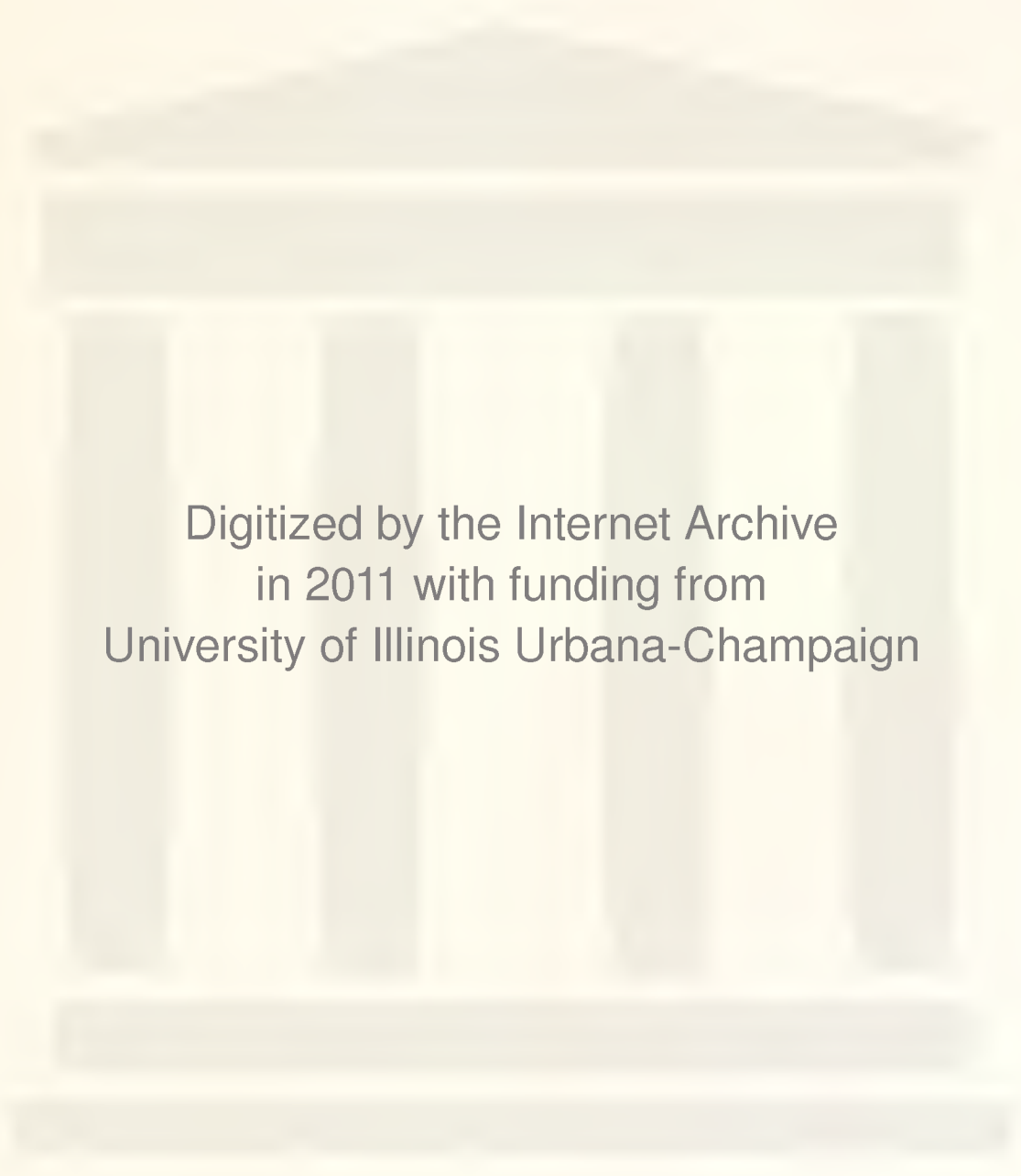


UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
STACKS



Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

<http://www.archive.org/details/onaccuracyofesti625fran>

Faculty Working Papers

ON THE ACCURACY OF ESTIMATING CUSTOMERS'
BILLS

Marvin Frankel, Professor, Department of
Economics

Walter J. Primeaux, Jr., Professor, Depart-
ment of Business Administration

#625

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign



FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

November 8, 1979

ON THE ACCURACY OF ESTIMATING CUSTOMERS'
BILLS

Marvin Frankel, Professor, Department of
Economics

Walter J. Primeaux, Jr., Professor, Depart-
ment of Business Administration

#625

Summary:

A large proportion of gas utilities face the need, whether continuous or occasional, of estimating their customers' consumption. The need is especially acute when the meter reading and billing cycles are not synchronous. This paper discusses the estimating problem in light of the authors' experience in reviewing the estimating methods of a major gas utility company. The article examines the alternative estimating methods available, their potential accuracy, and some of the factors to consider in seeking to improve accuracy. The problems of making estimates with minimal account data and of whether to include prior estimates in the data base also are discussed.



On the Accuracy of Estimating Customers' Bills

by

Marvin Frankel and Walter J. Primeaux, Jr.*

I. Introduction

A large proportion of gas utilities face the need, whether continuous or occasional, of estimating their customers' consumption. The need is especially acute when the meter reading and billing cycles are not synchronous. Thus, a company that reads bimonthly but bills monthly must estimate all of its accounts every other month.¹ Estimates also provide a means for testing the accuracy of meter readings, and they provide a basis for billing when, as often occurs, a reading cannot be made because of inclement weather or problems of gaining access.² The data in Table 1 show the large proportion of gas utilities involved in estimating customer consumption. The conditions which make estimated bills necessary or desirable are also faced by electric and water utilities. Although accurate estimation is crucial to utility companies, there has been a scarcity of treatment of the subject in the trade journals.

*Frankel is Professor of Economics and Primeaux is Professor of Business Administration at the University of Illinois at Urbana-Champaign. They would like to thank Angelo Cottini, Michael Reeves, W. E. Shipplett, William Terpstra, all of the Peoples Gas Light and Coke Company, Chicago, for their help and advice in developing the original study. Peoples Gas Light and Coke Company, of course, is not responsible for any of the views expressed herein.

¹The tendency in recent years for many utilities to change from monthly to bimonthly or quarterly meter readings is discussed by T. W. Sarowski in the session, "Estimating Customer Bills--Proven Methods and New Ideas," presented at A.G.A. and E.E.I. Accounting Conference, Houston, Texas, April 26-28, 1976, p. 1.

²Problems of meter reading are discussed in Samuel G. Hardy "Automatic Meter Reading: Panacea or Pandora's Box," Public Utility Fortnightly, December 6, 1973, p. 49.



Table 1

Estimating Frequency of Seventy-Three
Natural Gas Companies

	Number of of Companies	Size of Firm (Customers)			
		100,000 199,999	200,000 299,999	300,000 399,999	400,000 Over
<u>Residential Customers</u>					
YES they estimate consumption	69	33	10	6	20
NO they do not estimate consumption	1	--	1	--	--
NO response	3	3	--	--	--
<u>Commercial Customers</u>					
YES they estimate consumption	65	30	9	6	20
NO they do not estimate consumption	5	3	2	--	--
NO response	3	3	--	--	--
<u>Industrial Customers</u>					
YES they estimate consumption	32	12	5	5	10
NO they do not estimate consumption	32	21	6	1	10
NO response	3	3	--	--	--

Data are from a survey performed by Mr. L. Bramford of Washington Natural Gas Company in October 1974, cited by Jay R. Billingsley in a paper given at the session, "Estimating Customer Bills--Proven Methods and New Ideas," presented at A.G.A. and E.E.I. Accounting Conference, Houston, Texas, April 26-28, 1976, p. 6.



Events of the past several years have helped to focus attention on the estimating process and the problem of estimating accuracy. The increase in costs, including meter reading costs, has worked in favor of a reduction in the frequency of meter reading and a greater reliance on estimates. Inflationary pressures have given rise to periodic fuel cost adjustments and to a need by utilities to reflect them properly in customer billings, whether based on readings or estimates. These same pressures have made consumers more sensitive to their billings and more ready to complain about real or imaginary inaccuracies. In response, regulatory commissions have become alert to the problem and more disposed to raise questions about estimating methods.³ Accordingly, companies are more likely than in the past to be called on to explore and justify their procedures.

This article discusses the estimating problem in light of the authors' experience in reviewing the estimating methods of a major gas utility company. The article briefly considers the alternative estimating methods available, their potential accuracy, and some of the factors to consider in seeking to improve accuracy. Although individual utilities are variously situated with respect to their needs and to the opportunities and technical circumstances for making estimates, we believe that many of the points discussed will be of general relevance and interest.

³ See Sarowski, *op. cit.*, p. 2.

II. Methods of Estimation

Methods for estimating a customer's gas consumption may be variously classified. For our purposes, it is convenient to identify three main methods: the multi- or two-equation method; the regression method; and the projection method.⁴ In the two-equation approach, data from a customer's account history is identified for two separate time intervals. From these data, two equations are formed, each covering one of the intervals. In each equation, consumption for the interval is expressed as a function of the calendar days in the interval and its temperature (number of degree days). The two equations are solved to obtain a pair of estimating coefficients, one reflecting the base daily load and the other average temperature. These coefficients can then be applied to estimate consumption for any future period, given data on the days and degree days in the period.

The regression method, like the multi-equation method, treats consumption as dependent on a set of explanatory variables--days, degree days and any other factors deemed important. However, the estimating coefficients are derived through application of standard multiple regression techniques to the customer's account history. The coefficients are then used in the manner previously described to make estimates.

In the projection method, which may take different forms, data for a past period are projected forward with the aid of prior information or assumptions. Consider, for example, the base load-variable load

⁴Each of these methods presumes the existence of some consumption history for an account. Our discussion does not consider methods for making initial estimates for new customers.

approach. Experience might suggest that in a typical winter month, for a residential customer who uses gas for space heating as well as for cooking and hot water, 30% of consumption represents the base daily load and 70% is attributable to accumulated degree days. The assumption of a 30-70 split, together with information on days and degree-days for one or more past periods, permits calculation of the estimating coefficients. In a modification of this approach, it becomes essentially one of simple projection from data for a preceding period. The days and degree days for the period to be estimated can be ratioed to the corresponding figures for a past period, the two ratios combined into an overall average by applying weights, such as .30 and .70, to each, and the resulting average figure used to project the past period's consumption forward.

Each of the three methods is distinctive in its essentials. But each is subject to variation in the way it is actually applied.⁵ For example, with the regression method, different equational forms may be fitted to the data. With both this method and the two-equation method, different data bases may be relied on, with one or more periods excluded because they are thought to be unrepresentative or for other reasons. With any method a choice must be made of the explanatory variables to include. Thus, in estimating electricity or gas consumption, one may wish to include a variable for cooling days (as well as degree days).

⁵One study, examining one estimating procedure, is Billingsley, op. cit. The estimating method used, as well as the variables considered in the analysis, are affected significantly by the data bases available for use in the estimating model. Quality of the data, as well as the availability of certain series, is limited by economic considerations.

In estimating water consumption, a rainfall variable might be included for the summer months. There are options also in classifying accounts for estimating purposes. Thus, it ordinarily will be advantageous to treat residential and commercial accounts separately. Once the estimating coefficients have been obtained, further options are open for determining the procedures to be used for updating them and the standard to be used for deciding when that updating should occur.⁶ Finally, a complete estimating system may make provisions for testing the plausibility of meter readings, and different criteria can be used for this purpose.

III. A Case Study

The Peoples Gas Light and Coke Company, for whom our study was prepared, has a major need for consumption estimates, since it bills monthly but reads meters bimonthly. To facilitate our inquiry, the company provided data on 4000 accounts, half residential and half commercial, with each category divided into heating and non-heating groups. Each account contained approximately 13 periods of data, or all that were retained in active status. Each period averaged about one month in length, with considerable variation around this figure for some accounts. The information given for each account, for each period, included the beginning and ending dates, the number of days,

⁶For a brief discussion of one approach to updating, see the paper given by W. J. Cotting at the session, "Estimating Customer Bills-- Proven Methods and New Ideas," presented at the A.G.A. and E.E.I. Accounting Conference, Houston, Texas, April 26-28, 1976, pp. 3-4.

the cumulated percent season, and consumption or estimated consumption. The source of the consumption figure, such as meter reading, customer report, or company estimate, also was given. Other information included the type of account--whether residential, commercial, etc., the hi-lo category, and the geographic location of the account in terms of a grid map. The estimating coefficients most recently calculated for the account also were given.

These data were used to examine the efficacy of different estimating methods. Several versions of the two-equation approach and the regression approach were considered. Included in the review was a replication of the method--a variant of the two-equation approach--used by the company. In all instances only two explanatory variables were employed. One was a base load factor, represented by the number of days in a period. The other was a temperature factor, as represented by degree days or percent season. Allowance for the influence of a few other variables was accomplished through the classification and screening of account data. Thus, residential and commercial, and heating and non-heating, accounts were treated separately, and residential accounts with swimming pools were dropped from the eligible list.

The company's version of the two-equation method involves the following main steps for each account:

1. Separate the data into two time intervals, the first consisting of the entire time span for which data are available, e.g., October 3, 1977 through December 9, 1978, and the second consisting of the summer months only, with the exact span determined by the recorded meter reading dates.

2. For each interval, cumulate the number of days, number of degree days and consumption.

3. Form two equations, one for each interval, in which consumption as the dependent variable is related to the number of days and number of degree days as explanatory variables. To illustrate, the following two equations might result:

$$\begin{aligned} 366N + 111.60R &= 2664 \\ 124N + 1.61R &= 318 \end{aligned}$$

where N = base load or days coefficient
R = percent season coefficient

The first equation covers the entire time span and the second the summer interval. Solving these two equations yields $N = 2.355$ and $R = 16.148$. Given information for a subsequent period on days and percent season, these coefficients can be used to estimate the period's consumption.

Our review of the two-equation approach consisted of a replication of the company's basic procedure and of several alternative procedures. Of the latter, a few were designed to test the reliability of certain account entries and their usability for estimating purposes. Others involved variations in the estimating method. One such variation, for example, relied on only two periods of data, rather than all periods, in the account record. One of these periods contained the highest percent season per day, and the other the lowest percent season per day, of all periods available. The intent was to determine whether such a data base, though limited, might not in fact contain sufficient information to permit the calculation of acceptable estimating coefficients.

A second procedure tested whether, in instances where one or the other of the estimating coefficients turned out to be negative, estimates could be improved by eliminating the associated variable from consideration and recalculating the remaining coefficient using but one explanatory variable. A third procedure considered the effects of including in the data base information (where available) for two summer periods rather than only one. Other procedures explored yet other ways of determining the coefficients and tested the reliability and usefulness of certain account data.

Consider now the regression method. The basic procedure, the results of which we used for benchmark or reference purposes, involved the fitting to account data of an equation of the following form:

$$C = A + ND + RS$$

where C = gas consumption in the period (e.g., 30 days)
D = days in period
S = percent season in period

and A, N and R are coefficients to be determined. The variables C, D and S each represent sets of observations, one for each period, drawn from data in the individual account. The coefficients, once calculated, like the corresponding coefficients of the two-equation method, can subsequently be used to estimate consumption for any period, given the number of days and degree days it contains.

Other regression procedures examined included one that employed ratios of the variables, rather than the variables themselves; another that constrained the constant term to zero; and a third that incorporated a lagged effect to allow for the possibility that customers

may respond slowly, or with a delay, in adjusting their thermostats to seasonal temperature changes.

For each procedure, whether in the two-equation or regression category, the accuracy of the estimating coefficients was assessed through a three-stage process. First, for each account in the set of accounts under consideration (such as residential heating), the calculated coefficients were used to estimate consumption for the last recorded period. Second, the estimate was compared with actual consumption and the percentage difference, or error, calculated. Third a distribution of errors was formed and summarized by measures of the distribution's mean and its dispersion. The measures permit a comparative assessment to be made of the several procedures employed. They also serve to place the company's estimating procedure in technical perspective and provide a basis for determining whether particular changes in that procedure might contribute to improved estimates.

IV. Major Findings

The study produced a number of interesting results. A basic finding was that the estimating accuracy of procedures based on the regression method was, on the whole, inferior to the accuracy of procedures based on the two-equation method. The best, or strongest, of the two-equation procedures, when evaluated by one of the primary accuracy tests used in the study, proved marginally superior to the best of the regression procedures. When the same procedures were evaluated by a more stringent alternative test--one better suited to assessing forecasting accuracy--the two-equation method showed itself

superior by a wide margin. A probable reason for the poorer performance of the regression method, at least in the forms attempted by us, is its inability to reflect adequately the factors influencing summer consumption.⁷ In terms of causal factors, gas consumption during the summer is distinct, or disjoint, from consumption during the remainder of the year, since the temperature variable is above the threshold of influence in the summer season. The regression relationship does not properly reflect this fact, being weighted toward the more numerous months of the year when the seasonal force operates along a continuum. It may well be that other methods than those tried by us could effectively deal with this problem. It is possible also that the regression method might be satisfactorily used for the non-summer portions of the year, and a supplementary procedure used to estimate consumption for the summer months.

A second finding of interest was that a two-equation procedure using minimum data--one summer period and one winter period--produced an estimating accuracy equal to that of the corresponding procedure using all of the data in the account, amounting to thirteen periods in some instances. The two periods, drawn to reflect the maximum and minimum of seasonal influences, apparently contain virtually all of the information in the record relevant to the measure of seasonal and base load influences. This outcome is of particular interest for accounts

⁷As mentioned in footnote 5, an important limitation is the available data base which must be restricted for economic reasons. Moreover, there are practical limits as to what kind of data can be collected and maintained.

with limited histories, since it suggests that data gaps, even sizeable ones, may not compromise estimating accuracy. It suggests also, at least for gas consumption, that efforts to extend the account history beyond some modest time span, say twelve months, may not yield improvements in accuracy that are worth the extra costs of data storage.

A third finding, consistent with prior expectations, was that estimated consumption figures ought not to be used as part of the data base in calculating or updating the estimating coefficients. Such consumption figures contain no valid information beyond what is contained in the measured consumption data obtained from meter readings, and they may contain misinformation. Their use therefore cannot ordinarily improve the quality of the coefficients. In our experiments, the inclusion of estimated consumption data yielded poorer coefficients than when these data were excluded.

A fourth finding indicated greater difficulty in estimating commercial than residential accounts with comparable degrees of accuracy. Virtually all of the procedures attempted resulted in lower estimating accuracy for the former group, usually by substantial margins. This outcome was perhaps not surprising. Gas consumption by some business firms is sensitive to fluctuating levels of economic activity, as well as to base load and seasonal factors. None of the estimating procedures we attempted included an economic activity variable. Whether the inclusion of such a variable, or variables (e.g., retail sales or employment for the service region) would be useful depends upon whether firms, in their activity levels, respond in a sufficiently uniform way, to variations in business conditions.

One of the more important outcomes of the study for the sponsor was the discovery that its own estimating procedures, as applied in daily operations, produced accuracy levels equal to those of the best of the alternatives tested by us. This does not mean no improvement is possible, and the study suggested some avenues for exploration that might prove fruitful. Moreover, while the study examined what appeared to be the more promising of estimating alternatives, it was not exhaustive in its canvass. But it would appear that such improvements as might be achieved are more likely to be of a marginal than a major kind.

There are obvious limitations to the accuracy that can reasonably be expected of any estimating system. Since such systems typically must rely on but a few explanatory variables, they cannot reflect all of the influences on consumption. Moreover, there always will be some number of aberrant cases arising from unpredictable changes in a customer's behavior. But a good estimating system, in contrast to a less adequate one, can be expected to show a narrower error distribution, a smaller average error, a broader applicability to accounts containing limited data, and a capacity for minimizing the number of cases that must be relegated to manual review.



UNIVERSITY OF ILLINOIS-URBANA



3 0112 060296321