

**THE USE OF UTILITY DATA PROCESSING RECORDS  
AS A DATA MINING SOURCE ON LOW-INCOME CONSUMERS:  
CONVERTING INFORMATION TO KNOWLEDGE**

**By:**

**Roger D. Colton  
Fisher, Sheehan and Colton  
Public Finance and General Economics (FSC)  
34 Warwick Road, Belmont, MA 02178  
617-484-0597 / 617-484-0594 (fax)  
roger@fsconline.com \*\*\* <http://www.fsconline.com>**

**April 1999**

---

*We are drowning in information, but we are starved for knowledge.*

*James Naisbitt*

While most utilities view their data processing as a mere accounting system, in reality it is a valuable data base as well. The wealth of information maintained in existing utility systems can be tapped through simple analyses described herein to make educated determinations of: (1) who low-income consumers might be; and (2) what types of payment problems those consumers are experiencing. Using such analysis to target energy efficiency assistance to particular customers facing particular payment patterns would thus not only result in saved energy, but would help improve bill affordability as well. Such targeting would, in other words, help a utility use its low-income energy efficiency programs as a more refined tool to reduce utility expenses in the areas of bad debt, collection costs, working capital, and the like.

To more fully comprehend this notion of using existing internal data bases to "learn" about low-income customers, a basic understanding of a utility's customer data base is necessary. This paper introduces certain aspects of the customer data base for a public utility (electric or natural gas) and explains how that data base can be used to increase understanding of low-income payment patterns. It is necessary to keep in mind that while not all utility data bases will be identical, they will likely be substantially similar to the form discussed below.

The discussion is divided into two parts: The first part will describe particular information a utility is likely to maintain on each of its customers. The second part will suggest ways in which this information could be used in a utility data mining effort as to low-income consumers in particular.

The purpose of the discussion below is not to assert that particular payment patterns have particular meanings or can be attributable to particular causes. Instead, the purpose is simply to open up the possibility in the reader's mind that what has previously been considered a mere

accounting tool might, in fact, provide useful insights into the nature and causes of nonpayment. Accordingly, that data might be useful in helping to wield energy efficiency as a tool to address nonpayment problems.

## **THE DATA RECORDS**

It is not profound to recognize that a local utility keeps extensive records on each of its customers. Perhaps by better understanding the contents of those records, those records can be used to the advantage of low-income households in devising targeted energy efficiency programs to help address inability-to-pay problems. Utilities maintain three types of records that may well be helpful in targeting energy efficiency efforts. The first involves "treatment histories." The second involves "vintaging" of arrears. The third involves a series of dates that are important for each customer.

### ***Treatment History***

A utility's "account treatment" involves those actions which a utility takes to collect its bills each month. Accounts are often "treated" in a hierarchical fashion. A typical treatment hierarchy might involve the following four steps, with each step involving a more stringent collection technique:

Steps in the Account Treatment Process	
Step 1	Reminder notice
Step 2	Shutoff notice
Step 3	"Final" notice
Step 4	Service termination

Most utilities track a customer's "treatment history" in their data processing files. A utility might retain this history, depending on the utility, for 13 to 24 months or more. In its data base, the utility will record the highest treatment level experienced by the customer for a given month. Given the progressively more stringent nature of the treatment steps, a record of a high treatment level necessarily implies the presence of all "lesser included" treatment steps. Thus, in the above illustration, if a utility records a "final notice," that record necessarily implies that the household account has received the "treatment steps" of a reminder notice and a shutoff notice as well.

A utility will assign a numerical code to each treatment step for purposes of recordkeeping, with a higher number representing a more stringent treatment measure. So, for example, a typical utility data base code might look something like this:

Recording Steps in the Account Treatment Process		
Treatment Step	Treatment Action	Way Recorded in Data Processing System
	No treatment	0
Step 1	Sent reminder notice	1
Step 2	Sent shutoff notice	3
Step 3	Sent "final" notice	5
Step 4	Service was terminated	7

Treatment histories are kept on a rolling basis. Thus, in the instance where a utility keeps 13 months of history, in February of 1999, the company would add February to the file and delete the treatment history for January 1998. In each succeeding month, the most recent month is added and the oldest month deleted from the file.

An illustration might further help explain. Let's assume that we have a utility with three customers, which are cleverly named Customer 1, Customer 2 and Customer 3. The following treatment histories are for the calendar year 1998 (with January being in the left most field). In our hypothetical, the utility data records on treatment histories look like this:

Customer Name	Treatment History in Data Base
Customer 1	111105700011
Customer 2	301350333333
Customer 3	555571000013

If this type of field appears in a utility's data processing system, the way to read it is as follows:

	Account Treatments Directed Toward Customers											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Customer 1	1	1	1	1	0	5	7	0	0	0	1	1
Customer 2	3	0	1	3	5	0	3	3	3	3	3	3
Customer 3	5	5	5	5	7	1	0	0	0	0	1	3

0=No treatment; 1=sent reminder notice; 3=sent shutoff notice; 5=sent "final notice"; 7=service terminated.

The following discussion does two things with these records: (1) it explains what the records say on their face; and (2) it explains some of the conclusions/deductions one might reasonably draw from these particular records. What factual patterns lead to what deductions, of course, is highly dependent on the person making the deductions. In reaching conclusions, it is important that deductions are reasonably empirically based and not used simply to confirm predispositions or stereotypes.

### **Facial Information**

The following information appears on the face of the treatment histories which appear above.

**Customer 1:** This person received reminder notices in January through April. This customer received a "final notice" in June and was actually disconnected in July. It is not possible to know whether the household remained disconnected in August through October or whether the customer was reconnected and paid the bill in full and on time during those months. (Some utilities, however, will also keep a disconnect/reconnect date, as discussed below, so it is possible to separately check whether and when the household was reconnected). The treatment in November and December makes clear that the customer had service reconnected by that time and that the household received "reminder notices" in each of those months.

**Customer 2:** This customer received shutoff notices in every month but February, March, May and June. The household was paid in full and on time in February and received a reminder notice in May but was paid in full in June. We know that the "O" in February and June represents a paid bill (and not a month in which the household was disconnected and off of the system) because a shutoff would have been represented by a "7" in January and May.

**Customer 3:** This person received a reminder notice in June and November; a shutoff notice in December; and final notices in January through April. The household had service disconnected in May. Given the reminder notice in June and no further disconnections, we know the household was reconnected after the May disconnection and remained current on each monthly bill through October (assuming no voluntary disconnect during that time period).

### **Deductive information**

Beyond the information that is apparent on the face of the treatment histories, a number of deductions can be reached from those histories as well. It is up to each individual to decide which deductions are "reasonable":

**Customer 1:** This is a person who is not in chronic trouble with her energy bills. To

allow a payment to slide so as to get a reminder notice may be as much a personal money management technique as anything else. (In the alternative, it may indicate simply a mismatch in the date on which a social security check is received and the date that a utility bill payment is "due"). Customer 1 would appear to be a classic case of a catastrophic occurrence. It is impossible to tell whether it was an unexpected expense (such as an illness or the need for an appliance repair) or the temporary loss of a job or the like. Whatever the financial crisis, however, it is possible to conclude that it was serious enough to push a household that had experienced no prior significant payment troubles to the point of an actual disconnection for nonpayment. Once the crisis passed, the payments returned to normal.

**Customer 2:** This customer is in chronic payment trouble with her utility bill, receiving shutoff notices virtually every month. The household is likely a low-income household skating on the thin ice of economic viability. This household likely applied for, and received, LIHEAP. LIHEAP benefits are generally actually received in December and January. This household may have received an "emergency crisis" grant in May or June, thus allowing the June bill to be paid in full. This is a household for whom a payment plan or a budget plan (with leveled equal monthly payments) won't help. The household needs public assistance, not level payments. Energy efficiency directed toward this household may have one of two impacts, both of which are positive: (1) to reduce bills to the point of being affordable; or (2) reducing bills so as to at least reduce the financial exposure of the utility arising from nonpayment (assuming that an unpaid \$100 bill is less harmful than an unpaid \$150 bill).

**Customer 3:** This is a household with chronic winter payment troubles. There is no evidence of a LIHEAP or crisis payment. If such payments were obtained, the household's energy bill is so large (or the household's arrears are so substantial) that the benefits were swamped by the current bill or outstanding arrears. The household should be placed on a priority conservation/weatherization list in that event. The household likely took advantage of a winter moratorium to maintain winter service despite nonpayment and despite repeated shutoff notices. In May, after the moratorium lapsed, the high bills and arrears caught up with the household and service was disconnected. It is reasonable, also, to assume one of two alternative theories about this household's finances. First, the household could be sufficiently close to the economic edge that, while the low summer bills pose no problem, the higher winter bills are simply unaffordable. Second, the household could have seasonal employment, which unfortunately ends with the advent of winter. The loss of income combined with the high winter bills, pushes the household into a nonpayment situation. Household 3 may well be the type of household who could benefit from a leveled budget billing plan.

An agency or utility providing energy efficiency assistance, of course, in consultation with the local fuel assistance agency and consumer groups, may define "payment troubled" to meet its own criteria. There is no objective determination of "payment-troubled." Whether energy efficiency

assistance is to households with 10 shutoff notices, or with a certain sized arrears, or with some other characteristic in particular is not the point. Rather, the lesson to be learned is that utilities have an array of useful information that can be readily accessed for the benefit of targeting energy efficiency assistance.

***Arrears Vintaging***

All arrears are not alike on a utility bill. Most utilities keep track of their arrears by vintage. A typical utility places arrears into one of three categories:

- ∅ 31 - 60 days overdue
- ∅ 61 - 90 days overdue
- ∅ 91+ days overdue

Clearly, households with larger and older arrears are considered more problematic than households with smaller or newer arrears. A household with a 90-day arrears of \$200 is of more concern to a utility than a household with a 30-day arrears of \$200.

It is important to understand how a utility company bills and collects its accounts. A universal principle is that payments are applied to bills on a first-in, first-out basis. Again, perhaps an illustration can best help explain. Assume our Ratepayer --let's call her Customer A-- has received the following bills for the six months of January through June:

	Monthly Bills					
	Jan	Feb	Mar	Apr	May	Jun
Customer A	\$126.42	\$134.18	\$87.66	\$65.00	\$48.18	\$22.41

We will assume that Customer A began with a balance of zero dollars, so the only arrears are current arrears (she brought nothing forward from pre-January). Let's assume that Customer A made the following payments:

	Monthly Payments					
	Jan	Feb	Mar	Apr	May	Jun
Customer A	\$96.42	\$20.00	\$50.00	\$50.00	\$48.18	\$22.41

The arrears would be as follows:

	Monthly Arrears					
	Jan	Feb	Mar	Apr	May	Jun
Bills	\$126.42	\$134.18	\$87.66	\$65.00	\$48.18	\$22.41
Payments	\$96.42	\$20.00	\$50.00	\$50.00	\$48.18	\$22.41
Arrears	\$30.00	\$144.18	\$181.84	\$196.84	\$196.84	\$196.84

The household, however, is not considered "as well off" in June as in April, even though the amount of the arrears is the same. The *vintage* of arrears would be as follows:

	Total	30-60 Days	60-90 Days	90+ Days
January	\$30.00	\$30.00	\$00.00	\$00.00
February	\$144.18	\$134.18	\$10.00	\$00.00
March	\$181.18	\$87.66	\$94.18	\$00.00
April	\$196.84	\$65.00	\$87.66	\$44.18
May	\$196.84	\$48.18	\$65.00	\$83.66
June	\$196.84	\$22.41	\$48.18	\$126.25

As can be seen, while at first glance, this person may seem to be holding her own in recent months, in the eyes of the company, the debt is becoming older and, therefore, more in jeopardy of ultimate nonpayment. On the other hand, the customer *is* making payments each month. Indeed, given the customer's attempt to pay at least the current bill, the utility's response to this payment pattern might reasonably be to couple energy efficiency, with a levelized budget plan for current bills, and an extended payment plan for the arrears. In that fashion, the customer may end up paying no more (or not much more) than was currently being paid during the latter months (March through June), while still retiring the arrears.

Utilities may track the vintage of arrears in one of two ways:

**First:** The vintage of the oldest arrears might be given a code and tracked for 13 to 24 months or more (again, depending on the utility). Thus, a utility might code its arrears as follows: 30-60 days=1; 60-90 days=3; 91+ days=5. Given first-in, first-out accounting, with its policy of always applying payments to the oldest bills first, absent a disconnection of service, a household having an older arrears (e.g., 91+ days) will *always* have the newer arrears as well. Thus a 91+ days code will necessarily imply the presence of 30-60 day

arrears as well as 60-90 day arrears. Under such a coding, the hypothetical Ratepayer A would have the following arrears history:

<b>Error! Bookmark not defined.</b>	Arrears Vintage of Customer A					
	Jan	Feb	Mar	Apr	May	Jun
Customer A	1	3	3	5	5	5

**Second:** A utility might track the actual dollars of the most recent month's arrears by vintage. Thus, in our example, under this method, in July, the utility would record the arrears as being

	30 - 60 days	61 - 90 days	91+ days
July arrears	\$22.41	\$48.18	\$126.25

On occasion, there will be a utility which tracks vintages both ways in its customer data files. Most often, this utility will track vintages both ways for the past three months, but will use only a code for the arrears stretching back further than that.

The vintage of arrears would be helpful in directing the "problem" to be addressed (and thus the justification for) energy efficiency investments. Assume that a company has two customers. Customer A has high, but short-term arrears each year (*e.g.*, the customer incurs high winter arrears but always pays those arrears off by the next heating season). Customer B has marginally lower, but substantially longer-term, arrears. In the case of Customer A, energy efficiency used to generate non-energy financial benefits would rely upon savings in working capital expenses. To the extent that the high winter bills can be reduced, the utility would not need to carry the debt. In contrast, in the case of Customer B, energy efficiency used to generate non-energy financial benefits would rely not only on working capital savings, but on the possible reduction in bad debt as well.

In sum, the important data mining lesson for purposes here is simply that not all arrears are alike. Merely because two customers are both \$200 in arrears does not mean that they are "equal" in the eyes of a utility. The information is in the data processing system. The trick is to convert that information into knowledge.

***Important Dates.***

A utility will keep three dates in its data files, two of which are important for purposes of data mining in support of energy efficiency programs. (The third is discussed simply because of its propensity to cause confusion). The three dates are: (1) the meter date; (2) the in-service date;

and (3) the disconnect/reconnect date. Each will be briefly addressed below.

1. **Meter date:** This date is the date that the present meter was installed in a particular premises. It has significance for ratemaking, but not for any customer service purpose. The importance of this date for discussions here lies in what it is *not*. It does not indicate anything about how long the current customer has been taking service. The meter date, in other words, is tied to the premises, not to the customer. The meter date is not the same as the in-service date.
2. **In-service date:** This is the date on which the current customer first began service at the particular service location. Several important aspects of this date are worth noting. First, if the household is disconnected and reconnected, the in-service date does *not* change. The collection activity is reflected in the disconnect/reconnect date discussed below, as well as in the treatment history discussed above. Second, if the household is disconnected and then has service re-established in a different name (*e.g.*, a spouse, child, or other relative), there will be a new in-service date. The utility will consider it a new account with a new customer. Finally, if a household changes addresses, there will be a new in-service date. The in-service date, in other words, is for service at a particular location. It does not indicate the length of time a customer has been receiving service from a particular company, but rather the length of time the customer has been receiving service *at that address*.
3. **Disconnect/reconnect date:** This will likely be recorded as one date. If there has been no disconnection or reconnection in the past 12-months, the field will either be blank, or will be filled with zeroes. If there has been a disconnection and/or reconnection, the advocate must determine whether the customer is currently on the system. If the customer is active, the date will be the date of reconnection. If the customer is inactive, the date will be the date of disconnection. The date will only refer to the most recent disconnection and/or reconnection. If there has been more than one disconnection/reconnection in the immediately preceding twelve months, the older ones will not be reflected. (Remember, however, that the presence of disconnections will always appear in the treatment history discussed above.)

### ***Other Potential Data***

It is necessary for the energy efficiency planner to decide what other relevant pieces of information a utility's data base might record. One problem for many utilities, for example, is the mere process of identifying precisely who their low-income customers are. Many utilities do not have reason to record income information. It might be possible, however, to build on other knowledge of low-income consumers. For example, it is known that low-income households tend not to have bank accounts. Is it legitimate, therefore, to track customers who pay in cash? Is it possible to track customers who pay in cash at community pay stations in low-income neighborhoods? Is it possible to track customers who pay in cash at community pay stations

located at check-cashing stores in low-income neighborhoods? Is it possible to track customers who pay in cash at community pay stations located at check-cashing stores in low-income neighborhoods who have unpaid bills during the winter months? Is it possible to track customers who pay in cash at community pay stations located at check-cashing stores in low-income neighborhoods who have unpaid bills during the winter months and energy consumption more than 130% of the residential average? Would the process of making such inquiries help the energy efficiency program manager who wants to maximize his or her non-energy benefits to target the program's efforts? One might think so.

## **ANALYSIS OF THE DATA**

Utilities can use the information available in their present data files to engage in targeting energy efficiency to vulnerable populations. As can be seen, utilities have information available that allows them to identify households that might particularly benefit from the receipt of energy efficiency assistance. While utilities can always generally target energy efficiency assistance to low-income consumers, the following are examples of more sophisticated targeting.

1. **Winter treatment history**: Defining "winter" as October through April, a utility could target any household that received two or more disconnect notices in the past winter season. The utility would simply identify, in other words, those households with a code of "3" or above in two or more of the winter month treatment history records.
2. **Annual treatment history**: A utility could target any household with a minimum annual treatment history. For example, if the sum of the 12-months of codes exceeds 30 (or any other figure deemed appropriate), that household would receive outreach as to whether energy efficiency might be of help. A similar approach could be taken for the winter treatment history as well. If the sum of the treatment codes for the six winter months (November through April) exceeded a designated level, that household would receive special outreach efforts.
3. **Shutoff in treatment history**: A utility could target any household experiencing an actual disconnection of service within the past 12 months. The utility would simply search its treatment history records and extract those accounts which have a "7" appear.
4. **Annual 90-day arrears**: A utility could target any household that maintained a 90-days arrears in six or more months (or whatever other threshold is deemed appropriate). The utility would identify, in other words, what months have a "5" in the arrears vintaging fields. If there are six or more fields with a "5" or higher, that household would be contacted to see if energy efficiency might be of help.
5. **Pre-winter arrears**: A different way to use arrears involves utilities which track the dollars of arrears, by vintage, for the most recent month (as discussed above). This utility might look at its customers in October each year. Energy efficiency might then be

targeted to customers based on either of two types of criteria. On the one hand, the utility could target any customer with a 90-days arrears on the October bill. On the other hand, the utility may wish to place a dollar floor on the arrears. In this instance, for example, targeted assistance would go only to households with a 90-day arrears of \$100 or more.

6. **Spring shutoff:** A spring shutoff --defined as a shutoff in April or May-- can be determined from the "disconnection/reconnection" date in the utility data base. A spring shutoff may well indicate a household who relied upon a state's winter moratorium for protection during the heating months, but whose inability to pay could not be avoided when the moratorium lapsed. The offer of energy efficiency assistance might help this household bring its winter bills within an affordable range.

In each instance above, the means of targeting is intended to reach payment-troubled households. Engaging in the assumption that payment-troubled customers also tend to be low-income households, the above evaluation suggests that ways exist in which the utilities can use their own data bases to assist in addressing such problems by identifying them through their own data bases.

Public utilities can take an active role in targeted energy efficiency efforts. Not only has the need arisen for involvement in efforts to address low-income payment problems, but the utility program manager who is experiencing increasing pressure to cost-justify his or her program may want to engage in data mining to help expand the positive impacts to the company arising from the program.