NBER WORKING PAPER SERIES

DESIGNING NONLINEAR PRICE SCHEDULES FOR URBAN WATER UTILITIES TO BALANCE REVENUE AND CONSERVATION GOALS

Frank A. Wolak

Working Paper 22503
http://www.nber.org/papers/w22503

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016

I would like to thank Michael Miller, Ivan Korolev, and Onder Polat for outstanding research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Designing Nonlinear Price Schedules for Urban Water Utilities to Balance Revenue and Conservation Goals

Frank A. Wolak

NBER Working Paper No. 22503
August 2016
JEL No. L38,L5,L51,L95

ABSTRACT

This paper formulates and estimates a household-level, billing-cycle water demand model under increasing block prices that accounts for the impact of monthly weather variation, the amount of vegetation on the household's property, and customer-level heterogeneity in demand due to household demographics. The model utilizes US Census data on the distribution of household demographics in the utility's service territory to recover the impact of these factors on water demand. An index of the amount of vegetation on the household's property is obtained from NASA satellite data. The household-level demand models are used to compute the distribution of utility-level water demand and revenues for any possible price schedule. Knowledge of the structure of customer-level demand can be used by the utility to design nonlinear pricing plans that achieve competing revenue or water conservation goals, which is crucial for water utilities to manage increasingly uncertain water availability yet still remain financially viable. Knowledge of how these demands differ across customers based on observable household characteristics can allow the utility to reduce the utility-wide revenue or sales risk it faces for any pricing plan. Knowledge of how the structure of demand varies across customers can be used to design personalized (based on observable household demographic characteristics) increasing block price schedules to further reduce the risk the utility faces on a system-wide basis. For the utilities considered, knowledge of the customer-level demographics that predict demand differences across households reduces the uncertainty in the utility's system-wide revenues from 70 to 96 percent. Further reductions in the uncertainty in the utility's system-wide revenues in the, range of 5 to 15 percent, are possible by re-designing the utility's nonlinear price schedules to minimize the revenue risk it faces given the distribution of household-level demand in its service territory.

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
wolak@zia.stanford.edu

1. Introduction

There is a growing need for urban water utilities to manage periods with limited water supplies, particularly in arid parts of the United States. Because more that 85 percent of the total cost of a typical urban water utility does not vary with the volume of water produced, this has led to an increasing frequency of revenue shortfalls for these entities. According to the California Public Utilities Commission (CPUC), over the past 10 years as high as 50 percent of the largest water utilities it regulates have annual revenue shortfalls as large as 20 percent of their annual revenue requirement. These revenue shortfalls have resulted in a far greater use of ex post revenue adjustment mechanisms that increase water prices after periods with limited water availability to recover these revenue shortfalls. This has led to an increasing temporal mismatch between the retail price consumers are charged and their need to reduce to water consumption.[1]

Despite rapidly growing populations in the western states over the past 30 years, there has been no major water storage or delivery infrastructure investment west of the Continential Divide since the early 1970s. For example, the population of California in 1970, around the time the State Water Project was completed, was roughly half of the current value of 38.8 million. This hiatus in water infrastructure investments is partially responsible for the increasing frequency of shortfalls in water availability to urban water utilities in the West.

This set of circumstances suggests two possible approaches to meet the West's future water demand: (1) manage existing water resources, primarily through pricing, or (2) build and pay for additional water storage and/or transportation infrastructure. Both approaches argue for a significantly enhanced understanding of the customer-level demand for water. This argument is strengthened by the fact that nonlinear pricing is the standard approach used by water utilities to balance the competing goals of managing limited water resources and achieving sufficient revenues to recover their costs. Customers typically face schedules where the price charged for each additional unit, the marginal price, rises with the customer's monthly consumption. The marginal price is fixed for a block or range of values of monthly consumption, but it increases across these blocks with increases in the value of monthly consumption. For this reason these nonlinear price schedules are called increasing block price schedules.

[1] For example, the Water Revenue Adjustment Mechanism (WRAM) set by CPUC to recover past revenue shortfalls has temporarily increased future monthly water bills for the same level of consumption by more than 40%.

The form of the increasing block price schedule set by the utility impacts how much water each customer purchases and the revenues the utility receives from that customer. The form of the nonlinear price schedule also impacts the amount of uncertainty the utility faces in the quantity of water it sells and the revenues it receives from each customer. This uncertainty in customer-level water sales and revenues to the utility is aggregated across customers to create uncertainty in the utility-level water sales and revenues. If a utility can accurately predict the customer-level demand for water for any possible nonlinear price schedule it can then design increasing block price schedules to achieve any conservation or revenue goal while also minimizing utility-level water sales or revenue risk. Increased information about the distribution of customer-level demand directly translates into reduced water sales and revenue risk associated with any rate design goal.

This paper formulates and estimates a household-level demand for water under increasing block prices that accounts for the impact of weather variation within the household's billing cycle and customer-level heterogeneity in demand due to observable demographic characteristics and other unobserved factors that differ across customers. This model can be used to construct an estimate of the distribution of each customer's monthly demand and total amount paid for water for any arbitrary nonlinear price schedule. Combined with data on the distribution of observable customer-level heterogeneity in the utility's service territory, these household-level demand models can be used to compute the distribution of aggregate water demand for any possible price schedule.

This process also yields an estimate of the distribution of total utility-level revenues for any arbitrary nonlinear price schedule or set of nonlinear price schedules, which implies that the modeling results can be used to measure both the household-level and aggregate willingness to pay for a proposed water infrastructure investment. Specifically, it can be used to determine if there exists a nonlinear price schedule consistent with the utility's water pricing goals that recovers sufficient revenues to recover the cost of a given water infrastructure investment. In general, the estimated household-level water demand model can be used by the utility to design nonlinear prices for water to achieve a wide range of systemwide policy goals.

The model assumes that water demand depends on the price schedule faced by the household the characteristics of household (such as household income, the size of the dwelling, size of the property, number of adults living in dwelling), weather conditions (specifically, average daily temperature and rainfall during the customer's billing cycle), and a measure of the amount

of outdoor vegetation on the customer's property. Information on the amount of outdoor vegetation for each customer is obtained from satellite data compiled by the National Atmospheric Information Administration (NASA) on a bi-monthly basis.

The demand model is estimated for two water utilities charging increasing block price schedules using monthly billing cycle data for a sample of customers from each utility combined with data from the from United States (US) Bureau of Census Public Use Microdata Sample (PUMS) of American Community Survey on the distribution of household demographic characteristics in the United States Postal Service (USPS) Zip Code, the property-level vegetation index, and data from the National Oceanic and Atmospheric Administration (NOAA) on daily weather conditions in that Zip Code during the billing cycle.

Because there is some controversy about whether customers understand and are able to respond to nonlinear prices, a non-nested hypothesis test is performed comparing this model of household-level demand with nonlinear pricing to each of four competing models of household-level demand that embody alternative price measures that that household responds to.[2] For both of the utilities considered, the non-nested hypothesis tests find that the model of demand with nonlinear pricing provides a statistically superior description of the observed pattern of the household-level demand relative to each of the four alternative models.

Several summary statistics are compiled using the model to assess the impact of resolving uncertainty about customer-level demand and the distribution of these demands throughout utility's service territory on the systemwide sales and revenue uncertainty faced by the utility. The difference in the system-wide revenue risk between a scenario that assumes the utility only knows the prior distribution of demographic characteristics in the household's zip code and the scenario that assumes the utility knows each customer's demographic characteristics provides a metric for assessing the revenues and sales risk reduction benefits to the utility from collecting demographic information from each of its customers.

Counterfactual nonlinear price schedules are computed that yield no more than the same expected system-wide water sales and at least as much system-wide revenues as the utility's existing price schedules, but also minimize the uncertainty in the utility's annual revenues from

[2] Ito (2014) and Borenstein (2009) argue that consumers respond to the average price they face rather than to the nonlinear price schedule. Two of the alternative models of demand are based the assumption that the household's demand is a function of the average price.

water sales. Price schedules that reduce systemwide water consumption by 25 percent with a 95 percent probability while still obtaining at least a much expected sales, are also computed. These counterfactual price schedules are constructed under the assumption that the utility knows the posterior distribution of demographic characteristics for each household given its zip code and observed vector of billing cycle-level water consumption.

These experiments demonstrate several sources of economic benefits to the utility from having a more detailed knowledge of individual customers. First, knowledge of the customer-level demand can be used by the utility to design increasing block pricing plans that achieve any revenue or sales goals with less revenue or sales risk. Second, knowledge of how these demands different across customers based on observable characteristics of the customers can allow the utility to significantly reduce the utility-wide revenue or sales risk it faces for any pricing plan. Third, knowledge of how the structure of demand varies across customers can be used to design personalized (based on observable household demographic characteristics) increasing block price schedules to further reduce the risk the utility faces on a system-wide basis. Because it is relatively straightforward for the utility to prevent resale of residential water service, utilities can set different increasing block price schedules for each customer based on its observable demographic characteristics. Finally, with detailed knowledge of how demands different across customers based on observable demographic characteristics, the utility can more accurately assess the likely water sales and revenue impacts of changes in the number and types of customers in their service territory.

For the two utilities considered, knowledge of the customer-level demographics that predict demand differences across households reduced the uncertainty in the utility's system-wide revenues by 70 and 96 percent, respectively. Further reductions in the uncertainty in the utility's system-wide revenues in the range of 5 to 20 percent, are possible by re-designing the utility's nonlinear price schedules to minimize the revenue risk it faces given the distribution of household-level demand in its service territory. This household-level demand information is also particularly important for assessing the economic benefits of proposed water infrastructure projects and in designing the price schedules necessary for raising the revenue needed to pay for them with the least amount of water sales or revenue risk to the utility.

The remainder of the paper proceeds as follows. The next section discusses the design of nonlinear pricing plans. Section 3 describes the datasets used to estimate the demand model.

Section 4 presents the econometric model of demand, the specification tests performed, and estimation results. Section 5 describes how the model can be used to estimate the distribution of household-level and systemwide water sales and revenues. Section 6 presents the counterfactual experiments performed using the model results. Section 7 concludes.

2. Rate Design with Nonlinear Pricing

A major rationale for increasing block pricing by water utilities is that this form of nonlinear pricing balances two competing public policy goals. The first is to provide the "essential" amount of water a household needs for drinking, cooking, bathing, and other indoor use at a price that is affordable for virtually all households in the utility's service territory. The second goal is to provide a financial incentive for households using more than the "essential" amount to reduce their demand for water. By this logic, the higher-priced steps in the increasing block price schedule beyond the initial baseline or essential consumption level are designed to discourage less essential water consumption. For example, the second price step might be intended for the demand to fill the household's swimming pool. The third price step might be intended for the demand for watering the household's outdoor trees, bushes, and shrubs. The fourth price step might be intended for the demand for watering the household's lawn.

Another argument in favor of increasing block pricing of water is that it recovers an increasing amount of the utility's revenue from high demand customers, which tend to also be the high income customers. Because higher income consumers generally consume more water, the highest marginal price they pay is typically greater than the highest marginal price low income consumers pay. For this reason, increasing block pricing implies that high income consumers pay a higher average price (total monthly payments divided by total monthly consumption) for their water consumption than low income consumers.

Increasing block pricing can also create revenue adequacy challenges for the water utility if the utility makes the length of the baseline level of demand too large. High demand households might consume along the baseline marginal price step as opposed to consuming at a higher marginal price step. Figure 1(a) illustrates this case with $D_L(p)$, the demand curve for low-demand consumers, and $D_H(p)$, the demand curve for high income consumers. Both curves intersect the increasing block price schedule on the first price block, which raises significantly less revenue for the utility than would be the case if $D_H(p)$ intersected the price schedule on the higher-priced block. If the first block of the price schedule is too short, this can impose an excessive financial burden

on low-demand, low-income consumers by charging them the marginal price intended for high demand consumers. Figure 1(b) illustrates this case where both demand curves intersect the increasing block price schedule on the higher-priced block.

From the perspective of achieving enough revenues to recover the utility's costs, while selling no more than a certain amount of water to all customers, the design of a nonlinear pricing schedule amounts to choosing a length for each step that separates customers into distinct groups based on their willingness to pay for water. Moreover, if the utility has some uncertainty about the location and shape of each customer's demand, then reducing this uncertainty could help the utility determine where to set the baseline demand level, q_B, shown in Figure 1(c). This figure shows the range of possible uncertainty (from the perspective of the utility) in $D_L(p)$ and $D_H(p)$. This is indicated by the dotted lines to the left and right of each demand curve. Note that q_B has been chosen so that regardless of the realization of $D_L(p)$ and $D_H(p)$, each type of customer will continue to consume along the same step of the increasing block price schedule. Choosing the value of q_B in this manner limits the amount of revenue variability that the utility faces due to its uncertainty about the realized values of $D_L(p)$ and $D_H(p)$. The second part of the increasing block pricing design process must choose the levels of the first marginal price and the second higher marginal price to recover sufficient revenues to cover the utility's costs, while still achieving the goal of limiting the economic burden placed on low-income consumers in purchasing their essential water needs.

If the utility is able to sort households into different categories based on observable demographic characteristics, then it is also possible to assign separate increasing block price schedules to different households based on these characteristics. In this case, the utility would like to achieve the outcome in Figure 1(c) for each set of observable demographic characteristics that predict differences in the form of the demand.

One possible set of counterfactual pricing experiments would use the estimated household-level demand model to determine the extent to which it is possible for the utility to re-design its increasing block price schedule to achieve at least as much expected revenue and expected water sales no larger than it does under the current rate schedule while facing less risk to its total revenues. A second set of counterfactual pricing experiments could set separate increasing block prices schedules for households with different observable demographic characteristics to achieve at least as much expected revenue and no larger expected water sales than with the current

increasing block price schedules used by the utility while facing the utility with less risk to its total revenues. Both sets of counterfactual pricing experiments demonstrate that if a utility has more information about the demand for water of individual customers, it can significantly reduce the revenue or sales risk it faces in meeting a set of pricing goals.

3. Data Used in Analysis

Four datasets are used to estimate the customer-level demand model for each utility service territory. The first is billing cycle-level monthly water consumption data for a sample of households for at least one year in duration. The second dataset is composed of daily weather variables at the Zip Code level obtained from the National Oceanic and Atmospheric Administration (NOAA) for the utility's service territory. The third dataset is the distribution of household-level demographic characteristics within each Zip Code in the utility's service territory obtained from the US Bureau of the Census. The fourth dataset is composed of the value of the Normalized Vegetation Difference Index (NDVI) compiled by NASA for each household's property.

Monthly household-level water consumption is available from two utilities at the billing cycle-level, along with the customer's zip code, form of nonlinear price schedule faced by household, and other information necessary to compute customer's monthly water bill. Although utilities typically bill their customers on a monthly basis, customers receive their bills at different times during the month. The time between consecutive billing dates is called the customer's billing cycle and it depends on when the meter reader shows up at the customer's premises to read the meter each month. For example, one customer might be billed on the third day of every month, whereas another customer might be billed on the twentieth day of every month.

Having data available on each customer's billing cycle level is important for accurately modeling the impact of weather conditions on a household's demand for water. In terms of the above example, it might be the case that the first two weeks of July are extremely hot so the water demand is particularly high, whereas the last two weeks of July are mild and so water demand is significantly lower. The customer with a billing cycle that starts on the third day of the month will have much higher weather-related demand than customer whose billing cycle begins on the 20th day of the month. Only by knowing the customer's billing cycle is it possible to properly account for differences across customers in their weather-related demand for water.

The NOAA provides daily measures of rainfall and the maximum daily temperature at the Zip Code level for each utility service territory. The average value of the maximum daily temperature is computed as the average of the daily maximum temperature across all days in the billing cycle. The total amount of rainfall in that Zip Code during the billing cycle is also computed from this data. The inter-quartile range of the maximum daily temperatures and inter-quartile range of daily rainfall in the zip code during the billing cycle are also compiled.[3]

The distributions of household-level demographic variables for each Zip Code in each utility's service territory are obtained from the US Bureau of Census Public Use Microdata Sample (PUMS) of American Community Survey. The demographic characteristics for each household surveyed in each Public Use Microdata Area (PUMA) are compiled along with the sampling weight for that household. These PUMAs can be matched to zip codes so that a distribution of household-level demographic variables in the Zip Code is available for all Zip Codes in the service territory

The NDVI data is compiled by NASA from satellite data taken from the using NOAA's Advanced Very High Resolution Radiometer (AVHRR). An algorithm is applied to the wavelengths and intensity of visible and near-infrared light reflected by the land surface back up into space to quantify the concentrations of green leaf vegetation for 30 meter by 30 meter quadrants of the earth's surface. The NDVI lies on the interval [-1,1], with higher values indicating more green vegetation. Values close to -1 correspond to water, whereas values close to zero (-0.1 to 0.1) correspond to rock, sand, or snow. Small positive values, generally between 0.2 and 0.4, represent shrub and grassland, and values close to 1 indicate temperate and tropical rainforests.

Household-level, billing-cycle data is available from the Valley of the Moon (near Sonoma), California and Cobb County, Georgia water utilities. Daily weather data has also been compiled for the time period that the customer-level billing cycle data is available for each utility service territory. The Zip Code-level distribution of household demographic data has also been compiled for the time period that the customer-level billing cycle data is available for each utility service territory. The NDVI data is available on a bi-monthly basis for each household in both utility service territories. All nominal prices are converted to 2012 dollars using the Federal

[3] The inter-quartile range is the difference between the 75th percentile of the daily variables in the billing cycle and 25th percentile of this same distribution.

Reserve Economic Database Gross Domestic Product (GDP) deflator from St Louis Federal Resource Bank.[4]

4. Econometric Model

This section describes the specification of the econometric model of the billing cycle-level and household-level water demand under increasing block prices that accounts for the weather facing that household during its billing cycle, the differences in demographic characteristics across households and the differences in the NDVI value for the property over time and across households. This model is derived from the assumption that households choose their water consumption to maximize a utility function that depends on their demographic characteristics and NDVI value and unobserved heterogeneity.

This econometric model is used to derive the joint density of all billing cycle-level consumption choices for each household during the sample period conditional on the nonlinear price schedule the household faces, its demographic characteristics, the value of NDVI on its property, and the temperature and rainfall distributions it was exposed to each billing cycle during the sample period. Because the demographic characteristics of each household are unobserved, in order to arrive at the likelihood function used to estimate the parameters of the demand model, the conditional distribution of the household's monthly billing cycle-level consumption choices are integrated with respect to the distribution of demographic characteristics in the Zip Code that contains that household. This yields a likelihood function that depends on observable data—the household's vector of monthly water consumption choices, the values of the household's property vegetation index, the vector of the billing cycle-level monthly weather variables and the distribution of demographic characteristics for that household's Zip Code.

4.1. Water Demand Model

Let $U(x_i,w_i,A_i,Z_i,G_i,\varepsilon_i,\beta)$ equal the utility function for household i over the N-dimensional vector of goods, $x_i = (x_{i1},x_{i2},\ldots,x_{iN})$, where x_{ik} is household i's monthly consumption of good k, and w_i is the household i's monthly consumption of water. The utility function also depends on the household i's demographic characteristics, A_i; the vector of weather variables faced by household i, Z_i; the value of the NDVI index for household i, G_i; a vector of unobserved heterogeneity, ε_i. This utility function is parameterized by the vector β. Let p_k equal the price of the kth element of x_i, x_{ik}. Let $\theta_i(w)$ equal the increasing block price function that the household i

[4] Data available from https://research.stlouisfed.org/fred2/series/GDPDEF

faces for water. The value of this function at consumption level w is equal to, $\theta_i(w)$, the marginal price. Figure 1(a) to 1(c) shows several increasing block price schedules with two price blocks.

If household i purchases w+ units of water during the month then its total bill is equal to $R(\theta_i(w+)) = \int_0^{w+} \theta_i(s)ds$, which is equal to the area under the nonlinear price schedule up to the observed consumption level, w+. A household that consumes w units of water and the vector of other goods, x, has a monthly spending on water and the N other goods equal to $\sum_{k=1}^{N} p_k x_{ik}$ + $R(\theta_i(w))$. Under the assumption of utility-maximizing behavior, the household's observed choices of x and w are assumed to be the solution to the following optimization problem:

$$max_{x\geq 0, w\geq 0} U(x, w|A_i, Z_i, G_i, \varepsilon_i, \beta) \text{ subject to } \sum_{k=1}^{N} p_k x_{ik} + R(\theta_i(w)) = I_i, \quad (1)$$

where I_i is household i's monthly income. Solving problem (1) yields the household's utility-maximizing choices for x and w as a function of the vector of prices, $P = (p_1,p_2,\ldots,p_N)$ of the N other goods; the nonlinear price function, $\theta_i(w)$; and total monthly income, I_i.

Let $w(P,\theta_i,M_i,A_i, Z_i, G_i, \varepsilon_i, \beta)$ equal the solution to this household-level optimization problem. This function depends on the vector of prices of other goods, P; the nonlinear price schedule for water faced by household i, $\theta_i(w)$; household i's total monthly income, I_i; the vector of observed characteristics of household i, A_i; the vector of unobserved characteristics of the household, ε_i; the vector of weather variables faced by household i, Z_i; the household's vegetation index, G_i; and the parameters of the household's preference function, β.

Assuming a parametric joint density for ε, $f(\varepsilon|\delta)$, (where δ if the vector of parameters of this joint density) it is possible to derive the density of the household's vector of billing-cycle level observed water consumption, w, which I write $g(w|P,\theta,A, Z, G_i, \beta, \delta)$. This density is also equal to the conditional (on A_i) likelihood function for a single observation of monthly billing cycle-level consumption for household i.

4.2. Log-Likelihood Function

Let the subscript "t" denote the value of a variable for billing cycle t and T(i) equal the number of monthly consumption observations for household i in the sample and N is the total number of households in the sample. Let $W_i = (w_{i1},w_{i2},\ldots,w_{iT(i)})$' equal the T(i) dimensional vector of monthly water consumption observations for household i. Let $W = (W_1',W_2',\ldots,W_N')$' equal the vector of the N vectors of monthly water consumption observations for all households in the sample. The first step in computing the likelihood function for the econometric model is to compute the joint density of W_i for each household in the sample conditional on the household's

demographic characteristics and the T(i) realizations of monthly weather conditions that they faced. In terms of the above notation, this joint density takes the form:

$$\prod_{t=1}^{T(i)} g(w_{it} \mid P_{it}, \theta_{it}, A_n, Z_{it}, G_{it}, \beta, \delta) \quad (2)$$

The PUMS data from American Community Survey can be used to compute the probability density functions for the vector of demographic characteristics for each Zip Code in the utility's service territory. This dataset provides the sampling weights for each household in the American Consumer Survey and the vector of their demographic characteristics for each 5-digit Zip Code in the utility's service territory. Let (wt(i,n), A_n) for n=1,…,L(i) equal the values of these sampling weights and associated vector of demographic characteristics for each sampled household in the Zip Code that contains household i. In terms of this notation, the log-likelihood function for single observation is equal to:

$$L(W_i \mid \beta, \delta) = \ln[\sum_{n=1}^{L(i)} wt(i,n) \prod_{t=1}^{T(i)} g(w_{it} \mid P_{it}, \theta_{it}, A_n, Z_{it}, G_{it}, \beta, \delta)]. \quad (3)$$

Summing over all N households in the sample yields log-likelihood function for the entire sample:

$$L(W|\beta, \delta) = \sum_{i=1}^{N} \ln[\sum_{n=1}^{L(i)} wt(i,n) \prod_{t=1}^{T(i)} g(w_{it} \mid P_{it}, \theta_{it}, A_n, Z_{it}, G_{it}, \beta, \delta)], \quad (4)$$

Note that the joint distribution of $(w_{i1}, w_{i2}, \ldots, w_{iT(i)})$' is integrated with respect to the density of the vector of demographic characteristics, A_n, rather than the density of each w_{it} individually, in order to account for the persistence in household i's billing cycle level demand over time. If the consumption of household i is unexpectedly high in billing cycle t relative to what would be predicted based on the observable characteristics of this household, then it is likely that its consumption would be unexpectedly high in all other billing cycles. Integrating with respect to the density of A_n as is done in equation (4) is consistent with that logic.

4.3. Functional Forms

In order to implement the model empirically, it is necessary to choose functional forms for the household's utility function, $U(x_i, w_i, A_i, Z_i, G_i, \varepsilon_i, \beta)$, which yields the functional form for the household's demand function, $w(P, \theta_i, M_i, A_i, Z_i, G_i, \varepsilon_i, \beta)$. Because the distributions of monthly water consumption across both across households for the same month and for the same household over time are both positively skewed in the sense that many observations are just below the mean, but a few observations are far above the mean, the appropriate variable to model is the logarithm of the household's monthly demand for water.

This logic implies the following choice for the functional form for $w^*(\theta, M, A, Z, G, \varepsilon, \beta)$, the observable portion household's billing cycle-level monthly demand for water conditional on observing the household's vector of demographic characteristics, A:

$$\ln(w^*(p_w, V(A), A, Z, \beta)) = A'\beta_1 + Z'\beta_2 + G'\beta_3 + \alpha(A,G)\ln(p_w) + \rho(A,G)\ln(V(A,G)), \quad (5)$$

where $\alpha(A,G) = -\exp(A'\beta_4 + G\beta_5)$ and $\rho(A,G) = -\exp(A'\beta_6 + G\beta_7)$. Define $\beta = (\beta_1', \beta_2', \beta_3, \beta_4', \beta_5, \beta_6' \beta_7)'$ as the vector of parameters of the demand function. This functional form implies that the coefficient determining the price responsiveness of demand, α(A,G), is minus 1 times the exponential function of a linear combination of some of the elements of the vector of the household's demographic variables and its vegetation index, and ρ(A,G) is an exponential function of a linear combination of some of the elements of the vector of the household's demographic variables and its vegetation index. V(A) is the household's monthly virtual income and it is written as a function of this vector of demographic characteristics to denote the fact that the household's income is one the elements of the vector of household characteristics that we "integrate" out with respect to in computing the likelihood function for household i in equation (4).

This functional form allows for substantial differences in both the price responsiveness and income responsiveness of water demand across households in each utility service territory. Both the price and income coefficients depend on the value of the vegetation index for the household's property and a subset of the vector of demographic characteristics to allow for differences in both the income and price elasticities across households and over time for the same household.

There are two sources of unobservables for each month and household $\varepsilon = (\eta, \nu)$, where $\eta \sim N(0, \sigma_\eta^2)$ and $\nu \sim N(0, \sigma_\nu^2)$ are independent random variables distributed independently across households and over time for the same household. This implies that $\delta = (\sigma_\eta^2, \sigma_\nu^2)'$ in the notation of the likelihood function (2). Constructing the conditional (on demographic characteristics) likelihood function (2) for household i, requires computing the density of the observed value of $\ln(w_{it})$, using the deterministic portion of the demand function and joint distribution of ε. The elements of $\varepsilon = (\eta, \nu)'$ are called the unobserved household-level heterogeneity, η, and the household-level optimization or technological uncertainty error, ν. The former is assumed to be observed by the household, but the latter is assumed to be unobserved by the household. Both elements of ε are unobserved by the econometrician.

To understand the determination of the household's virtual income, V(A), and the mapping from $\ln(w^*(p_w, V(A), A, Z, G, \beta))$ to the logarithm of observed consumption of the household,

consider a four-tier increasing block price schedule with a fixed charge. This implies $p_1 < p_2 < p_3 < p_4$. As shown in Figure 2(a), $(0,w_1^*)$ is the range of consumption where marginal price is p_1, (w_1^*,w_2^*) is the range of consumption for the marginal price p_2, (w_2^*,w_3^*) is the range of consumption for the marginal price p_3, (w_3^*,∞) is the range of consumption for the marginal price p_4, and FC equals the household's fixed charge for the billing cycle. The increasing block price schedule implies the piece-wise linear budget set composed of the segments, BS_1, BS_2, BS_3, and BS_4 shown in Figure 2(a). For each segment of the increasing block price schedule, the segment of the household's budget set becomes increasingly steep because the slope of each segment is equal to $-p_k/p_o$, where p_k is the marginal price for the kth price block and p_o is the price of all other goods.

Define d_k as the difference between the cost of consumption level w (in the kth block of the increasing block price schedule) when all units are purchased at price p_k, the marginal price for this block, and the actual cost purchasing w^+ under the increasing block price schedule, so that $d_k = p_k w^+ - R(\theta(w^+)) = -FC - \sum_{j=1}^{k-1}(p_j - p_{j+1})w_j^*$. For example, if the household is purchasing along the first price tier, then d_1 = - FC and if the household is purchasing along the second price tier, then d_2 = - FC – $(p_1 - p_2)w_1^*$. A household with a vector of demographic characteristics, A, purchasing along the k^{th} price tier has virtual income of $V_k(A) = I(A) + d_k$, where I(A) is the household's income. Income is written as a function of the vector of demographic characteristics to denote the fact that monthly income is one of the elements of A, the vector of demographic characteristics obtained from PUMS data.

Figure 2(a) shows a point of tangency between the household's indifference curve and the nonlinear budget constraint. If η was the only unobservable in the household's demand function then the following statements would hold. If there is a point of tangency between the household's indifference curve and one of the piecewise linear budget set segments, then there should be one value of η that yields this observed value of water consumption. However, as Figure 2(b) demonstrates, it is also possible that a point of tangency could occur at a kink point of the piecewise linear budget set. In Figure 2(b), the kink point is at the consumption level w_2^*. In this case, there would be a set of values of η such that the household consumes w_2^* because there are a number of possible values of η that shift and rotate the household's indifference curves which yield this point as the household's utility maximizing consumption choice.

It is virtually impossible for a household to manage its water consumption with so much precision as to end up exactly at a kink point on the piecewise linear budget set. Virtually all water services demanded by the household involves uncertainty in the actual amount of water consumed that is observed by the household. A household member demands water services such as taking a bath or shower, filling their swimming pool, washing their car, or watering their plants or lawn. Despite the individual's best intentions to use only a certain amount of water, each water service demand has technological uncertainty in the exact amount of water consumed. For example, running water for a hot shower on a cold day takes longer than on a warm day and therefore uses more water. For this reason, a second stochastic unobservable, ν, the so-called optimization error, is introduced into the demand model to account for uncertainty in the actual amount of water consumed by the household relative to their intended water service consumption level based on only η.[5]

The mapping from the realized values of the unobservables (η, ν) to the observed value of the logarithm of the household's monthly billing cycle-level consumption, ln(w), for a K-step increasing block price schedule takes the form

$$
\begin{aligned}
&\ln(w) = \ln(w^*(p_1, V_1(A), A, Z, G, \beta)) + \eta + \nu \\
&\quad \text{if } \eta < \ln(w_1^*) - \ln(w^*(p_1, V_1(A), A, Z, G, \beta)) \\
&\ln(w) = \ln(w_1^*) + \nu \\
&\quad \text{if } \ln(w_1^*) - \ln(w^*(p_1, V_1(A), A, Z, G, \beta)) < \eta < \ln(w_1^*) - \ln(w^*(p_2, V_2(A), A, Z, G, \beta)) \\
&\ln(w) = \ln(w^*(p_2, V_2(A), A, Z, \beta)) + \eta + \nu \\
&\quad \text{if } \ln(w_1^*) - \ln(w^*(p_2, V_2(A), A, Z, G, \beta)) < \eta < \ln(w_2^*) - \ln(w^*(p_2, V_2(A), A, Z, G, \beta)) \\
&\ln(w) = \ln(w_2^*) + \nu \\
&\quad \text{if } \ln(w_2^*) - \ln(w^*(p_2, V_2(A), A, Z, G, \beta)) < \eta < \ln(w_3^*) - \ln(w^*(p_2, V_2(A), A, Z, G, \beta)) \\
&\qquad \ldots \qquad (6) \\
&\ln(w) = \ln(w_{K-1}^*) + \nu \\
&\quad \text{if } \ln(w_{K-1}^*) - \ln(w^*(p_{K\text{-}1}, V_{K\text{-}1}(A), A, Z, G, \beta)) < \eta < \ln(w_{K-1}^*) - \ln(w^*(p_K, V_K(A), A, Z, G, \beta)) \\
&\ln(w) = \ln(w^*(p_K, V_K(A), A, Z, G, \beta)) + \eta + \nu \\
&\quad \text{if } \ln(w_{K-1}^*) - \ln(w^*(p_K, V_K(A), A, Z, G, \beta)) < \eta
\end{aligned}
$$

where $V_k(A) = I(A) + d_k$ for k=1,2,…,K is equal to:

[5] A similar error structure is employed by Hewitt and Hanemann (1995) and Olmstead, Hanemann, and Stavins (2007) to derive the likelihood function for their demand models.

$$\sum_{k=1}^{K} \frac{\phi(s_k)}{\sqrt{\sigma_\eta^2 + \sigma_\nu^2}} (\Phi(r_k) - \Phi(n_k)) + \sum_{k=1}^{K-1} \frac{\phi(u_k)}{\sigma_\nu^2} (\Phi(m_k) - \Phi(t_k)) \quad (7)$$

where $t_k = [\ln(w_k^*) - \ln(w^*(p_k, V_k(A), A, Z, G, \beta))]/\sigma_\eta$, $r_k = (t_k - \rho s_k)/\sqrt{1-\rho^2}$, $\rho = \frac{\sigma_\eta^2}{\sqrt{(\sigma_\eta^2)(\sigma_\eta^2+\sigma_\nu^2)}}$

$s_k = (\ln(w_{it}) - \ln(w^*(p_k, V_k(A), A, Z, G, \beta))/\sqrt{\sigma_\eta^2 + \sigma_\nu^2}$, $n_k = (m_{k-1} - \rho s_k)/\sqrt{1-\rho^2}$

$m_k = (\ln(w_k^*) - \ln(w^*(p_{k+1}, V_{k+1}(A), A, Z, G, \beta))/\sigma_\eta$, $u_k = (\ln(w_{it}) - \ln(w_k^*))/\sigma_\eta$.

The multiplying this likelihood for billing cycle t for observation i by this same likelihood for all T(i) months for household i yields the likelihood function for observation i given in equation (2).

Maximizing this likelihood with respect to (β',δ')' yields the maximum likelihood estimates of this parameter vector. Two sets of standard errors for the parameter estimates are computed. The first set uses the inverse of the matrix of the sum of the outer products of the gradient of the log-likelihood for each household evaluated at the maximum likelihood parameter estimates. The second set uses the White (1982) quasi-maximum likelihood estimate covariance matrix which is equal to the inverse of the matrix of the second partial derivatives evaluated at the maximum likelihood parameter estimates pre- and post-multiplied by the matrix of the sum of the outer products of gradient of the log-likelihood function.

4.2. Estimation Results

Table 1 contains the estimation results for Valley of the Moon (VoM). Table 2 contains the estimates for Cobb County. The coefficient estimates and the two sets of standard errors described in the previous section are reported for each region. The number of households in the sample is also reported for each region. There are different numbers of months of data for each household because of differences in billing cycles across households during the sample period for each utility.

The following variables make up Z_{it}, the vector of weather characteristics that customer i was exposed to during billing cycle t.[6]

Average high temperature: The average of the daily maximum temperature values in household i's Zip Code during household i's billing cycle.

Inter-quartile range of maximum daily temperatures: The 75th percentile of the daily maximum temperature values in household i's Zip Code during household i's billing cycle minus the 25th

[6] All of the Zip Code-level weather data for each utility was obtained from the www.wunderground.com.

percentile of the daily maximum temperature values in household i's Zip Code during household i's billing cycle

Total precipitation in billing cycle: Sum of daily precipitation in inches during the billing cycle for the Zip Code containing household i.

Interquartile range of daily precipitation: The 75^{th} percentile of the daily precipitation in household i's Zip Code during household i's billing cycle minus the 25^{th} percentile of the daily precipitation in household i's Zip Code during household i's billing cycle

Vegetation—Value of NDVI for household i as of the start of billing cycle t.

Figures 3(a) and 3(b) present the histogram of the NDVI index for households in VoM and Cobb, respectively. Consistent with the hotter and wetter climate in Georgia versus Northern California, the average value of the NDVI in Cobb is higher than in VoM, and the spread of the distribution of the NDVI is significantly larger in Cobb relative to VoM.

The household-level demographics variables, the vector A, all come from the PUMS data set. A subset of the available demographic variables most likely to predict differences in water demand across households are included in A.

Monthly income of household: Monthly household income in 2012 dollars. (Annual number reported in PUMS data divided by 12)

Number of people over 18 years-old living in the household

Number of people under 18 years-old living in the household

House Size Indicators--House acreage between 1 and 10 acres. House acreage above 10 acres.

Number of bedrooms in the house

As discussed earlier, for each household sampled by the US Bureau Census in a given Zip Code, this demographic information is reported along with a sampling weight indicating the number of households in the Zip Code estimated to have the same demographic characteristics vector as the sampled household. Dividing each sampling weight by the sum of the sampling weights for all households sampled in that zip code yields the weight, wt(i,n), used in the construction of the likelihood function.

The price coefficient differs across households in the utility service territory because the coefficient on the logarithm of price depends on A_{in} and $G_{it.}$ Nonlinear pricing of water and the assumed stochastic structure described in the previous subsection that gives rise to the joint density of W_i (the vector of billing cycle-level consumption values for household i) implies that the

coefficient on the logarithm of price for a given household cannot be interpreted as a price elasticity of demand. The same logic applies to the coefficient on logarithm of household-level income. Nevertheless, as shown in the following section, analogues to price and income elasticities can be computed with respect to the expected water demand of the household.

Parameter estimates of the model can be used to compute the posterior probability that household s has the vector of demographics A_{sn} given its vector of billing cycle-level consumption W:

$$p\,(A_{sn}|W) = \frac{wt(s,n)\prod_{t=1}^{T(s)} g(w_{st}\,|P_{st},\theta_{st},A_{sn},Z_{mt},G_{st},\beta,\delta)}{\sum_{n=1}^{N(s)} wt(s,n)\prod_{t=1}^{T(s)} g(w_{st}\,|P_{st},\theta_{st},A_{sn},Z_{mt},G_{st},\beta,\delta)}. \quad (9)$$

For each household in the sample, compute the L(s) values of $p\,(A_{sn}|W)$ for s=1,2,…,L(s). The value of A_{sn} that has the highest posterior probability for that household is assigned that vector of demographics for the purposes of computing the distrbution of systemwide sales and revenues assuming that the utility knows each household's demographic attributes.

4.3. Specification Tests for Non-Nonlinear Pricing Model

This section presents the results of the specification tests of the model household-level demand subject to nonlinear pricing. These tests uses four alternative models of the household-level demand for water where households respond to different price and income measures and compares the optimized value of the log-likelihood function from each of these models to the optimized value of the log-likelihood function from the model of household-level demand subject to nonlinear pricing. From the results of Vuong (1989), the appropriately normalized difference between these optimized log-likelihood functions has an asymptotic N(0,1) distribution under the null hypothesis that both models are equidistant (according to the Kullback-Leiber criteria) from the true unknown data generation process. The direction of rejection of the two-sided test indicates which of the two competing models provides a statistically superior description of the distribution of the observed endogeneous variables given the observed conditioning variables.

Four alternate "price" and "income" demand response models are considered for the same functional form and distribution of unobservables. The functional form for each of the four demand functions is:

$$\ln(w^*(p_w,V(A),A,Z,\beta)) = A'\beta_1 + Z'\beta_2 + G'\beta_3 + \alpha(A,G)\ln(\text{price}) + \rho(A,G)\ln(\text{income}), \quad (10)$$

where $\alpha(A,G) = -\exp(A'\beta_4 + G\beta_5)$ and $\rho(A,G) = -\exp(A'\beta_6 + G\beta_7)$. The four models differ only in terms of what variables are substituted for "price" and "income" in equation (10). Given

the assumed distribution for $\varepsilon = (\eta, \nu)$', each of these models gives rise to a log-likelihood function which is then optimized with respect to (β',δ')'. The four models considered are:

1) Actual price tier—"price" = tier price at actual consumption level and "income" = actual income less the fixed connect charge (This model ignores utility-maximizing choice of the price step.)

2) Average variable price—"price" = (Variable Cost of Bill)/(Actual Consumption) and "income" = actual income less the fixed connect charge

3) Alternative actual price tier—"price" = tier price at their actual consumption and "income" = actual income less the fixed connect charge plus additional income due to nonlinear price schedule (This model also ignores utility-maximizing choice of price step)

4) Total Average Price—"price" = (Total Bill)/(Actual Consumption) and "income" = actual income

Let ln(f(Y|X,θ)) denote the log-likelihood function for an observation from the demand model with non-linear pricing and ln(g(Y|X,γ) the log-likelihood function for one of four competing price response models. Vuong (1989) proposed the following non-nested test between two competing parametric models for the conditional density of Y given X

H: $E(\ln(f(Y|X,\theta^*)) = E(\ln(g(Y|X,\gamma^*)$ versus K: $E(\ln(f(Y|X,\theta^*)) > E(\ln(g(Y|X,\gamma^*)$ (11)

where E(.) is expectation with respect to true joint distribution of Y and X, θ^* and γ^* are plims of ML estimates of θ and γ. The null hypothesis is that the expected value of the log-likelihood functions for both models with respect to h(Y,X), the true joint density of Y and X, are equal versus the alternative that the expected value for one model is greater than the other. Failure to reject the null hypothesis implies that both models are equidistant from the true data generation process, whereas a rejection implies that the model with the log-likelihood $\ln(f(Y|X,\theta^*)$ has a statistically superior average log-likelihood function value.

To implement the hypothesis test, estimate one of the four alternative models, g(Y|X,γ) and compute $W_i = \ln(f(Y_i|X_i,\hat{\theta})) - \ln(g(Y_i|X_i,\hat{\gamma}))$, the difference between the maximized log-likelihood function value for i^{th} observation for each model where $\hat{\theta}$ is the maximum likelihood estimate of θ^* and $\hat{\gamma}$ is the maximum likelihood estimate of γ^*. Vuong (1989) shows that under null hypothesis, $Z = \sqrt{N}\bar{W}/S$ is asymptotically N(0,1) where N = number of customers $\bar{W} = \frac{1}{N}\sum_{i=1}^{N} W_i$ and S = $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(W_i - \bar{W})^2}$.

Table 3 shows the results of these hypothesis tests for VoM and Cobb for each of the four alternate price response models. In all cases, the null hypothesis is overwhelmingly rejected against the alternative that the nonlinear price model has the highest value average log-likelihood. This is consistent with the conclusion that it provides a statistically superior description of the conditional density of Y given X relative for the four alternative models considered.

5. Using Model to Reduce Revenue and Quantity Risk

The estimates of the parameters of the household-level demand model given in Tables 1 and 2 make it possible to compute an estimate of the distribution of a household's water consumption and monthly bill for any nonlinear price schedule either conditional on the household's assigned demographic characteristics or without conditioning on the household's demographic characteristics.

The expected value and variance of these magnitudes can be computed as follows. For a given price schedule that could depend on the household's demographic characteristics, $\theta^C(w,A^*)$, a household with demographics A* has expected consumption and the variance in this consumption equal to:

$$E[w(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)] = \int_{-\infty}^{\infty} w(P,\theta^C,M,A^*,Z,G,s,\beta)f(s,\delta)ds, \quad (10)$$

$$V[w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)] = \int_{-\infty}^{\infty}(w^*(P,\theta^C,M,A^*,Z,G,s,\beta) - E[w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)])^2 f(s,\delta)ds \quad (11)$$

where β and δ in the above expression are evaluated at the maximum likelihood estimates given in Tables 1 and 2. The expectations in the above expression are to be taken with respect to the distribution of ε given A* assigned by the rule based on equation (9). A household with assigned demographic characteristics A^* has an expected monthly water bill and variance of its monthly water bill equal to:

$$E[R(\theta^C(w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta),A^*)] = \int_{-\infty}^{\infty} R(\theta^C(w^*(P,\theta^C,M,A^*,G,s,\beta)\ A^*)f(s,\delta)ds, \quad (12)$$

$$V[R(\theta^p(w^*(P,\theta^C,M,A^*,G,\varepsilon,\beta),A^*] = \int_{-\infty}^{\infty} R(\theta^C(w^*(P,\theta^C,M,A^*,Z,G,s,\beta),A^*) - E[R(\theta^C(w(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta),A^*)])^2 f(s,\delta)ds. \quad (13)$$

For the case that the household i's demographics are assumed to be unknown, the household's expected monthly water consumption and bill and the variance in its monthly water

consumption and bill for the demographic characteristics-dependent increasing block price schedule, θ^C(w,A*), are equal to:

$$E[w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)] = \sum_{n=1}^{L(i)} wt(i,n) \int_{-\infty}^{\infty} w^*(P,\theta^C,M,A_{in},Z,G,s,\beta) f(s,\delta) ds, \quad (14)$$

$$V[w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)] = \sum_{n=1}^{L(i)} \int_{-\infty}^{\infty} (w^*(P,\theta^C,M,A^*,Z,G,s,\beta) - E[w^*(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)])^2 f(s,\delta) ds, \quad (15)$$

and

$$E[R(\theta^C(w(P,\theta^C,M,A,G,\varepsilon,\beta),A)] = \sum_{n=1}^{L(i)} \int_{-\infty}^{\infty} wt(i,n) R(\theta^C(w^*(P,\theta^C,M,A_{in},G,Z,s,\beta),A_{in}) f(s,\delta) ds, \quad (16)$$

$$V[R(\theta^p(w^*(P,\theta^C,M,A,G,\varepsilon,\beta),A)] = \sum_{n=1}^{L(i)} \int_{-\infty}^{\infty} wt(i,n)(R(\theta^C(w^*(P,\theta^C,M,A_{in},Z,G,s,\beta)) - E[R(\theta^C(w(P,\theta^C,M,A,Z,G,\varepsilon,\beta),A])^2 f(s,\delta) ds. \quad (17)$$

The expectations in (14) to (17) are taken with respect to the distribution of $\varepsilon = (\eta, \nu)$' and the distribution of the demographic characteristics within the household's Zip Code. The expectations in (10) to (13) are taken with respect to the distribution of $\varepsilon = (\eta, \nu)$' for the value of the household's demographic characteristics assigned using the approach described above. Consequently, comparing the variance of water consumption and total revenues, given the assigned value of A and the variance with respect to the distributions of ε and A, provides a measure of the value of demographic information to utility. It is also possible to substitute the posterior probabilities computed from equation (9) into equations (14) to (17) and compute the expected values and variances of sales and revenues based on these distributions of the demographic characteristics for household i.

These expressions in (10) and (14) can also be used to compute analogues to the price elasticity and income elasticity of the demand for water. For the case of the price elasticity this is computed as

$$\{E[w(P,\theta,M,A^*,Z,G,\varepsilon,\beta)] - E[w(P,\theta+,M,A^*,Z,G,\varepsilon,\beta)]\}/\{0.05*E[w(P,\theta^C,M,A^*,Z,G,\varepsilon,\beta)]\} \quad (18)$$

where θ is the actual nonlinear price schedule charged by the utility and θ+ is the actual nonlinear price schedule with each price step multiplied by 1.05. This "price elasticity" is the percent change in household i's expected water consumption as a result of a 5 percent increase in all prices on the nonlinear price function divided by 0.05. Computing an "income elasticity" as the percent change in expected consumption from a 5 percent increase in household i's income divided by 5 percent yields the coefficient on logarithm of income. Consequently, the model of demand with nonlinear

pricing and demographic characteristics in the price coefficient implies a different "price elasticity" for each household, but the same income elasticity for each household. However, because demographic characteristics and the vegetation index are included in the income coefficient, the income "elasticity" also differs across households.

The "price elasticities" can be computed conditional on the vector of the household's demographic characteristics or unconditional on the household's vector of demographic characteristics. The only differences in the two "price elasticities" is whether the expectations in (18) are taken with respect to the distribution of A or assume a fixed value of A. Figures 4(a) and 4(b) compute the joint distribution of income and price elasticities for VoM and Cobb, respectively, using the posterior distribution of A, given in equation (9) for each observation in the sample. There is considerable heterogeneity in these elasticity estimates for both utilities. However, the majority of the probability mass is concentrated on price and income elasticities that are less than one in absolute value.

It is also possible to compute the distribution of water consumption for all households in the utility's service territory and the analogous aggregate demand elasticity estimates. Suppose there are J types of households, where households of type j have a vector of observed attributes, A_j, and H_j is the number of type j customers in the utility's service territory. This implies that the expected sales of water by the utility (summed across all customers) associated with rate schedule $\theta^C(w,A)$ is:

$$\text{Expected System-wide Water Sales} = \sum_{j=1}^{J} \mathrm{E}[w^*(\mathrm{P}, \theta^C, \mathrm{M}, A_j, \mathrm{G}, \varepsilon, \beta] H_j \quad (19)$$

$$\text{Variance in System-wide Water Sales} = \sum_{j=1}^{J} \mathrm{Var}[w^*\left(\mathrm{P}, \theta^C, \mathrm{M}, A_j, \mathrm{G}, \varepsilon, \beta\right)] H_j. \quad (20)$$

Following the same procedure for system-wide revenues yields:

$$\text{Expected System-wide Revenues} = \sum_{j=1}^{J} \mathrm{E}[\mathrm{R}(\theta^C(w^*\left(\mathrm{P}, \theta^C, \mathrm{M}, A_j, \mathrm{G}, \varepsilon, \beta\right), \mathrm{A}] C_j \quad (21)$$

$$\text{Variance in System-wide Revenues} = \sum_{j=1}^{J} \mathrm{Var}[\mathrm{R}(\theta^C(w^*\left(\mathrm{P}, \theta^C, \mathrm{M}, A_j, \mathrm{G}, \varepsilon, \beta\right), \mathrm{A}] C_j. \quad (22)$$

Given the estimated distribution of $\varepsilon = (\eta, \nu)$' and the distribution of demographic attributes in each Zip Code within the utility's service territory, other functions of the distribution of system-wide sales and revenues can be computed. The water utility or its regulatory body might be interested in the probability that system-wide sales or revenues exceed or fall below a pre-specified value for a prospective rate schedule. The model estimates can be used to compute that probability.

The aggregate or system-wide "price elasticity of demand" can be computed by finding the percentage increase in expected system-wide demand as a result of a 5 percent increase in all price steps faced by all customers divided by 5 percent. The aggregate "income" elasticity is the percentage increase in expected system-wide demand as a result of a 5 percent increase in all customer incomes divided by 5 percent.

6. Counterfactual Price Schedules

This section first quantifies the revenue risk reduction that is possible for the utility simply from gathering information on the demographic characteristics of its customers. It then reports on the computation of counterfactual price schedules to achieve two policy goals to demonstrate potential used of the model of demand.

For VoM, I compute a counterfactual price schedule that is consistent with California's current water demand reduction goals and unlikely to run afoul of Proposition 218, which requires that municipal water customers only pay for the cost of the water that they consume. Specifically, I compute a price schedule which yields 25 percent less system-wide water sales than the existing price schedule with 95 percent probability and achieves the same system-wide expected revenue goals as the existing price schedule, while minimizing a measure of the financial burden of achieving these water consumption reduction goals across all classes of customers. The price schedule chosen minimizes the weighted sum of the squares of the difference between expected payments by each household under the counterfactual schedule and the current price schedule weighted by the inverse of that household's expected payments under the current price schedule. This objective function places the greatest burden to achieve water consumption reductions on households that currently have the largest water bills. In solving this problem, the monthly fixed charge, FC, was reduced to achieve the goal of maintaining system-wide expected revenues equal to those under the current price schedule.

For Cobb, I first compute a price schedule which yields the same or superior sales and revenue outcomes for the utility but minimizes aggregate revenue risk. This price schedule minimizes the standard deviation of utility-wide water revenues subject to the constraints that the utility expects to sell no more water than it does under the current price schedule, and raises at least as much total revenue for the utility as the existing rate. I then solve the same optimization problem subject to the same constraints, but now allowing the utility to set two price schedules that depend on value of the household's NDVI. Specifically, the utility is allowed to set a schedule

for households with an NDVI value less than 0.35 and one for households for an NDVI value greater than 0.35. For each of these counterfactual price schedules I impose two additional constraints. First, the lowest marginal price in the counterfactual price schedule cannot be higher than the lowest marginal price in the actual price schedule. Second, the highest marginal price in the counterfactual price schedule cannot be higher than the highest marginal price in the actual price schedule. Each counterfactual price schedule is allowed to have as many marginal price steps as the actual price schedule subject to these two constraints. For Cobb, the counterfactual price schedules did not change, FC, the monthly fixed charge that the household faces.

Three main conclusions emerge from this counterfactual price schedule design exercise:

1) The model of the household-level demand for a water utility can be used to reduce the system-wide revenue or sales risk associated with achieving any water pricing goal.

2) By compiling information on the demographic characteristics of their customers and building this information into the utility's model of household-level water demand, utilities can significantly reduce (up to 96% for two utilities considered) both the water sales and revenue risk associated with any expected water sales and revenue goals.

3) The customer-level model of demand incorporating demographic characteristics can be used to design a menu of price schedules that can be offered to households (that allows them to select which specific price schedule they would like to be on based on their NDVI index) to achieve a given water pricing goal for the utility.

The price schedule optimization framework can readily incorporate constraints such as the majority of customers having the same or lower monthly water bills under the optimal price schedules compared to the current schedules.

For both Cobb and VoM, the distribution of sales and revenues for each household is computed under three different assumptions about the distribution of demographic characteristics (in each zip code) in order to quantify the impact of compiling information on the demographic characteristics of each household in the utility's service territory. The first case assumes the demographic characteristics of each household are unknown, but drawn from the prior distribution for that zip code obtained from the PUMS data. The second case assumes the household's demographic characteristics are unknown, but drawn from the posterior distribution for that zip code obtained from the model estimated and equation (9). The final set assumes the household's demographic characteristics are known and set equal to the value of A_n with the highest posterior probability from equation (9).

Table 6(a) and 6(b) report the expected revenue and consumption per month per customer, the standard deviation of systemwide revenues, and the standard deviation of systemwide consumption per month for each of the three distributions of demographic characteristic for each household. Two conclusions emerge from this table. First, for both VoM and Cobb, the standard deviation in systemwide revenues for the case of known demographics is between 0.30 = (1711/5555) and 0.04 = (2143/59091) of the value for case that the demographics are drawn from the prior distribution of A_n. Second, using the posterior distribution instead of the prior distribution of A_n yields an estimate of systemwide standard deviation of revenues that is slightly larger than the value for the case where the vector of demographic characteristics is assumed to be known. Taken together, these results emphasize that even imperfect knowledge of the value of A_n, which is obtained from estimating the demand model and computing the posterior distribution of A_n, can significantly reduce the sales and revenue risk faced by each utility. For this reason, all of our counterfactual computations are based on this posterior distribution for each utility.

6.1. VoM—Demographics Drawn from Posterior Distribution

This section considers a set of counterfactual price schedule choices that reflect policy goals and constraints relevant to California during the summer of 2015, the fourth consecutive summer of low water availability in the state. As a consequence, in the spring of 2015, Governor Jerry Brown issued an executive order requesting a 25 percent reduction in state-wide urban water consumption relative to 2013. A pre-existing legal constraint has further complicated the ability of municipal utilities to achieve this goal.

Proposition 218 (The Right to Vote on Taxes Initiative) requires that municipal utility consumers only pay what it costs to provide them with the water that they consume. AB 2882 (Allocation-based conservation water pricing), signed into law in 2008, attempts to clarify how nonlinear pricing of water can be implemented to avoid running afoul of Proposition 218. However, a recent lawsuit filed by customers of the municipal utility in San Juan Capistrano and the resulting decision which struck down the utility's increasing block rate structure has led to considerable uncertainty over the use of nonlinear pricing of water in California (Stephens, 2015).

One possible solution to this problem is to determine a system-wide average cost of delivering a thousand gallons of water for the utility and then setting a nonlinear price schedule so that the revenues recovered from each type of household (as determined by their demographic characteristics) equal this average cost times the amount of water they consume. Because this

average cost information is not available for VoM, an aggregate revenue constraint is imposed so that households in the utility service territory do not pay more under the new schedule than they were under the existing price schedule. (The constraint implicitly assumes that the utility was only recovering the cost of the water supplied under the existing schedule.) The other constraint on the counterfactual price schedule is that it reduces system-wide water consumption by 25 percent relative to expected consumption under the existing schedule with at least a 95 percent probability.

The objective function assumed for the optimal tariff design problem is to minimize the weighted sum of squared differences between each household's expected monthly bill under the current price schedule and the household's expected monthly bill under the counterfactual price schedule (where the weight applied to each household-level squared difference is the inverse of that household's expected monthly bill under the current price schedule). This objective function is designed to obtain the largest revenue increases from households with the largest current water bills and the smallest revenue increases from households with the smallest current water bills. Finding this price schedule requires solving the following optimization problem:

$$\min_{\theta(w)} \sum_{h=1}^{H} [\ E(R_h(\theta(w)) - E(R_h(\theta_e(w))\ ]^2 / E(R_h(\theta_e(w)) \tag{23}$$

$$\text{subject to Prob}(\ \sum_{h=1}^{H} q_h(\theta(w)) < Q(water))\ \geq 0.95$$

$$\text{E}(\ \sum_{h=1}^{H} R_h(\theta(w)) - R_h(\theta_e(w))) \leq 0$$

where $\theta(w)$ is the price schedule being solved for, $\theta_e(w)$ is the existing price schedule, $R_h(\theta(w))$ is the revenue received from household h under the price schedule $\theta(w)$, $q_h(\theta(w))$ is the quantity demanded by household h under the price schedule $\theta(w)$, and E(.) is the expectation operator.

While there are many other possible objective functions one can optimize to obtain Governor Brown's desired 25 percent reduction in system-wide water consumption with a high probability, this one has the desirable property of putting less of the burden on households that are presently spending less on water.

Figure 8(a) plots the actual price schedule and price schedule that solves (23). This optimization problem also required reducing FC by 13 percent in order to achieve the expected revenue constraint. Figure 8(b) plots the simulated distribution of household-level consumption under the current price schedule set by VoM and simulated distribution of household-level consumption under the counterfactual price schedule.

6.2. Cobb—Demographics Drawn from Posterior Distribution

Figure 6(a) plots actual price schedule faced by customers in Cobb County. This figure also plots two optimal counterfactual NDVI-based price schedules. These price schedules are computed by minimizing the standard deviation of systemwide revenues subject to the constraints that expected revenues are at least as large as under the current price schedule and expected water sales are no larger than under the current price schedule. Additionally, one price schedule will be assigned to customers with a NDVI value of less than 0.35 and the other is assigned to customers with a value less than 0.35. Recall that the constraint that the lowest step of each of these schedules is constrained from below by the lowest price on the actual price schedule and the highest price step is constrained from above by the highest price on the actual price schedule. This form of NDVI-based pricing reduces the standard deviation of systemwide revenues relative to the actual price schedule by 2 percent.

Figure 6(b) plots the two optimal NDVI-based price schedules and the single price schedule that minimizes the standard deviation of systemwide revenues subject to achieving at least as much expected revenues and no larger expected water sales than the actual price schedule. This single price schedule achieves slightly more than a 1 percent reduction in the standard deviation of systemwide revenues relative to the actual price schedule.

A number of other counterfactual price schedules can be computed that depend on demographic characteristics, elements of Z, and combinations of these variables. Reductions in the standard deviation of systemwide revenues of over 10 percent are possible with greater differentiation of price schedules based on household characteristics.

6.3. Assessing the Relative Size of Optimization Error

This section provides an ex post check on the reasonableness of the assumption of the existence of optimization error or technological uncertainty associated with water use. As shown in Section 5, the demand model can be used to compute the distribution of annual water consumption for each household in the VoM sample. There are three sources of randomness in annual water consumption (AWC). First, is uncertainty in the vector of demographic characteristics, A. The second is uncertainty in unobserved heterogeneity, η. The third is the optimization error or technological uncertainty, ν. Let ν(annual) equal the vector of values of ν for all billing cycles in the last year of the sample, and η(annual) equal the vector of values of η for all billing cycles in the last year of the sample. For each household in the sample, compute the

variance in AWC integrating with respect to the distributions of ν(annual), η(annual) and prior distribution of A. Conditional on the value of ν(annual) compute the variance of AWC for each household. The ratio of Var(AWC|ν(annual)), the conditional variance of annual consumption given ν(annual), to, Var(AWC), the unconditional variance of annual consumption, quantifies the impact of optimization error or technological uncertainty on the variance of annual consumption.

Figure 7 plots the histogram of values of Var(AWC|ν(annual))/Var(AWC) for each household in the VoM sample. The average value of this ratio is 0.91, which implies less than 10 percent of the variance in annual consumption can be attributed to optimization error. This histogram ranges from 0.85 to 0.95. Values in this range are plausibly consistent with ν representing optimization error or technological uncertainty in the household's water demand.

7. Conclusions

The model of demand can be used to simulate the distribution of the customer-level billing cycle level household demand for water for any increasing block price schedule. This model can then be used to simulate the distribution of the system-wide demand for water for any nonlinear price schedule. The model can then be used to set price schedules that achieve a wide range of water supply risk or revenue risk management goals in the utility's rate design process.

An important implication of this modeling and simulation exercise is to demonstrate the tremendous reduction in revenue risk the utility faces if it is has the information data on the demographic characteristics of its households. For the case of Cobb, the measure of the variance of system-wide revenue conditional on the assumed knowledge of the vector of demographic characteristics was roughly 4% of the measure of the variance in system-wide revenues, assuming only the distribution of demographic in each Zip Code in the utility's service area was known.

The model was used to show that further revenue variance reductions could be achieved by demographics-based price schedules. The household-level water demand model was used to solve for the optimal (minimum system-wide revenues) demographic-based price schedules. Again, significant variance reductions were possible without be used to compute the distribution expected demand and variance in demand conditional on demographics. The model can even be used to assist the utility in managing water shortfall and potential revenue shortfalls.

The results presented here demonstrate that there is significant value to be had for the utility from understanding the distribution of household level demand in designing price schedules to achieve competing policy goals. In particular, by compiling demographic data on customers and

using such data in customer-level models of demand, utilities can significantly reduce the variance in both the system-wide revenues and the amount of water sold in achieving any price schedule design process. This results implies up to a roughly 96% reduction in the revenue risk that the utility faces if demographic characteristics of its customers is known, suggesting significant economic benefits to water utilities from collecting demographic data on its customers and formulating household-level demand models for price schedule design.

References

Borenstein, Severin (2009) "To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing," working paper available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.4825&rep=rep1&type=pdf

Hewitt, Julie A. and Hanemann, W. Michael, (1995) "A Discrete/Continuous Choice Approach to Residential Water Demand under Block Rate Pricing," *Land Economics*, 75, 173-194.

Olmstead, Sheila M., Hanemann, W. Michael, and Stavins, Robert N. (2007) "Water Demand under Alternative Price Structures," *Journal of Environmental Economics and Management*, 54, 181-198.

Ito, Koichiro (2014) "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," American Economic Review, 104(2): 537–563

Sato, Kazuo (2015) "Additive Utility Functions with Double-Log Consumer Demand Functions," Journal of Political Economy, 80(1), 102-124.

Stephens, Matt (2015) "In blow to water conservation, court rejects San Juan Capistrano's tiered rates," Los Angeles Times, April 20, 2105, at http://www.latimes.com/local/lanow/la-me-ln-water-rates-case-20150405-story.html.

Vuong, Quang (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," Econometrica, 57(2), 307.333.

White, Halbert (1982) "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50(1), 1-25.

Table 1: Model Parameter Estimates and Standard Errors—Valley of The Moon, California

Parameter Name	Estimate	Standard Error (Outer Product of Gradients)	Standard Error (White (1982) Formula)
Constant in the income elasticity formula	-0.31125	0.01867	0.09857
Constant in the price elasticity formula	-0.83170	0.10444	0.31342
Std.dev. of household heterogeneity, η	0.27898	0.01190	0.08486
Std.dev. of optimization error, ν	0.26800	0.01179	0.08305
Constant	-6.33831	0.38147	0.55999
Average high temp in billing cycle	0.00782	0.00169	0.00111
75th - 25th percentile of temperature in billing cycle	-0.01843	0.00203	0.00239
Total precipitation in billing cycle	0.00670	0.00459	0.00236
75th - 25th percentile of precipitation in billing cycle	-0.87861	0.37849	0.13993
Number of people over 18 in house	0.11379	0.02726	0.01570
Number of people under 18 in house	0.71123	0.02683	0.08927
House acreage above 1 acre	0.00088	0.03371	0.00106
Number of bedrooms in house	0.48981	0.07595	0.04090
Price*temp	-0.03615	0.00169	0.00862
Price*precip	0.00792	0.00222	0.00108
Price* # of adults	0.04054	0.03586	0.01603
Price* # of children	0.42252	0.02051	0.02822
Price* # of bedrooms	0.25201	0.01033	0.02526
Income* # of bedrooms	-0.02861	0.01493	0.00560
Vegetation Index	-0.15522	0.40903	0.00989
Income*Vegetation Index	0.02361	0.01166	0.00903
Price*Vegetation Index	0.05847	0.02878	0.00668
Number of customers	**2001**		

Table 2: Model Parameter Estimates and Standard Errors—Cobb County, Georgia

Parameter Name	Estimate	Standard Error (Outer Product of Gradients)	Standard Error (White (1982) Formula)
Constant Term in income elasticity formula	-1.76995	0.06094	0.10967
Constant Term in price elasticity formula	-0.21491	0.09163	0.17331
Std.dev. of household heterogeneity, η	0.03006	0.02165	0.01576
Std.dev. of optimization error, ν	0.35782	0.00149	0.00559
Constant	0.40180	0.08892	0.13906
Average high temp in billing cycle	-0.00574	0.00113	0.00123
75th - 25th percentile of temperature in billing cycle	0.00318	0.00156	0.00137
Total precipitation in billing cycle	0.00448	0.00644	0.01009
75th - 25th percentile of precipitation in billing cycle	-0.40481	0.16279	0.16473
Number of people over 18 in house	0.54096	0.02075	0.05288
Number of people under 18 in house	-0.33908	0.01233	0.04119
House acreage above 1 acre	-0.11647	0.03392	0.04544
Number of bedrooms in house	0.28470	0.03759	0.08463
Price*temp	-0.00569	0.00039	0.00081
Price*precip	-0.00105	0.00140	0.00433
Price* # of adults	0.08484	0.00425	0.00732
Price* # of children	0.01378	0.00292	0.01437
Price* # of bedrooms	0.50440	0.02261	0.02493
Income* # of bedrooms	0.57176	0.02174	0.02191
Vegetation Index	-0.02372	0.06132	0.08541
Income*Vegetation Index	0.51465	0.07720	0.19072
Price*Vegetation Index	0.41746	0.07557	0.21559
Number of customers	**1034**		

Table 3: Non-Nested Test of Nonlinear Pricing Model Versus Alternative Price Model

Utility	**Alternative Model**			
	Actual Price Tier	Average Variable Price	Alternative Actual Price Tier	Total Average Price
Valley of the Moon	17.83	16.56	16.78	11.66
Cobb County	15.24	14.90	15.79	18.11

Note: Test statistic is asymptotically distributed as a N(0,1) random variable under null hypothesis. All tests overwhelmingly reject null hypothesis in favor of nonlinear pricing model.

Table 4(a): Revenue Uncertainty and Demographic Information--VoM				
	E[Revenue per month per bill]	sd[System-wide revenue]	E[Consumption per month per bill]	sd[System-wide consumption]
Demographics drawn from prior	$32.78	$5555	7.00 TGAL	1422 TGAL
Demographics drawn from posterior	$34.93	$2025	7.67 TGAL	485 TGAL
Demographics predicted	$35.29	$1711	7.76 TGAL	395 TGAL

Table 4(b): Revenue Uncertainty and Demographic Information—Cobb				
	E[Revenue per bill]	sd[System-wide revenue]	E[Consumption per bill]	sd[System-wide consumption]
Demographics drawn from prior	$73.86	$59091	6.82 TGAL	4560 TGAL
Demographics drawn from posterior	$55.01	$3434	5.34 TGAL	344 TGAL
Demographics predicted	$55.20	$2143	5.36 TGAL	213 TGAL

Note: TGAL = Thousands of gallons

Figure 1(a)

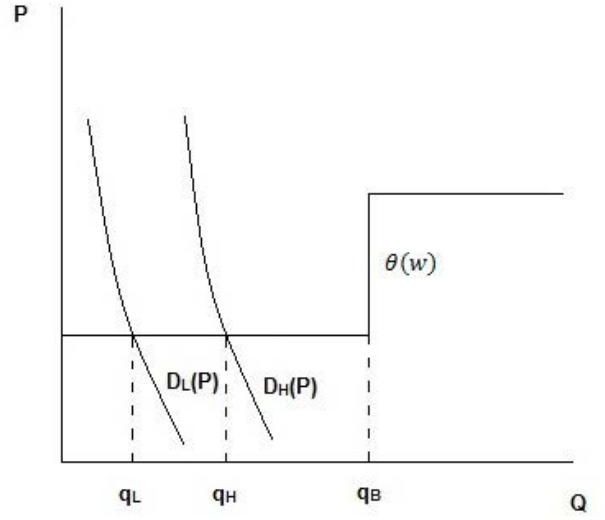

Figure 1(b)

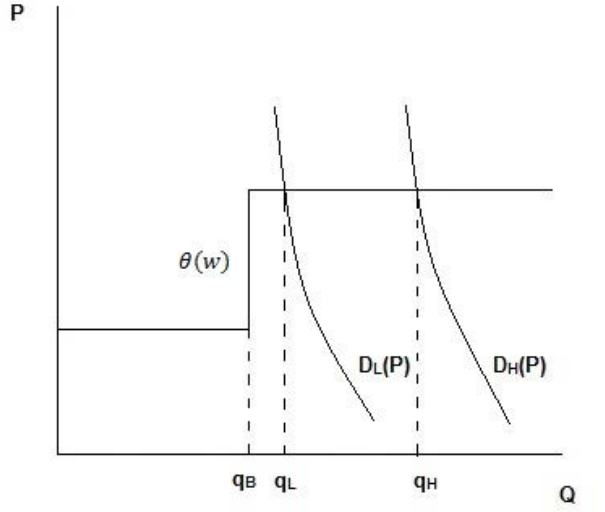

Figure 1(c)

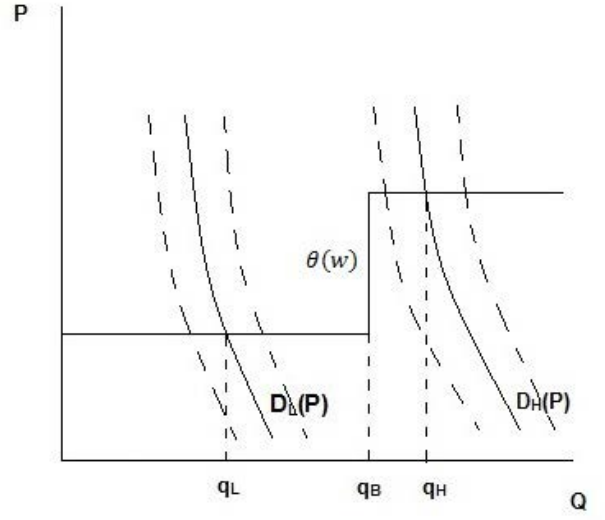

Figure 2(a)

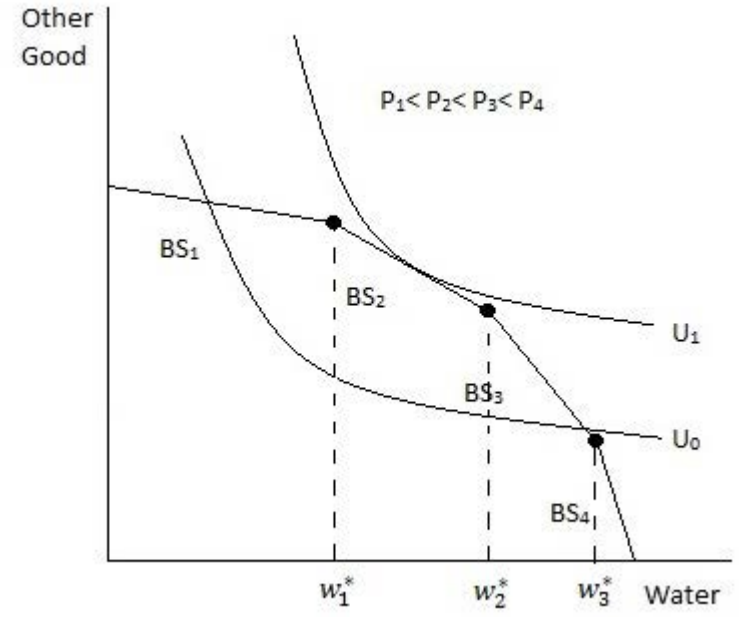

Figure 2(b)

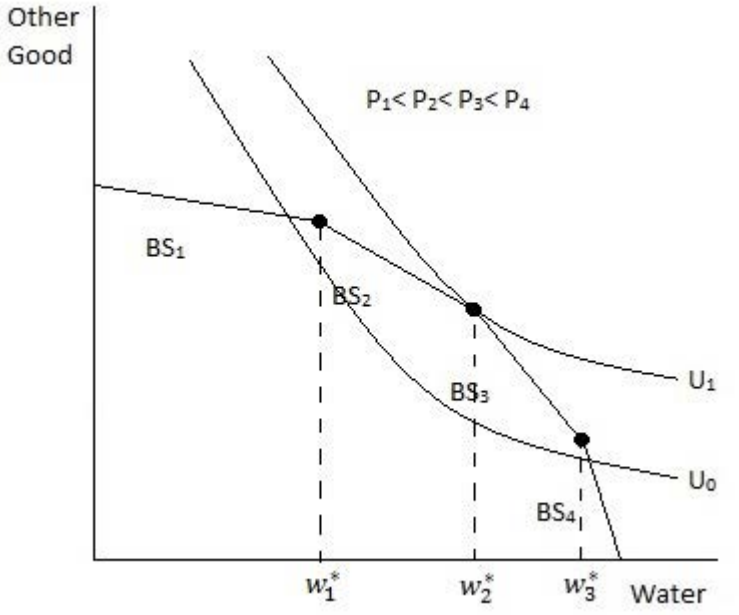

Figure 3(a): Histogram of NDVI for Valley of the Moon

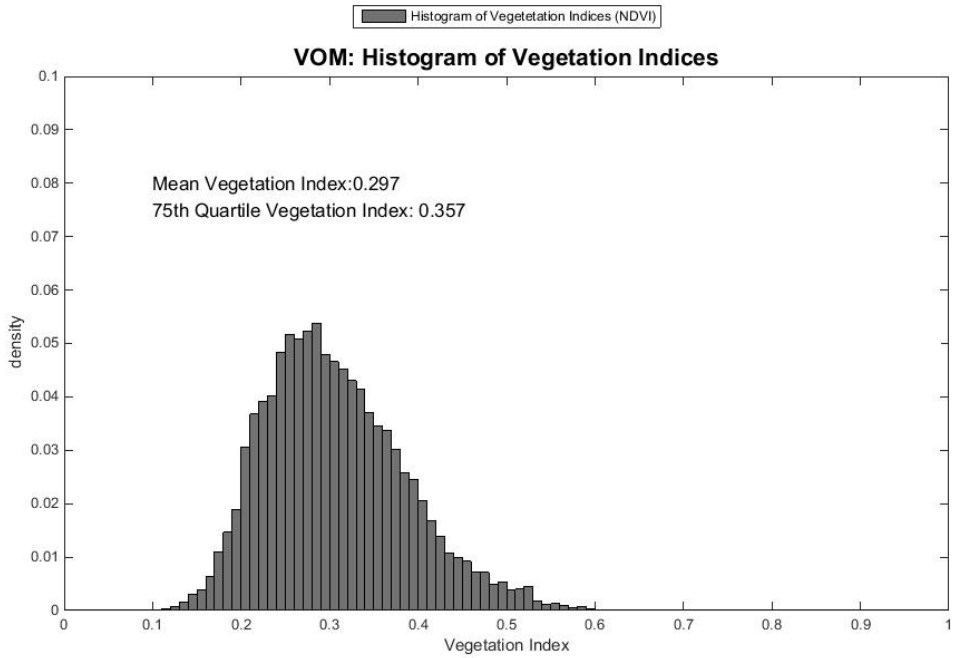

Figure 3(b): Histogram of NDVI for Cobb County, Georgia

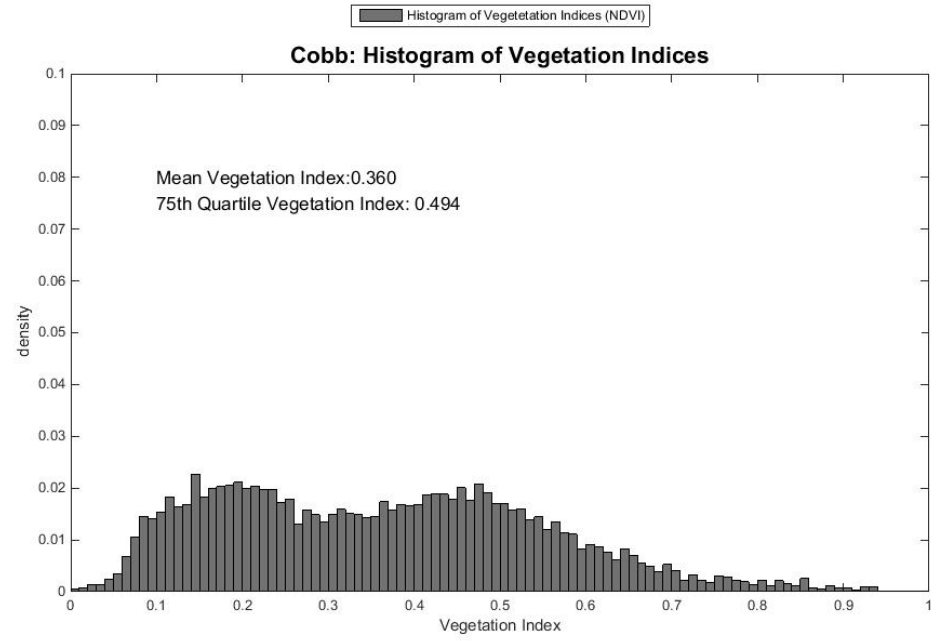

Figure 4(a): Histogram of Price and Income Elasticities for VoM
(Demographics Drawn from Posterior Distribution)

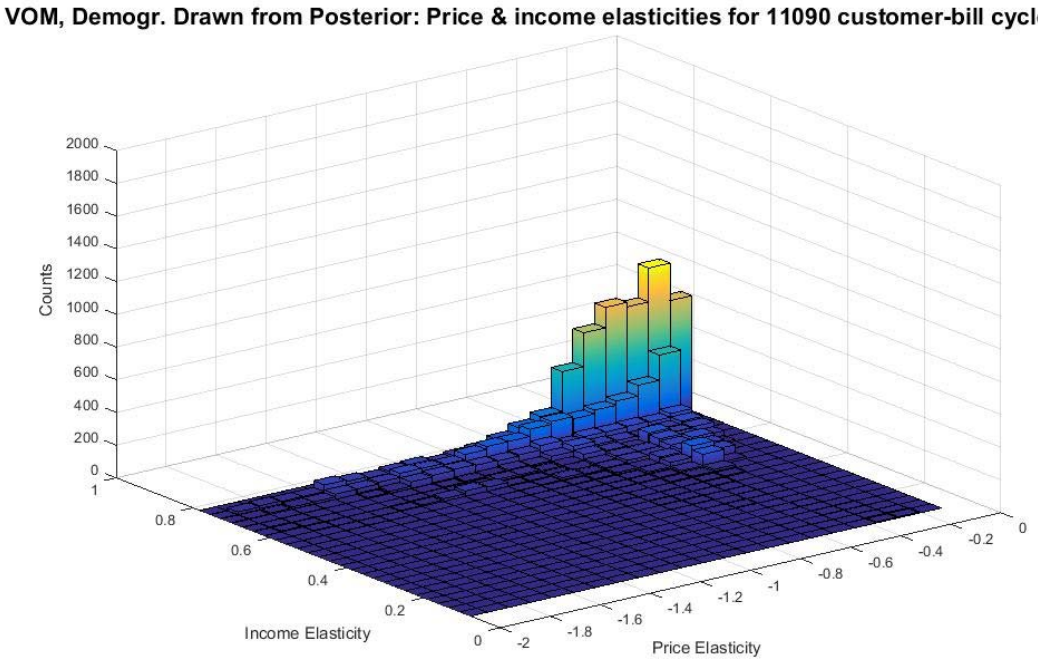

Figure 4(b): Histogram of Price and Income Elasticities for Cobb
(Demographics Drawn from Posterior Distribution)

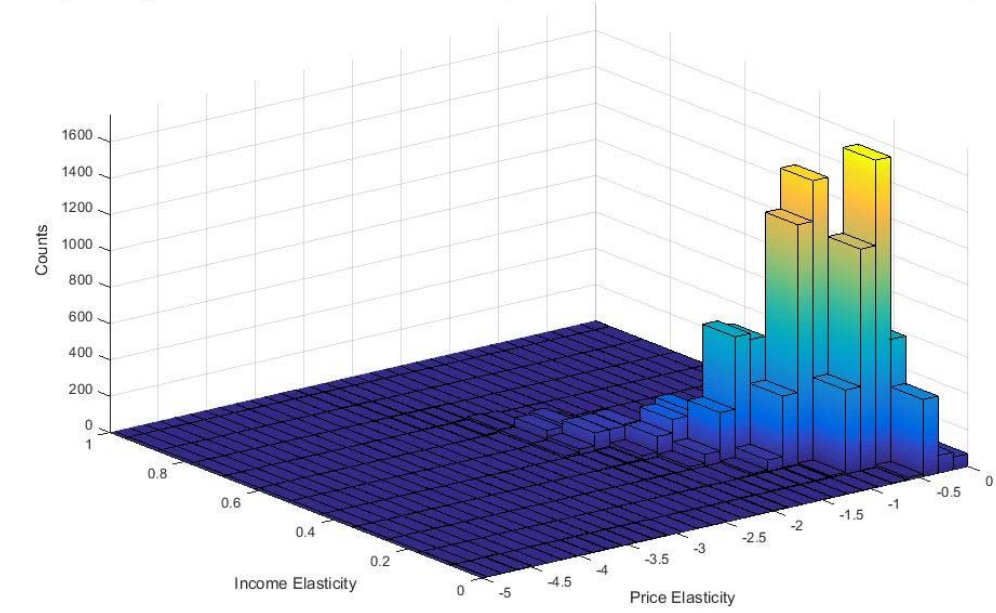

Figure 5(a): Minimum Loss Conservation Price Schedules
(Demographic Drawn from Posterior Distribution)

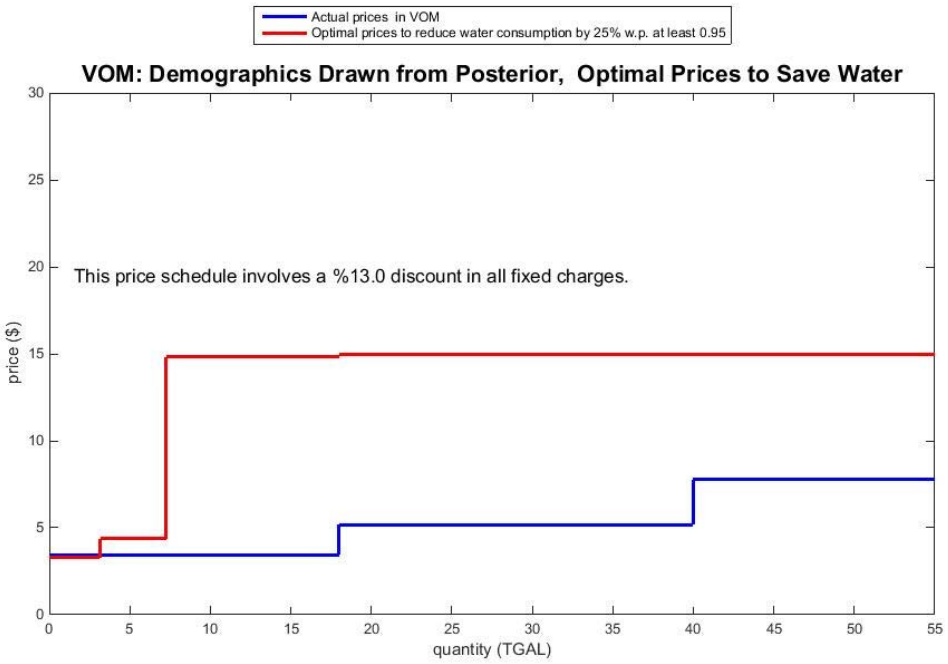

Figure 5(b): Histogram of Consumption Under Conservation Price Schedules
(Demographics Drawn from Posterior Distribution)

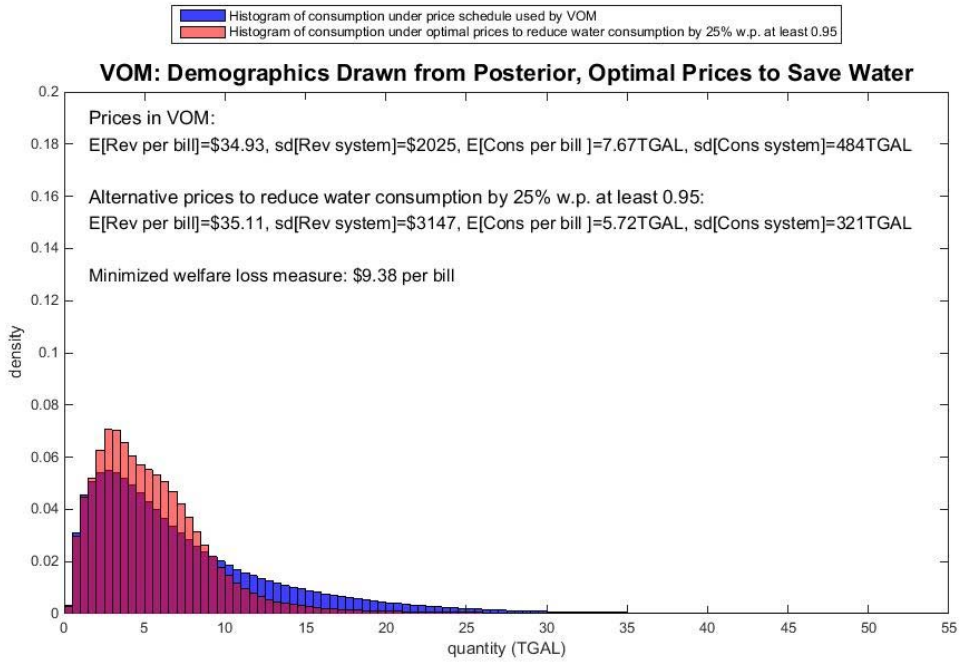

Figure 6(b): Vegetation Index-Based Price Schedules
(Demographics Drawn from Posterior Distribution)

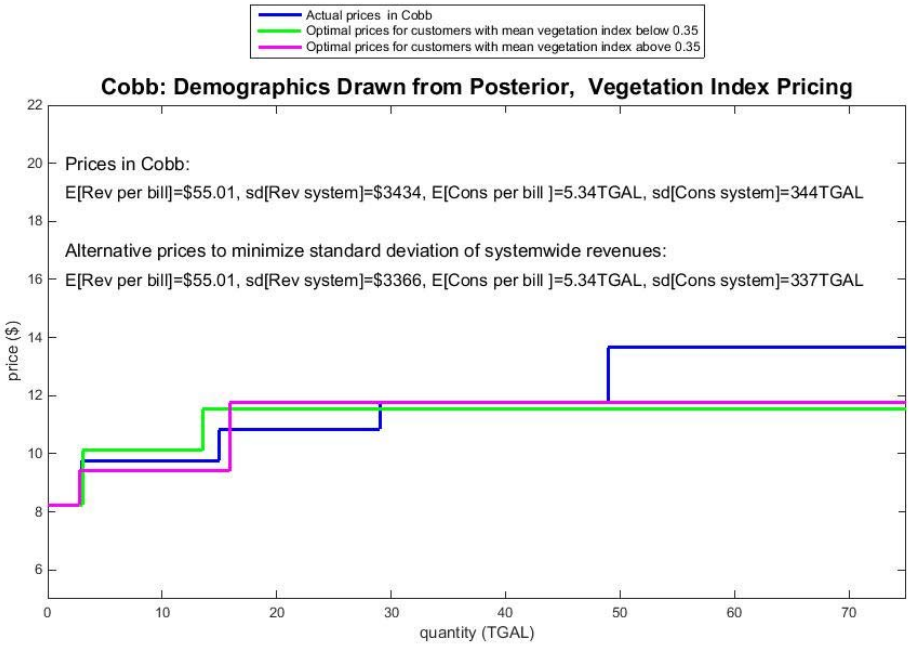

Figure 6(b): Vegetation Index-Based Price Schedules
(Demographics Drawn from Posterior Distribution)

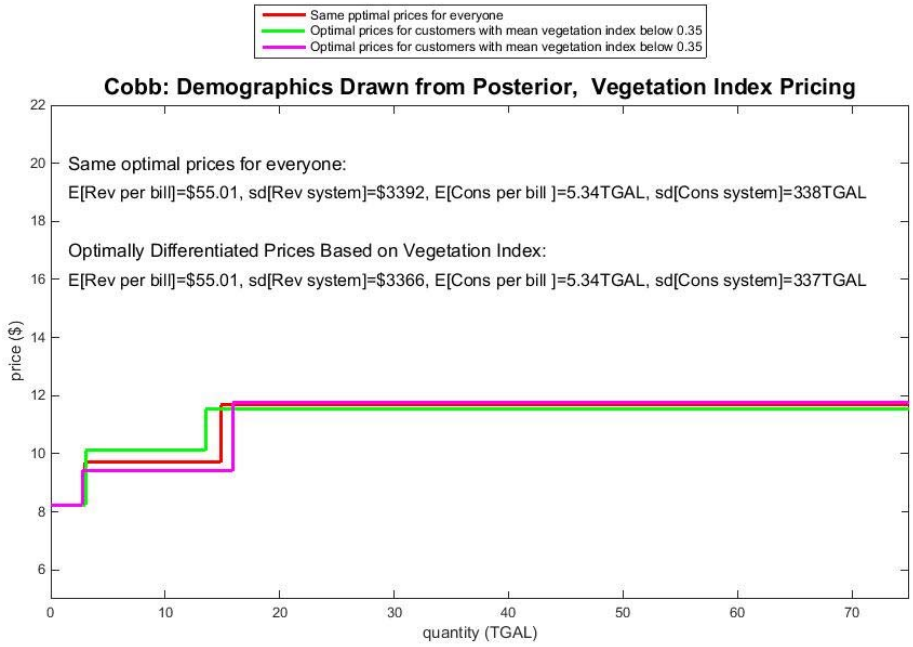

Figure 7: Decomposition of Variance of Annual Consumption—VoM
(Demographics Drawn from Prior Distribution)