Caveat emptor -This is editing work in progress. Please check back for updates – essentially it is accurate – just needs editing a bit B. Browne brian@h2oecon.com 6.15.2023

Statistical Analysis of San Francisco Public Utilities Commission (SFPUC) Retail Water System and Bay Water Supply and Conservation Agency Wholesale Purchase.

Part I – San Francisco Retail System

Figure 1 employs an indexing system to outline the historical trajectory of crucial SFPUC retail water parameters, with 1985 values set as the base at 100 (Nuffield Foundation, undated). The data in Figure 1 highlights growth patterns from 1985 to 2022 for Aggregate Retail Demand in Hundred Cubic Feet (CCF/748 U.S. gallons), San Francisco Population, and Per Capita Use (calculated by dividing Aggregate Demand by Population). The historical data is accompanied by linear regression trend lines, each with their respective equations and estimated R2 (R-square) goodness-of-fit measure for the three linear regression models. A score of 1.0 indicates that the equation fully explains the growth. The estimated R-squared values of 0.94 (San Francisco Population), 0.79 (SFPUC Aggregate Water Demand), and 0.88 (SFPUC Per Capita Use) suggest that these linear regression trend lines (Yale, 1997-98) have a high explanatory value. The San Francisco population trend (U.S. Census Bureau, 2023) shows a gradual increase(except post Covid-19), while both Aggregate SFPUC Water Demand and SFPUC Per Capita Demand exhibit a steady decline. The causes for these changes are examined below.

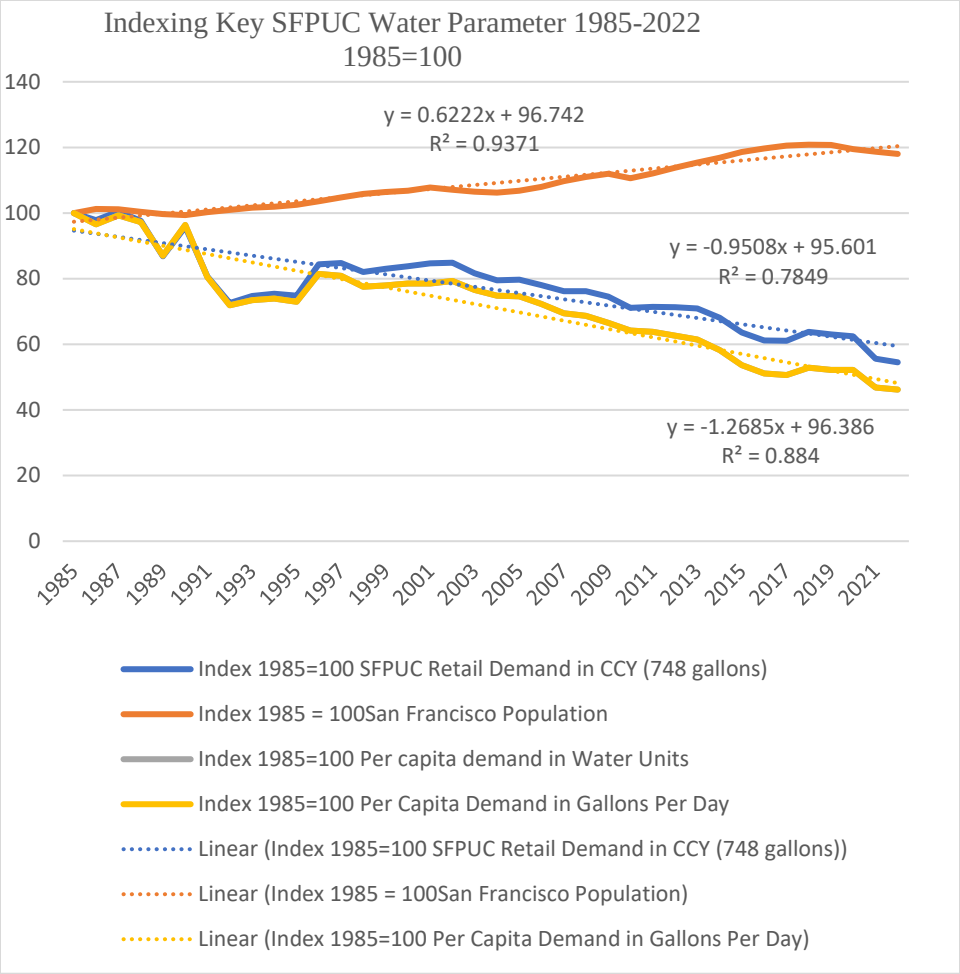
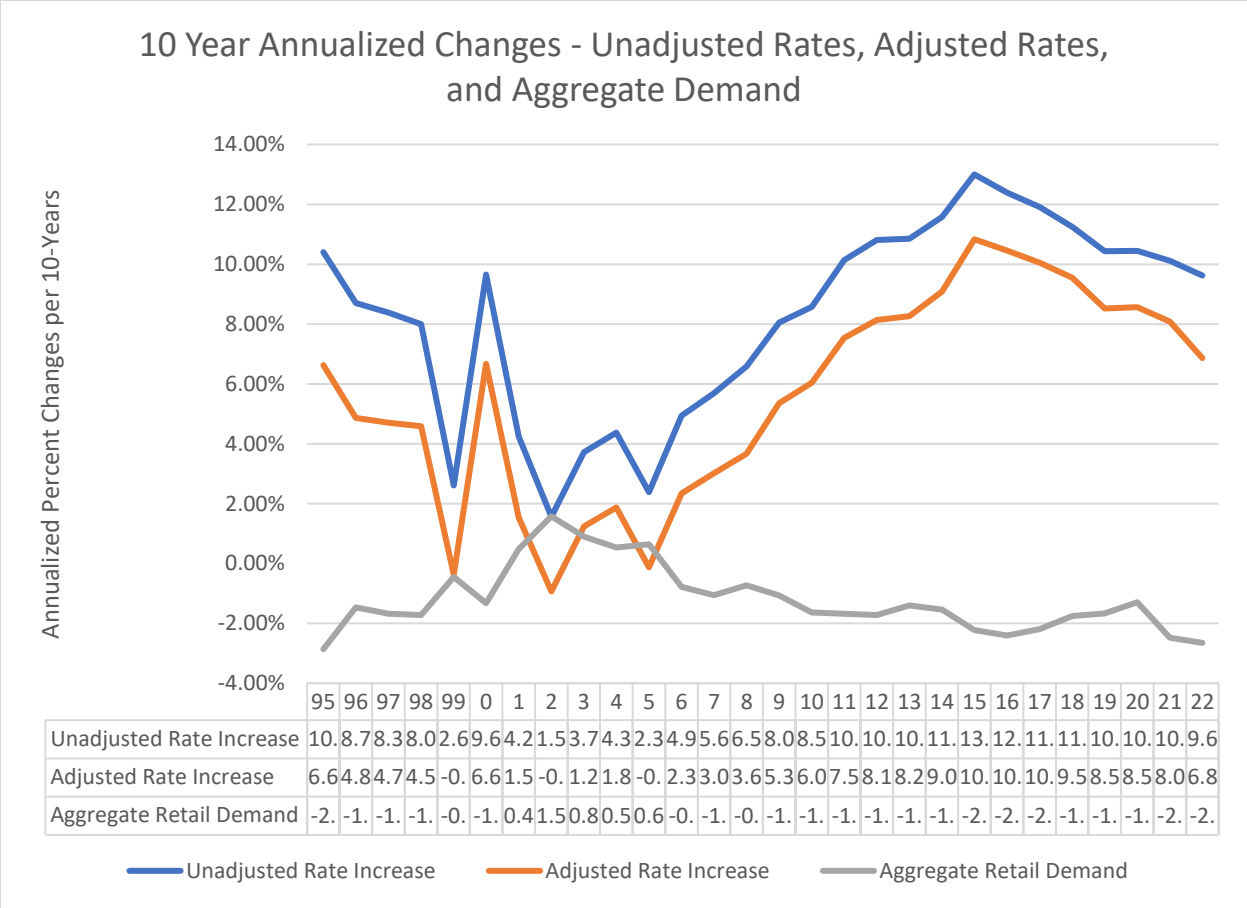


Figure 1

Figure 2 presents a 10-year analysis of annualized Unadjusted Average Rates (in current dollars), Adjusted Rates (using the CPI deflator with 1985 as the base year), and Aggregate Retail Demand. The period spans from 1995 to 2022. A visual correlation between nominal and adjusted rate increases is evident. Aggregate Demand, rather than Per Capita Demand, was utilized in the context of the SFPUC's need to establish a long-term relationship between revenue requirements as a function of pricing and quantities sold or demanded.



	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Unadjusted Rate Increase	10.	8.7	8.3	8.0	2.6	9.6	4.2	1.5	3.7	4.3	2.3	4.9	5.6	6.5	8.0	8.5	10.	10.	10.	11.	13.	12.	11.	11.	10.	10.	10.	9.6
Adjusted Rate Increase	6.6	4.8	4.7	4.5	-0.	6.6	1.5	-0.	1.2	1.8	-0.	2.3	3.0	3.6	5.3	6.0	7.5	8.1	8.2	9.0	10.	10.	10.	9.5	8.5	8.5	8.0	6.8
Aggregate Retail Demand	-2.	-1.	-1.	-1.	-0.	-1.	0.4	1.5	0.8	0.5	0.6	-0.	-1.	-0.	-1.	-1.	-1.	-1.	-1.	-1.	-2.	-2.	-2.	-1.	-1.	-1.	-2.	-2.

Figure 2

Additional analysis – Model Building (Multiple Regression)

The subsequent stage in analyzing the SFPUC's Aggregate Retail Water Demand involved employing multiple regression techniques to determine the correlation between Aggregate SFPUC Retail Demand (the dependent variable) and other influencing factors (such as rates, population, etc.). The general equation for a multiple regression model can be expressed as $Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn$, where $\beta 0, \beta 1, \beta 2, \ldots, \beta n$ denote coefficients representing the impact of each independent variable (Xi) on the dependent variable (Y). A key advantage of multiple regression is its applicability in intricate relationships. Two significant limitations include the assumption of linear relationships and the potential omission of crucial driving variables.

Figure 3 shows the steps followed to develop an econometric model to analyze the determinants of demand of the SFPUC's Aggregate Retail Demand sector. These steps may be described:

Phase 1 - Database, Economic Theory, Forecasts of Available Explanatory Variables

The database was compiled using data obtained from Freedom of Information Act (FOI) requests (SFPUC/CCSF), reviewing available SFPUC data, and sources such as the Federal Reserve Bank of Saint and other governmental sources. Excel updates provided by former GM Harlan Kelly and former Acting GM Michael Carlin (et al) were particularly helpful. Orthodox economic analysis tools were strictly

followed. Various forecasting entities were considered, including SFPUC forecasts and extrapolations of developed models with scenarios from banks and chambers of commerce. The model scenarios were designed for scenario testing with assigned probabilities that sum to 1. Since 2000, data changes have influenced the preselection of models. This preselection was customized to establish conditions specific to the observed relationships between SFPUC retail water demand in San Francisco and the available explanatory parameters.

Tests for model and variable significance

R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. The R-squared statistic shows how well the data fit the regression model (the goodness of fit).

Adjusted R- is a corrected version of R-squared that increases when adding a predictor improves the model more than expected by chance and decreases when adding a predictor does not improve the model much. Adjusted R by subsuming the number of variables decreases and is considered a more accurate metric for evaluating the equation.

The significance F value for the equation is the p-value associated with the overall F statistic for the regression model. This value indicates if the regression equation is statistically significant. The equation is statistically significant if the value is less than the chosen statically level criteria ($p<.05$)

The p-value is used in hypothesis testing to help decide whether to reject the null hypothesis. The null hypothesis (Ho) assumes no relationship exists between two variables or phenomena. Hypothesis testing checks the validity of Ho. Sufficient evidence to reject the Ho and accept Ha (alternative hypothesis) for this study was a p-value of less than 5 percent (<.05). The p-value Ho/Ha testing is performed for the overall equation and all variables.

The Durbin-Watson test for autocorrelation or serial correlation between variables has been incorporated into the Excel output. This test produces scores ranging from 0 to 4. Values between 0 and less than 2 indicate positive autocorrelation, while those between 2 and 4 signify negative autocorrelation. Acceptable levels of serial or autocorrelation are represented by values in between.

Forecast Availability: Certain macroeconomic systems offer predictions derived from data accessibility and intricate mathematical associations. These are utilized when suitable. Comparisons with SFPUC forecasts are conducted. Scenario testing, employing both subjective and randomly-selected (roulette) methods, is available. The evaluated probabilities span from 0 to 1

Model development involves examining numerous relationships to elucidate the aggregate retail demand of SFPUC. This extensive process has evolved over several years, starting with the Mayor's Infrastructure Task Force, the unsuccessful Revenue Bond Oversight Committee (RBOC), addressing a GM request to validate the 2018 rate case, and serving academic and editorial purposes.

Phase 2 – Statistical analysis and model selection and Phase 3 - Validation are merged in this analysis.

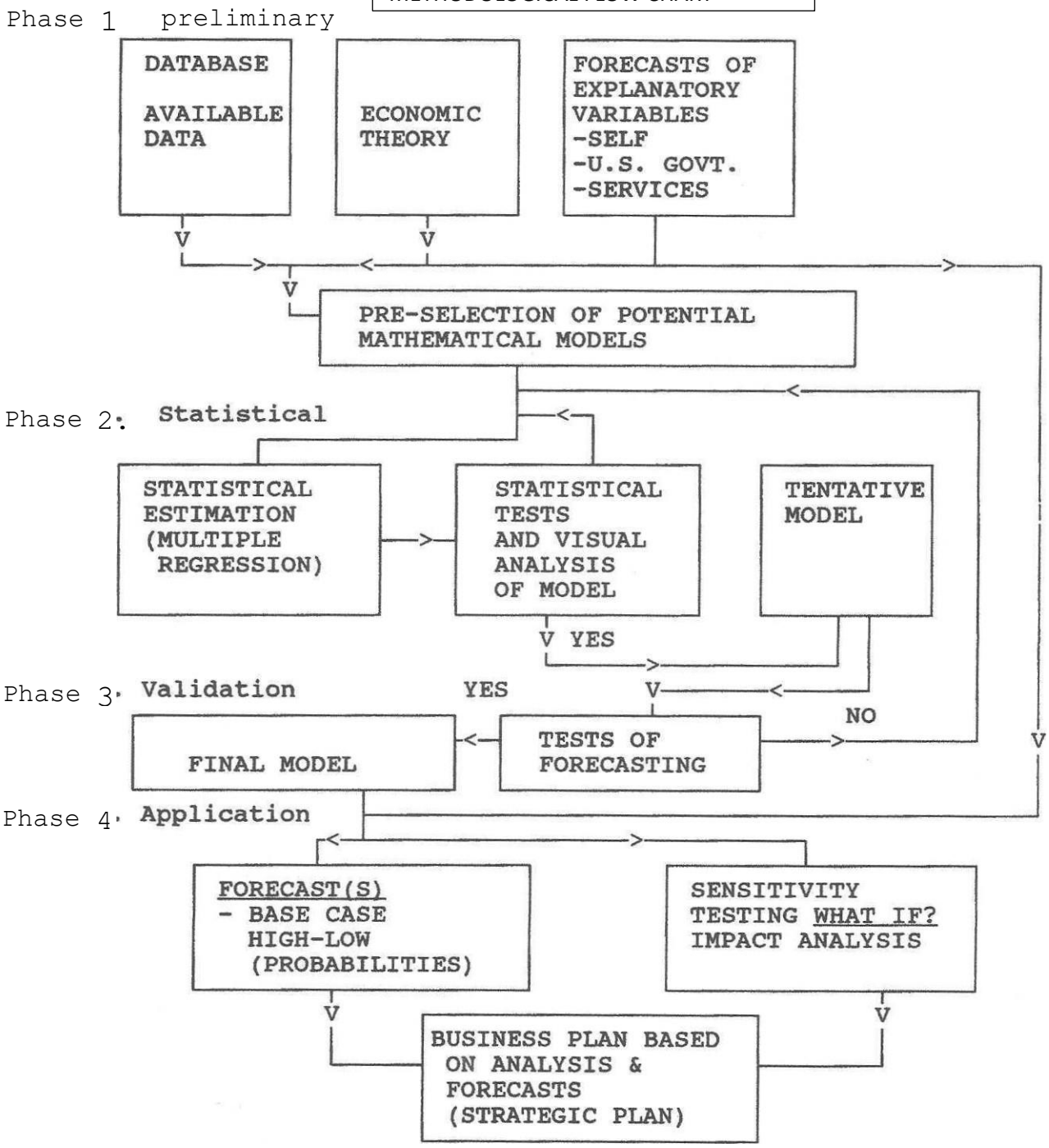
Appendix 1 presents samples of models in progress. These models varied in complexity and data availability.

The database primarily consists of continuous variables, except quantifying drought sequences as discrete (1 or 0) for the periods 1987-1992, 2007-2009, 2012-2016, and 2020-2021. An ordinary least square analysis was conducted using a combination of continuous and discrete variables. Converting data to logs in regression analysis addresses issues related to distributional shape, heteroscedasticity, and coefficient interpretation (displaying percent changes as beta coefficients versus absolute changes when using non-log data).

Data are difficult to obtain. An era of data obstruction has descended on the SFPUC. This phenomenon validates the thesis that process does determine output and narrative control is a vital element in this control. There is no real regulatory oversight (just the official narrative) and current practices under 1996 Proposition 218 are deliberately not being properly implemented. The need to control the narrative is synonymous with such behavior. These data shortages for the period 1985 to 2022 necessitated only accessible and validated data available. These are shown in Figure 4. These data are from an author-developed Excel workbook containing macro-sectoral, socioeconomic, and system variables. Sources are cited.

Part 2 BAWSCA – Wholesale – City Gate Demand in CCF (HCF)

METHODOLOGICAL FLOW CHART



ECON. 395 – APL BRIAN BROWNE (Y Aureille, 1973)

Data Received from the SFPUC

Col. 1 FYE	Col 2 Total Demand Water Delivery Volumes (MGD)	Col 3 Wholesale Water Demand Million Gallons Per Day (MGD)	Col 4 Retail Demand (San Francisco) Million Gallons per Day (MGD)	Col 5 Total Revenues Collected from Aggregate Demand	Col 6 Wholesale Revenues Collected from Peninsula Wholesale Customers
1985	275	170	105	$55,434,000	$28,921,699
1986	273	170	103	$56,961,000	$29,099,711
1987	290	185	106	$59,582,000	$27,602,484
1988	284	181	103	$59,741,000	$26,559,965
1989	229	138	91	$71,153,000	$22,034,190
1990	261	160	101	$62,914,000	$33,879,912
1991	217	132	85	$76,214,512	$36,243,421
1992	202	125	76	$97,794,281	$48,039,270
1993	212	133	79	$106,888,550	$60,280,877
1994	228	149	79	$90,994,023	$44,670,576
1995	224	146	79	$105,251,495	$51,892,697
1996	250	162	89	$112,799,128	$57,448,521
1997	260	171	89	$120,394,213	$60,043,848
1998	245	158	86	$116,281,737	$56,106,018
1999	256	169	87	$112,883,893	$52,117,271
2000	261	173	88	$135,950,428	$72,140,428
2001	264	175	89	$139,719,486	$76,156,486
2002	261	171	89	$144,304,220	$76,388,220
2003	255	169	86	$148,975,443	$75,589,443
2004	264	181	84	$174,933,601	$99,987,601
2005	251	167	84	$164,089,742	$92,098,742
2006	246	164	82	$167,381,602	$84,477,602
2007	256	176	80	$201,268,555	$106,915,555
2008	253	173	80	$219,767,046	$113,932,046
2009	242	164	78	$236,476,615	$118,129,615
2010	224	149	75	$241,390,322	$118,193,322
2011	219	144	75	$273,038,624	$132,212,624
2012	219	144	75	$341,980,341	$182,609,341
2013	223	148	75	$389,598,954	$211,147,954
2014	221	150	72	$370,987,619	$178,953,619
2015	195	128	67	$369,742,524	$174,654,524
2016	175	111	64	$412,145,303	$203,005,303
2017	180	116	64	$466,279,497	$233,356,497
2018	196	129	67	$520,133,000	$262,764,000
2019	191	125	66	$520,485,444	$250,454,444
2020	197	132	66	$595,509,000	$303,340,000
2021	193	135	58	$562,160,885	$275,113,885
2022	186	128	57	$566,342,712	$261,187,561
1. Received from SFPUC 9/21/2022				Received from SFPUC 9/21/2022	
2. Source Compliance Audit Statement					
3. Source 1985-1999 Financial Statements, Customer Billing records; 2000-2021 Annual Comprehensiv					
4. Source Customer billing records					
5. Suspect - as of when in 2022					

Figure 4 (SFPUC, 2022)

The Aggregate Retail San Francisco Econometric Model and Tests for Statistical Significance – Excel plus Durbin Watson Statistics (Durbin, J., & Watson, G. S., 1951).

	SUMMARY OUTPUT								
							Key P-Values		
	Regression Statistics					Ho = Null Hypthesis	P Values		
	Multiple R	0.964915832				Ha = Alternative Hypothesis	Regression Model	Seek Ha	
	R Square	0.931062563					$/CCF Constant SF Retail log	Seek Ha	
	Adjusted R Square	0.924979848					San Francisco Population log	Seek Ha	
	Standard Error	0.042850568					San Francisco Population log	Seek Ha	
	Observations	38					Drought 1 or Zero no Drought	Not Reject Ho	
	ANOVA								
		df	*SS*	*MS*	*F*	*Significance F*			
	Regression	3	0.843171304	0.281057101	153.0669376	8.24505E-20			
	Residual	34	0.062429821	0.001836171					
	Total	37	0.905601124						
		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
	Intercept	-1.378166736	4.925163401	-0.279821526	0.781309815	-11.38730301	8.630969542	-11.38730301	8.630969542
	$/CCF Constant SF Retail log	-0.378798907	0.038060104	-9.95265047	1.31713E-11	-0.456146344	-0.301451471	-0.456146344	-0.301451471
	San Francisco Population log	1.395451803	0.363314248	3.840894794	0.000509659	0.657108418	2.133795188	0.657108418	2.133795188
	Drought 1 or Zero no Drought	0.024287041	0.014407862	1.685679693	0.10101533	-0.004993258	0.05356734	-0.004993258	0.05356734

Figure 5 (Excel output, 2022)

The estimated equation

Y = -A (1.378166736) - 0.378798907* (X1) + 1.395451803* (X2) + 0.024287041 *(X3) + e

Where

A (constant) =1.378166736 -vertical intercept

B1 $/CCF log using constant (inflation-adjusted) rate dollars for rate estimation. The estimated value of B1 is -0.378798907. This means that for every 1 percent rate increase in the rate coefficient for retail water, there is a nearly 0,4 percent decrease in demand for retail water.

B2 San Francisco Population log is estimated at 1.395451803. This means that for every 1 percent rate increase in the City's population, there is a nearly 1.4 percent increase in demand for retail water.

The relative effects of these changes must be considered in the context of growth. Rates in nominal terms for the period 1985 to 2022 increased by a factor of 21.11 and in constant terms by a factor of 7.7 while population growth increased by a factor of 1.18. These statistics show that price (Alchian, A. A., & Allen, W. R. (1967)) or rate increases statistically were significantly more causal than both population and more so using per capita growth

Drought 1 or zero, no Drought, is estimated at a positive 0.024287041. This binary statistic suggests that there is a small increase in aggregate demand for every drought period. The impact on the overall model was inconsequential and counterintuitive, suggesting the need for more disaggregated data (not available) and the use of sophisticated lagged functions. As with the San Francisco population variable, this partial regression coefficient is "outweighed" by the negative effect on demand by the influence of escalating rates.

e = Is the error term or difference between the Yi observed between observed and Yp predicted outcomes using the model.

Key comments for the statistical output

The R-Square of 93 percent indicates that the equation explains 93 percent of variations in the dependent variable (Demand for San Francisco retail water).

Test Analysis The p-Value tests for the overall regression fit, rate partial regression coefficient, and income partial regression were all less than 0.05 significance level which meant that Ho (Null hypothesis) was rejected and Ha (Alternative hypothesis) was sought. The partial regression coefficient for the drought partial regression coefficient was 0.10 or greater than 0.05. Meaning that Ha was not rejected. However, at a 90-percent significance test, it would have been and an alternative explanation investigated (see above discussion).

The Durbin-Watson (DW) test result of 1.1 indicates potential positive autocorrelation in the residuals, suggesting that the errors in the model may not be independent and could be correlated. Addressing this issue may involve using lagging variables (i.e., distributed lagged function); however, this was not performed due to insufficiently disaggregated data. The DW test should be considered alongside other controlling variables. Future analyses ought to utilize more disaggregated data and lag essential variables, such as rates and droughts, which do not immediately impact demand.

Visual Test

The visual test is where the estimated regression model is used to try and forecast or rather backcast the observed data. The predicted and observed lines track well in terms of physical differences and variations.

Summary

The high Adjusted R-square and low P-values (<.05) validate the use of this equation as a predictive model assuming historical changes are well correlated with future events. The strength of the price elasticity coefficient (as subsumed by relative annual changes in population) indicates the main explanatory factor, based on these results, for decreasing demand, has been the continual and increasing increase in water charges to its customers by the SFPUC Retail Division.

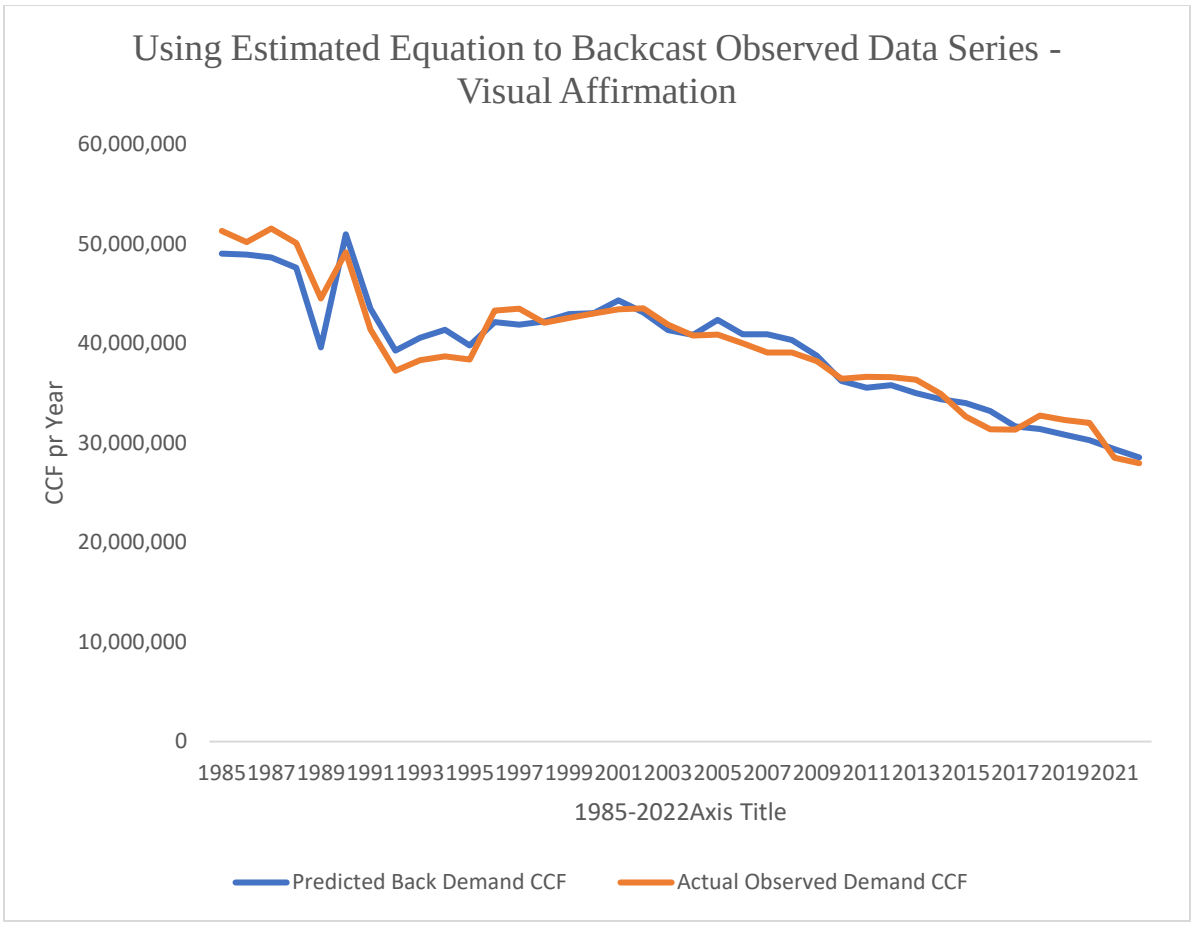

Using Estimated Equation to Backcast Observed Data Series - Visual Affirmation
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
CCF pr Year
1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019 2021
1985-2022Axis Title
Predicted Back Demand CCF
Actual Observed Demand CCF

Figure 6

Part 2 – City Gate or BAWSCA Model

The Aggregate Wholesale BAWSCA Econometric Model and Tests for Statistical Significance – Excel plus Durbin Watson Statistics (Durbin, J., & Watson, G. S., 1951).

Excel Output – BAWSCA – City Gate Statistical Analysis

SUMMARY OUTPUT Log analysis ln(Bi)								
Regression Statistics								
Multiple R	0.815840744							
R Square	0.66559612							
Adjusted R Square	0.636089895							
Standard Error	0.083143864							
Observations	38							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	0.467820016	0.155940005	22.55782049	3.21541E-08			
Residual	34	0.235038672	0.006912902					
Total	37	0.702858688						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-13.79615384	6.184354169	-2.230815614	0.032397	-26.36427364	-1.228034037	-26.36427364	-1.228034037
LOG Dollars per AF in Constant Terms 1985=1 BAWSCA	-0.419919614	0.064274362	-6.533236611	1.76325E-07	-0.550540833	-0.289298395	-0.550540833	-0.289298395
LOG BAWSCA Number of customers on HH System (nonlog)	1.92038537	0.44327184	4.332297242	0.000123734	1.019548607	2.821222134	1.019548607	2.821222134
Dummy 1/0 for C19 X3	-0.014689865	0.056489049	-0.260048011	0.796395118	-0.129489424	0.100109695	-0.129489424	0.100109695

Figure 7 (Excel output 2022)

The estimated equation

Y = -A (-13.796) - -0.419919614* (X1) + 1.92038537* (X2) -0.014689865 *(X3) + e

A (constant) --13.796 vertical intercept

B1 $/CCF log using constant (inflation-adjusted) rate dollars for rate estimation. The estimated value of B1 is 0.419919614. This means that for every 1 percent rate increase in the rate coefficient for retail water, there is a nearly a -.41 percent decrease in physical demand from BAWSCA for water.

B2 the BAWSCA customer base log is estimated at 1.92038537. This means that for every 1 percent increase in the BAWSCA customer base, there is a nearly a 2 percent increase in demand for water from BAWSCA.

The relative effects of these changes must be considered in the context of growth. Rates in nominal terms for the period 1985 to 2022 increased by a factor of 1.11 and in constant terms by a factor of 7.7 while population growth increased by a factor of 1.18. These statistics show that price (Alchian, A. A., &

Allen, W. R. (1967) or rate increases statistically were significantly more causal than both population and more so using per capita growth

Null hypothesis is a statement that assumes there is no significant difference between two variables or groups being compared. In statistical analysis, the null hypothesis is tested against an alternative hypothesis to determine whether the observed results are statistically significant or occurred by chance. setting up a null hypothesis using P scores involves determining an alpha value, calculating a P-value, and comparing it to the alpha value to determine whether to reject or fail to reject the null hypothesis.

Ho for B1 was 0.05 (95 percent certainty) and the estimated P-value of 1.76E-07 caused the rejection of the Ha "no significant difference" hypothesis and acceptance of Ha for B1 (alternative) as a determining equation factor.

Ho for B2 was 0.05 (95 percent certainty) and the estimated P-value of 0.000123734 caused the rejection of the Ha "no significant difference" hypothesis and acceptance of Ha for B2 (alternative) as a determining equation factor.

Ho for B2 was 0.05 (95 percent certainty) and the estimated P-value of 0.796395118 caused the non-rejection of the Ha "no significant difference" hypothesis and questioned B2 (alternative) as a determining equation factor.

Ho for B3 was 0.05 (95 percent certainty) and the estimated P-value of 0.796395118 caused the non-rejection of the Ha "no significant difference" (never accepted) hypothesis and questioned B2 (alternative) as a determining equation factor.

Visual affirmation of estimated regression equation

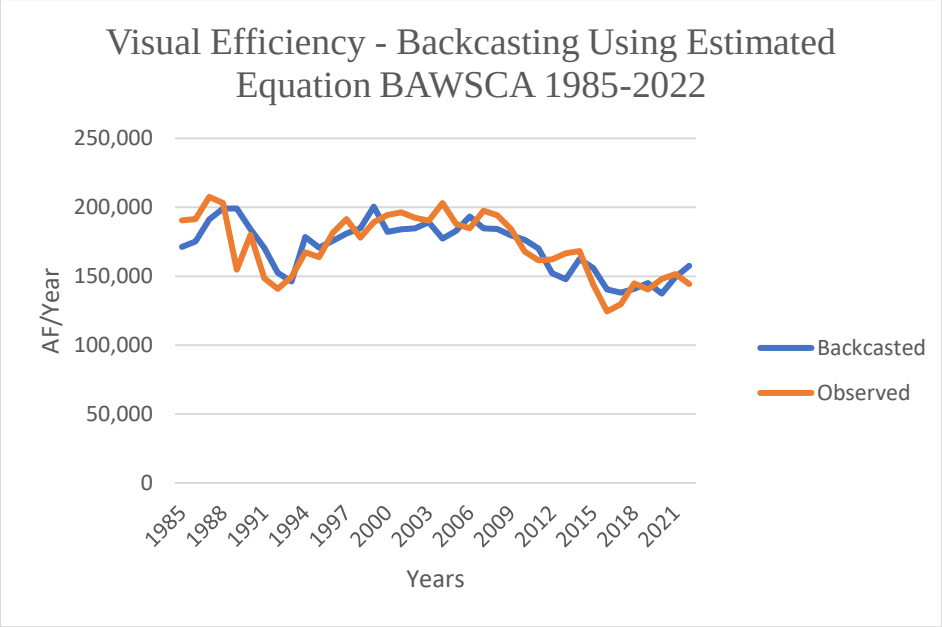


Figure 8

Footnotes

1. Alchian, A. A., & Allen, W. R. (1967). University economics: Elements of inquiry. Belmont, Calif: Wadsworth Pub. Co.
   2. Mankiw, N. G., & Taylor, M. P. (2014). Economics. Andover: Cengage Learning.
   3. Samuelson, P. A., & Nordhaus, W. D. (2010). Economics. New York: McGraw-Hill/Irwin.
2. https://www.nuffieldfoundation.org/sites/default/files/files/FSMQ%20Average%20earnings.pdf
3. http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm
4. 1. "Introduction to Linear Regression Analysis" by Douglas C. Montgomery was first published in 1982 by John Wiley & Sons, Inc. The book provides an introduction to the theory and application of linear regression analysis, including simple and multiple regression models, model building, diagnostics, and remedial measures.

   2. "Linear Regression Analysis: Theory and Computing" by Xin Yan and Xinyu Song was first published in 2009 by the World Scientific Publishing Company. The book focuses on the theoretical foundations of linear regression analysis, including estimation, hypothesis testing, and model selection. It also covers computational methods for implementing linear regression models.

   3. "Statistical Methods for Psychology" by David C. Howell was first published in 1987 by Duxbury Press. The book provides an introduction to statistical methods commonly used in psychology research, including descriptive statistics, inferential statistics, correlation analysis, and regression analysis.