Appendix B
Demand Forecasting and Revenue Modeling

Contents

Introduction

Demand forecasting serves as a critical step in the planning, design and evaluation of a rate structure. In order to ensure that revenue collected will cover costs, water suppliers need to anticipate how much water they expect to sell. As water rates are typically reviewed and revised every few years, it is also important that water suppliers forecast future demand several years in advance to ensure that sufficient funds are collected.

Robust evaluation of efficiency-oriented rate structures requires an additional layer of forecasting impacts. Since variations in demand tend to create revenue volatility, forecasting models must consider the impact of block rate structures not just on demand, but also on sales revenue. Accurately forecasting long-term sales volume lies at the heart of establishing a correct rate level. Analysts must consider water supply availability, future water demand, and the effect of different types of rates on revenue.

This appendix covers approaches, concepts and methods for water demand and water sales/ revenue modeling.[1]

Demand Modeling

Different types of models can be used to incorporate the various forces driving water demand. Models can be classified as aggregate (total water demand for an entire service area or customer class) or disaggregate (demand by individual customer or individual end uses). In principle, disaggregate models can answer a wider range of questions; they also require more detailed data, more data manipulation, and more data validation effort. For readers interested in pursuing a modeling effort using disaggregated data, an example of disaggregate models applied to the prediction of revenue uncertainty can be found in Chesnutt, et al. (1995b). For the heuristic purposes of this handbook, aggregate data are used to illustrate models of water demand.

Most methods used to predict the effect of rate changes on demand response look at average water demand. Customer billing records provide a good tool for seeing these demand distributions, which tend to be very skewed. Figure B.1 depicts the parametric demand distribution from a random sample of single family customers using recent year data—with values for the mean and standard deviation (logarithm of bimonthly use) of 3.4 and 0.7 respectively. The distribution is notably skewed to the right; relatively few customers use a large amount of water.

1 Parts of this appendix were adapted from *"Designing, Evaluating, and Implementing Conservation Rate Structures",* July, 1997, California Urban Water Conservation Council.

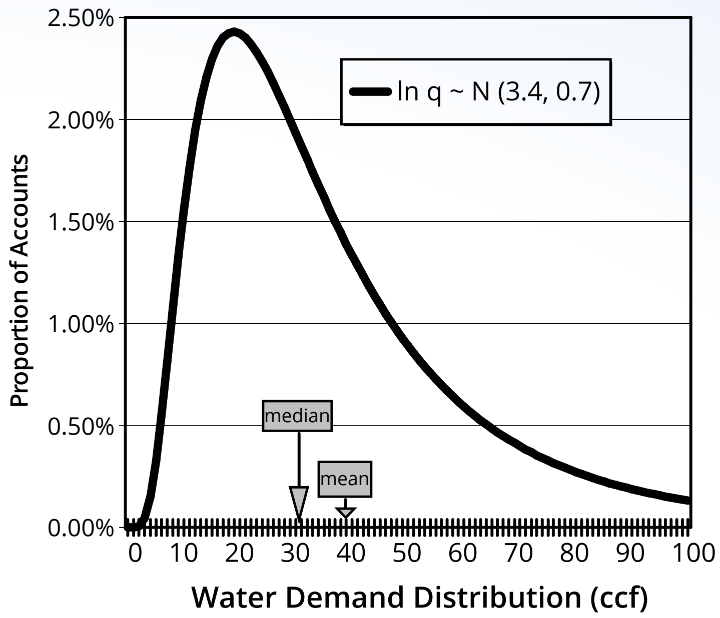

Figure B.1 Distribution of Single Family Water Demand

A right-skewed distribution complicates the design of block rate structures. Revenue prediction for block rate structures requires a more nuanced model than average water demand alone; it requires a model of the entire demand distribution (Chesnutt, et al. 1995b). The rate analyst might want to know how many customers and how much water would be affected by a second block. To estimate revenue, the analyst must know how much water is affected by the higher price in the second block. The next complication in estimating revenue from blocks is accounting for the fact that not all of the water used by customers in the second block is priced at the higher second price.

Model Zero: Water Requirements Model

The date is July 14, 2010. The location is West Anywhere. The water manager of the West Anywhere Water Authority (WAWA) has asked the in-house expert to forecast water sales for next year. The rate analyst hit the recalculate button on the computer and results were produced: water sales in the next year will be about 35 thousand acre feet. The manager inquired as to the model behind this prediction. "Well, the city planning department has projected next year's population of around 170 thousand people and water requirements for the last twenty years have averaged about 183 gallons per capita per day. The rest is algebra (170,000 people x 183 gallons per capita per day x 365 days per year / 325900 gallons per AF = 34,842 AF)".

The astute manager, worried that recent trends might change the result, asked the analyst to repeat the calculation using water requirements only from the last three years. The rate analyst hit the recalculate button on the computer, the lights dimmed, and results were produced: water sales in 2011 will be 35,600 thousand acre feet. "Are you sure," demanded the manager? "Yes, I hit the recalculate button three times before the lights blew. I got the same answer each time."

Confident in this sensitivity analysis, the water manager adopted 35 thousand acre feet as the 2011 sales forecast. The water sales in the next year ended up being less than 27 thousand acre feet. Later the manager and the analyst compared notes in the unemployment line: "Where did we go wrong?" They decided to think more formally about their mistaken analysis. They began with a formal statement of the water requirements model:

Sales Quantity = f (Population), or
Q=Pop·μ, where
μ ≡ mean *use per person*

They decided that this explanatory model should be expanded to account for more determinants of water demand. They decided that to conduct their retrospective analysis in a spreadsheet (demand. xls) using a monthly data set compiled by the previous year's summer intern.

Model 1: A Very Simple Water Use Model

The first step to improving the water requirements model requires adding additional explanation to the sales model. A simple possibility would be to add measures of weather— temperature and precipitation—to the model. In functional notation, the model would be described as:

Sales Quantity = ***f*** *(Population, Temperature, Rainfall)*

To make the model explicit, one must specify exactly how these determinants relate to sales, that is, the form of the function ***f***. A simple possibility would be a linear equation for monthly water sales:

$$Q_t = \beta_0 + \beta_1 \cdot Pop_t + \beta_2 \cdot Temp_t + \beta_3 \cdot Rain_t$$

Through the miracles of modern statistical technology, the four coefficients (β0 - β3) can be estimated to "fit" this surface to observed data. Each of the β parameters also has an interpretation—β0 is the "intercept" that represents a constant level of sales each month and the other β's are the "slope" coefficients of the determinants. These slope coefficients represent how water sales would change if one determinant changes by a small amount while all other determinants remained unchanged. For example, we expect use to decrease if rainfall increases in a given month; therefore β3 should be a negative quantity. The reverse holds for temperature and population, so β1 and β2 should be positive values.

If life were simple, the systematic determinants in the model described by the equation above would fit the data perfectly: all water sales on record would lie on the plane defined by f. Clearly, this will not be a problem that many analysts need lose sleep over. The vertical distance from any point to this plane defines the nonsystematic error in the model. Defining this quantity by ε, the very simple model of water use (Model 1 in the spreadsheet) can be described as

$$Q_t = \beta_0 + \beta_1 \cdot Pop_t + \beta_2 \cdot Temp_t + \beta_3 \cdot Rain_t + \varepsilon_t$$

In general, modelers are happier when they can minimize the unexplained random error of a model while maximizing a model's explanatory power. The most popular regression method, Least Squares, derives its coefficient estimates so as to minimize the (squared) error around the equation.

Current generation spreadsheets implement Least Squares regression as an analysis option. The reader may visit the accompanying spreadsheet (demand.xlw) to produce estimates of the four coefficients in the model described by equation above.

These estimates imply the following water use equation:

$$Q_t = -84.7 + 0.31 \cdot Pop_t + 1.61 \cdot Temp_t - 53.4 \cdot Rain_t + \varepsilon_t$$

The coefficients may be interpreted as the effect, with everything else constant, of a one unit change in the determinant upon the dependent variable, water sales. Thus, a one unit increase in Anywhere population (one thousand people) would result in a .31 acre foot per day increase in monthly water sales. Similarly, a one unit increase in precipitation (inches per day) would result in a 53.4 acre foot per day decrease in monthly water sales and a one unit increase in monthly average maximum daily air temperature (degrees Fahrenheit) would result in a 1.6 acre foot per day increase in monthly water sales. Note that our unemployed research team was careful to standardize all measures for the number of days in the month to ensure comparability.

Critique: The main strength of this model is ease of explanation. Technically, the model has more than a few shortcomings. The "fit" of the model is not terrific for a trending dependent variable. The R^2 statistic (R^2=.77) refers to the proportion of the variation in water sales explained by the model; Model 1 explains about 77 percent of the variation in water sales. The estimated error, implied by the estimated coefficients, is far from random. (This can be verified by plotting the estimated error, or a 12 month moving average, over time.) The functional form of the model asserts that the estimated effects remains the same in each month throughout the year; one inch of rain in January, for instance, would produce the same drop in sales as an inch in July. The rainfall and temperature measures are also highly (negatively) correlated; when rainfall increases, air temperature tends to decrease. This makes it difficult for any amount of statistical magic to discern the independent effect of each. The functional form of the model only allows for seasonal movement in sales through seasonal movements in temperature. Last, this model implies that changes in the price of water have no effect upon the level of water sales. None. Zero.

Model 2: A Simple Water Demand Model

To improve upon Model 1, several changes are adopted. First, a different functional form is specified—a logarithmic transformation—for the dependent variable and the independent variables. A different functional form illustrates a different connection between water demand and its determinants, a set of coefficients having different statistical properties, and a different set of coefficient interpretations. Second, Model 2 permits a separate intercept term for each month to better capture a constant seasonal pattern. Third, the climatic measures are expressed somewhat differently. Instead of the absolute amount of rainfall that fell in a month, the model uses the amount of rain minus the average rainfall for the month. Similarly, temperature is expressed as its deviation from monthly mean temperature. Logarithmically transformed population is expressed as its deviation from sample mean. Last, a measure of real (inflation-adjusted) marginal price is added to the model. Because all measures are logarithmically transformed, the estimated coefficients can be interpreted as elasticities: the percentage effect that a one percent change in the determinant will have on water sales. (Because the mean monthly amount of daily rainfall is fractional, a scaling factor of one is added prior to logarithmic transformation.)

Re-expressed, the second attempt at model improvement results in the following model specification:

$$lnQ_t = \beta_1 \cdot (lnPop_t - \mu_{lnPop_t}) + \beta_2 \cdot (lnPrice_t - \mu_{lnPrice_t}) + \beta_3 \cdot (lnTemp_t - \mu_{lnTemp_t}) + \beta_4 \cdot ((lnRain+1)_t - \mu_{lnRain+1_t}) + \beta_5 \cdot (mo1 \equiv Jan. = 1) + \beta_6 \cdot (mo2 \equiv Feb. = 1) + \cdots + +\beta_{16} \cdot (mo12 \equiv Dec = 1) + \varepsilon_t$$

Table B.1 provides the regression estimates of the Model 2 coefficients that imply the following monthly water demand equation.

$$lnQ_t = \beta_1 \cdot (lnPop_t - \mu_{lnPop_t}) + \beta_2 \cdot (lnPrice_t - \mu_{lnPrice_t}) + \beta_3 \cdot (lnTemp_t - \mu_{lnTemp_t}) + \beta_4 \cdot ((lnRain+1)_t - \mu_{lnRain+1_t}) + \beta_5 \cdot (mo1 \equiv Jan. = 1) + \beta_6 \cdot (mo2 \equiv Feb. = 1) + \cdots + +\beta_{16} \cdot (mo12 \equiv Dec = 1) + \varepsilon_t$$

Table B.1 A simple Model of Water Demand

SUMMARY OUTPUT, MODEL 2

Regression Statistics	
Multiple R	0.919
R Square	0.844
Adjusted R Square	0.834
Standard Error	0.081
Observations	264

Date	Time
29-Sep-91	1:12:20 AM

ANOVA

	df	SS	MS	F
Regression	16	8.897	0.556	83.765
Residual	248	1.646	0.007	
Total	264	10.543		

	Coefficients	Standard Error	t Stat	P-value
ln_pop (demeaned)	0.988	0.174	5.695	0.000
ln_price (demeaned)	-0.090	0.048	-1.887	0.060
dlr_mean	-0.805	0.111	-7.251	0.000
dlt_mean	0.914	0.127	7.173	0.000
mo1	4.160	0.017	239.247	0.000
mo2	4.170	0.017	239.923	0.000
mo3	4.183	0.017	240.467	0.000
mo4	4.317	0.017	248.343	0.000
mo5	4.422	0.017	254.440	0.000
mo6	4.529	0.017	260.709	0.000
mo7	4.601	0.017	264.786	0.000
mo8	4.614	0.017	264.996	0.000
mo9	4.496	0.017	258.401	0.000
mo10	4.389	0.017	252.493	0.000
mo11	4.285	0.017	246.666	0.000
mo12	4.196	0.017	241.501	0.000

The coefficients can be interpreted as the percentage effect on water demand associated with a one percent change in the determinant, with everything else constant. Thus, a one percent increase in Anywhere population results in almost one percent increase in monthly water demand. A one percent increase in the real price of water results in less than a tenth of one percent decline in water demand. A one percent increase in precipitation (over its monthly mean) results in a .8 percent in monthly water demand and a one percent increase in monthly average maximum daily air temperature (over its monthly mean) results in a .9 percent increase in monthly water demand. The intercept term has been given a seasonal dimension; each month has its own intercept. This monthly intercept represents an estimate of the normal water use pattern over the historical period.

Critique: The strength of Model 2 is that it carefully separates a constant seasonal pattern from the climatic measures. Because any constant seasonal pattern has been removed from the climatic measures (the monthly averages of climate are estimated via "interim" regressions in the spreadsheet), the "departure-from-mean" form of rainfall and temperature are independent of the seasonal effect (the 12 monthly indicator or dummy variables). The logarithmic transformation of water use results in a less skewed dependent variable and a better fitting equation. The improved fit implies that the model is leaving less unexplained; the random error that remains is 1) smaller in magnitude, 2) more normally distributed, and 3) has less "structure" in it. The model, though straightforward and parsimonious, is still far from perfect:

- The effect of a drought in earlier historical periods has been ignored.
- The effects of WUE (conservation) programs) have been ignored.
- The effect of climate is constant through the year (a one percent increase in temperature above normal has the same percentage effect in July and January.)
- The effect of climate has no memory (last month's or year's climate does not effect the current month.)
- The equation error still contains information that could be used to improve the (efficiency of) statistical estimation of the structural coefficients.

Further caveats are in order for the statistical estimate of the response of water demand to price. Even the single (mean) price elasticity produced by aggregate demand models tends to be very sensitive to model specification and the period of time over which the model is estimated. The fact that aggregate time-series models tend to produce unstable estimates of price response can be attributed to several factors: 1) insufficient variation in historical water rates, 2) measurement error in the price measure (any single price measure used—modal, median, or some melded average—does not reflect the true marginal price faced by each customer), and 3) omitted long-run determinants due to lack of measures or an insufficiently long time period. Empirical investigators interested in estimating the determinants of demand tend to favor customer-specific models to bring the weight of more data to bear on these difficult questions.

Model 2 does illustrate, however, that careful data construction can produce a simple model with plausible estimates of short and long-run determinants of demand. Other useful ingredients for rate evaluation also can be derived from this model. One construct that will prove very handy in the next section is a measure of the percentage effect of climate on aggregate demand. Since climatic uncertainty is a very important driver of demand uncertainty in the short-run, quantifying the magnitude of this uncertainty allows the analyst to construct a measure of revenue risk. Similarly, an estimate of the demand pattern under normal weather conditions—a constant seasonal pattern—can be derived. Estimating an average seasonal pattern permits empirical testing for changes to the pattern of seasonal peaking. Last, the model provides an explicit method for addressing how changes in rates can affect water demand.

Revenue and Sales Modeling

Most methods used to predict the effect of rate changes on demand response look at average water demand. Block rate structures, however, require a more nuanced model than average water demand alone; they require a model of the entire demand distribution (Chesnutt, et al. 1995b). Customer billing records provide a good tool for seeing these demand distributions. The distribution of customer use tends to be very skewed. To illustrate, a random sample of single family customers was taken from the WAWA billing system, with data including meter read date, meter read amount (in one hundred cubic feet, CCF), and the number of days in the billing system. For the purposes of illustration, Figure B.2 depicts the parametric demand distribution using the recent year data—with values for the mean and standard deviation (logarithm of bimonthly use) of 3.4 and 0.7 respectively. The distribution is notably skewed to the right; relatively few customers use a large amount of water.

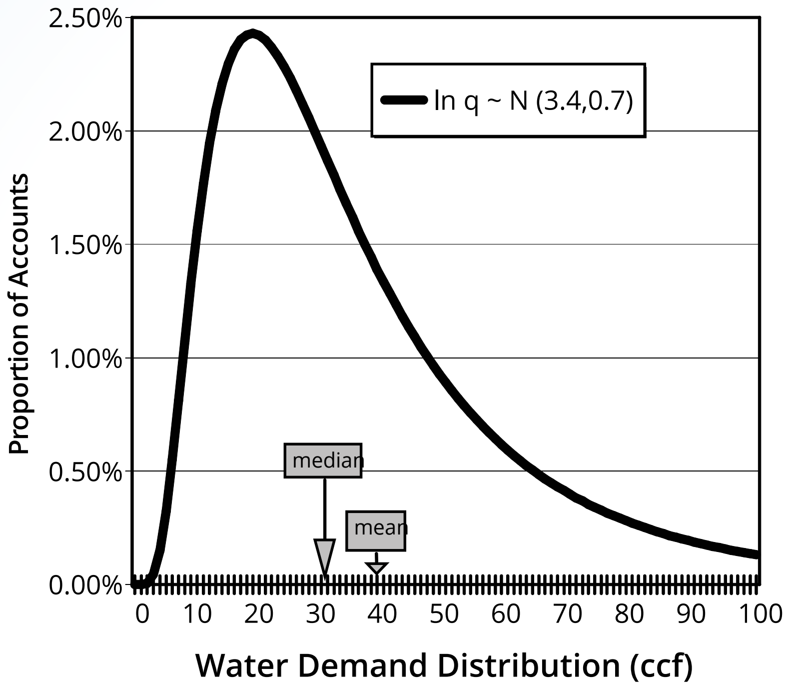

Figure B.2 Distribution of Single Family Water Demand

A right-skewed distribution characterizes water use in most water utilities and complicates the design of block rate structures. Suppose the WAWA rate analyst wants to design an increasing-block rate structure with two blocks. It directly follows from Figure B.2 that if the switch point-where the first block ends and the second begins-were set to median water use (about 31 CCF per bimonthly), then half of the customers would see the lower price in block 1 and half of the customers would face the higher price in block 2. Does this mean half of all water consumption is facing price 1 and the other half faces price 2? No. The mean of the distribution in Figure B.2 (about 40 CCF) would be the switch point where water consumption is split in half.

Customer Accounts versus Water Use in an Upper Block

The WAWA rate analyst might want to know how many customers and how much water would be affected by a second block. Figure B.3 plots the proportion of customer accounts falling into the second block as the block switch point changes. To estimate revenue, the analyst must know how much water is affected by the higher price in the second block. Figure 8.3 also illustrates that the proportion of water use falling into the upper block is greater than the proportion of accounts. This fact is directly implied by the right skewed distribution of water consumption.

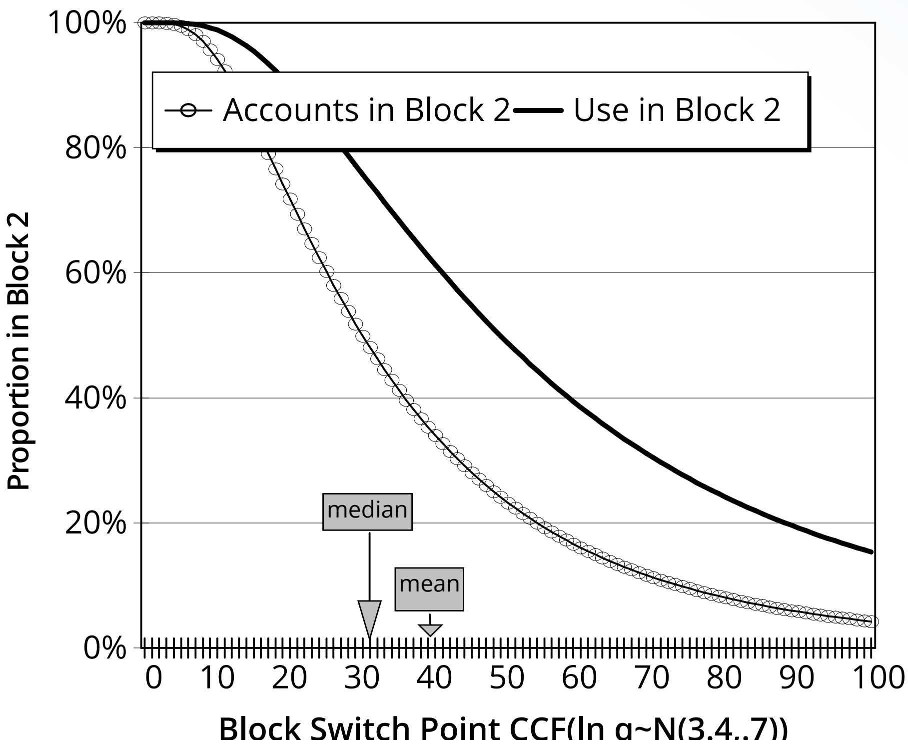

Figure B.3 Customer Accounts versus Water Use in an Upper Block

The next complication in estimating revenue from blocks is accounting for the fact that not all of the water used by customers in the second block is priced at the higher second price. The first k units (where k is the number of units in the first block) are priced at the lower first price. Figure B.4 removes the first k units for each customer in the second block to arrive at the line (with the crosses) that depicts, for any switch point, the proportion of total water use priced at the higher, block 2 price.

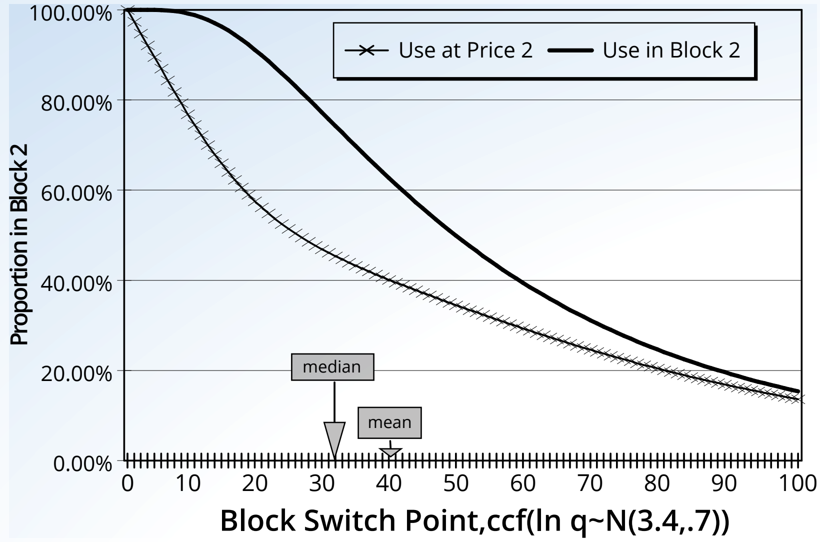

Figure B.4 Proportion of Total Water Use Priced at Price 2

Revenue prediction for block rates requires a deeper understanding of demand distributions. Household demand models can generate the expected (mean) demand in a given time period, E[Qt], and a measure of the dispersion about that mean, the variance V[Qt]. These two parameters are sufficient, for any given rate structure, to determine system revenue. For a uniform rate structure, a model of system revenue requires only expected demand:

$$SystemRevenue = E[\sum_{i=1}^{N} Bill_i = N \cdot E[Bill] = N \cdot (\alpha + P \cdot E[Q])$$

where N is the number of accounts in this customer class and α is a fixed charge.

Alternatively, system revenue can be expressed as the combination of fixed revenue and variable revenue:

$$SystemRevenue = FixedRevenue + VariableRevenue = N \cdot (\alpha + N \cdot P \cdot E[Q])$$

A seasonal rate structure requires the addition of a time index to this equation. Block-rate structures require knowledge of both the mean and the dispersion. The model of system revenue from block rate structures uses the demand models to predict the proportion of accounts (n/N) and the proportion of water use (ρ) that fall within a consumption block. For example, the variable revenue from accounts falling entirely within the first block (from 0 consumption units to q1 consumption units) is:

$$Variable\ Revenue_1 = P_1 \cdot E[WaterUseinBlock1] = P_1 \cdot \rho_1 \cdot E[Q] \cdot N$$

Variable revenue from accounts in the second block will be broken into two parts: (1) revenue from the first block (the quantity of water in the first block times the first block price), and (2) the additional revenue from the second block (the amount of water in the second block times the second block price):

$$\begin{aligned} Variable\ Revenue_2 &= P_1 \cdot E[WaterUseinBlock1] + P_2 \cdot E[AdditionalWaterUseinBlock2] \\ &= P_1 \cdot q_1 \cdot n_2 + P_2 \cdot (E[WaterUseinBlock2] - q_1.n_2) \\ &= P_1 \cdot q_1 \cdot n_2 + P_2 \cdot (\rho_2 \cdot E[TotalWaterUse] - q_1.n_2) \end{aligned}$$

The key to carrying out this kind of calculation is arriving at (1) the proportion of accounts that fall within each block ρ (= n/N), and (2) the proportion of total use falling within a block (ρi).

Simulation Methods — A Monte Carlo Example of Revenue Volatility

The propensity of a rate structure to generate revenues that exactly match the revenue requirements of a water utility is subject to risks involving both supply and demand. These risks can produce revenue instability in the form of both revenue surpluses and revenue shortfalls. These risks are associated with changes in the number of customers, changes in customer mix [e.g., the loss of a large user], changes in usage patterns, changes in weather, changes in conservation ethic, changes in the price elasticity of water demands, and changes in rate structure [Beecher and Mann, 1991].

An important additional source of risk comes from supply or drought-driven curtailments. These sources of risk need to be assessed in the process of determining revenue requirements and mechanisms such as contingency funds and automatic rate adjustments put in place for coping with the unanticipated revenue changes [Chesnutt, et al, 1995b]. One of the other important drivers of short-term revenue uncertainty is climatic uncertainty. This exercise uses the estimated historical effect of climate from the aggregate model of demand.

Analysts are encouraged to aggregate the swings in revenue over multiple months or even multiple years. The estimated risk of revenue surplus (or deficit) will be greater over a multiple year period due to streaks of hot and dry (or cool and wet) weather. The magnitude of the increase in multiple year risk depends directly on the ability of utilities to adjust their rates over time to cope with revenue swings. Chapter 4 of Building Better Rates for an Uncertain World on Financial Policies discusses some of dynamic rate adjustment strategies to cope with revenue risks. In an ideal world, rate analysts would calculate revenue risks for each rate alternative. For an impression of how different rate structures can vary in terms of revenue risk, readers are referred to Chesnutt, et, al. 1996.