihood. The LISREL model's most general format consists of two sets of linear relationships: a structural equation model and a measurement model. The structural equation model describes the relationship between a random vector \( \eta' = (\eta_1, \eta_2, \ldots, \eta_m) \) of independent variables and a random vector \( \xi' = (\xi_1, \xi_2, \ldots, \xi_n) \) of latent dependent variables:

\[
B\eta = \Gamma\xi + \zeta, \tag{E.1}
\]

where \( B(m \times m) \) and \( \Gamma(m \times n) \) are coefficient matrices and \( \xi' = (\xi_1, \xi_2, \ldots, \xi_n) \) is a random vector of residuals. Some assumptions about these matrices and vectors are that \( B \) is nonsingular, \( \zeta \) is uncorrelated with \( \xi \), and \( E[\xi] = 0 \).

By including a measurement model we assume in general that we do not actually observe the vectors \( \eta \) and \( \xi \); rather, we observe vectors \( y' = (y_1, y_2, \ldots, y_p) \) and \( x' = (x_1, x_2, \ldots, x_q) \). These vectors are related to \( \eta \) and \( \xi \) by the following linear models:

\[
y = \Lambda_y \eta + \omega, \tag{E.2}
\]

\[
x = \Lambda_x \xi + \delta. \tag{E.3}
\]

\(^1\)For a detailed description of the model and its applications, see Joreskög and Sörbom (1978) and Jöreskog (1979).

\(^2\)In Eq. (E.2), we have denoted the error by \( \omega \), instead of the usual LISREL notation \( \epsilon \), to avoid a conflict with the notation used in our formulation of the calling rate model.
The vectors $\omega$ and $\delta$ are uncorrelated errors of measurement in $y$ and $x$, respectively. They are assumed to have expectation zero and to be uncorrelated with $\eta$, $\xi$, and $\zeta$. $\Lambda_y$ and $\Lambda_x$ are simply regression matrices of $y$ on $\eta$ and $x$ on $\xi$, respectively.

The covariance matrices for the random vectors $\xi$ and $\zeta$ are $\Phi(n \times n)$ and $\Psi(n \times n)$, respectively. Also, let $\Theta_\omega(p \times p)$ and $\Theta_\delta(q \times q)$ be covariance matrices for the respective measurement errors. These covariance matrices together with the matrices $B$, $\Gamma$, $\Lambda_y$, and $\Lambda_x$ contain the parameters of the model. The parameters may be fixed (equal to a constant), constrained (equal to one another), or free. Parameters that are not fixed are called independent parameters.

To estimate the independent parameters of a LISREL model by maximum likelihood, the usual distributional assumption is that the observed variables $y$ and $x$ have a joint multivariate normal distribution so that the first and second moments of $z = (y', x')'$ are sufficient to describe this distribution. Because the LISREL model leaves the mean vector of $z$ unconstrained and the covariance matrix $\Sigma$ of $z$ is a function of the independent parameters in $B$, $\Gamma$, $\Lambda_y$, $\Lambda_x$, $\Phi$, $\Psi$, $\Theta_\omega$, and $\Theta_\delta$, the model is estimated by fitting the sample covariance matrix $S$ of $z$ to $\Sigma$.

---

**FORMULATION OF THE CALLING RATE MODEL AS A LISREL MODEL**

As presented in the text, the calling rate model we estimated has the formulation of a linear regression model:

$$C_{it} = \beta_0 + Z_i\beta + (a_0 + Z_i\alpha)T_t + \mu_i + v_iT_t + \varepsilon_{it},$$

(E.4)

for $i = 1, 2, \ldots, 641; t = 1, 2, \ldots, 6$. As we shall see below, reformulating this model as a LISREL model allows direct estimation of all coefficients and stochastic parameters.

---

1. The matrix $S$ to be analyzed can also be a moment matrix or a correlation matrix; however, a matrix of moments about zero must be used when a constant term is included in the structural equations.
Define the following latent variables:

\[ F_i = \beta_0 + \beta'Z_i + \mu_i , \]

and

\[ M_i = \alpha_0 + \alpha'Z_i + \nu_i. \]

In matrix notation:

\[
\begin{bmatrix}
F_i \\
M_i
\end{bmatrix}
= \begin{bmatrix}
\beta_0 & \beta' \\
\alpha_0 & \alpha'
\end{bmatrix}
\begin{bmatrix}
1 \\
Z_i
\end{bmatrix}
+ \begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix}. \quad (E.5)
\]

We identify this equation with the structural equation \( Bn_i = \Gamma \xi_i + \xi_i \), where \( B \) is in this case a \( 2 \times 2 \) identity matrix. Also, we write the six equations of (E.4) for \( t = 1, \ldots, 6 \) in matrix form using the definition of \( F_i \), \( M_i \), and \( T_t \) to get:

\[
\begin{bmatrix}
C_{i1} \\
C_{i2} \\
C_{i3} \\
C_{i4} \\
C_{i5} \\
C_{i6}
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 1 \\
1 & 1 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
F_i \\
M_i
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{i1} \\
\varepsilon_{i2} \\
\varepsilon_{i3} \\
\varepsilon_{i4} \\
\varepsilon_{i5} \\
\varepsilon_{i6}
\end{bmatrix} \quad (E.6)
\]

This equation we identify with the measurement model for \( \eta_i \) above:
\( Y_i = \Lambda Y + \eta_i + \omega_i \). Finally, we assume that we measure \( \xi_i \) without error so that \( x_i = \xi_i \).

Our previous assumptions about the error structure of the \( \mu_i \), \( \nu_i \) and \( \varepsilon_{it} \) imply that:
\[ \theta_{\omega} = \begin{bmatrix} \sigma^2_e & \rho \sigma^2_e & \rho^2 \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & \rho \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & \rho \sigma^2_e & \rho^2 \sigma^2_e & 0 & 0 \\ \sigma^2_e & \rho \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & \rho \sigma^2_e & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 & 0 \\ \sigma^2_e & 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ \theta_{\delta} = 0 \]

\[ \Psi = \begin{bmatrix} \sigma^2_\mu & \sigma_{\mu \nu} \\ \sigma_{\mu \nu} & \sigma^2_\nu \end{bmatrix} \]

\[ \phi = \Sigma_{XX}, \text{ the sample covariance matrix over} \]
\[ \text{all households computed from } x_{i1} = \xi_{i1} = [1, z_{i1}]'. \]

Because of the nonlinear relationship among the elements of \( \theta_{\omega} \), we had to estimate the model using an iterative procedure. We fixed a trial value for \( \rho \sigma^2_e \) and estimated values for \( \sigma^2_e \) and \( \rho \sigma^2_e \). The implied value of \( \rho \) was used to fix a new trial value for \( \rho \sigma^2_e \), and the process was iterated to convergence.

**AN ALTERNATIVE LISREL FORMULATION OF THE MODEL**

The following alternative LISREL formulation of our calling rate model allows for a direct estimation of the autoregressive parameter \( \rho \). This formulation is a specific application of the approach taken by Lillard and Weiss (1979). The trick is to treat six additional equations—equations that transform the autocorrelated errors \( \epsilon_{it} \) into iid random variables \( \epsilon_{it} \)—as a part of the structural equation.
model. The transformation is:

\[
\begin{bmatrix}
(1-\rho^2)^{\frac{1}{2}} & 0 & 0 \\
-\rho & 1 & 0 \\
0 & -\rho & 1
\end{bmatrix}
\begin{bmatrix}
e_{i1} \\
e_{i2} \\
e_{i3}
\end{bmatrix}
= 
\begin{bmatrix}
e_{i1} \\
e_{i2} \\
e_{i3}
\end{bmatrix}
\]

and similarly for \([e_{i4} e_{i5} e_{i6}]\). We denote the \(3 \times 3\) transformation matrix above by \(R\) and stack these six equations together with the other two structural equations (E.5) to get the following structural equation model:

\[
\begin{bmatrix}
I_2 \\
0_{3 \times 2} & R \\
0_{3 \times 2} & 0_{3 \times 3}
\end{bmatrix}
\begin{bmatrix}
F_1 \\
M_i \\
\varepsilon_{i1} \\
\varepsilon_{i2} \\
\varepsilon_{i3} \\
\varepsilon_{i4} \\
\varepsilon_{i5} \\
\varepsilon_{i6}
\end{bmatrix}
= 
\begin{bmatrix}
\beta_0 & \beta' \\
0 & a' \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
1 \\
z_i' \\
0_{6 \times q}
\end{bmatrix}
+ 
\begin{bmatrix}
\mu_i \\
\nu_i \\
\varepsilon_{i1} \\
\varepsilon_{i2} \\
\varepsilon_{i3} \\
\varepsilon_{i4} \\
\varepsilon_{i5} \\
\varepsilon_{i6}
\end{bmatrix}
\]

(E.7)

where \(I_2\) is a \(2 \times 2\) identity matrix, the 0s are zero matrices of the indicated dimensions, and \(q\) is equal to the dimension of \([1, z_i]\).

The measurement model must of course be written in terms of the same vector of latent variables \(\eta_i = [F_1 M_i \varepsilon_{i1} \varepsilon_{i2} \varepsilon_{i3} \varepsilon_{i4} \varepsilon_{i5} \varepsilon_{i6}]\) as follows:

\[\text{See, for example, Johnston (1963) or Park and Mitchell (1980).}\]
where $I_3$ and $0_3$ are $3 \times 3$ identity and zero matrices. Because the $\varepsilon_{it}$ are now included in the $\eta$ vector, the residual vector $\omega$ is identically zero. Thus, we have the following covariance matrices for this formulation:

\[
\begin{align*}
\begin{pmatrix}
C_{i1} & C_{i2} & C_{i3} & C_{i4} & C_{i5} & C_{i6}
\end{pmatrix}
&= 
\begin{pmatrix}
1 & 0 & I_3 & 0_3 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
F_1 & M_1 & \varepsilon_{i1} & \varepsilon_{i2} & \varepsilon_{i3} & \varepsilon_{i4} & \varepsilon_{i5} & \varepsilon_{i6}
\end{pmatrix}
\end{align*}
\]

(E.8)

\[
\begin{align*}
\theta_{\omega} &= 0 \\
\theta_{\delta} &= 0 \\
\psi &= 
\begin{pmatrix}
\sigma^2_{\mu} & \sigma_{\mu \nu} & 0_{2 \times 6} \\
\sigma_{\mu \nu} & \sigma^2_{\nu} & 0_{6 \times 2} \\
0_{6 \times 2} & \sigma^2_{\nu} & I_6
\end{pmatrix}
\end{align*}
\]

\[
\phi = S_{xx}.
\]

As in the previous LISREL formulation, iterative estimation is necessary. We fixed a trial value for $(1 - \rho^2)^{\frac{1}{2}}$ in the $B$ matrix,
and estimated \( \rho \) (together with the other free parameters). This value of \( \rho \) was used to fix a new trial value of \( (1 - \rho^2)^{1/2} \), and the process was iterated to convergence.

The estimates of the systematic coefficients obtained using this alternative formulation are identical to those reported in Table 5. A slightly different set of stochastic parameters is estimated under the alternative formulation, as shown in Table E.1. The parameter estimates differ from those in Table 6 only slightly, and probably only because our criterion for stopping the iterative estimation processes was not sufficiently fine.

THE MODEL FOR ANALYSIS BY GROUPS

As formulated above this model does not allow for estimating separate parameters for each telephone exchange. Fortunately, the LISREL computer program allows the user to apply his basic model to subsets of his observations, called groups. The basic formulation of the model for analysis of the data by groups differs little from that shown above. In general, the structural and measurement models have the same forms in each group; however, each parameter may be estimated independently. Thus, if there were \( k \) independent parameters in the original model, there would be a maximum of \( g \times k \) parameters when the model applies to each of \( g \) groups.

For each group a separate (moment, covariance, or correlation) matrix is analyzed. Although we assumed that the error structure of the various group models were independent of each other, we did consider several ways the parameters could be constrained across the groups. Because the LISREL program can handle the joint estimation of the parameters for all groups together with equality constraints among parameters across groups, we were able to explore the effects of constraining the parameters in several ways.

EXAMPLE OF INPUT TO THE LISREL COMPUTER PROGRAM

The input "statements" listed below completely describe for LISREL the structural and measurement models for each group together
Table E.1
ESTIMATES OF THE STOCHASTIC PARAMETERS USING THE ALTERNATIVE LISREL MODEL FORMULATION

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>( H_2 ) With Demographics</th>
<th>Estimate</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_e^2 )</td>
<td>iid random error</td>
<td></td>
<td>0.067</td>
<td>22.4</td>
</tr>
<tr>
<td>( \rho )</td>
<td>first order auto-correlation of ( \varepsilon )</td>
<td></td>
<td>-0.235</td>
<td>4.9</td>
</tr>
<tr>
<td>( \sigma_e^2 )</td>
<td>implied within-household error</td>
<td></td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>( \sigma_\mu^2 )</td>
<td>household-specific error</td>
<td></td>
<td>0.307</td>
<td>15.9</td>
</tr>
<tr>
<td>( \sigma_\nu^2 )</td>
<td>household-specific reaction to measured service</td>
<td></td>
<td>0.055</td>
<td>5.8</td>
</tr>
<tr>
<td>( \sigma_{\mu\nu} )</td>
<td>covariance of ( \mu ) and ( \nu )</td>
<td></td>
<td>-0.043</td>
<td>4.8</td>
</tr>
<tr>
<td>( \delta )</td>
<td>implied correlation of ( \mu ) and ( \nu )</td>
<td></td>
<td>-0.346</td>
<td></td>
</tr>
</tbody>
</table>

The parameters \( \sigma_e^2 \) and \( \delta \) were not estimated directly, but calculated as \( \sigma_e^2 = \sigma_e^2/(1 - \rho^2) \) and \( \delta = \sigma_{\mu\nu}/(\sigma_\mu^2 \sigma_\nu^2)^{1/2} \).

with the starting values that should be used in the iterative maximum likelihood optimization procedure. The moment matrix for each group was provided in a separate input file (file code 10).
LISREL Job Input for Calling Rate Model

// EXEC LISREL,REGG=350K,SIZE=100,MIN=0,SEC=40
//F707F001 DD SYSOUT=A,DCB=(RECFM=FA,BLKSIZE=133)
//SYSIN DD *
LINEAR MODEL FOR TELEPHONE CALLING RATE AT CLINTON
DA NG=3 NI=18 NO=175 MA=MM
LABELS
*
'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MVY'
'AGE' 'AGESQ' 'MVA' 'UNDR12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=1D C
GA=FU,FR C
PS=SY,FR C
LY=FU,FI C
TE=SY,FR
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
   TE(2,4) TE(2,5) TE(2,6) C
   TE(3,4) TE(3,5) TE(3,6)
MA GA
*
1.5882 0.5974 1.1288 -0.0631 0.1778
0.3080 -0.1623 0.0000 0.0000 0.0000 0.1944
0.3482 -0.1143 -0.2042 0.0371 0.0690
-0.1838 -0.0044 0.0000 0.0000 0.0000 -0.0123
MA PS
*
0.3342 -0.0588 0.0859
MA LY
*
1 0
1 0
1 1
1 1
1 1
1 1
ST 0.0580 TE(1,1)
ST 0.0 TE(1,4)
ST 0.0125 TE(1,2)
ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
LINEAR MODEL FOR TELEPHONE CALLING RATE AT JACKSONVILLE
DA NI=18 NO=374 MA=MM
LABELS
*
'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MVY'
'AGE' 'AGESQ' 'MVA' 'UND12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=ID C
GA=FU,FR C
PS=SY,FR C
LY=FU,FI C
TE=SY,FR
EQ TE(1,1,1) TE(1,1)
EQ TE(1,1,2) TE(1,2)
EQ TE(1,1,3) TE(1,3)
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
   TE(2,4) TE(2,5) TE(2,6) C
   TE(3,4) TE(3,5) TE(3,6)
EQ GA(1,1,2) GA(1,2)
EQ GA(1,1,3) GA(1,3)
EQ GA(1,1,4) GA(1,4)
EQ GA(1,1,5) GA(1,5)
EQ GA(1,1,6) GA(1,6)
EQ GA(1,1,7) GA(1,7)
EQ GA(1,1,8) GA(1,8)
EQ GA(1,1,9) GA(1,9)
EQ GA(1,1,10) GA(1,10)
EQ GA(1,1,11) GA(1,11)
EQ GA(1,2,2) GA(2,2)
EQ GA(1,2,3) GA(2,3)
EQ GA(1,2,4) GA(2,4)
EQ GA(1,2,5) GA(2,5)
EQ GA(1,2,6) GA(2,6)
EQ GA(1,2,7) GA(2,7)
EQ GA(1,2,8) GA(2,8)
EQ GA(1,2,9) GA(2,9)
EQ GA(1,2,10) GA(2,10)
EQ GA(1,2,11) GA(2,11)
EQ PS(1,1,1) PS(1,1)
EQ PS(1,2,1) PS(1,2)
EQ PS(1,2,2) PS(2,2)
MA GA
* 1.5882 0.5974 1.1288 -0.0631 0.1778
  0.3080 -0.1623 0.0000 0.0000 0.0000 0.1944
  0.3482 -0.1143 -0.2042 0.0371 0.0690
  -0.1838 -0.0044 0.0000 0.0000 0.0000 -0.0123
MA PS
* 0.3342 -0.0588 0.0859
MA LY
*
1 0
1 0
1 0
1 1
1 1
1 1
ST 0.0580 TE(1,1)
ST 0.0 TE(1,4)
ST 0.0125 TE(1,2)
ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
LINEAR MODEL FOR TELEPHONE CALLING RATE AT TUSCOLA
DA NI=18 NO=92 MA=MM
LABELS
* 'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MVY'
'AGE' 'AGESQ' 'MVA' 'UNDR12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=ID C
GA=FU,FR C
PS=SY,FR C
LY=FU,FI C
TE=SY,FR
EQ TE(1,1,1) TE(1,1)
EQ TE(1,1,2) TE(1,2)
EQ TE(1,1,3) TE(1,3)
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
   TE(2,4) TE(2,5) TE(2,6) C
   TE(3,4) TE(3,5) TE(3,6)
EQ GA(1,1,2) GA(1,2)
EQ GA(1,1,3) GA(1,3)
EQ GA(1,1,4) GA(1,4)
EQ GA(1,1,5) GA(1,5)
EQ GA(1,1,6) GA(1,6)
EQ GA(1,1,7) GA(1,7)
EQ GA(1,1,8) GA(1,8)
EQ GA(1,1,9) GA(1,9)
EQ GA(1,1,10) GA(1,10)
EQ GA(1,1,11) GA(1,11)
EQ GA(1,2,1) GA(2,1)
EQ GA(1,2,2) GA(2,2)
EQ GA(1,2,3) GA(2,3)
EQ GA(1,2,4) GA(2,4)
EQ GA(1,2,5) GA(2,5)
EQ GA(1,2,6) GA(2,6)
EQ GA(1,2,7) GA(2,7)
EQ GA(1,2,8) GA(2,8)
EQ GA(1,2,9) GA(2,9)
EQ GA(1,2,10) GA(2,10)
EQ GA(1,2,11) GA(2,11)
EQ PS(1,1,1) PS(1,1)
EQ PS(1,2,1) PS(2,1)
EQ PS(1,2,2) PS(2,2)
MA GA
* 1.5882 0.5974 1.1288 -0.0631 0.1778
   0.3080 -0.1623 0.0000 0.0000 0.0000 0.1944
   0.3482 -0.1143 -0.2042 0.0371 0.0890
   -0.1838 -0.0044 0.0000 0.0000 0.0000 -0.0123
MA PS
* 0.3342 -0.0588 0.0859
MA LY
* 1 1
   1 1
   1 1
   1 1
   ST 0.0580 TE(1,1)
   ST 0.0 TE(1,4)
   ST 0.0125 TE(1,2)
   ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
/*
//GO.FT10F001 DD DSN=W.W3137.A2616.LRTEMP,DISP=SHR,
// UNIT=USER, VOL=SER=USER51
//
//W3137BMW JOB (2616,60,133),' B.WETZEL ',CLASS=N
// EXEC LISREL,REGR=350K,SIZE=100,MIN=0,SEC=40
//FTO7F001 DD SYSOUT=A,DCB=(RECFM=FA,BLKSIZE=133)
//SYSIN DD *
LINEAR MODEL FOR TELEPHONE CALLING RATE AT CLINTON
DA NG=3 NI=18 NO=175 MA=MM
LABELS
* 'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MYV'
'AGE' 'AGESQ' 'MVA' 'UNDR12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=ID C
GA=FU,FI C
PS=SY,FR C
LY=FU,FI C
TE=SY,FR
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
TE(2,4) TE(2,5) TE(2,6) C
TE(3,4) TE(3,5) TE(3,6)
FREE GA(1,1) GA(2,1)
MA GA
*
1.5882 0.0000 0.0000 +0.0000 0.0000
0.0000 +0.0000 0.0000 0.0000 0.0000 0.0000
0.3482 +0.0000 +0.0000 0.0000 0.0000
+0.0000 +0.0000 0.0000 0.0000 0.0000 +0.0000
MA FS
*
0.3342 -0.0588 0.0859
MA LY
*
1 0
1 0
1 1
1 1
1 1
ST 0.0580 TE(1,1)
ST 0.0 TE(1,4)
ST 0.0125 TE(1,2)
ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
LINEAR MODEL FOR TELEPHONE CALLING RATE AT JACKSONVILLE
DA NI=18 NO=374 MA=MN
LABELS
*
'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MXY'
'AGE' 'AGESQ' 'MVA' 'UND12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=ID C
GA=FU,FI C
PS=SY,FR C
LY=FU,FI C
TE=SY,FR
EQ TE(1,1,1) TE(1,1)
EQ TE(1,1,2) TE(1,2)
EQ TE(1,1,3) TE(1,3)
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
TE(2,4) TE(2,5) TE(2,6) C
TE(3,4) TE(3,5) TE(3,6)
FREE GA(1,1) GA(2,1)
EQ GA(1,1,2) GA(1,2)
EQ GA(1,1,3) GA(1,3)
EQ GA(1,1,4) GA(1,4)
EQ GA(1,1,5) GA(1,5)
EQ GA(1,1,6) GA(1,6)
EQ GA(1,1,7) GA(1,7)
EQ GA(1,1,8) GA(1,8)
EQ GA(1,1,9) GA(1,9)
EQ GA(1,1,10) GA(1,10)
EQ GA(1,1,11) GA(1,11)
EQ GA(1,2,2) GA(2,2)
EQ GA(1,2,3) GA(2,3)
EQ GA(1,2,4) GA(2,4)
EQ GA(1,2,5) GA(2,5)
EQ GA(1,2,6) GA(2,6)
EQ GA(1,2,7) GA(2,7)
EQ GA(1,2,8) GA(2,8)
EQ GA(1,2,9) GA(2,9)
EQ GA(1,2,10) GA(2,10)
EQ GA(1,2,11) GA(2,11)
EQ PS(1,1,1) PS(1,1)
EQ PS(1,2,1) PS(2,1)
EQ PS(1,2,2) PS(2,2)

MA GA
*
1.5882 0.0000 0.0000 +0.0000 0.0000
0.0000 +0.0000 0.0000 0.0000 0.0000
0.3482 +0.0000 +0.0000 0.0000 0.0000
+0.0000 +0.0000 0.0000 0.0000 0.0000

MA PS
*
0.3342 -0.0588 0.0859

MA LY
*
1 0
1 0
1 0
1 1
1 1
1 1

ST 0.0580 TE(1,1)
ST 0.0 TE(1,4)
ST 0.0125 TE(1,2)
ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
LINEAR MODEL FOR TELEPHONE CALLING RATE AT TUSCOLA
DA NI=18 NO=92 MA=MM
LABELS
*
'C1' 'C2' 'C3' 'C4' 'C5' 'C6' 'WT' 'H1' 'LOGH' 'LOGY' 'MVY'
'AGE' 'AGESQ' 'MVA' 'UND12' 'TEENS' 'OVER19' 'FRIEND'
SS SY UNIT=10 FO
(10A8)
SELECTION
1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18/
MO NY=6 NX=11 NE=2 NK=11 FIXEDX C
BE=ID C
GA=FU,FI C
PS=SY,FR C
LY=FU, FI C
TE=SY, FR
EQ TE(1,1,1) TE(1,1)
EQ TE(1,1,2) TE(1,2)
EQ TE(1,1,3) TE(1,3)
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
EQ TE(1,2) TE(2,3) TE(4,5) TE(5,6)
EQ TE(1,3) TE(4,6)
FIX TE(1,3) TE(1,4) TE(1,5) TE(1,6) C
   TE(2,4) TE(2,5) TE(2,6) C
   TE(3,4) TE(3,5) TE(3,6)
FREE GA(1,1) GA(2,1)
EQ GA(1,1,2) GA(1,2)
EQ GA(1,1,3) GA(1,3)
EQ GA(1,1,4) GA(1,4)
EQ GA(1,1,5) GA(1,5)
EQ GA(1,1,6) GA(1,6)
EQ GA(1,1,7) GA(1,7)
EQ GA(1,1,8) GA(1,8)
EQ GA(1,1,9) GA(1,9)
EQ GA(1,1,10) GA(1,10)
EQ GA(1,1,11) GA(1,11)
EQ GA(1,2,1) GA(2,1)
EQ GA(1,2,2) GA(2,2)
EQ GA(1,2,3) GA(2,3)
EQ GA(1,2,4) GA(2,4)
EQ GA(1,2,5) GA(2,5)
EQ GA(1,2,6) GA(2,6)
EQ GA(1,2,7) GA(2,7)
EQ GA(1,2,8) GA(2,8)
EQ GA(1,2,9) GA(2,9)
EQ GA(1,2,10) GA(2,10)
EQ GA(1,2,11) GA(2,11)
EQ PS(1,1,1) PS(1,1)
EQ PS(1,2,1) PS(2,1)
EQ PS(1,2,2) PS(2,2)
MA GA
*
1.5882 0.0000 0.0000 +0.0000 0.0000
0.0000 +0.0000 0.0000 0.0000 +0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.3482 +0.0000 +0.0000 0.0000 0.0000 +0.0000 +0.0000 0.0000 0.0000 0.0000 +0.0000 0.0000 0.0000
MA PS
*
0.3342 -0.0588 0.0859
MA LY
*
1 0
1 0
1 0
1 1
1 1
1 1
ST 0.0580 TE(1,1)
ST 0.0 TE(1,4)
ST 0.0125 TE(1,2)
ST 0.003703 TE(1,3)
OU PT PM MR SE PC TV FD ND=4
/*
// GO.FT10F001 DD DSN=W.W3137.A2616.LRTEMP,DISP=SHR,
// UNIT=USER, VOL=SER=USER51
// */
Appendix F

ESTIMATES OF FIRST AND SECOND MOMENTS OF FITTED DISTRIBUTIONS
OF MONTHLY CALLING RATES

To provide estimates of the mean and standard deviation of untransformed calling rate C based on specific values of the mean and variance of the fitted distribution of transformed calling rate C', we used approximation formulas reported by Huang and Kelingos (1979) and Huang and Grawe (1980).

In this report we work with fitted models of the general form:

\[(\text{CALLS})^\lambda = \mu + \xi, \text{ where } \xi \sim N(0, \sigma^2) \text{ approximately},\]

and \(\mu, \sigma^2\) are conditional on \(\lambda (= 0.27)\). Using this notation we write below a straightforward generalization of formulas (13) and (14) reported by Huang and Grawe (1980) for the conditional mean of untransformed calls C. Our generalization yields the following approximation formulas for the pth moment of \(C(p \geq 1)\) conditional on \(\mu, \sigma^2\) and \(\lambda:\)

\[
\mathbb{E}[C^p | (\mu, \sigma^2, \lambda)] = \mu^p / \lambda \left[ 1 + \sum_{n=1}^{\infty} \frac{c_n^{(p)}}{2n} K_{2n} \right],
\]

where

\[
c_n^{(p)} = \prod_{j=1}^{2n} \frac{p - (j - 1)\lambda}{j\lambda},
\]

\[
K_{2n} = (\sigma/\mu)^2 \left[ -2\mu h(\mu) + (2n - 1)K_{2(n-1)} \right], \text{ with } K_0 \equiv 1,
\]

\[
h(t) = f(t)/\int_{-\mu}^{\mu} f(x)dx, \quad f(t) = \exp[-(t/\sigma)^2/2]/\sigma(2\pi)^{1/2},
\]

for \(-\mu \leq t \leq \mu\).
Notice that the number of terms in the formula needed to assure a specified degree of precision will depend importantly on the ratio \( \sigma/\mu \) and \( \lambda \). For "large" \( \mu \) and "small" \( \sigma \) we would expect that only a "few" terms of the formula will need to be computed. For our work \( \sigma/\mu \) was less than 1/4 for nearly all the computations we made. For derivations and a detailed discussion of the existence of the conditional mean (i.e., \( p = 1 \) above) consult the two papers cited above.
REFERENCES


