A COMPARISON OF RESIDENTIAL WATER AND SEWER RATES IN GEORGIA

Mark Horowitz

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Science and Engineering.

Chapel Hill
2007

Approved by:

David Moreau

Jeffrey Hughes

Philip Singer

ABSTRACT

MARK HOROWITZ: A Comparison of Residential Water and Sewer Rates in Georgia
(Under the direction of David Moreau)

When setting rates, many utilities use rate surveys – regional compilations of utilities' rates – to gauge a fair price increase. However, each utility has a unique set of factors that affect its rate, so simple comparisons between two utility rates may lead to the wrong conclusion. This thesis describes regression models which provide better comparisons by incorporating factors that influence rates. Two types of bills – water only and combined water and sewer – are modeled at four consumption levels: 3000, 6000, 9000, and 12000 gallons per month. The models use the data from all the utilities in the sample to provide an estimated average bill, with a 95% confidence interval, for each utility. Then, each utility can compare its actual bill with this estimate. The models also show that high bills (both types) are associated with source water, recent rate changes, large grants, and large connection fees at most consumption levels.

ACKNOWLEDGEMENTS

I want to thank David Moreau, Jeffrey Hughes, and Philip Singer for being on my committee. I also want to thank Mr. Hughes for sponsoring the research assistantship through the Environmental Finance Center (EFC). The EFC team – Shadi Eskaf, Tim Healy, Rich Thorsten, and Andrew Westbrook – was also a big help in collecting and organizing the primary data. The Odum Institute helped me greatly in understanding the nuances of multiple regression and Stata programming. The following people provided secondary data, and I am greatly indebted to them: Ashley Meggitt of the Association County Commissioners of Georgia; Deatre Denion and Irette Patterson of the Georgia Department of Community Affairs; Jason Bodwell and Melanie Manning of the Georgia Environmental Facilities Authority; Doug Davenport, Bill Frechette, Alice Keyes, and Ade Oke of the Georgia Environmental Protection Division; and Jack Stanek of the United States Department of Agriculture. Finally, I want to thank Em Chuang, Lindsay Dubbs, Kevin Ivarsen, Joon-Suk Lee, Nina Yamanis, and William Zieburtz for ideas and feedback on the report.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CBO	Congressional Budget Office
CF	cubic feet
CI	confidence interval
CR	Commercial rates
CWNS	Clean Watersheds Needs Survey
CWS	community water system
DVA	dummy variable adjustment
EFC	Environmental Finance Center
EPA	Environmental Protection Agency
FHA	Farmers Home Administration
Ga EPD	Georgia Environmental Protection Division
GEFA	Georgia Environmental Facilities Authority
GM	gallons per month
GU	groundwater under the influence of surface water
GUP	purchased groundwater under the influence of surface water
GW	groundwater
GWP	purchased groundwater
MWDOC	Metropolitan Water District of Orange County
NCDC	National Climatic Data Center
No.	number
Ohio EPA	Ohio Environmental Protection Agency
OLS	ordinary least squares

OR	outside rates
PDF	portable document format
Pop.	population
Prob.	probability
PS	primary water source
PWS	public water system
RS	rate structures
SDWIS	Safe Drinking Water Information System
SPLOST	special purpose local option sales tax
SW	surface water
SWP	purchased surface water
UO	utility ownership
USDA	United States Department of Agriculture
W&S	water and sewer
WIFA	Water Infrastructure and Finance Authority of Arizona

CHAPTER 1

INTRODUCTION

The primary objective of this report is to help utilities compare their rates to those of other utilities. The utilities in the sample (publicly-owned water and sewer systems in Georgia), and other utilities, make comparisons in order to gauge fair price increases for residential customers and to measure their performance. Another objective of the report is to define the factors associated with the rates in order to uncover potential policy implications. In addition, future organizations conducting rate surveys – regional compilations of utility's rates – in Georgia may want to include information on these factors instead of performing a full statistical analysis using the methods in this report. The following paragraphs discuss the problem statement, report objectives, and contents of the report.

Problem Statement

For the past several decades, water and sewer utilities in the United States have faced increasing costs due to aging systems and increasing regulation to protect public health and the environment. Revenue comes from a variety of sources, including federal loans and subsidies, state loans and grants, local taxes, and ratepayers' bills. The majority of revenue comes from the latter (Congressional Budget Office 2002). While utilities set their rates to recover as much expenses as possible, they also attempt to make rates affordable. Thus, as costs have risen, utilities have paid more attention to different rate-setting techniques and have turned to consultants as well as manuals, such as the "Principles of Water Rates, Fees, and Charges (M1)," (AWWA 2000) for assistance.

Utilities have also used rate surveys, which generally list the monthly residential water and/or sewer bills of other utilities in a particular region, usually a state. The main purpose of a survey is to allow utilities to compare their rates to other utilities, which can be useful during the rate-setting process. If a utility has relatively low rates, it can use a survey to help justify rate increases to their customers as well as the city council or board of directors. Also some utilities with relatively low rates use surveys to promote themselves as high-quality, low-cost service providers (City of Cartersville 2007; Dalton Utilities 2007). If a utility has relatively high rates, the surveys may prompt it to improve efficiency and management. However, if a utility with high rates is already efficient, it may feel less inclined to raise rates for political reasons and/or to reduce the burden on consumers.

While surveys are useful for portraying the general trends in a region, it is difficult to make accurate rate comparisons as they are not adjusted to reflect influential factors. At the least, utilities can use a survey for ballpark comparisons with neighboring utilities that have a similar amount of customers. However, it is very easy for utility managers, boards of directors, and the public to compare the rates directly, which may lead to the wrong conclusions about a utility. Accurate comparisons should account for different characteristics of the utility and its service area, such as source water, population served, and utility finances.

Report Objectives

This report incorporates those and other characteristics through statistical models based on ordinary least squares (OLS). Several benefits are attained through this technique. First, given the data and consumption level, the statistical models can estimate the expected monthly residential bill for an average utility with specific characteristics. Thus, a utility may directly compare its actual bill to the expected bill for an average utility with the same traits. Not only can a utility incorporate several influential factors more accurately, but the analysis includes utilities in the entire sample and not just nearby. Second, the hypothesized factors that influence the bill can be tested for statistical significance. In other words, given the data, OLS can show which factors demonstrate a close relationship with the utility rates as well as the magnitude of that relationship. Future Georgia rate surveys may choose to include the most influential factors along with the rates so utilities can develop better comparisons. Also, this analysis, combined with future statistical analyses in other regions, could dispel or reinforce commonly held notions about the factors influencing the bill.

To illustrate the first objective, say Utility X wants to know how the bill for its residents at average monthly consumption (6000 gallons) compares to other utilities in the state. With a normal rate survey, Utility X may compare its rates to the state average. It may also find utilities with similar population and make a simple comparison of the bills. However, the statistical models account for differences in characteristics for *all* utilities in the state and provide an estimate of the water bill with a 95% confidence interval. In other words, the model can produce an estimate of an average bill for a utility with the same characteristics as Utility X. For example, the model could estimate that an average utility like Utility X bills $20 for 6000 gallons of water per month with the 95% confidence interval ranging from $17.50 to $22.50. One interprets the confidence interval as a 95% probability that, given the sample and variables, the average estimated bill is within that range[1]. If Utility X's actual bill is $15, then the bill is below the confidence interval for the average bill. If Utility X's actual bill is $18, then it is considered average. And if it is $25, then it is above the confidence interval for the average bill. Keep in mind that, while the estimates from the statistical models are more accurate than rule-of-thumb comparisons from rate surveys, the estimates are not 100% accurate. The models cannot account for all the factors that affect the bills. The key for each utility is to gauge the difference between its actual bill and the estimated range of the bill and determine the acceptability of that difference.

Report Contents

The following chapters discuss the background, literature review, methodology, results, and conclusions. The background and literature review (Chapter 2) provides a review of past rate surveys, the conventional wisdom on rate determinants, a discussion of past OLS studies on rate determinants, and an overview of water and sewer utilities in Georgia. Chapter 3 describes the general form of the models as well as the independent variables and their predicted effect on the bill. The chapter further describes how the data for the independent variables were collected and organized. In addition, the sample criteria and regression diagnostics are discussed. Chapter 4 illustrates the comparison of

[1] Note that confidence intervals are not to be confused with prediction intervals where the latter estimate the 95% probability that, given the data, Utility X's bill is within the range predicted by the model. The confidence interval is the 95% probability that the *average* bill for a utility like Utility X is within that range. Since the confidence interval bounds the average bill, it is inherently smaller than the prediction interval, which bounds a single utility's bill.

the actual bills with the estimated bills for a small sample of utilities. It also provides an analysis of the statistically significant variables and rationalization of variables that did not meet predictions. Chapter 5 discusses the interpretation of the average bills, policy implications of the statistically significant factors, limitations, and the pros and cons to using statistically significant variables to augment future rate surveys.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter provides a review of past rate surveys, the conventional wisdom on rate determinants, a discussion of past OLS studies on rate determinants, and an overview of water and sewer utilities in Georgia. The first two sections provide background material while the last two sections discuss the factors influencing water and sewer rates identified in the literature. These factors are the foundation of the statistical models discussed in Chapter 3.

Review of Rate Surveys

This section discusses a brief timeline of rate surveys in the US and their general contents. Please note the timeline is limited to surveys found in library stacks, scholarly databases, and internet searches and is by no means comprehensive. It is likely that many regional surveys were used locally and probably became "lost" over time.

The New Mexico Agricultural Experiment Station developed one of the first rate surveys for 302 utilities within the state during the 1970 calendar year (Randall and Dewbre 1972). Only a few more rate surveys from the 1970's could be found through scholarly databases and internet searches. The 1980's, however, saw the beginning of annual rate surveys from the Ohio Environmental Protection Agency (Ohio EPA 2006) for utilities in its state (1983-present), Massachusetts Water Resources Authority (MWRA Advisory Board 2006) for the utilities it serves (1987-present), and Draper

Aden Associates (DAA 2006) for Virginia public water and sewer utilities (1989-present). One of the earliest national rate surveys was conducted for 62 cities (Dallas Water Utilities Department 1975). However, a national biennial rate survey did not begin until 1996 (Raftelis Financial Consultants and American Water Works Association 2007). Many more rate surveys have been developed in the past decade by consulting firms (Black & Veatch 2006; Tighe & Bond 2004), universities (Jordan 1998; Manning, Barefield and Mays 2005), state agencies (New Hampshire Department of Environmental Services 2005; Water Infrastructure Finance Authority of Arizona 2006), non-profits (National Association of Clean Water Agencies 2006; Western Resource Advocates 2006), water wholesalers (Metropolitan Water District of Orange County 2006; San Diego County Water Authority 2006), and councils of governments (Southeast Michigan Consortium for Water Quality 2004; Triangle J Council of Governments 2007).

For Georgia, the earliest survey found was for fiscal year 1994 (Jordan 1996). The same author also conducted a wastewater rate survey for Georgia in 1998. Other organizations have created water rate surveys for Georgia (Georgia Municipal Association 2005; Metropolitan North Georgia Water Planning District 2006; Zieburtz 2004), which are typical of surveys around the country.

The heart of any rate survey is the list of utilities with residential water and/or sewer bills at one or more consumption levels. A number of surveys provide this information only, while some also provide summary statistics. Many surveys provide additional information from each utility, such as rates outside of the utility's jurisdiction, commercial rates, senior citizen rates, connection fees, billing cycle, effective date of rates, population served, and type of rate structure. Other surveys go further by providing a breakdown of the entire rate structure (MWRA Advisory Board 2006).

While the surveys are designed to compare utility rates, some survey providers warn against direct comparisons and suggest consideration of factors that influence the rates. For example, MWDOC (2006) states:

> Water rates can differ substantially among the 31 retail water utilities in Orange County. An uninformed reader might jump to the conclusion that the higher rates are unreasonable. But, care must be exercised when making direct comparisons among water utilities' rates due to the variation in conditions affecting the utilities' revenue-developing structures.

MWDOC goes on to a lengthy discussion of perceived influential variables in general and specific to its area. MNGWPD and WIFA also list the factors that may affect rates.

Conventional Wisdom on Rate Determinants

The independent variables – the factors influencing the rates – in the OLS models for this report are based on the writings of George Raftelis who has produced a number of surveys over the past few decades and wrote a textbook on rate setting (Raftelis 1988; Raftelis 2005). In the textbook, the author describes the rate determinants. One determinant is the type of rate structure (i.e., increasing, decreasing, and uniform blocks), but most are based more or less on the factors that influence the cost of water and sewer service:

- **Geography** – A utility located far from water sources and sewer discharge points generally has higher costs, and thus, higher rates. Topography also plays a role in pumping costs. And, utilities in less densely populated areas pay more for transmission and piping costs.

- **Peak demand** – The capacity of the system is based on the ability to meet peak demand at any time of the day or year. Thus, as the capacity increases, so do the charges.
- **Commercial customers** – If a relatively large number of high-volume customers are served, then the utility has relatively low costs for administration. However, if commercial customers discharge high strength wastewater, then operating costs could be relatively high.
- **Treatment** – The more treatment of water and sewer, the higher the costs.
- **Government subsidization** – Some county and municipal utilities receive subsidies to their budgets and/or in-kind services from other government departments, which lower expenses. These systems can also subsidize other government departments.
- **Grants** – The more grants received, the lower the costs.
- **Age of system** – Generally, older systems face high maintenance and upgrade costs than newer systems. However, new systems may be more expensive because of debt service payments while older systems are usually paid off. New systems also may have increased costs due to initial adjustments in design and operations.
- **Infiltration and inflow of stormwater and groundwater** – Sewer systems with high infiltration and inflow have higher costs for treatment and conveyance.
- **Rate-setting approach** – How a utility allocates its costs to rates at various consumption levels and customer classes can affect the total bill.
- **Other factors** – Demographics, political issues, management, and similar issues have an impact on costs.

OLS Studies on Rate Determinants

Several of the factors listed above have been tested for statistical significance through multiple regression by a few researchers (Hollman and Boyet 1975; Mann 1972; Thorsten, Eskaf and Hughes 2007) and many agree with Raftelis's hypotheses. However, these articles do not discuss using the statistical models to determine average rates for water and sewer utilities. The following paragraphs cover contents and findings from these articles.

Hollman and Boyet model the residential water bill at consumption levels of 6,000 and 10,000 gallons per month (GM) for 86 rural water systems in Mississippi. The next table lists the variables in both models and the expected effect on the bill. The table is followed by summarized explanations of the variables.

Table 1: Description of Independent Variables in Hollman and Boyet (1975)

Variable	Expected Effect	Reason
Population with water service	–	Costs are spread to more people, so bills drop
Water source change in past 10 years	+	Change is probably to more expensive source
Grants and subsidies	–	Other utility income may lower bill
Production and distribution expenses	+	As costs increase, so should bill
Loans (FHA* financing)	+	FHA mandates bills to be tied to debt
Connection fees	+	High charges signify high debt, so bills increase

*Farmers Home Administration

Hollman and Boyet use the *population with water service*[2] variable as a proxy for the system size (Raftelis's "peak demand" variable). And with increased production capacity comes greater economies of scale, thus the bills should decrease. Regarding source water, a rational utility would first develop the least expensive sources first. If a utility switches sources, then this could indicate that costs, and the bill, would increase. Utilities

[2] From this point forward, complete, and sometimes abbreviated, variables are italicized in the body of the report.

receiving *grants and subsidies* are expected to pass the savings to their customers through lower bills. Increased *production and distribution expenses* would naturally increase the residential bill. Also, *loans* require higher rates to pay the debt service. Finally, a high *connection fee* suggests that a utility has high debt and would raise its rates to pay this debt.

In both models, all variables were statistically significant at the 5% level. The signs for most coefficients, except *water source change*, met expectations. Hollman and Boyet assume that changing sources may also entail equipment upgrades, and thus, improved efficiency, lower costs, and reduced bills.

Thorsten et al. model the combined water and sewer residential bill at 6000 GM for 211 public water and sewer utilities in North Carolina. It is assumed that cost factors influence the total bill, but they hypothesize that factors associated with demand, institutions, and location are also important. Through several iterations, they develop a log-log regression model that includes all cost factors and only statistically significant demand and institutional variables. The next table, modified from Thorsten et al., lists the factors in the model of the total bill and is followed by explanations of the variables.

Table 2: Independent Variables Used in Thorsten et al. (2007)

Independent Variable	Expected Effect	Reasoning
Cost		
Annual production	–	Economies of scale
Long-term debt	+	Higher debt means higher bills
Source water*	+ or –	Groundwater expected to be least costly
Treats own wastewater	–	Not paying premium over treatment costs
Population density	–	Reduced pipeline costs
Demand		
Median household income	+	Wealthy persons can pay more
Percent impoverished	–	Pressure to make rates affordable
Median year homes built	–	Older towns resist increasing rates
Percent elderly	–	Pressure to make rates affordable
Percent homes owned	–	More households directly billed
Average annual temperature	+	Higher temperatures reduces water supply
Average annual rainfall	–	Higher rain increases water supply
Expected customer growth rate	+	System expansion requires more revenue
Institutional		
Ownership type*	+ or –	Municipalities expected to be least costly
Higher rates for residents outside jurisdiction** (Outside rates)	–	Other utility income may lower bills for "inside" residents
Higher rates for non-residents**	–	Other utility income may lower bills
State infrastructure grant	–	Other utility income may lower bills
Operating ratio	+	Higher ratio means higher bills
Location		
River basin*	+ or –	Polluted basins require funds for treatment
Avg. bill of neighboring utilities	+	Utilities may influence each other's bills

*Dummy variables with greater than two categories
**Dummy variables with two categories

Beginning with the cost variables, *annual production*, like Hollman and Boyet's *population with water service*, is the proxy for the system size. For *long-term debt*, utilities with high loans and bonds have large interest payments, which may be recovered through increased rates. The model includes dummy variables for the following source waters, starting from the least expensive: groundwater, surface water, purchased surface water. For *treats own wastewater*, some utilities in North Carolina must pay other utilities for wastewater treatment, which includes a surcharge that may be passed to residents. And the higher the population density, the lower the pipeline and distribution costs, and thus, lower residential bills. For *expected customer growth rate*, the utilities were requested to estimate their future growth rate, and the authors hypothesize that utilities may increase today's rates to pay for tomorrow's system expansion.

Regarding the demand variables, the *median household income*, *percent impoverished*, and *percent elderly* variables represent the ability of customers to pay the bill, and utilities may adjust their rates accordingly. For *median year homes built*, the authors expect that utilities with a low influx of new customers, usually located in established towns, would have difficulty raising their rates. Also, utilities could recover more costs, and thus reduce bills, if more customers own their homes (*percent homes owned*), which are usually metered. Apartment complexes usually do not have meters, so residents are not encouraged to limit consumption. Utilities in high *temperature* regions could have higher costs because water is scarcer. Conversely, utilities in high *rainfall* areas could have lower costs because water is more plentiful.

For institutional factors, utilities in North Carolina can be owned by authorities, counties, sanitary districts, and municipalities with the latter expected to have the lowest rates. The subsequent three variables in Table 2 are assumed to subsidize the bill for customers inside a utility's jurisdiction. And, if the *operating ratio* – the ratio of revenue to expenses – is greater than one, this may be due to a relatively high bill.

Location factors are used mainly to control for spatial autocorrelation. The authors expect that a utility's bill would be like neighboring utilities' bills. Since many utilities request rate surveys, their bills are probably influenced by them. Spatial effects are controlled through separately using river basin dummy variables, spatial regression, and a variable for the average bill for all utilities in neighboring counties.

The basic log-log model shows that *annual production*, *debt*, *median household income*, and *outside rates* are statistically significant. Also, several river basin dummy variables were statistically significant as well as the average bill of utilities in neighboring counties. The spatial regression analysis showed that autocorrelation exists, but the factor used to control this was not statistically significant.

Mann's (1972) report on using multiple regression to find determinants of water bills was unavailable electronically or in local libraries. However, the author did publish a related article in a readily accessible academic journal (Mann 1970). This article discusses using principal components analysis to determine the relationship of six measures of water prices in 113 urban areas in the US during 1960: the price at 500 cubic feet (CF), 1,000 CF, 10,000 CF, and 100,000 CF along with the average price per unit (revenue per million gallons produced and revenue per million gallons sold). The first vector explained 74% of the variation in the relationships between all six measures with the coefficients ranging from 0.35 to 0.45. This suggests that utilities with relatively high rates for one level of price generally have high rates at all price levels, and vice-versa. The second vector explained 14% of the variation with only two of the coefficients being notably high: the price at 100,000 CF and water revenue per million gallons sold (0.74 and -0.65, respectively). Mann concludes that a relatively high price for large industrial consumers (100,000 CF) is associated with a relatively low average price (water revenue per million gallons sold) and that further research could explain this outcome. Mann ends the analysis at the second vector as the third and fourth vectors explain a total of only 10% of the variation. Lastly, when the component values were computed for each utility, the 25 lowest price utilities are near lakes or major rivers, and the 25 highest price utilities are privately owned.

Overview of Public Water and Sewer Utilities in Georgia

The Georgia utilities considered in this report provide water and/or sewer service. Few provide sewer service only, and these utilities are not included in the analysis (EPA 2003). Technically, the utilities in the sample are public water systems (PWS), which are defined by the Environmental Protection Agency (EPA) as providing "water for human consumption through pipes or other constructed conveyances to at least 15 service connections or serves an average of at least 25 people for at least 60 days a year" (EPA). They are labeled *public* because they are open to the public, not because of ownership. All utilities in the sample are community water systems (CWS), as opposed to transient non-community water systems and non-transient non-community water systems. All are owned by local governments as opposed to private, federal government, or state government owners. For more on the different types of ownership, please see the "GEFA-EFC Survey" section in the next chapter.

Of Georgia's 9.1 million residents (2005 US Census estimate), approximately 7.4 million people, or 82%, receive water service from PWS's. Most of the remaining fraction receives water from on-site, private wells. Nearly all of the PWS population is served by CWS's. The majority of CWS's are either owned by private companies (64%) or by local governments (35%), but the latter serve approximately 96% of the CWS population, or 7.1 million people. The utilities in the sample for this report serve about 6.9 million people, or about 93% of all Georgia residents using PWS's.

Unlike other states, Georgia does not regulate the rates of any water and sewer utilities. And, according to Georgia Code § 36-1-26, § 36-30-3, and § 36-80-17, the utilities have few restrictions for setting residential rates, except renewing the rates at least every 10 years. Rates are formulated by a team of utility staff members and/or consultants, which consists of engineers and/or financial professionals (Donahue 1996). The utility board then advises the team and approves the rates. Even if costs increase, the board may decide not to raise rates and obtain revenue from other sources, such as grants, revenue/in-kind services from other government departments (municipalities and counties only), and taxes. Since board members from municipalities and counties are elected – unlike their counterparts at utility commissions and authorities who are appointed – they may feel less inclined to raise rates because this may cause voters not to reelect them.

CHAPTER 3

METHODOLOGY

This chapter is divided into five sections which describe the regression models, data from a survey conducted by the Georgia Environmental Facilities Authority (GEFA) and the Environmental Finance Center (EFC) at the University of North Carolina-Chapel Hill (GEFA and EFC 2007), secondary data, sampling procedure, and tests for troubleshooting the regressions. The following bullets summarize the contents of each section.

- **Description of the models** – The first section describes the general form of the models, an example model showing how the average bills are estimated, the functional form of the models, the reasons for estimating two different types of models, the reasons for estimating at different consumption levels, and the basis and predicted effects of the independent variables.
- **GEFA-EFC survey data** – This survey contains the primary data for this report, and it was conducted in late 2006. The section describes the data collected from the GEFA-EFC survey, how they are processed, and resolutions for complex data.
- **Secondary data** – This data came from the following sources: Georgia Environmental Protection Division (Ga EPD), GEFA, United States Department of Agriculture (USDA), National Climatic Data Center (NCDC), SDWIS, and US Census. The context of the data and how they were processed is discussed.

- **Sampling procedure** – The fourth section covers the criteria for including utilities in the sample, the summary statistics for each model, and the distribution for each variable.
- **Regression diagnostics** – The fifth section explains how the models were determined through transforming the variables, running specification tests, examining the variance inflation factor for collinearity, and checking the adjusted R^2 value for best model fit. This section also shows tests for heteroscedasticity and normality of the residuals along with distribution plots of the residuals and scatter plots of the residuals versus the fitted values.

Description of the Models

This section describes the general form of the models, an example model showing how the average bills are estimated, the functional form of the models, the reasons for estimating two different types of models, the reasons for estimating at different consumption levels, and the basis and predicted effects of the independent variables.

General Model

Two types of OLS models were developed for Georgia: one for the water bill only ("water model"), and the other for the combined water and sewer bill ("W&S model"). Each model type estimates the bill at four different consumption levels: 3000, 6000, 9000, and 12000 GM. The modeled bills are for residential customers inside a utility's jurisdiction only. However, some utilities do not distinguish between residential and commercial classes, so the latter group can also be represented, especially at the 9000 and 12000 GM levels. The general form of each model is:

$$P = f\ (\text{cost of service factors, control factors})$$

where P is the monthly bill at a specific consumption level. The cost of service factors are defined as variables assumed to directly influence the cost, and thus, the rates. The control factors are defined as characteristics of the service area or variables that indirectly influence cost. Both models are created and analyzed through Stata 8.2.

Example

As stated earlier, one of the objectives of the models is to estimate the average bill for each utility. To illustrate how the models produce an estimate, let us examine a simple version of the general model where we assume only three factors influence the bill:

$$P_{model} = \alpha + \beta_1\,SOURCE + \beta_2\,POP + \beta_3\,POPDENS + \varepsilon$$

where P_{model} is the average water bill at 6000 GM, α is the constant (y-intercept) for the regression equation, β_n (n = 1, 2, 3) are the coefficients for the independent variables, SOURCE is a dummy variable representing two types of source water (groundwater = 1, surface water = 0), POP is the population served with water, POPDENS is the population density in persons per square mile, and ε represents the error term for the regression equation. After determining the coefficients and constant through OLS, the example equation becomes:

$$P_{model} = 9 + (4) * SOURCE + (0.01) * POP + (0.2) * POPDENS$$

Now, we can input the characteristics (source water, population, and population density) for any utility and calculate the average bill. If we input any utility's information into the equation, say Utility X (source water = 1; population = 1,000; population density = 50), we can find the average bill for a utility with the same characteristics as Utility X. Then, we can compare this estimate to Utility X's actual bill. After inputting Utility X's information, the previous equation looks like:

$$P_{model} = 9 + (4) * (1) + (0.01) * (1000) + (0.2) * (50)$$

$$P_{model} = \$33 \ (\pm \$4)$$

$$P_{actual} = \$40$$

$$P_{state} = \$25$$

For P_{model}, the ± $4 term is the 95% confidence interval based on the standard error of the estimate. P_{actual} is the actual bill for Utility X, and P_{state} is the simple (unconditional) average of 6000 GM water bills throughout the state. Compared to the model estimate, the actual bill is not only above the average for a utility with the same characteristics, but it is also above the upper bound of the confidence interval. However, the model estimate is between the state average bill and the actual bill. So, prior to using OLS, the comparison may have been between the state average and the actual bill, which is a difference of $15. But, when we compare Utility X to utilities with the same characteristics, the difference is actually $7. In fact, the actual bill is only $3 from the upper range of the average bill. Thus, Utility X could feel less anxious about having a relatively high bill.

Functional Form

Since few models have been developed for estimating the water and/or sewer bill, the functional forms of both models were determined through an iterative process. Continuous independent variables with skewed distributions were transformed to a variety of functions, and the models were changed to conform to tests for specification error, collinearity, and linearity between the predictors and the dependent variable at the highest possible R^2. More on regression diagnostics can be found in the final section of this chapter. In addition, controlling for spatial autocorrelation was beyond the scope of this report.

All variables in the models, regardless of significance, were used to estimate bills for each utility. This is a conservative approach because the 95% confidence interval for each utility's bill is larger than if it was produced from significant variables only. This increases the likelihood that the actual bill falls within the confidence interval and is considered "average."

Reasons for Two Models

To keep the analysis simple, it would have been preferable to use the W&S model only, but this would leave out a significant number of utilities (118) that use water only. Average bills for these utilities cannot be estimated with that model. Alternatively, the water model could have been used instead, but then significant information on the utilities with sewer service would be missing. Hence, both models are used. The water model is intended for all utilities that provide at least water service, but the W&S model is intended for utilities that provide both water and sewer services only. A model for the sewer bill only was not developed because only six of the 415 utilities in the sample provide sewer service only.

Reasons for Estimating at Multiple Consumption Levels

Even though average household consumption in Georgia[3] is 6000-7500 GM, it is useful to estimate the bills at different consumption levels for the following reasons. First, varying the consumption levels allows testing the sensitivity of the predictors and each utility's confidence interval for the average bill. It is important to know if a statistically significant factor at one consumption level remains significant at other consumption levels. Furthermore, a utility may have a bill that is, say, below the confidence interval at one consumption level and within the confidence interval at all other levels. Second, estimating at different consumption levels allows comparison by class, assuming that upper classes consume more than lower classes. However, a weakness in the models is that household income is recorded at the median and not at upper and lower percentiles. Third, consumption varies widely over the year with summer amounts sometimes reaching twice the amount consumed during winter. So, 3000 and 6000 gallons could be considered 'winter' amounts and 9000 and 12000 gallons could be considered 'summer' amounts. One limitation, however, is that the variables for temperature and rain remain at average annual amounts instead of estimated at their seasonal rates.

[3] The average per capita consumption in the US is estimated at 80-100 gallons per day Water Q&A: Water use at home, 2007, <http://ga.water.usgs.gov/edu/qahome.html#HDR3>.over the course of one year. Since the average household in both samples contains 2.5 persons per household (US Census 2000), the average consumption per household per month is approximately 6000-7500 gallons.

Basis and Predicted Effects of the Independent Variables

As stated in the previous chapter, the independent variables in the regression models are based on Raftelis (2005). Table 3 lists proxy variables in both models corresponding to the variables in Raftelis. Table 4 and the subsequent discussion provide more details on these and other variables.

Table 3: Variables in Both Models Corresponding to Raftelis (2005)

Variables in Raftelis (2005)	Variables in Both Models
Rate structures	Rate structures
Peak demand	Population with water
Treatment	Primary source
Government subsidization	Utility ownership
Grants	Grants, outside rates, commercial rates, connection fees
Age of system	Median year homes built

Some independent variables were excluded because the data were unavailable. This includes geography, number of commercial customers, infiltration and inflow, and rate-setting approach. In addition, other factors believed to be similar to, or in the spirit of, the variables mentioned in Raftelis were included in the models. While some variables from Hollman and Boyet (1975) and Thorsten et al. (2007) were not statistically significant, they were still included in the models because Georgia bills may have a different relationship to these variables. More research is needed in order to effectively conclude if a relationship exists between certain variables and the bills.

Table 4 lists all the variables in the water model, shows each variable's expected effect on the bill (positive or negative), and briefly explains this effect. The table also lists many, but not all, of the variables used in the W&S model (see Table 5 for the remaining variables). The W&S model includes most of the independent variables in the water model with a few adjustments to accommodate sewer factors. Following the table is a discussion of the interpretation of the dummy variables in the table and the predicted effects of the variables on the water and combined bill.

Table 4: Predicted Effects of the Independent Variables on the Water and Combined Bills

Independent Variables	Effect on Bill	Reasoning
Cost of Service Factors		
Interbasin transfer?	+	Expensive to transport water from different basins
Median year homes built	+ or –	Old util. need repair & replacement; New util. have high start-up cost
Utility ownership – Authority?	+	Board has less political motivation to keep rates low
Utility ownership – County?	+ or –	Organizational structure is similar to municipal system
Utility ownership – Municipality	N/A	N/A - Reference variable. Not in model.
Utility ownership – Utility commission?	+	Board has less political motivation to keep rates low
Population growth rate (% per year)	–	Costs are spread to more people (connections), so bills drop
Population density (persons per square mile)	–	Reduced pipeline and distribution costs
Population with water service	–	Costs are spread to more people (connections), so bills drop
Primary source – Groundwater (S. of Fall Line)	N/A	N/A - Reference variable. Not in model.
Primary source – Groundwater (N. of Fall Line)?	+	Groundwater from fractured rock is more expensive
Primary source – Purchased surface water?	+	Surface water needs more treatment than groundwater
Primary source – Surface water?	+	Surface water needs more treatment than groundwater
USDA & GEFA loans (1997-2006)?	+	Financing costs increase with loans
USDA & GEFA grants (1997-2006)?	–	Other utility income lowers bills for all residents
High charge for residents outside jurisdiction? (Outside rates)	–	Other utility income lowers bills for residents inside jurisdiction
Higher charge for taps > ¾"? (Commercial rates)	–	Other utility income lowers bills for all residents
Total water & sewer connection fees	–	Other utility income lowers bills for all residents
Sewer service?*	–	Shared utility administrative costs can lower bills
Control Factors		
Average annual rainfall (inches per year)	–	High rainfall could increase water supplies and reduce costs
Average annual temperature (°F)	+	High temperatures could lower water supplies & increases costs
Median household income ($ per year)	+	Wealthy persons consume more water and demand higher quality
Rate structure – Decreasing block?	–	Lower price than uniform block at the same consumption
Rate structure – Flat?	+ or –	Lower price at high consumption; but vice-versa
Rate structure – Increasing block?	+	Higher price than uniform block at the same consumption
Rate structure – Uniform block	N/A	N/A - Reference variable. Not in model.
Year of last rate change	+	Recent rate changes can keep up with inflation

*Not used in the W&S model

In the above table, all dummy variables are considered as "yes" or "no" observations (coded 1 or 0, respectively) and are identified by question marks at the end of the name. For example, *interbasin transfer?* is a dummy variable that records whether or not a utility imports water from another basin. In addition, categorical variables, such as *utility ownership*, are split into individual dummy variables with one variable removed to avoid perfect collinearity in the model. The table lists the reference variables left out of the model. For dummy variables not part of a categorical set, the predicted effects are relative to the dummy variable equaling zero. For example, regarding *interbasin transfer?*, the bill is likely to be more expensive (positive sign) if a utility imports water (*interbasin transfer?* = 1) than if it does not (*interbasin transfer?* = 0). For categorical sets, all predicted effects for the dummy variables in the model are in relation to the reference dummy variable. For instance, regarding *utility ownership*, authorities and utility commissions are expected to have higher bills than municipalities (the reference variable).

The following paragraphs discuss the predicted effects of the independent variables on the water and combined water and sewer bill from Table 4. For a detailed discussion on the manipulation of these variables and their sources, please see the "GEFA-EFC Survey Data" and "Secondary Data" sections.

Starting with the cost variables, *interbasin transfer* is a straight cost because any imported water would probably be more expensive than water within the basin because of transmission costs. The *median year homes built* variable is used as a proxy for the age of the system, which is a different hypothesis from Thorsten et al. (see Table 2). Since many utility distribution networks are constructed along with homes in the area, this variable was viewed as a suitable approximation. The sign is uncertain because older systems are expected to be expensive because of up-keep, but newer systems could be expensive because of high start-up costs. The *utility ownership* variable controls for both politics and government subsidization. Voters may express their disapproval of a rate hike by not reelecting the city councilmen or county commissioners. Regarding government subsidies, municipalities and counties have more opportunity to receive subsidies than authorities and utility commissions. Unlike Thorsten et al., who use the expected customer growth rate, the Georgia models use the historical *population growth rate* from the US Census's annual estimates for 2000-2005. It is assumed that utilities with recent high growth rates would have expanded their facilities to meet a long planning horizon, so as the population grows, the bills decrease due to economies of scale. *Population density* (used in Thorsten et al) and *population with water service* (used in Hollman and Boyet) represent the cost of distribution, pipelines, and other system assets. As these variables increase, the cost will also increase, but the bill should decrease due to economies of scale. The *primary source* variables attempt to capture differences in treatment levels, and thus cost, with groundwater south of the Fall Line[4] being the least expensive. Debt is another variable in Thorsten et al.; however, not all debt information was available in Georgia, so *loans* from USDA and GEFA were used instead. *USDA & GEFA grants*, *outside rates*, *commercial rates*, and *connection fees* are all potential sources of subsidization for residential bills inside a utility's jurisdiction. Raftelis (2005) only discusses the effect of grants on the rates and does not mention other

---

[4] The area separating the Coastal Plane from the Piedmont is known as the Fall Line, which is about 20 miles wide and was the shoreline of the Atlantic Ocean during the Mesozoic Era. See the section "Secondary Data" for details. This section also describes why the groundwater south of the Fall Line is expected to be the least expensive source.

sources of revenue. However, it is assumed that *outside rates*, *commercial rates*, and *connection fees* would all follow suit. Note that the expected effect for *connection fees* listed in Table 4 is different from Hollman and Boyet's hypothesis, even though they found that the variable was significant with the opposite sign. Since the study focused on rural Mississippi and contained fewer variables, the sign may be different for Georgia, and the alternative hypothesis will be tested. The *sewer service* variable is used only in the water model and represents the potential administrative cost savings for combining both water and sewer services.

Concluding with the effects of the control factors, it is assumed that high *temperature* regions would have less water available because of high yearly usage and evapotranspiration, which would increase the costs and rates. Conversely, areas with high *rainfall* would have more water available, which would decrease the costs and rates. Also, utilities serving areas with a large *median household income* may feel less inhibited to maintain low rates. The *rate structure* variables control for the effects that different block rates would have on the overall bill. Finally, the *year of last rate change* is important to record because more recent rate changes are likely to be keeping up with inflation.

The W&S model includes most of the variables listed in the previous table. However, some variables are modified to account for the interaction of water rate structures, water rates for customers outside the utility's jurisdiction, and commercial water rates with their sewer counterparts. In addition, the *sewer service* dummy variable is dropped from the W&S model because all the utilities in this model have sewer service. For a discussion on the interactions, see the section “GEFA-EFC Survey Data”.

The following table lists the modified variables for the W&S model.

Table 5: Predicted Effects of the Modified Independent Variables on the Combined Bill

Independent Variables	Effect on Bill	Reasoning
Rate structure		
Decreasing block - Water & sewer?	–	Lower price at the same consumption
Decreasing for water & uniform for sewer?	–	Lower price at the same consumption
Flat - Water & sewer?	+ or –	Lower price at high consumption; but vice-versa
Increasing block - Water & sewer?	+	Higher price at the same consumption
Increasing for water & uniform for sewer?	+	Higher price at the same consumption
Uniform block - Water & sewer?	N/A	N/A - Reference variable. Not in model
Uniform for water & decreasing for sewer?	–	Lower price at same consumption
Uniform for water & flat for sewer?	+ or –	Lower price at high consumption; but vice-versa
Uniform for water & sewer cap?	+ or –	Lower price at high consumption; but vice-versa
Outside rates		
Water only?	–	Other utility income lowers bill for residents inside jurisdiction
Water & sewer?	–	Other utility income lowers bill for residents inside jurisdiction
None	N/A	N/A - Reference variable. Not in model
Commercial rates		
Sewer only?	–	Other utility income lowers bill for residents inside jurisdiction
Water only?	–	Other utility income lowers bill for all residents
Water & sewer?	–	Other utility income lowers bill for all residents
None	N/A	N/A - Reference variable. Not in model

GEFA-EFC Survey Data

The primary data for this report comes from a rate survey conducted during October 2006-January 2007 by GEFA and EFC. The purpose of the GEFA-EFC survey was to provide Georgia's public utilities a general idea of how their rates and connection fees compare to their fellow public utilities across the state. A cover letter requesting this information (Appendix A) was sent to approximately 570 public utilities found in databases from SDWIS, Clean Watersheds Needs Survey (CWNS), National Pollutant Discharge Elimination System, 2002 Census of Governments, Georgia Municipal Association, and Georgia Association of Water Professionals. About 110 did not respond, 40 were no longer in operation or had consolidated, and 415 provided their rate and connection fee schedules. Along with each utility's bill at 6000 GM, the survey provides summary statistics of the minimum charges by number of accounts, different types of rate structures, amount of connection fees, bills per month over various ranges of consumption for residential and commercial consumers, and bills by utility ownership and river basin. The following table shows the participating utilities by type of service and utility ownership.

Table 6: Number of Utilities Participating in GEFA-EFC Study by Service and Utility Ownership

Utility Ownership	Service Type			Total
	Both Water & Sewer	Water Only	Sewer Only	
Authorities	19	18	1	38
Counties	18	12	0	30
Municipalities	245	86	5	336
Utility Commissions	9	2	0	11
Total	291	118	6	415

As demonstrated in the table, a maximum of 409 out of 415 (six utilities provide only sewer service) can be analyzed for the water model. Similarly, a maximum of 291 utilities (124 provide only one service) can be analyzed for the W&S model. As will be seen in the section, "Sampling Procedure," the total number of utilities analyzed in both models will decrease slightly because of missing data or a dummy variable with 1 or 2 observations was dropped.

The survey provided data on all of the dependent variables – the water bills and combined water and sewer bills – as well as the following independent variables: *outside rates*, *commercial rates*, *sewer service*, *connection fees*, *utility ownership*, *rate structures*, and *year of last rate change*. The dependent variables were based on the components of the rate schedules, which were entered in a Microsoft Access database, and an Excel macro calculated the bills. Some utilities charge on a bimonthly and/or cubic feet basis, but for simplicity, the bills for all utilities were standardized to gallons per month. The following subsections discuss the independent variables and the logic behind the manipulation of the data.

Outside Rates, Commercial Rates, Sewer Service, & Connection Fees

Outside rates is a variable that records if a utility charges a high rate for residents outside a utility's jurisdiction compared to those inside the jurisdiction. Some utilities charge higher rates because those areas cost more to serve and/or the residents do not pay local taxes which subsidize the system. For simplicity, *outside rates* is a dummy variable instead of the actual value of the outside bill, which would need to be calculated at every consumption level. The variable *commercial rates* is a dummy for whether a utility charges different rates for commercial consumers. The variable records utilities that use separate rate structures for different consumer classes or charge different rates based on tap size, of which 3/4" is very common for residential consumers. The *sewer service* variable is used in the water model only and simply records if a utility provides sewer in addition to water. The *connection fees* variable accounts for charges to homes that tap into the water and/or sewer system for the first time. The fees include the installation, or tap, fee as well as the non-installation charge – also known as system development charge, impact fee, assessment fee, etc – which offsets the cost of new system assets, such as water towers and treatment plants. Sewer connection fees were included in the water model because the total revenue from both services may subsidize residential water rates.

Utility Ownership

The *utility ownership* categorical variable contains four types of local government systems: authorities, counties, utility commissions, and municipalities. All systems have a board of directors and technical staff. In general, the staff manages the day-to-day operations of the utility and reports to the board of directors for advice and approval of plans, rates, etc. W&S rates are unregulated across the state.

County and municipal systems have similar organizational structures where the county commissioners and city councilmen, respectively, act as the boards of directors. The board members may feel less inclined than board members for authorities and utility commissions – their appointed counterparts – to raise rates because voters may not reelect them. These systems can take advantage of administrative services from other departments within their governments. It is also possible that some revenue from the system could subsidize the government departments (Raftelis 2005). County and municipal systems can collect revenue from property taxes and sales taxes like the special purpose local option sales tax (SPLOST).

Authorities and utility commissions are the semi-autonomous counterparts to the county and municipal systems, respectively. The board members for authorities and utility commissions are appointed by the local county commissioners and city councilmen, respectively. Unlike their counterparts, the only taxes available to authorities and utility commissions are from SPLOST. Authorities are created by acts of the Georgia state legislature, while utility commissions are created by municipalities. One key difference between authorities and utility commissions is that the latter usually provides other utility services, such as power, gas, and/or cable. Authorities provide water and/or sewerage only.

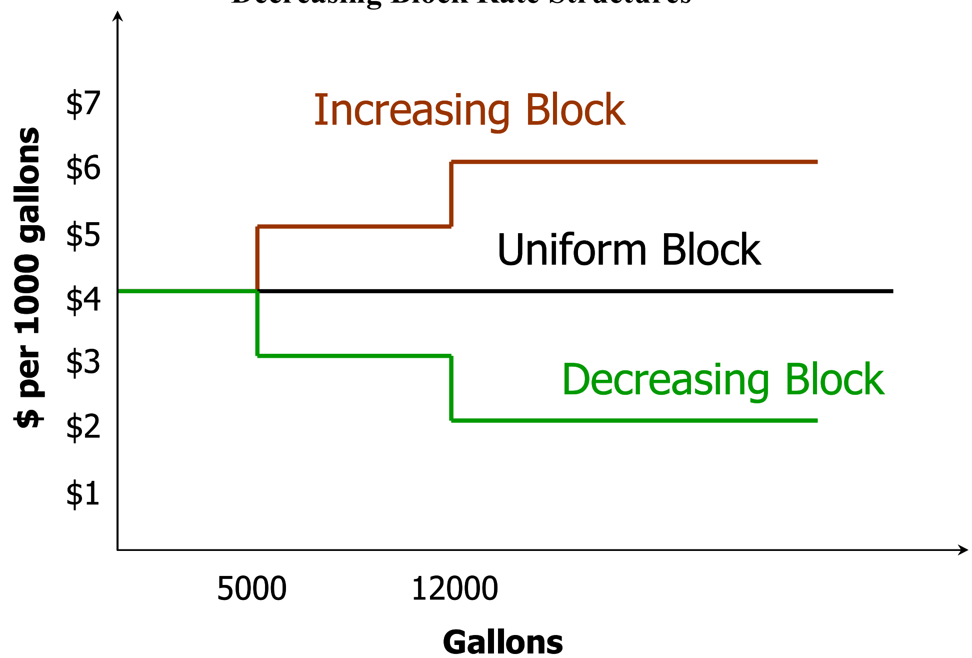
Ownership was classified primarily through the name of the utility. For example, the ownership of the following utilities was easy to identify: Alcovy Shores Water and Sewer *Authority*, Bartow *County* Water Department, *City* of Atlanta Department of Watershed Management, and Fitzgerald Water Light and Bond *Commission*. Utilities with ambiguous names were called to verify their ownership type.

Rate Structures

The typical residential bill has two parts: 1) a fixed minimum fee with zero or a small consumption allowance, and 2) a charge per thousand gallons. The latter is represented by the *rate structures* categorical variable, and it notes, as the consumption increases, if the charge per thousand gallons increases (increasing block), decreases (decreasing block), or stays the same (uniform block). For example (refer to Figure 1 below), if a utility charges $4 per thousand gallons for consumption between 0 and 5000

GM, and then charges $5 per thousand gallons for consumption between 5000 and 12000 GM, this would be an increasing block rate. If the second block was charged at $3 per thousand gallons instead of $5 per thousand gallons, then this would be a decreasing block rate. And if the second block remained at $4 per thousand gallons, then this would be a uniform block rate.

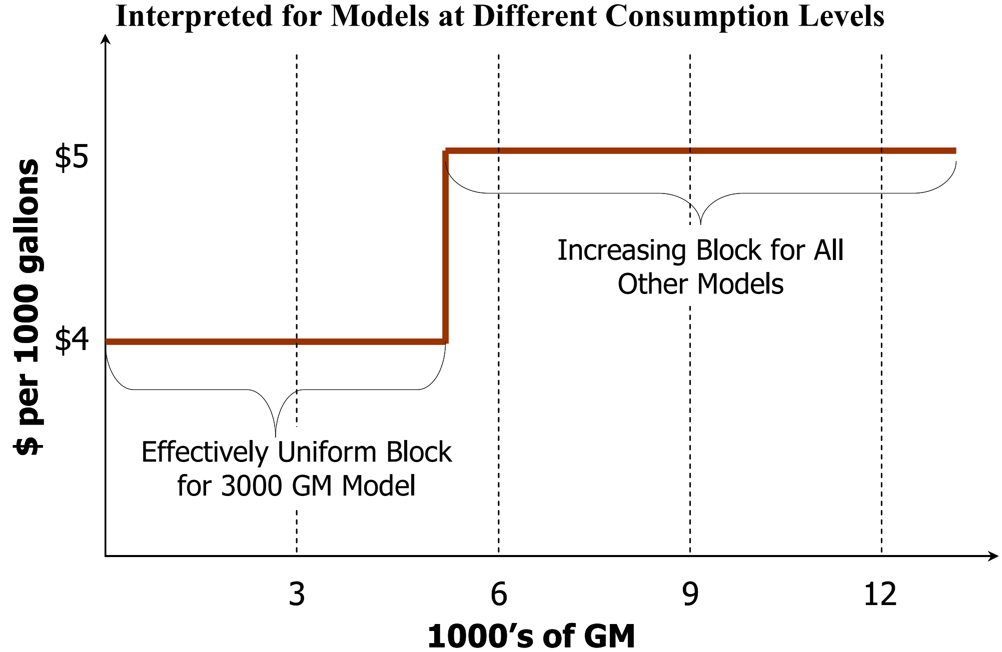
Figure 1: Graphic Example of Increasing, Uniform, and Decreasing Block Rate Structures



The *rate structures* variable also records if a utility charges a basic, or flat, fee for unlimited consumption. In addition, a number of utilities cap their charge for sewer service at a specific amount, so these utilities were given their own category. In other words, some utilities have a rate structure applied to sewer service, but residents are not assessed a per thousand gallon charge above a certain amount.

There are two ways to manipulate the data for the water and sewer *rate structures* categorical variables: use only the <u>effective</u> rate structure up to the consumption modeled or use the rate structure for infinite consumption. For example (refer to Figure 2 below), say Utility X has an increasing block rate structure where the first block is $4 per thousand gallons of water between 0 and 5000 GM, and the second block is $5 per thousand gallons for consumption above 5000 GM. For the model at 3000 GM, Utility X's <u>effective</u> rate structure is uniform block because the charge per thousand gallons does not change from 0 to 3000 GM. But, for the model at 6000, 9000, or 12000 GM, Utility X's effective rate structure is increasing block.

Figure 2: Graphic Example of How Rate Structures May be Interpreted for Models at Different Consumption Levels

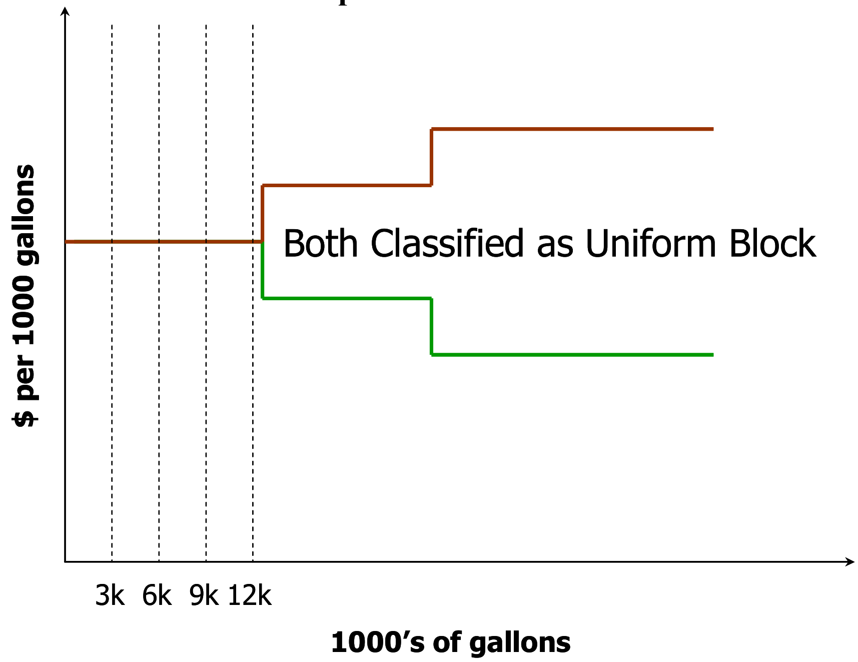


The hypothesis for the effective rate structure is that the bill may not be affected if the blocks do not change in the modeled consumption range. This is especially true as some utilities define blocks over very large consumption ranges as a way to price discriminate between residential, commercial, and industrial classes under one, instead of many rate, structures. So, the second block may start at consumption levels far beyond residential consumption. An alternative hypothesis is that a utility with an increasing block rate would have higher overall bills at all blocks; and vice-versa for decreasing blocks. This is supported by Mann's (1970) principal component analysis.

Using effective rate structures causes a significant problem as the number of observations for each rate structure dummy variable would change at different consumption levels. Comparing the variables' coefficients and p-values at all consumption levels would not be as "clean" as comparing the variables with consistent observations, especially when the observations for other variables are not changing. For example, for the water model at 3000 GM, most utilities effectively have uniform rates and a few others have increasing, decreasing, or flat rates. As the modeled consumption level increases, the number of utilities with uniform blocks would decrease and those with the other three structures would increase. Meanwhile, the other independent variables in the models would remain the same at all consumption levels.

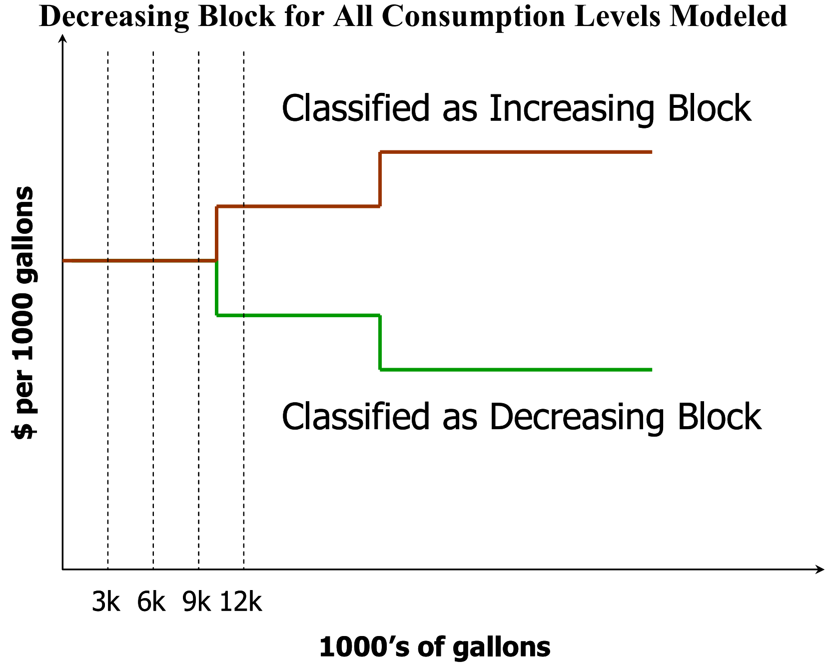
To simplify the analysis, yet account for price discrimination by some utilities, the models use a blend of the two methods for manipulating rate structure data. As shown in Figure 3, if a utility's second block begins at a consumption level greater than 12000 GM – the maximum residential consumption level modeled – then the rate structure is defined as uniform.

Figure 3: Rate Structures Classified as Uniform Block for All Consumption Levels Modeled



However, as Figure 4 shows if the second block begins at a level lower than 12000 GM, then the rate structure is defined as increasing or decreasing regardless of the consumption level modeled.

Figure 4: Rate Structures Classified as Increasing or Decreasing Block for All Consumption Levels Modeled



From the figure above, we see that, even though the block rate first changes after 9000 GM, the rate structures in the 3000, 6000, and 9000 GM models are classified as either decreasing or increasing, even though they are effectively uniform. As an effective rate structure, this is false, but the alternative hypothesis mentioned above may hold and is tested in this report. Future researchers may want to use effective rates structures at all consumption levels modeled, but this is beyond the scope of this report. In all, the rate structures for 47 utilities changed from increasing or decreasing to uniform because the second block was greater than 12000 GM. Specifically, 38 utilities had the water and/or sewer rate structure changed to uniform while nine utilities had only the sewer rate structure changed.

This same process is applied to caps on sewer charges. If a utility's sewer cap is above 12000 GM, then it is recoded as an increasing, decreasing, or uniform block. In all, 26 utilities in the sample had sewer caps, but only 10 had caps below 12000 GM.

Year of Last Rate Change

The *year of last rate change* variable records the year, up to 2007, in which the rate schedule became effective. Many of the utilities list the exact day the rate schedule went into effect; however, several others list only the month and some list just the year, so the effective year was used in the models. Approximately 28 utilities did not identify the effective year. Considering there are, at most, 409 utilities in the water model and 291 in the W&S model, this is a non-trivial amount of missing data (7% and 10% missing, respectively). Dropping the utilities with missing effective dates from the analysis may bias the results. An alterative is to impute the missing values.

Traditional methods of imputation, beginning with the simplest and least precise, include mean substitution, dummy variable adjustment, and multiple regression (conditional mean substitution). For mean substitution, the average of the variable containing the missing data replaces the missing values. Dummy variable adjustment (DVA) goes one step further by including a dummy variable coded 1 for missing data and 0 otherwise. This variable provides more information on how missing years, in relation to the mean year, effects the bill. Conditional mean substitution uses the effective year as the dependent variable in a regression equation with the same independent variables in the water and W&S models. Then, the equation can be used to predict the missing data in *year of last rate change*. All three methods provide values for missing data, but they bias the coefficients on the predictors and underestimate the standard errors in the ultimate regression models. More advanced methods, like maximum likelihood estimation and multiple imputation, are available, but they are beyond the scope of this report.

If the options are removing the sets of observations (the utilities) or using the traditional methods, Allison (2001) implicitly recommends the former. However, for this study, mean substitution was used because there was little difference in the coefficients and standard errors between the models with imputed data and the models without them. And since one goal of this report is to estimate the average bills for as many utilities as possible, mean substitution was preferable. The other traditional methods can probably close the gap, but it is negligible and not worth either the cost in efficiency due to the additional dummy variable in DVA and the time spent developing eight regression models for conditional mean substitution.

The most likely reason there was little difference in the estimates between the models with imputed data and the models without them is that the effective year for 90% of utilities was in the last six years. Since the range is narrow, substituting the mean for the missing values seems to be an adequate approximation. The following table shows the percent of utilities that last changed their rates in the corresponding year or later. Note that half of all rate schedules were effective for 2006 and 2007, and the average effective year was 2004.

Table 7: Percent of Utilities with Rate Change in Corresponding Year or Later

Percent Utilities that Changed Rates in Effective Year or Later	Effective Year
99%	1989
95%	1997
90%	2001
75%	2004
50%	2006
5%	2007

Interactions for the W&S Model

There are two ways to include the sewer outside charges, large tap charges, and rate structures for the W&S model: create a duplicate set of sewer dummy variables that mirror the water dummy variables, or interact the set of water variables with the set of sewer variables. Both methods make the W&S model less efficient as the first produces collinearity and the second would add more variables than the first method, thus increasing the length of the confidence intervals. In addition, with the second method, the sample would decrease due to the sample criteria requiring rare combinations of water and sewer dummy variables to be excluded from the analysis (see the section "Sampling Procedure"). Ultimately, the second method was chosen because the coefficients of the interactions are more informative on the overall utility operation. For example, if water and sewer variables for increasing blocks were separately included, both variables might not be significant. However, if a single variable represented utilities that had increasing blocks for both water and sewer, then it might be significant. This was observed in trial runs for both methods, and an F-test on both the dummy variable representing increasing water blocks and a dummy representing increasing sewer blocks was not significant.

Secondary Data

The secondary data came from the following sources: Georgia Environmental Protection Division (Ga EPD), GEFA, USDA, National Climatic Data Center (NCDC), SDWIS, and US Census. The following table lists the variables from each source.

Table 8: Data Sources for Independent Variables

Sources	Variables
Ga EPD	Interbasin transfer?
GEFA & USDA	Grants Loans
NCDC	Average annual rainfall Average annual temperature
SDWIS (Ga EPD)	Population with water service Primary water source
US Census	Median year homes built Median household income Population growth rate Population density

The next subsections discuss the context of the data and how they were processed.

Interbasin Transfers

The information on interbasin transfers came from Ga EPD, which permits and tracks withdrawals of surface water and groundwater for all purveyors. Their data tables showed the amount of water transferred by specific utilities and then listed utilities receiving the transfer. However, the amount of water transferred was not broken down by receiving utilities, so a dummy variable represents them. Also, data on permitted and actual withdrawals were not used because the amount of water sold would be a better indicator for modeling the water bill, and the municipal withdrawals are for both commercial and residential consumption.

Loans and Grants

Loans and grants were limited to those from USDA and GEFA. The USDA Rural Development Utilities Program provides grants and loans to publicly-owned water and sewer utilities that serve less than 10,000 people. Of the 415 utilities in the GEFA-EFC survey, 318 have less than 10,000 people. GEFA maintains the Drinking Water State Revolving Fund, Clean Water State Revolving Fund, and other construction and emergency grant and loan programs for publicly-owned water and sewer utilities. The data used in the models were based on loans and grants executed over the past 10 years (1997-2006) for water and sewer projects. While some funding was specifically for one service, many loans and grants were listed for both services. Thus, loans and grants used in the water model include those for sewer projects. Since a large share of utilities did not receive loans and grants from either agency (96 received grants and 201 received loans), the information was represented as dummy variables.

While GEFA and USDA provide a large share of funding in Georgia, non-current liabilities – the amount of long-term debt remaining – considers other sources of financing, such as bonds. Non-current liabilities are recorded on local government audits, which are collected by the Georgia Department of Audits and Accounts. Unfortunately, each audit is in portable document format (PDF) and not in one convenient database. Moreover, not all local governments have submitted their audits for fiscal year 2005. The 10-year timeframe was chosen in order to capture loans still in payment. Naturally, loans older than 10 years will still be in payment; however, it was decided not to extend the timeframe because a large fraction of each loan would already be paid.

Rainfall and Temperature

Rainfall and temperature data were based on climate normals, or averages, for the years 1971 to 2000 (NCDC 2002), which is the most recent period for processed climate data. For the climate summary, NCDC recorded average rainfall and temperature for each month in the period, fixed inconsistencies in the observations due to equipment malfunction or other reasons, cross-checked the data for one weather station against stations in the vicinity, and calculated missing values. Considering the timeframe and the standard of accuracy, of the 200 or so active weather stations in Georgia, rainfall normals were calculated for 154 weather stations and temperature normals were calculated for 96 weather stations. All of the weather stations in the climate summary that recorded temperature data also recorded rainfall data. NCDC makes available recorded, raw climate data for all weather stations, but it was decided to use NCDC's climate summary as this is more accurate.

Since not all of Georgia's 530 cities and towns have a weather station, each city with a utility was assigned to the closest station using the great circle distance formula, which is the shortest distance between two points on a sphere (Meeus 1991):

$$d = 3963 * \arccos[\sin(lat1*\pi/180) * \sin(lat2*\pi/180) + \cos(lat1*\pi/180) * \cos(lat2*\pi/180) * \cos(lon2*\pi/180 - lon1*\pi/180)]$$

where *d* is the distance between a city and a weather station, *lat1* and *lon1* are the latitude and longitude for a city, *lat2* and *lon2* are the latitude and longitude for a weather station, and arccos, sin, and cos represent the trigonometric functions arccosine, sine, and cosine, respectively. The value 3963 is the radius of the earth in miles, and π/180 is the conversion from degrees to radians. NCDC lists the coordinates of each weather station and the Geographic Names Information System from USGS provides the coordinates for each city. Since authorities and counties serve a number of different cities, the coordinates chosen were based on the cities in their mailing addresses.

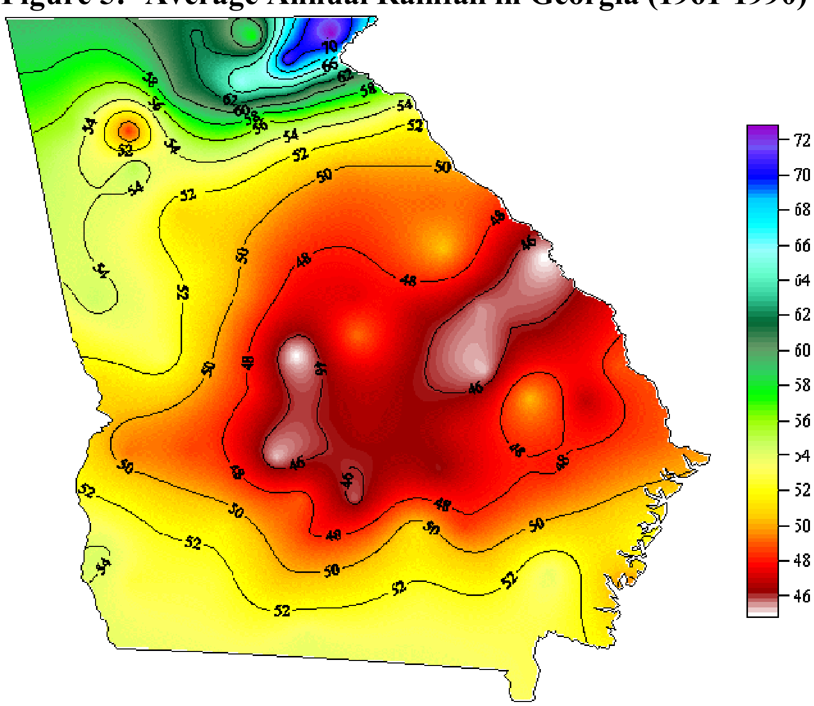
The mean and median distance between each utility and its closest weather station with rain data was 7.3 miles. The mean distance between each utility and its closest station with temperature data was 9.6 miles, and the median distance was 10.3 miles. The following table shows the distribution for the distances between each utility and its closest weather station. For instance, ninety percent of all utilities are within 14.2 miles of a weather station with rainfall data and 16.6 miles from a weather station with temperature data.

Table 9: Distribution of Distances between each Utility and Its Closest Weather Station

Percent Utilities within X Miles	Rainfall (miles)	Temperature (miles)
1%	0.2	0.4
5%	0.6	0.9
10%	0.9	1.5
25%	2.3	4.2
50%	7.3	10.3
75%	11.1	13.8
90%	14.2	16.6
95%	16.0	18.1
99%	19.9	22.5

Unsurprisingly, utilities are closer to the weather stations with rainfall data than stations with temperature data. However, the real question is the accuracy of using climate data from weather stations relatively far from utilities. A better measure would be to assign utilities to the closest weather station in the same area between two rainfall/temperature isograms. The following figure is an example of such a graph.

Figure 5: Average Annual Rainfall in Georgia (1961-1990)



Source: Georgia Automated Environmental Monitoring Network

Unfortunately, no data was available for assigning utilities to areas between isograms. At any rate, it is noted that 90% of the utilities are within approximately 15 miles of a weather station, and the data from those stations may be adequate for the purposes of this report.

Population and Source Water

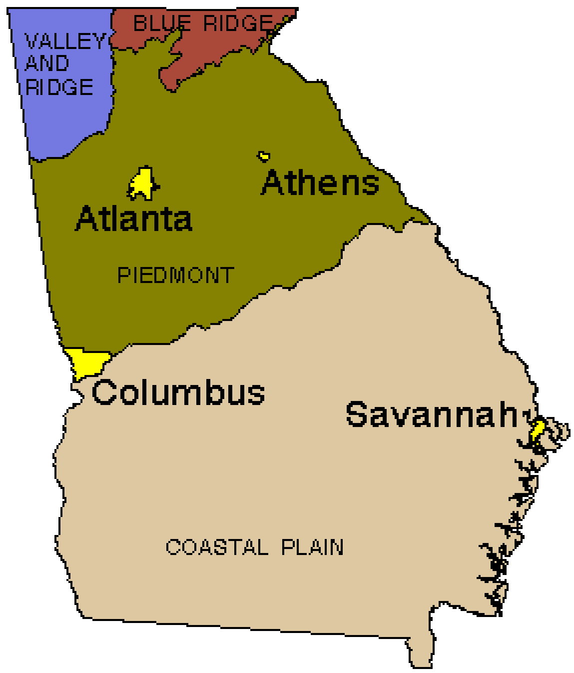
The SDWIS database provided information on the number of people served with water and the primary sources of water for each utility, which are different varieties of groundwater and surface water. While the EPA maintains SDWIS for the whole country, each state's environmental division collects and submits the data to EPA. This report uses a more up-to-date version (November 2006) of Georgia's SDWIS information from Ga EPD.

Regarding the data, the population with water service recorded in SDWIS was used to estimate the entire service population, even though the number of people with sewer service may be different. The only central database found that recorded population with sewer service was CWNS; however, much of this data was last updated in 1999. Alternatively, the US Census population estimates could have been used, but not all service areas match the political boundaries of counties and cities.

SDWIS records the following types of primary source water: groundwater (GW), purchased groundwater (GWP), groundwater under the influence of surface water (GU), purchased groundwater under the influence of surface water (GUP), surface water (SW), and purchased surface water (SWP). However, as will be described in the next section, GWP, GU, and GUP were not used in the models because they had few observations.

GW was split into two dummies because groundwater from the sediments in the Coastal Plane is easier, and less costly, to extract than groundwater from the crystalline rock in northern Georgia, which is predominantly in the Piedmont. SW and SWP did not need modifications because these sources are located in northern Georgia. The following figure illustrates the major physiographic regions in Georgia. The yellow areas identify a few major cities.

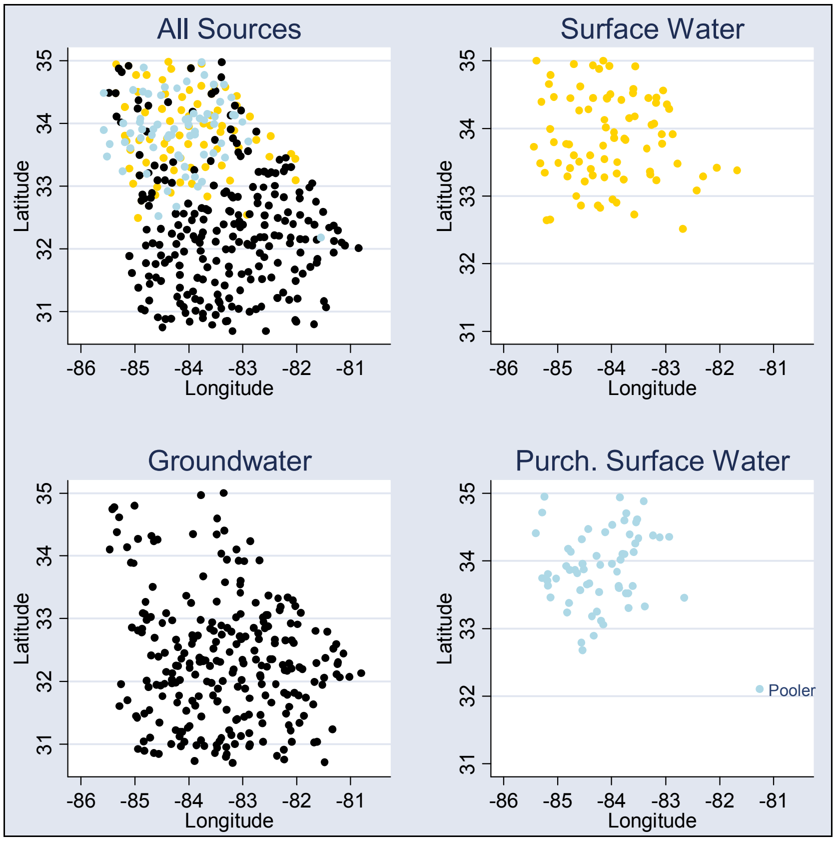
Figure 6: General Physiographic Regions of Georgia



Source: United States Geological Survey

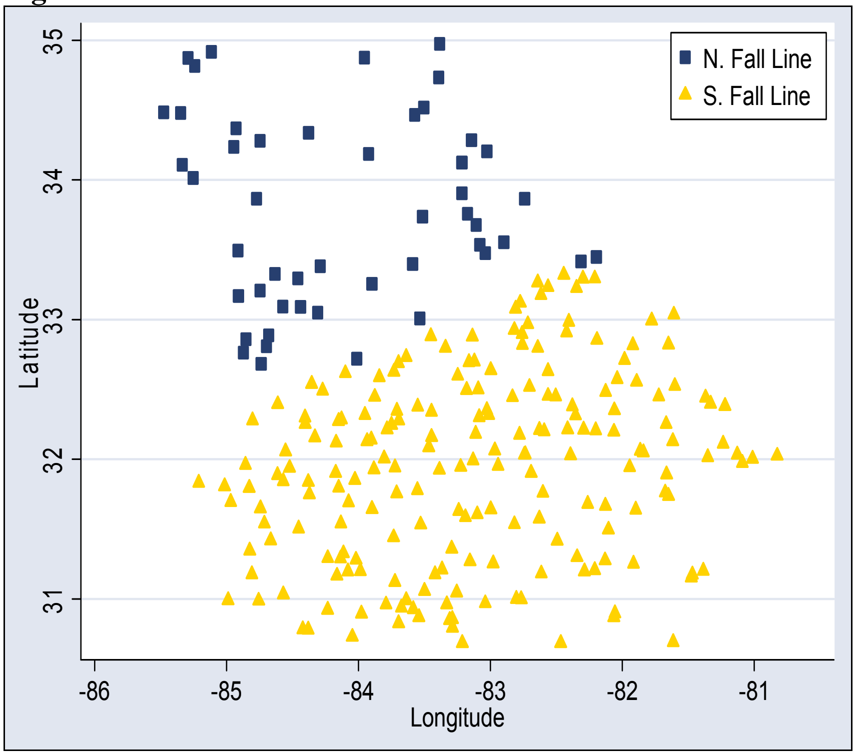
The line that separates the Coastal Plane from the Piedmont is known as the Fall Line, which is about 20 miles wide and was the shoreline of the Atlantic Ocean during the Mesozoic Era. The following figure shows the relative location of utilities by source water.

Figure 7: Location of Utilities According to Primary Source Water



Notice that nearly all utilities with SW and SWP are located above the Fall Line. A few utilities with SW lie just below the Fall Line, but the major exception is Pooler, which is labeled in the graph for SWP. Pooler is on the coast, and other utilities there use surface water, such as Savannah, but not as a primary source. Also notice that utilities with GW are primarily in southern Georgia. The following figure shows the location of utilities with GW north and south of the Fall Line.

Figure 8: Groundwater Utilities North & South of the Fall Line



Considering there are several utilities with GW north of the Fall Line, it was decided to split the variable for GW into two dummies.

Census Data

The variables *median household income*, *median year homes built*, *population growth rate*, and *population density* came from the US Census Bureau's 2000 survey and its estimates for years up to 2005. All of the data are based on the political boundaries for each city and county. While the boundaries do not precisely match the utilities' service areas, the data was the best approximation available. All the variables, except population growth, came from the 2000 US Census. The population growth rate was based on the US Census estimates of annual population[5] for 2000-2005.

[5] The 2000 US Census survey records population up to 4/1/2000. The US Census population estimates cover fiscal years begining 7/1/2000; therefore, the two population values for 2000 will differ.

Sampling Procedure

This section describes the sampling criteria, summary statistics, and distributions for each variable.

Criteria

The general criteria for the sample are that:

1) **Utilities must provide at least water service** – The water model requires utilities with water and/or sewer service, and the W&S model requires utilities with both services. Utilities with sewer service only (n = 6) cannot be modeled.
2) **Categorical variables must have enough observations for each category (n ≥ 2)** – Variables with one or two observations add little information to the model and increase the size of the confidence intervals for the model estimates.
3) **Utilities must have data for all variables in the models** – Estimates from the model cannot be generated from a sample where some utilities have missing information. The software package, Stata 8.2, automatically excludes utilities from the sample if their data is missing for at least one variable.
4) **Rate structures should be consistent throughout the modeled consumption levels and throughout the year** – Rate structures should be convenient to model and not change (i.e., increasing block to decreasing block) at different consumption levels between 3000 and 12000 GM. Also, the models use bills that are consistent throughout the year, so the rate structures should not change according to season.

Based on the criteria, the sample of utilities dropped from a maximum of 415 to 391 for the water model and 269 for the W&S model. The following paragraphs discuss how the sampling criteria were applied. Table 10 summarizes the number of utilities removed from the model samples and the respective reasons.

For the first criterion, six utilities provide sewer only, so they were removed from both models. For the W&S model, 118 utilities provide water only and they were removed from the model sample.

With the second criterion, one utility was dropped because it was unique in having "incremental rates" – blocks with flat charges regardless of consumption (e.g., the utility charges a flat $4.56 for any consumption between 2000-8000 gallons). For the W&S model, six utilities were dropped because they had an unusual combination of water and sewer rate structures not shared by others (i.e., increasing water block rates and flat sewer rates).

Since there were few utilities in the sample with GWP, GU, and GUP (2, 4, and 1, respectively), the utilities with GWP and GUP were dropped from the analysis and GU was combined with GW. GU is still filtered through the ground, so the treatment cost might be comparable to regular groundwater. Most utilities use one type of source water. However, SDWIS breaks down some utilities into multiple records that show different source waters. This occurred for 10 utilities, and the source water with the highest service population was chosen.

For the third criterion, utilities with no source water data or census data were removed from the sample. Also, as mentioned in the previous section, the variable *year of last rate change* had a significant portion of missing data for which the values were imputed.

Regarding the last criterion, some utilities use seasonal rates where the charge per thousand gallons increases during the summer. Additionally, some utilities base their summer rates on a multiple of the household's winter consumption. Since the bill at a specific consumption level varies throughout the year, utilities with seasonal rates were not modeled. Three other utilities were dropped because their water rate structures had both increasing and decreasing blocks within the consumption range of 3000-12000 gallons per month, and the rate structures could not be conveniently modeled.

The following table summarizes the number and reason for decreasing the sample.

Table 10: Number of Utilities Removed from Model Samples

Criterion	Number of Utilities Excluded from Sample	Reason
Water Model		
1	6	Serve sewer only
2	1	Uses incremental rates
2	2	Use GWP
2	1	Uses GUP
3	4	No source water data
3	1	No census data
4	6	Use seasonal rates
4	3	Use both increasing and decreasing rates
	24	TOTAL
Water & Sewer Model		
1	6	Serve sewer only
1	118	Serve water only
2	1	Uses incremental water & sewer rates
2	1	Uses increasing water rate & flat sewer rate
2	1	Uses uniform water rate & increasing sewer rate
2	1	Uses increasing water rate & sewer cap
2	1	Uses decreasing water rates & sewer cap
2	2	Use decreasing water rate & flat sewer rate
2	2	Charge higher sewer outside rates
2	2	Use GWP
2	1	Uses GUP
3	2	No source water data
3	1	No census data
4	5	Use seasonal water & sewer rates
4	2	Use both increasing & decreasing water rates
	146	TOTAL

The following table shows summary statistics of the data for the water model.

Table 11: Summary Statistics – Water Model

Independent Variables	No. of Utilities	Mean or % of Sample	Standard Deviation
Water Bills			
Bill @ 3000 GM	391	$14.22	$5.68
Bill @ 6000 GM	391	$21.39	$8.19
Bill @ 9000 GM	391	$28.75	$11.40
Bill @ 12000 GM	391	$36.22	$14.96
Cost of Service Factors			
Interbasin transfer?	32 of 391	8.2%	27.4%
Median year homes built	391	1972.8	8.6
Utility ownership – Authority?	35 of 391	9.0%	28.6%
Utility ownership – County?	26 of 391	6.6%	24.9%
Utility ownership – Municipality?	320 of 391	81.8%	38.6%
Utility ownership – Utility commission?	10 of 391	2.6%	15.8%
Population growth rate (% per year)	391	1.5%	3.1%
Population density (persons per square mile)	391	598	540
Population with water service	391	15,153	58,238
Primary source – Groundwater (S. of Fall Line)?	197 of 391	50.4%	50.1%
Primary source – Groundwater (N. of Fall Line)?	52 of 391	13.3%	34.0%
Primary source – Purchased surface water?	62 of 391	15.9%	36.6%
Primary source – Surface water?	80 of 391	20.5%	40.4%
USDA & GEFA loans (1997-2006)?	92 of 391	23.5%	42.5%
USDA & GEFA grants (1997-2006)?	193 of 391	49.4%	50.0%
Outside rates?	168 of 391	43.0%	49.6%
Commercial rates?	174 of 391	44.5%	49.8%
Total water & sewer connection fees	391	$1,566	$1,834
Sewer service?	277 of 391	70.8%	45.5%
Control Factors			
Average annual rainfall (inches per year)	391	50.3	4.8
Average annual temperature (°F)	391	62.7	3.1
Median household income ($ per year)	391	$31,009	$9,474
Rate structure – Decreasing block?	25 of 391	6.4%	24.5%
Rate structure – Flat?	10 of 391	2.6%	15.8%
Rate structure – Increasing block?	81 of 391	20.7%	40.6%
Rate structure – Uniform block	275 of 391	70.3%	45.7%
Year of last rate change	391	2004.3	3.5

The most common utilities in the water model are owned by municipalities, serve groundwater south of the Fall Line, do not receive water from outside their basin, provide sewer service, charge uniform block rates, and do not charge higher rates to outside and commercial customers. Also, an average utility serves 15000 customers (median =

2500), and the service area generally has 600 persons per square mile (median = 470) with a population growth rate of 1.5% (median = 0.67%). For households, median income is $31,000 (median = $29,000), and 1973 is the median year homes were built. As far as climate, from 1971 to 2000, the average utility had rainfall of 50 inches per year and an average temperature of 63 degrees.

Regarding finances, over the last 10 years, 24% of the utilities have received grants and 49% have received loans. The average total connection fees for both water and sewer are about $1600 (median = $800). Also, the average water bill at average consumption (6000 GM) is $21. The *year of the last rate change* variable includes imputed values, and the average year was 2004.

The following table shows summary statistics of the collected data for the W&S model. The table includes interactions between water and sewer rate structures, outside charges, and commercial rates.

Table 12: Summary Statistics – Water & Sewer Model

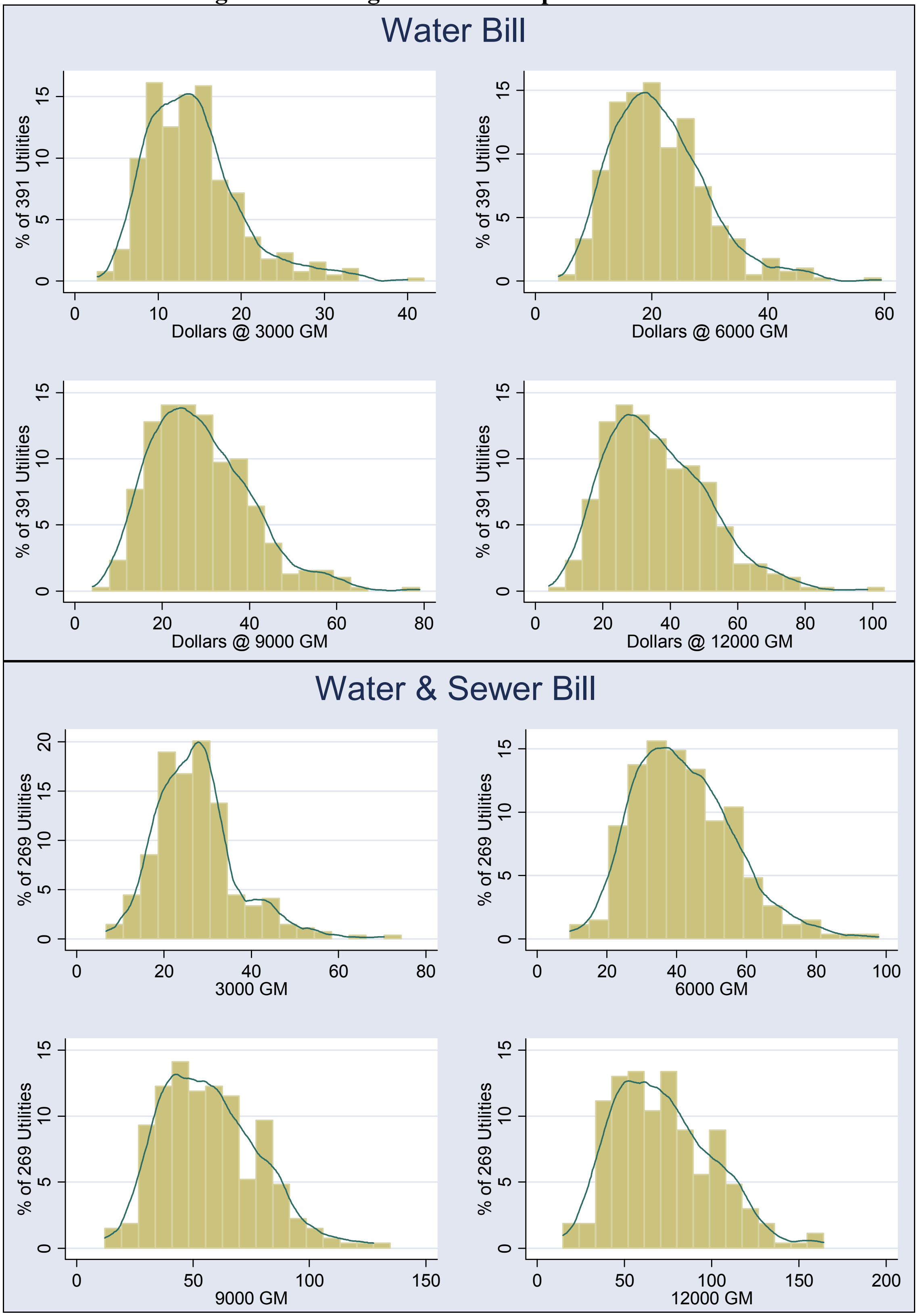
Independent Variables	No. of Utilities	Mean or % of Sample	Standard Deviation
Water & Sewer Bills			
Bill @ 3000 GM	269	$27.50	$9.49
Bill @ 6000 GM	269	$42.13	$14.33
Bill @ 9000 GM	269	$57.09	$20.90
Bill @ 12000 GM	269	$72.18	$28.16
Cost of Service Factors			
Interbasin transfer?	25 of 269	9.3%	29.1%
Median year homes built	269	1972.8	8.0
UO: Authority?	17 of 269	6.3%	24.4%
UO: County?	14 of 269	5.2%	22.3%
UO: Municipality?	230 of 269	85.5%	35.3%
UO: Utility commission?	8 of 269	3.0%	17.0%
Population growth rate (% per year)	269	1.8%	3.4%
Population density (persons per square mile)	269	733	569
Population with water service	269	19,513	68,070
PS: Groundwater (S. Fall Line)?	134 of 269	49.8%	50.1%
PS: Groundwater (N. Fall Line)?	25 of 269	9.3%	29.1%
PS: Purchased surface water?	41 of 269	15.2%	36.0%
PS: Surface water?	69 of 269	25.7%	43.8%
USDA & GEFA loans?	155 of 269	57.6%	49.5%
USDA & GEFA grants?	60 of 269	22.3%	41.7%
OR: Water only?	36 of 269	13.4%	34.1%
OR: Water & sewer?	108 of 269	40.1%	49.1%
OR: None	125 of 269	46.5%	50.0%
CR: Sewer only?	4 of 269	1.5%	12.1%
CR: Water only?	16 of 269	5.9%	23.7%
CR: Water & sewer?	117 of 269	43.5%	49.7%
CR: None	132 of 269	49.1%	50.1%
Total water & sewer connection fees	269	$1,995	$2,039
Control Factors			
Average annual rainfall (inches/year)	269	50.3	4.7
Average annual temperature (°F)	269	62.7	3.2
Median household income ($/year)	269	$30,590	$9,457
RS: W&S-Decreasing?	9 of 269	3.3%	18.0%
RS: W-Decreasing, S-Uniform?	5 of 269	1.9%	13.5%
RS: W&S-Flat?	3 of 269	1.1%	10.5%
RS: W&S-Increasing?	38 of 269	14.1%	34.9%
RS: W-Increasing, S-Uniform?	22 of 269	8.2%	27.5%
RS: W&S-Uniform?	175 of 269	65.1%	47.8%
RS: W-Uniform, S-Decreasing?	3 of 269	1.1%	10.5%
RS: W-Uniform, S-Flat?	7 of 269	2.6%	15.9%
RS: W-Uniform, S-Cap?	7 of 269	2.6%	15.9%
Year of last rate change	269	2004.6	3.0

Like the water model, the most common utilities in the W&S model are owned by municipalities, serve groundwater south of the Fall Line, do not receive water from outside their basin, charge uniform block rates for both services, and do not charge higher rates to outside and commercial customers for either service. Also, an average utility serves 19500 customers (median = 4000), and the service area generally has 730 persons per square mile (median = 630) with a population growth rate of 1.8% (median = 0.85%). Values for *median household income*, *median year homes built*, *rainfall*, *temperature*, *year of last rate change*, are almost identical in both models. Regarding finances, over the last 10 years, 22% of utilities have received grants and 58% received loans. The average total connection fees for both water and sewer are about $2000 (median = $1100). Also, the average water bill at average consumption (6000 GM) is $42.

The changes between the samples for the water and W&S models are slight, but there are a few noticeable differences. First, the sample in the W&S model has a higher population (19,500 versus 17,000) and population density (730 versus 600 persons per square mile), which signifies this sample is more urban. This is not a surprise since the water model includes utilities that serve water only, which are usually located in rural areas. Most rural households use septic systems, so sewerage is not needed. Second, the sample in the W&S model has larger connection fees ($2000 versus $1600), which is again due to the water model having water-only utilities. Since these utilities have only one service, they do not need the additional funds.

The following figure shows the distribution for the dependent variables, which show the percent of utilities with a specific bill value. For example, in the first graph for the water bill at 3000 GM, about 13% of the 409 utilities charge $10.

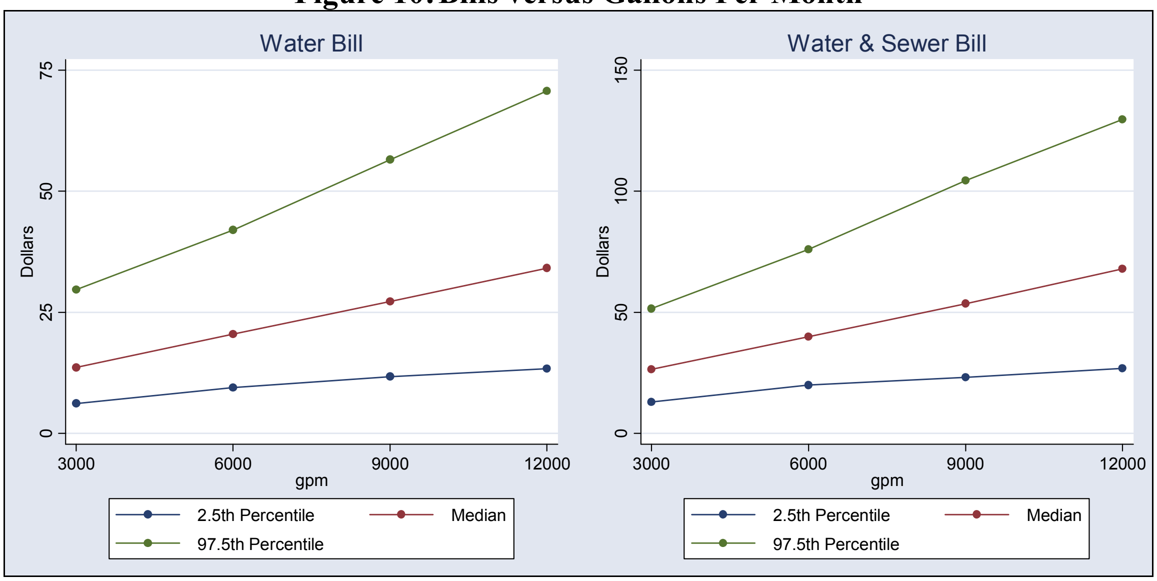
Figure 9: Histograms of the Dependent Variables



The distributions have a significantly thick tale on the right side representing a number of utilities charging high prices.

The following figure shows a different way of looking at the distribution of the bills. The top line is the bill for the 97.5 percent of utilities that charge that amount or less, the middle line represents the median bill value, and the bottom line represents the 2.5 percent of utilities that charge that amount or less.

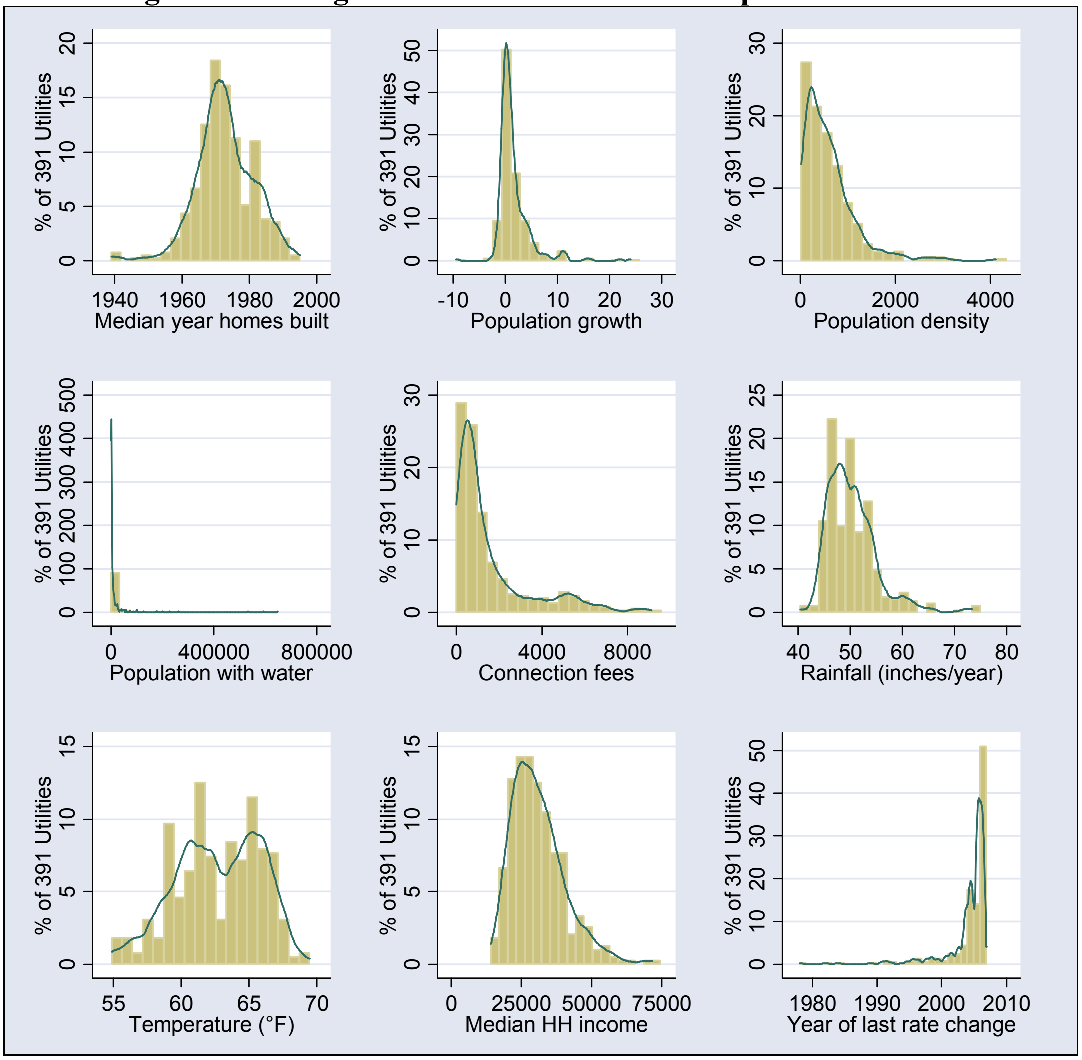
Figure 10: Bills versus Gallons Per Month



Note that the variability of the bill increases as consumption increases.

The following figure shows individual distributions of the continuous variables used in the water and W&S models. Since the general shapes are the same in both models, only the distributions for the water model are produced.

Figure 11: Histograms of the Continuous Independent Variables



Regression Diagnostics

Regression diagnostics were performed using Stata 8.2. Developing the models at each consumption level included the following steps:

1) Transforming continuous independent variables to other functions if they have skewed distributions or if they have a recognizable functional relationship, other than linear, with the dependent variable.

2) For each consumption level, running various regressions with different combinations of raw and transformed variables.

3) Running three specification tests on each version and keeping those that pass at least two of the tests.
4) Examining the variance inflation factor (VIF) for each independent variable. If any variable is above 10, then throw out the model with that variable.
5) Of the remaining versions, selecting the model with the highest adjusted R^2.

Along with these procedures, the normality of the residuals was checked using the Shapiro-Wilk test, Shapiro-Francia test (StataCorp 2003d), and a distribution plot of the residuals. Normality is important for making accurate t-tests and correctly interpreting the coefficients, but OLS can be robust in the face of non-normal residuals. Regardless of the test results, the models were not changed. The Breusch-Pagan / Cook-Weisberg test, White test (StataCorp 2003b), and plots of the residuals versus the fitted values were used to check for heteroscedasticity. If heteroscedasticity exists, then the standard errors and hypothesis tests could be flawed (Hamilton 2004). The following sections provide details on the procedures outlined above.

Transformations

Three types of variables were transformed: 1) those that showed normality in ladder of powers transformations (explained in the next paragraph), 2) those with highly skewed distributions, and 3) those with nonlinear relationships of the dependent and independent variables. For the first type, converting a variable's distribution to normality could improve model specification and the normality of the model residuals as well as reduce collinearity. Even if a normal transformation does not exist, a variable with a highly skewed distribution could become closer to normal through a logarithmic transformation. The last type is required because OLS assumes a linear relationship between the dependent and independent variables; otherwise, the interpretation of coefficients will be incorrect.

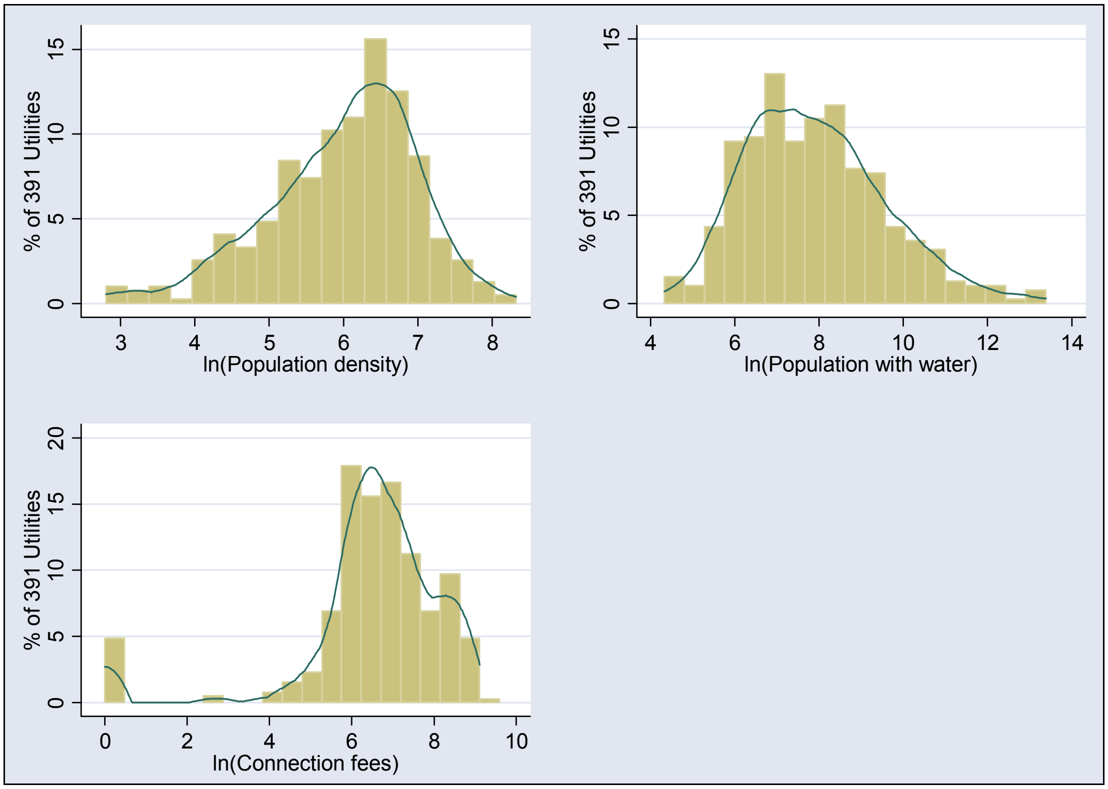
The ladder of powers (cubic, square, square root, log, reciprocal root, reciprocal, reciprocal square, and reciprocal cubic) transformations for each variable was tested using a chi-square statistic for skewness and kurtosis. Only *average annual rainfall* and *median household income* had transforms where the null hypothesis (normality) was not rejected at the 10% level. The following table lists the variable, its transform, the chi-square statistic, and the probability.

Table 13: Transformations of Variables to Normality

Variable	Transformation	Chi-square	Prob. > Chi-square
Average annual rainfall	reciprocal cubic	3.47	0.176
Median household income	reciprocal root	2.32	0.313

In addition to these transforms, the natural log transformation improved the distribution of the following positively skewed variables: *population density*, *population with water service*, and *connection fees*. The following figure shows the log transformed distributions for the water model only. The distributions for the W&S model are similar. Please see Figure 11 to view the raw distributions.

Figure 12: Histograms of Log Transformed Independent Variables for the Water Model

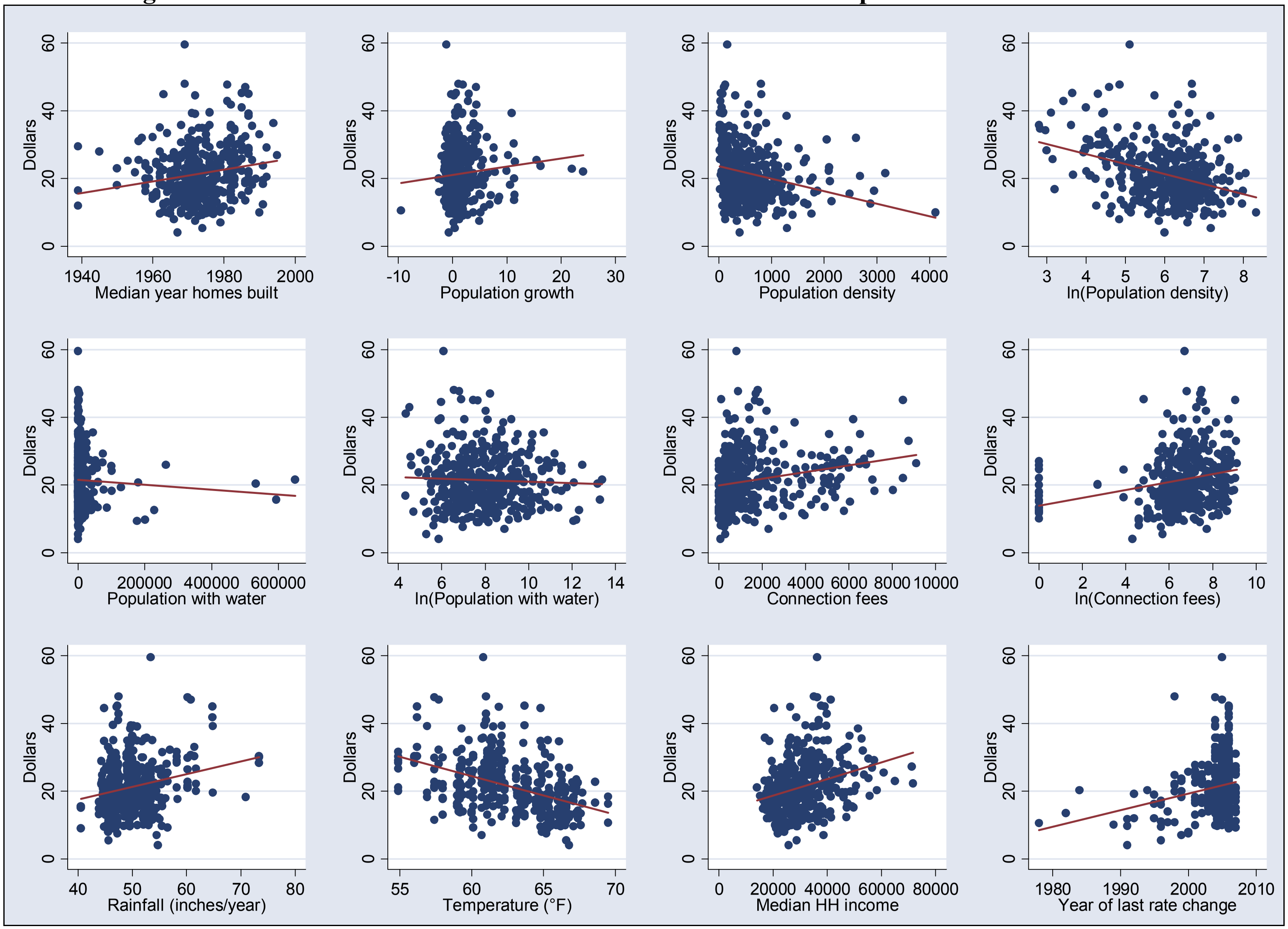


As shown, the variables have more normal distributions than before. Note that 5% of the utilities in the water model have zero logged connection fees. The raw data for these points are actually zero dollars, but as the natural log of zero is undefined, a "1" was added to each zero value in order to keep those utilities in the analysis. And the logarithm of one is simply zero.

Not all transforms are included in the models because some may actually reduce the models' explanatory power, even though the transform improves the relationship between the dependent variable and itself. This is because other independent variables interact with the transformed variables. In fact, while developing the models, none of the variables with statistically significant transforms to normality (*average annual rainfall* and *median household income*) were found to improve the models.

Before running the regressions, scatter plots of the dependent variables versus each independent variable (raw and transformed) were scanned for recognizable functions other than linear. The following figure shows the scatter plots for the water model at 6000 GM, which are similar to the scatter plots at other consumption levels and in the W&S model.

Figure 13: Scatter Plots of the Water Bill vs. the Continuous Independent Variables at 6000 GM



All scatter plots had no discernable shape in the data, except temperature in the water model. The relationship was slightly curvilinear and the square of temperature was significant at the 1% level at all consumption levels. However, the improvement in R^2 was only as high as 4% in the 3000 GM model and as low as 2% in the 12000 GM model. Thus, the square of temperature was not included in the water model.

Logged Dependent Variable vs. Raw Dependent Variable

While developing the models, it was discovered that, at most consumption levels, a log transformation of the dependent variable was needed to pass most of the specification tests. However, as this subsection shows, a logged bill requires a few extra steps to convert the model estimates to real bills. In addition, this subsection demonstrates how to compare the explained variation of a model with a logged dependent variable to a model with a raw dependent variable. This will show that all of the models with raw bills had a slightly better explanatory power. The subsection will conclude with the reason why the logged versions were kept for all consumption levels.

To find the real estimate of the bills from the logged fitted values, one might assume that the antilog would be sufficient. However, this consistently underestimates the bill for each utility (Wooldridge 2003). The following equation can estimate the raw fitted values if the error term is independent of the explanatory variables:

$$\hat{y} = \acute{\alpha}_0 * \exp(\log\hat{y})$$

where $\hat{y}$ is the raw fitted values, $\acute{\alpha}_0$ is a multiplier discussed below, and $\exp(\log\hat{y})$ is the antilog of the logged fitted values ($\log\hat{y}$). The multiplier is found through the following steps:

1) Regress the logged model and find the fitted values, which is $\log\hat{y}$.

2) Calculate the antilog of the fitted values, which is $\exp(\log\hat{y})$.

3) Regress the raw dependent variable on only $\exp(\log\hat{y})$ and with no intercept. The coefficient on $\exp(\log\hat{y})$ is $\acute{\alpha}_0$.

4) Multiply $\exp(\log\hat{y})$ by $\acute{\alpha}_0$ to find the raw estimates of the water bills.

These estimates are slightly biased, but this is the best method available for the retransformation of the logged dependent variable.

Estimating the 95% confidence interval also changes from the normal routine. The standard equation is:

$$CI = \hat{y} \pm c * \hat{e}_s$$

where *CI* is the confidence interval; *c* is the 97.5th percentile in a t_{n-k-1} distribution, which is approximately 1.97 for these models; and $\hat{e}_s$ is the standard error of the fitted values. However, for logged dependent variables, this changes to:

$$CI = \exp(\log\hat{y} \pm c * \log\hat{e}_s)$$

where $\log\hat{e}_s$ is the standard error for the logged fitted values (Nelson 1973). Note that the interval is not symmetric around $\hat{y}$ because converting from $\log\hat{y}$ to $\hat{y}$ skews the distribution.

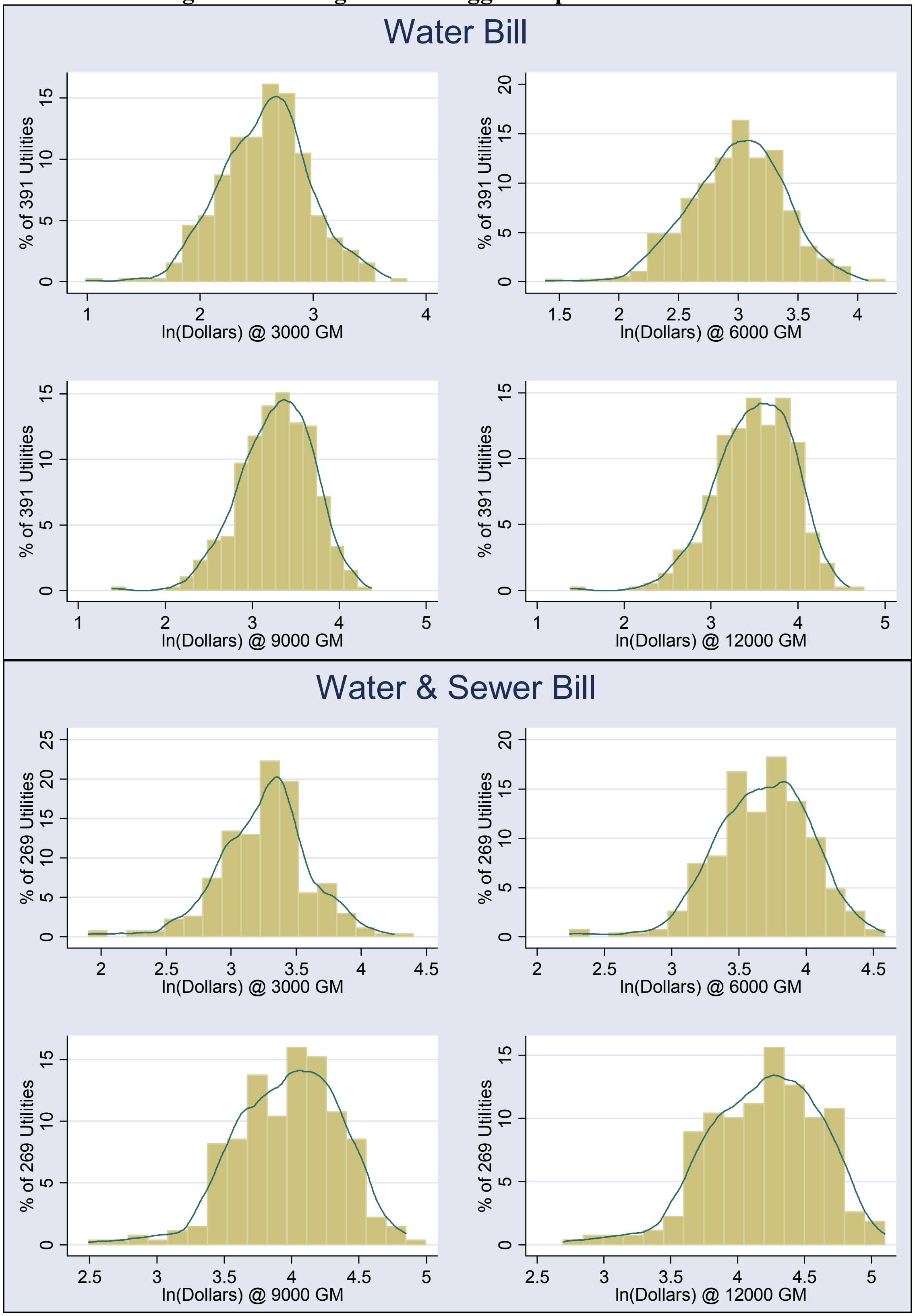
Regarding the method for comparing the explained variation of the models, Wooldridge (2003) discusses a way to extract a goodness-of-fit measure from the model with a logged dependent variable that is comparable to R^2 in the model with a raw dependent variable. One cannot directly compare R^2 or adjusted R^2 from both models because the different functional forms of the dependent variables have different amounts of variation. Continuing from step 4 above, the measure can be calculated by squaring the sample correlation between ŷ and the actual y, which is the raw dependent variable (i.e., the bill). The following table compares these measures.

Table 14: Goodness-of-Fit Test – Logged Bill versus Raw Bill

Quantity	Water Model		W&S Model	
	Log bill	Raw bill	Log bill	Raw bill
3000 GM	0.377	0.379	0.296	0.297
6000 GM	0.477	0.485	0.437	0.442
9000 GM	0.530	0.537	0.516	0.520
12000 GM	0.562	0.568	0.561	0.563

As shown, the models with the raw dependent variable have a slightly better goodness-of-fit measure. The logged bills are kept because the models with the raw bills fail a few specification tests at some consumption levels. Plus, the difference in explanatory power between the two models is relatively small. Also, dependent (and independent) variables with normal distributions increase the likelihood that the residuals are normally distributed, which in turn, makes the t-tests valid (Chen et al. 2003). In fact, the distribution of the residuals in the models with logged bills appeared more normal than in the distribution of the residuals in the models with raw bills. The following figure shows the distributions of the logged bills. Compared to the distributions of the raw dependent variables in Figure 9, the logged bills are more normally distributed.

Figure 14: Histograms of Logged Dependent Variables



Specification Tests

The first specification test, the link test, regresses the fitted values and the squared fitted values only on the dependent variable (StataCorp 2003a). If the model is properly specified, then a power of the fitted values should have no predictive capability. The regression specification error test (RESET) is similar but uses the original model along with the second, third, and fourth powers of the fitted values as independent variables (Wooldridge 2003). Another type of RESET uses the original regression with the second, third, and fourth powers of each continuous independent variable (StataCorp 2003b). Again, none of these should be statistically significant if the model is properly specified. The following table lists the test statistics and the probability. The null hypothesis is that some powers are statistically significant predictors.

Table 15: Specification Tests

Model	Link Test		RESET - Fitted Values		RESET - Ind. Vars.	
	t	Prob > |t|	F	Prob > F	F	Prob > F
Water - 3000 GM	-0.151	0.880	0.189	0.904	1.884	0.006
Water - 6000 GM	-0.979	0.328	1.045	0.373	2.163	0.001
Water - 9000 GM	-0.949	0.343	1.922	0.126	2.166	0.001
Water - 12000 GM	-0.787	0.432	1.893	0.130	2.084	0.002
W&S - 3000 GM	0.273	0.785	0.199	0.897	1.333	0.135
W&S - 6000 GM	-0.443	0.658	0.214	0.887	1.287	0.165
W&S - 9000 GM	-0.426	0.671	0.758	0.519	1.235	0.205
W&S - 12000 GM	-0.293	0.769	1.150	0.330	1.197	0.239

Note that the water model passes the link test and RESET for fitted values, but not RESET for independent variables. No other combination of raw and transformed variables could change the results for the latter test. This indicates that some important variables need to be included in the model. However, for the most part, the model specification is adequate. The W&S model passes all three specification tests.

Variance Inflation Factor

After each model was run, the VIF was computed for all independent variables. VIF is a ranking of collinearity with no firm cut off point. Some researchers suggest anything above 10 could have an adverse impact on the model while others suggest 30 as the cut off point. In addition, if the mean VIF is considerably greater than one, collinearity may be a problem (StataCorp 2003c). The following table lists the VIFs for the water model.

Table 16: Variance Inflation Factor for the Water Model

Variable	VIF		Variable	VIF
Temperature	3.74		Grants	1.52
PS: SW?	3.20		Rainfall	1.50
PS: SWP?	2.93		Population growth	1.40
PS: GW N. of Fall Line?	2.44		Population	1.40
Median household income	2.23		Outside rates	1.40
Population density (log)	2.21		Interbasin transfer	1.38
Sewer service?	1.68		RS: Increasing block	1.22
Connection fees (log)	1.65		Commercial taps	1.16
UO: County?	1.64		Year of last rate change	1.16
UO: Authority?	1.62		RS: Flat fee	1.13
Median year houses built	1.61		UO: Utility commission	1.13
Loans	1.58		RS: Decreasing block	1.11
			Mean VIF	1.75

As shown above, no VIF is greater than ten, and the mean VIF is not considerably greater than one. The following table lists the VIF's for the W&S model.

Table 17: Variance Inflation Factor for the W&S Model

Variable	VIF	Variable	VIF
Temperature	4.31	RS: W-Inc, S-Uni	1.42
PS: SW?	4.26	Loans	1.39
PS: SWP?	3.53	OR: Water only	1.35
Population (log)	3.30	UO: Utility commission	1.34
Median household income	2.87	Year of last rate change	1.23
Population density (log)	2.68	CR: W&S	1.22
UO: Authority?	2.40	CR: Water only	1.21
PS: GW N. of Fall Line?	2.32	CR: Sewer only	1.21
UO: County?	2.07	RS: W&S-Inc	1.18
Median year houses built	1.87	RS: W-Uni, S-Flat	1.17
Rainfall	1.71	RS: W-Uni, S-Dec	1.16
Connection fees (log)	1.60	RS: W-Uni, S-Cap	1.15
Population growth	1.58	RS: W&S-Dec	1.14
OR: W&S	1.56	RS: W-Dec, S-Uni	1.12
Grants	1.50	RS: W&S-Flat	1.06
Interbasin transfer	1.48	Mean VIF	1.85

As in the water model, the independent variables in the W&S model do not have a VIF greater than ten, and the mean VIF is not considerably greater than one. Now, examining the correlation of the regression coefficients reveals that *temperature*, *SW*, and *SWP* are all highly correlated to each other (0.6-0.7 in both models). When removing *temperature* from the models, the regression coefficients and size of the confidence intervals did not change much. The same was true when removing only the primary source water variables. Therefore, the models were not changed.

Analysis of the Residuals

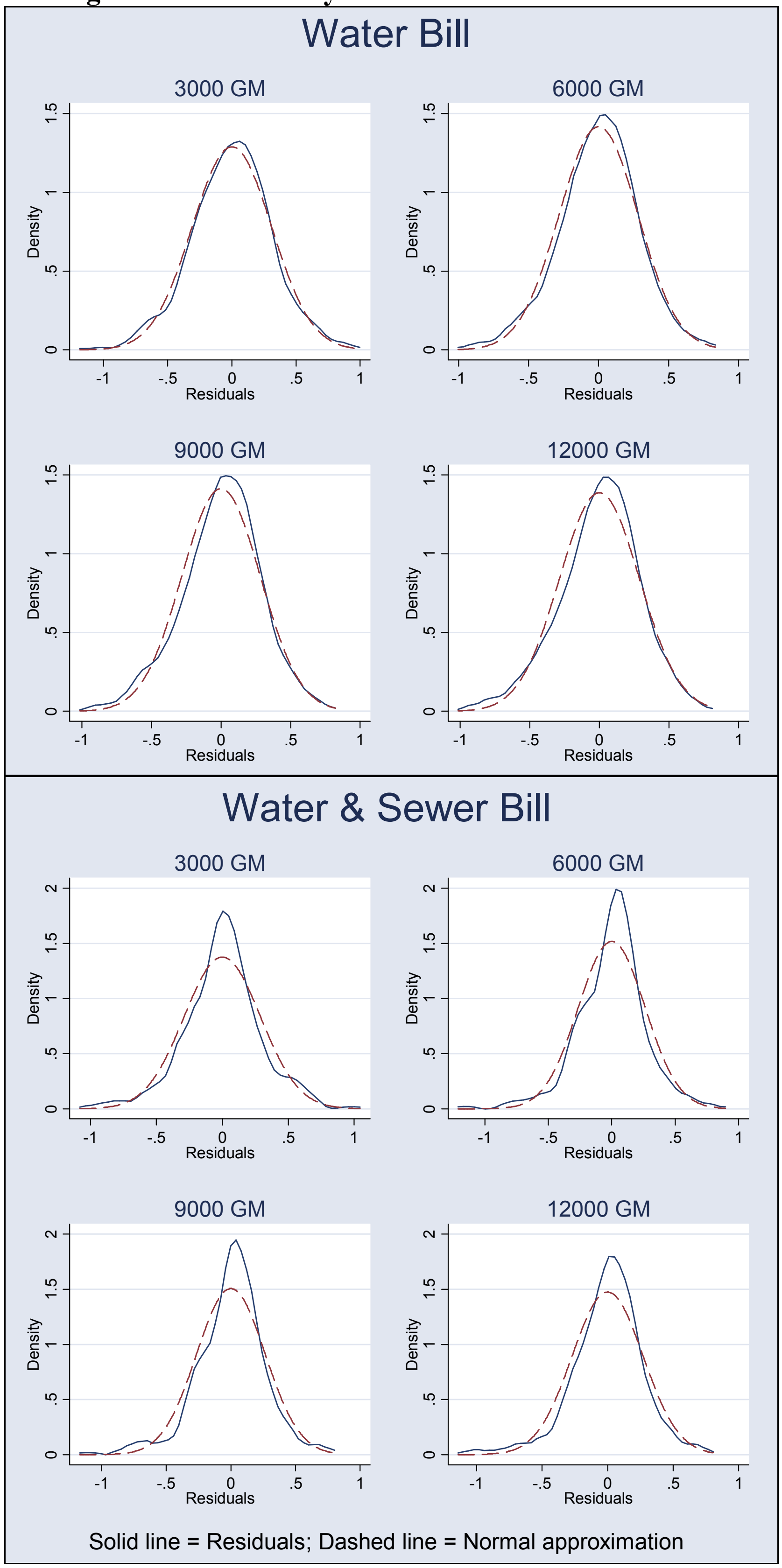
As stated previously, normality of the residuals is important for accurate interpretation of the coefficients, but OLS is fairly robust to non-normality. The Shapiro-Wilk and Shapiro-Francia tests were used to check the normality of the residuals. The null hypothesis for both is that a normal distribution is valid. The following table shows the results of those tests.

Table 18: Tests for Normality of the Residuals

Model	Shapiro-Wilk Test		Shapiro-Francia Test	
	z	Prob > z	z	Prob > z
Water - 3000 GM	0.689	0.245	1.021	0.154
Water - 6000 GM	1.615	0.053	1.633	0.051
Water - 9000 GM	2.371	0.009	2.297	0.011
Water - 12000 GM	2.730	0.003	2.592	0.005
W&S - 3000 GM	3.147	0.001	3.184	0.001
W&S - 6000 GM	3.954	0.000	3.913	0.000
W&S - 9000 GM	4.252	0.000	4.097	0.000
W&S - 12000 GM	4.416	0.000	4.210	0.000

Notice that the null hypothesis is rejected for most models, except the water models at 3000 and 6000 GM at the 5% level. The following figure shows the distribution plots of the residuals.

Figure 15: Probability Distributions of the Residuals



Notice that the residuals for the water model look close to normally distributed. The residuals for the W&S bill are less convincing.

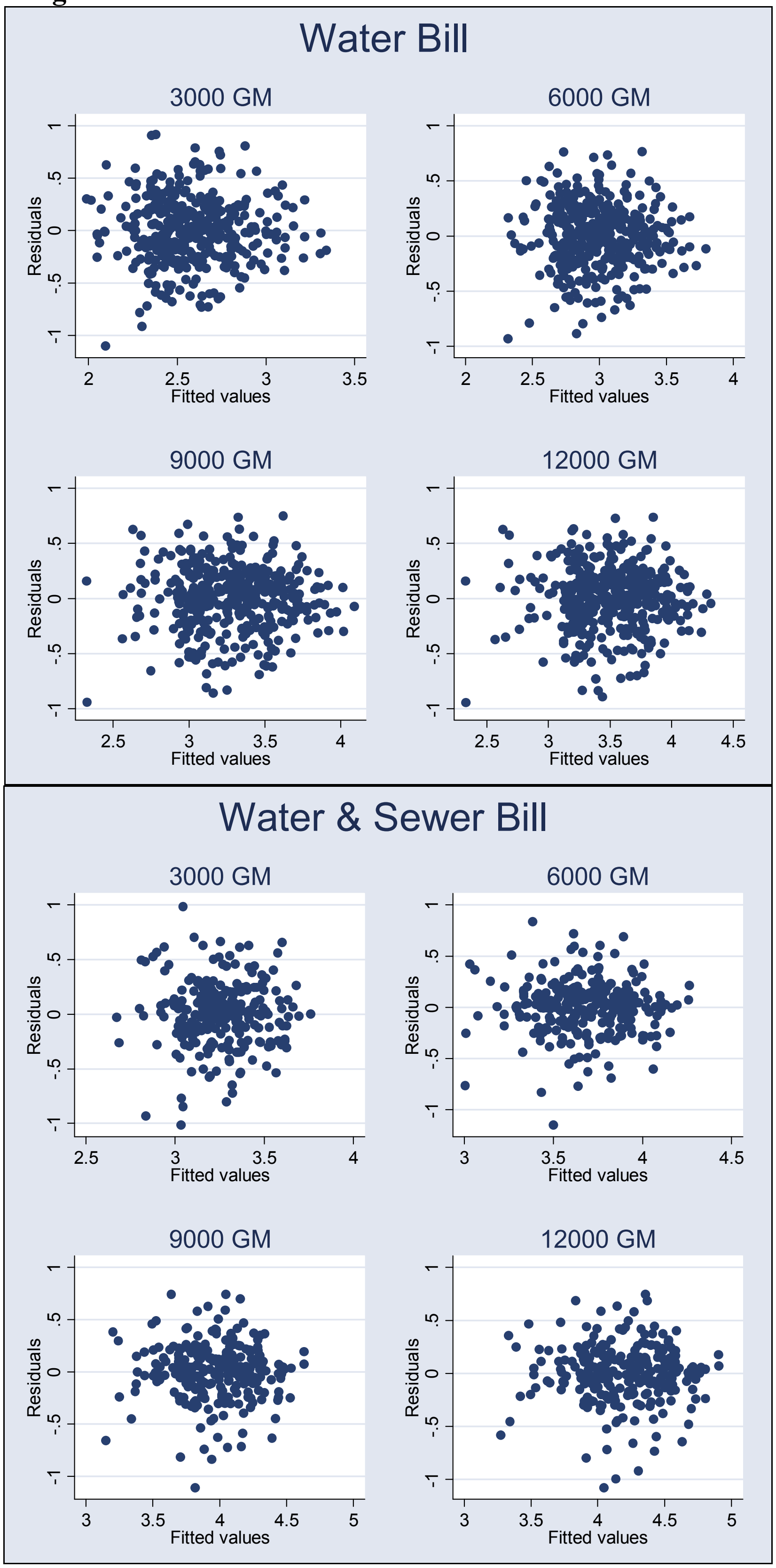
We test for heteroscedasticity in order to ensure valid hypothesis tests. The following table shows the Breusch-Pagan / Cook-Weisberg and White tests for heteroscedasticity. The null hypothesis for both tests is that the error variance is constant.

Table 19: Tests for Heteroscedasticity

Model	Breusch-Pagan / Cook-Weisberg test		White test	
	chi-squared	Prob > chi-squared	chi-squared	Prob > chi-squared
Water - 3000 GM	5.02	0.025	284	0.331
Water - 6000 GM	8.75	0.003	290	0.238
Water - 9000 GM	10.4	0.001	292	0.218
Water – 12000 GM	11.2	0.001	283	0.335
W&S - 3000 GM	7.89	0.005	269	0.471
W&S - 6000 GM	6.48	0.011	269	0.471
W&S - 9000 GM	3.39	0.065	269	0.471
W&S - 12000 GM	2.00	0.158	269	0.471

For the Breusch-Pagan / Cook-Weisberg test, the null hypothesis is rejected for all models at all consumptions except the W&S model at 9000 and 12000 GM. However, the White test shows the opposite. To help us decide if heteroscedasticity is a problem, we can examine the plot of the residuals versus the fitted values (Figure 16), which is a standard plot for examining the variance of the residuals. As one can see, the scatter in both models appears to be uniform. In fact, it is more uniform than the scatter in the models with raw bills. We conclude that the models do not need to be changed due to heteroscedasticity.

Figure 16: Plots of the Residuals versus the Fitted Values



CHAPTER 4

RESULTS

This chapter presents the results from the regressions of the water and W&S models. The first section summarizes the estimates of the water bills and combined water and sewer bills for each utility. Since the estimates cover 42 pages, the output is made available online only (see Appendix B for the website address). For both models at all consumption levels, the workbook contains each utility's actual bill, estimated bill, upper and lower bounds of the confidence interval, and the length of the confidence interval. The workbook contents are illustrated in this chapter through a small sample of utilities. The second section discusses the sign and statistical significance of the coefficients in the regression models.

Estimates of Average Bills

Five utilities were chosen to demonstrate the estimated bills from the regressions. The following table shows the results from the water model at 6000 GM.

Table 20: Average Water Bills for 6000 GM for Select Utilities

Utility	Bill for 6000 GM		95% Confidence Interval		
	Actual	Estimate	Low	High	Difference
Maxeys	$32.00	$31.12	$25.68	$35.38	$9.70
Mount Zion	$25.00	$33.16	$27.30	$37.80	$10.50
Cadwell	$16.25	$12.83	$10.56	$14.62	$4.05
Uvalda	$14.00	$14.05	$11.64	$15.92	$4.28
Habersham County WSA	$47.70	$31.70	$26.70	$35.31	$8.62

The "actual bill" is the bill calculated from each utility's rate schedule. The next column shows the estimated water bill for an average utility with the same characteristics as the utility for that row. The 95% confidence interval is the likelihood that, given the data, the average bill is within that range. The last column shows the length of the estimated confidence interval. Notice that the confidence interval varies widely. A short interval indicates that many of the utility's factors are close to the mean, and vice-versa for utilities with longer intervals. For instance, most utilities are municipalities that use groundwater (south of the Fall Line), which describes Cadwell. But, Habersham County WSA is an authority that uses purchased surface water. Since this utility is relatively unusual, the uncertainty of the point estimate is reflected by a larger confidence interval ($8.62 vs. $4.05).

The selected utilities demonstrate a situation where a utility's actual bill, when compared to the arithmetic mean in the sample, can be mistakenly identified as below average or above average. Starting with Maxeys, notice that its actual bill of $32 is much higher than the state average for 6000 GM in Table 11 of $21.39. However, the model estimate is nearly identical to the actual bill. In other words, Maxeys' water bill is nearly the same as the average bill for utilities in its unique class. Looking at Mount Zion's actual bill of $25, it is slightly higher than the state average bill. However, the model predicts a bill of $33, and the actual bill falls below the confidence interval. While the actual bill is above the state average bill, it is considered below the average bill for utilities like Mount Zion. A utility like Cadwell with an actual bill of $16 may be mistaken for below average, but the model shows that its bill is above the average bill for utilities in its class. Uvalda's actual bill of $14 is well below the state average bill, but the model shows that its bill is the same as the average bill for utilities in its unique class. A fair number of other utilities in the sample follow the same pattern of being misidentified.

Many more utilities are correctly assumed to be above or below the average, but the model estimates are closer to the actual bills. This should reduce the perception that these utilities are at the extremes. For example, Habersham County WSA charges $47.70, which is above the upper bound of the confidence interval. Without the model estimate of $32, one might assume that Habersham County WSA charges a very large amount compared to the average bill in the sample of $21.39. By accounting for various factors, the model shows that Habersham County WSA's bill is not too far from the average bill in its unique class.

In fact, by consulting a rate survey alone, Habersham County WSA might be considered grossly overpriced because utilities in the same county charge much less. Both Clarksville and Cornelia use groundwater, charge ~$22 for 6000 GM, and serve 4000 and 6000 customers, respectively. Habersham County WSA, on the other hand, purchases its own surface water and serves 1000 customers. These utilities seem similar in the important factors, yet Habersham County WSA's bill is more than double. However, the models show that Habersham County WSA's actual bill is closer to the average bill of its peers throughout the state ($32 vs. $21.39 for the simple state average of all utilities).

The following table summarizes the number of utilities according to standing for each model.

Table 21: Number of Utilities by Standing

Bill	Below CI	Within CI	Above CI	Total
Water - 3000 GM	120	152	119	391
Water - 6000 GM	109	160	122	391
Water - 9000 GM	110	164	117	391
Water - 12000 GM	109	158	124	391
W&S - 3000 GM	60	149	60	269
W&S - 6000 GM	63	151	55	269
W&S - 9000 GM	60	152	57	269
W&S - 12000 GM	55	152	62	269

Roughly 40% of the utilities in the water model are considered average, and almost 60% are considered average in the W&S model. This discrepancy is due to the higher number of variables in the W&S model. With more predictors in the model comes increased variability in the estimation. And, there are 122 fewer utilities in the W&S model than in the water model.

Keep in mind that a utility can have an actual bill within, above, or below the confidence interval for one consumption level, but not all others. The next table provides an example of five utilities with each showing a general pattern for each utility's actual bill in relation to the confidence interval at each consumption level.

Table 22: Examples of Standing at Each Consumption Level

Utility	Standing				
	3000 gpm	6000 gpm	9000 gpm	12000 gpm	Same Standing
Utility A	Within CI	Within CI	Within CI	Within CI	4 of 4
Utility B	Above CI	Above CI	Above CI	Within CI	3 of 4
Utility C	Below CI	Below CI	Within CI	Within CI	2 of 4 (2 sets)
Utility D	Below CI	Within CI	Within CI	Above CI	2 of 4

Utility A's actual bill falls within the confidence intervals estimated at each consumption levels, so its bills have the same standing for four of four levels. Utility B's actual bills are above the confidence intervals for three consecutive consumption levels but not the highest level ("3 of 4"). Utility C has below average bills for 3000 and 6000 GM and average bills for 9000 and 12000 GM, so its bills have two sets of the same standings ("2 of 4 (2 sets)"). Utility D has a two consumption levels at the same standing only. The next table shows the number of utilities with a certain amount of equivalent standing across the consumption levels.

Table 23: Number of Utilities with Equivalent Standing Across Consumption Levels

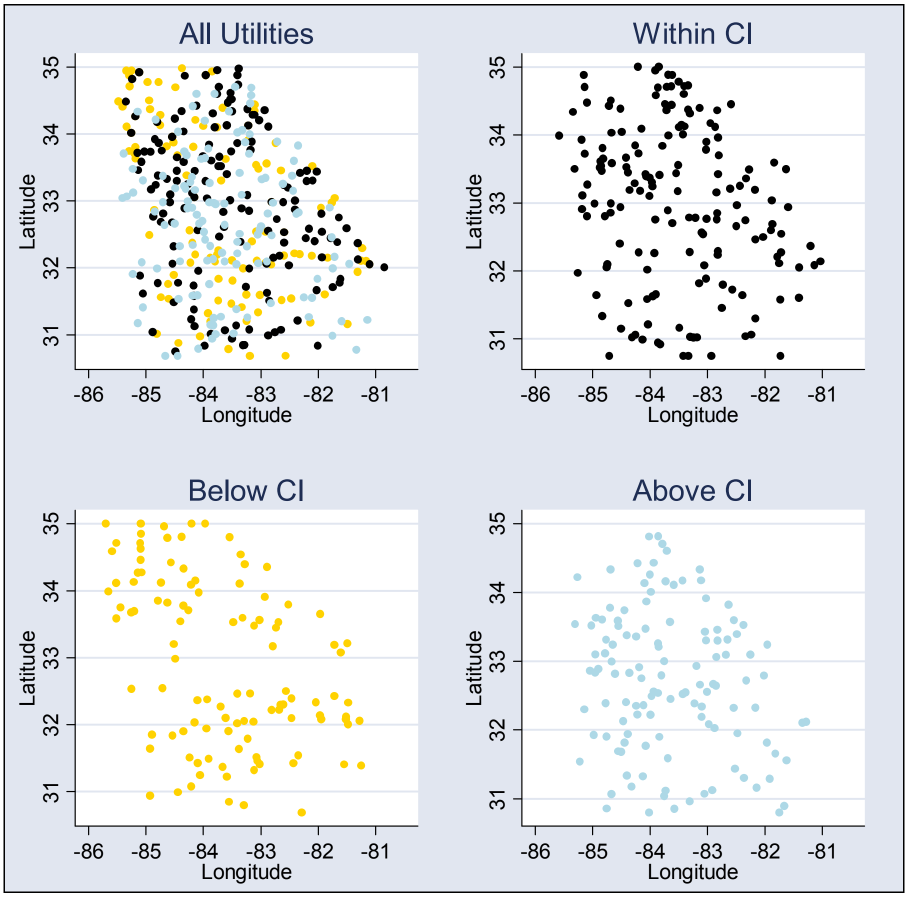
Same Standing	Utilities	
	Water Model	W&S Model
4 of 4	246	178
3 of 4	101	73
2 of 4 (2 sets)	35	16
2 of 4	9	2
Total	391	269

*One utility each from both models (different utilities) had standings that alternated across the consumption levels

The water and W&S models have 246 and 178 utilities, respectively, with the exact same standing across all consumption levels. And, 101 and 73 utilities have the same three standings at consecutive consumption levels in the water and W&S models, respectively. Only a few utilities in both models have two consecutive standings at 6000 and 9000 GM ("2 of 4"). So, for most utilities, the actual bill in relation to the estimated confidence interval is consistent across most of the consumption levels.

The following figure shows the location of utilities with below average, average, and above average bills for the water model at 6000 GM. The scatter of utilities in each graph is more or less homogenous across Georgia, and this is similar for other models.

Figure 17: Location of Utilities by Standing for the Water Model at 6000 GM



Description of the Statistically Significant Variables

The regression results for each model are described in nine subsections. The first subsection discusses how to interpret the coefficients. The second focuses on the variables in the water model statistically significant at the 5% level for three or four of the consumption levels ("highly significant"). The third examines the variables in the water model significant at the 10% level at one or more consumption levels ("moderately significant"). The fourth subsection discusses the variables in the water model with no significance at any consumption level ("non-significant"). The fifth subsection describes the results from regressions with alternative reference variables for each set of categorical variables in the water model. For example, the reference variable for *utility ownership* is *municipality*, but this subsection examines the coefficients of the other *utility ownership* variables if *county* is the reference variable. The remaining subsections repeat the 2nd-5th subsections for the W&S model.

The next four tables identify the coefficients, standard errors, and statistical significance of the variables in the water and W&S models at 3000, 6000, 9000, and 12000 GM. The first table for each model contains cost variables, while the second contains control factors. Also, the second table for each model records the number of observations and the adjusted R^2. Appendix C shows the analysis of variance tables along with the F statistic, R^2, and root mean square error.

Table 24: Regression Results for Cost Factors in the Water Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
Interbasin transfer?	0.108	0.114*	0.112*	0.113*
	[0.069]	*[0.063]*	*[0.063]*	*[0.064]*
Median year homes built	-3.16E-05	-1.21E-03	-1.72E-03	-2.03E-03
	[2.398e-03]	*[2.182e-03]*	*[2.187e-03]*	*[2.228e-03]*
UO: Authority?	0.108	0.125*	0.132**	0.143**
	[0.072]	*[0.065]*	*[0.066]*	*[0.067]*
UO: County?	0.112	0.09	0.084	0.086
	[0.083]	*[0.075]*	*[0.076]*	*[0.077]*
UO: Utility commission?	-0.196*	-0.212**	-0.230**	-0.242**
	[0.109]	*[0.099]*	*[0.099]*	*[0.101]*
Population growth	-0.021***	-0.014**	-0.011*	-0.008
(% per year)	*[0.006]*	*[0.006]*	*[0.006]*	*[0.006]*
Population density	-0.084***	-0.064***	-0.051**	-0.042*
(log persons/mi2)	*[0.024]*	*[0.022]*	*[0.022]*	*[0.022]*
Population with water	-7.096e-07**	-6.874e-07**	-7.184e-07**	-7.619e-07**
	[3.292e-07]	*[2.996e-07]*	*[3.002e-07]*	*[3.059e-07]*
PS: GW (N. of Fall Line)?	0.037	0.151**	0.205***	0.242***
	[0.074]	*[0.068]*	*[0.068]*	*[0.069]*
PS: SWP?	0.149*	0.273***	0.331***	0.364***
	[0.076]	*[0.069]*	*[0.069]*	*[0.070]*
PS: SW?	0.06	0.182***	0.235***	0.266***
	[0.072]	*[0.065]*	*[0.065]*	*[0.067]*
Loans?	0.063	0.068*	0.071*	0.073*
	[0.041]	*[0.037]*	*[0.037]*	*[0.038]*
Grants?	0.124***	0.100**	0.086**	0.079*
	[0.047]	*[0.043]*	*[0.043]*	*[0.044]*
Outside rates?	-0.064*	-0.054	-0.048	-0.045
	[0.039]	*[0.035]*	*[0.035]*	*[0.036]*
Commercial tap rates?	1.75E-03	-7.38E-03	-7.58E-03	-9.04E-03
	[3.506e-02]	*[3.190e-02]*	*[3.197e-02]*	*[3.257e-02]*
Connection fees (log $)	0.020*	0.026**	0.026**	0.026**
	[0.011]	*[0.010]*	*[0.010]*	*[0.011]*
Sewer service?	-0.107**	-0.106**	-0.110***	-0.108**
	[0.046]	*[0.042]*	*[0.042]*	*[0.043]*

Notes: 1) Standard errors in brackets
2) * significant at 10%; ** significant at 5%; *** significant at 1%
3) Dependent variables are logarithmically transformed
4) Reference variables: UO - Municipality; PS - GW (S. of Fall Line)

Table 25: Regression Results for Control Factors in the Water Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
Rainfall (inches/year)	-4.41E-03	-4.76E-03	-4.34E-03	-3.71E-03
	[4.144e-03]	*[3.771e-03]*	*[3.779e-03]*	*[3.850e-03]*
Temperature (°F)	-0.020**	-0.020**	-0.020**	-0.019**
	[0.010]	*[0.009]*	*[0.009]*	*[0.009]*
Median HH income ($/year)	-2.90E-07	1.02E-06	1.70E-06	2.38E-06
	[2.551e-06]	*[2.322e-06]*	*[2.327e-06]*	*[2.371e-06]*
RS: Decreasing block?	0.027	0.06	0.048	0.021
	[0.070]	*[0.063]*	*[0.064]*	*[0.065]*
RS: Flat fee?	0.163	-0.164*	-0.417***	-0.615***
	[0.109]	*[0.099]*	*[0.099]*	*[0.101]*
RS: Increasing block?	0.087**	0.081**	0.104**	0.134***
	[0.044]	*[0.040]*	*[0.040]*	*[0.041]*
Year of last rate change	0.023***	0.022***	0.022***	0.023***
	[0.005]	*[0.005]*	*[0.005]*	*[0.005]*
Constant	-42.123***	-37.827***	-36.114***	-36.365***
	[11.258]	*[10.245]*	*[10.266]*	*[10.460]*
Observations	391	391	391	391
Adjusted R-squared	0.332	0.441	0.501	0.542

Notes: 1) Standard errors in brackets
2) * significant at 10%; ** significant at 5%; *** significant at 1%
3) Dependent variables are logarithmically transformed
4) Reference variables: RS - Uniform block

Table 26: Regression Results for Cost Factors in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
Interbasin transfer?	0.088	0.088	0.084	0.082
	[0.079]	*[0.072]*	*[0.072]*	*[0.074]*
Median year homes built	-8.68E-04	-2.33E-03	-2.70E-03	-2.97E-03
	[3.216e-03]	*[2.916e-03]*	*[2.937e-03]*	*[3.000e-03]*
UO: Authority?	0.096	0.135	0.148	0.164
	[0.120]	*[0.109]*	*[0.109]*	*[0.112]*
UO: County?	0.056	0.036	0.032	0.027
	[0.122]	*[0.110]*	*[0.111]*	*[0.114]*
UO: Utility commission?	-0.069	-0.127	-0.155	-0.168
	[0.128]	*[0.116]*	*[0.117]*	*[0.119]*
Population growth	-1.986e-02***	-1.052e-02*	-6.74E-03	-4.51E-03
	[6.921e-03]	*[6.276e-03]*	*[6.322e-03]*	*[6.456e-03]*
Population density	-0.037	-0.04	-0.04	-0.039
(log persons/mi2)	*[0.037]*	*[0.034]*	*[0.034]*	*[0.035]*
Population with water (log)	-0.055**	-0.03	-0.019	-0.014
	[0.023]	*[0.021]*	*[0.021]*	*[0.021]*
PS: GW (N. of Fall Line)?	0.094	0.197**	0.245***	0.280***
	[0.099]	*[0.090]*	*[0.090]*	*[0.092]*
PS: SWP?	0.252**	0.373***	0.428***	0.459***
	[0.098]	*[0.089]*	*[0.090]*	*[0.092]*
PS: SW?	0.174*	0.268***	0.307***	0.330***
	[0.089]	*[0.081]*	*[0.081]*	*[0.083]*
Loans ($)	0.066	0.053	0.053	0.052
	[0.045]	*[0.041]*	*[0.041]*	*[0.042]*
Grants ($)	0.166***	0.136***	0.119**	0.112**
	[0.055]	*[0.050]*	*[0.050]*	*[0.052]*
OR: Water & sewer?	-0.086*	-0.086**	-0.089**	-0.089**
	[0.048]	*[0.043]*	*[0.044]*	*[0.045]*
OR: Water only?	-0.153**	-0.148**	-0.146**	-0.147**
	[0.064]	*[0.058]*	*[0.059]*	*[0.060]*
CR: Sewer only?	0.02	-0.035	-0.05	-0.06
	[0.171]	*[0.155]*	*[0.156]*	*[0.159]*
CR: Water & sewer?	0.058	0.013	-0.002	-0.012
	[0.042]	*[0.038]*	*[0.038]*	*[0.039]*
CR: Water only?	-8.65E-04	3.96E-03	9.51E-03	1.12E-02
	[8.735e-02]	*[7.921e-02]*	*[7.979e-02]*	*[8.148e-02]*
Connection fees (log $)	0.040***	0.048***	0.051***	0.052***
	[0.014]	*[0.013]*	*[0.013]*	*[0.013]*

Notes: 1) Standard errors in brackets
2) * significant at 10%; ** significant at 5%; *** significant at 1%
3) Dependent variables are logarithmically transformed
4) Reference variables: UO - Municipality; PS - GW (S. of Fall Line); OR - None; CR - None

Table 27: Regression Results for Control Factors in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
Rainfall (inches/year)	3.38E-03	2.38E-03	2.40E-03	2.59E-03
	[5.219e-03]	*[4.733e-03]*	*[4.767e-03]*	*[4.869e-03]*
Temperature (°F)	1.39E-02	7.06E-03	3.38E-03	1.86E-03
	[1.230e-02]	*[1.115e-02]*	*[1.123e-02]*	*[1.147e-02]*
Median HH income ($/year)	-2.21E-07	-1.36E-06	-1.85E-06	-1.74E-06
	[3.373e-06]	*[3.059e-06]*	*[3.081e-06]*	*[3.146e-06]*
RS: W&S-Decreasing?	0.057	0.087	0.087	0.077
	[0.111]	*[0.101]*	*[0.102]*	*[0.104]*
RS: W-Decreasing,S-Uniform?	0.131	0.132	0.114	0.116
	[0.147]	*[0.133]*	*[0.134]*	*[0.137]*
RS: W&S Increasing?	0.058	0.099*	0.144***	0.181***
	[0.059]	*[0.053]*	*[0.054]*	*[0.055]*
RS: W-Increasing,S-Uniform?	0.161**	0.117	0.103	0.109
	[0.082]	*[0.074]*	*[0.075]*	*[0.076]*
RS: W&S-Flat?	0.320*	-0.038	-0.299*	-0.500***
	[0.184]	*[0.167]*	*[0.169]*	*[0.172]*
RS: W-Uniform,S-Decreasing?	-0.251	-0.153	-0.089	-0.118
	[0.193]	*[0.175]*	*[0.176]*	*[0.180]*
RS: W-Uniform,S-Flat?	-0.193	-0.327***	-0.413***	-0.474***
	[0.128]	*[0.116]*	*[0.117]*	*[0.119]*
RS: W-Uniform,S-Cap?	0.07	0.09	0.071	0.003
	[0.126]	*[0.115]*	*[0.115]*	*[0.118]*
Year of last rate change	0.024***	0.022***	0.021***	0.022***
	[0.007]	*[0.006]*	*[0.006]*	*[0.007]*
Constant	-44.063***	-36.202**	-33.964**	-33.801**
	[15.828]	*[14.353]*	*[14.458]*	*[14.765]*
Observations	269	269	269	269
Adjusted R-squared	0.217	0.394	0.485	0.538

Notes: 1) Standard errors in brackets
2) * significant at 10%; ** significant at 5%; *** significant at 1%
3) Dependent variables are logarithmically transformed
4) Reference variables: RS - W&S-Uniform

Interpretation of the Coefficients

In general, the coefficients are interpreted as the unit change in the dependent variable per unit change in the predictor when all other predictors are constant. Since the bills are logarithmically transformed, the coefficients are interpreted instead as an approximate percent change of the bill per unit change in the predictor. To find the actual percent change, the following adjustment is needed:

$$\%\Delta y = 100 * [\exp(\tilde{\beta}_j) - 1]$$

$$j = 1, 2, \ldots, J$$

where $\%\Delta y$ is the percent change in the bill and $\tilde{\beta}_j$ is the coefficient for the *j*th variable. For small values, the coefficient is almost the same as the actual percent change. If the independent variables are also logged, than the coefficient is an elasticity: the percent change in the bill per percent change in the independent variable. For example, if the coefficient is 0.11, then the bill increases by 0.11% per 1% increase in the independent variable.

The following examples help illustrate the interpretation of the coefficients. All the percentages mentioned are from the unadjusted coefficients (Tables 24 and 25) in the water model. The actual percent change is shown in Tables 28 and 29. The first example is of a raw, continuous predictor, such as *year of last rate change* in Table 25. In the 12000 GM model, an increase in this variable by one year is associated with an increase in the water bill of 2.3%. In other words, a bill with a rate schedule effective in any year (say 2005) is 2.3% higher than a bill with a rate schedule effective in the previous year (say 2004). With dummy variables, the coefficients are interpreted as the percent change of the bill if the dummy variable is true versus false. For example, when all independent variables are constant in the model at 12000 GM, utilities with sewer service (Table 24) generally have water bills that are about 10.8% lower than the water bills for utilities with no sewer service. With categorical variables, the coefficient is relative to the reference variable. So, for example, controlling for all other factors, authorities generally have water bills that are about 14.3% higher than the water bills for municipalities. When predictors are also logarithmically transformed, or if the units are in percent, the coefficients are interpreted as elasticities. In the model for 12000 GM, the bill increases by approximately 0.042% as the *population density* variable increases by 1%.

Highly Significant Variables in the Water Model

Highly significant variables in this report are defined as being significant at the 5% level for at least three consumption levels. The following table shows the actual percent change in the bill for variables that are statistically significant at three or four consumption levels.

Table 28: Marginal Effects of Highly Significant Variables – Water Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
Population with water (10^4)	-0.71%**	-0.69%**	-0.72%**	-0.76%**
Sewer service?	-10.1%**	-10.1%**	-10.4%***	-10.2%**
Temperature (°F)	-2.0%**	-2.0%**	-2.0%**	-1.9%**
RS: Increasing block?	9.1%**	8.4%**	11%**	14.3%***
Year of last rate change	2.3%***	2.2%***	2.2%***	2.3%***
UO: Utility commission?	-17.8%*	-19.1%**	-20.5%**	-21.5%**
Population density (log)	-0.084%***	-0.064%***	-0.051%**	-0.042%*
PS: SWP?	16.1%*	31.4%***	39.2%***	43.9%***
Grants?	13.2%***	10.5%**	9%**	8.2%*
Connection fees (log $)	0.02%*	0.026%**	0.026%**	0.026%**
PS: GW (N. of Fall Line)?	3.8%	16.3%**	22.8%***	27.4%***
PS: SW?	6.2%	20%***	26.5%***	30.5%***

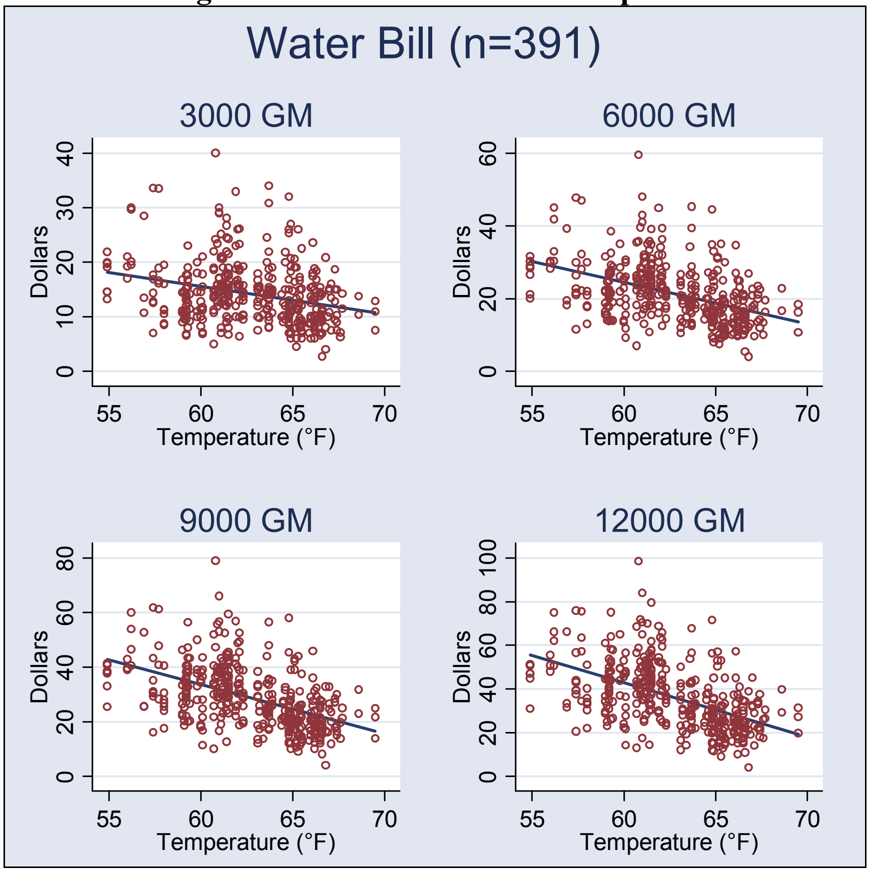
Notes: 1) * significant at 10%; ** significant at 5%; *** significant at 1%
2) Dependent variables are logarithmically transformed

The signs for most coefficients meet the predictions in Table 4, except *grants*, *connection fees*, *utility commission*, and *temperature*. Grants were predicted to have a negative effect on the bill. Since a grant displaces some loans and/or bonds, it was expected that utilities would pass the savings to their customers. A possible explanation is that GEFA and USDA provide assistance to utilities that already have progressive financial practices. Another explanation is that utilities may be attempting to maximize all of their revenue sources in order to pay for capital improvements. This reason may also be true for the *connection fees*, which Hollman and Boyet's hypothesize for this variable (see Chapter 2).

Regarding *utility commissions*, it was expected that the appointed board of directors could charge high rates without fear of being voted out of office. In other words, it was predicted that utility commissions would have higher rates than their counterpart: municipalities. However, the opposite is true. A possible explanation is that utility commissions may subsidize the water bill through other utility services, such as power, gas, and/or cable. Seven out of the 10 utility commissions in the sample provide these services. On the other hand, many of the municipally-owned utilities could be subsidizing other government departments, which is why the bill is higher. However, many W&S utilities do not have positive operating ratios, so this hypothesis is unlikely (Thorsten et al., 2007).

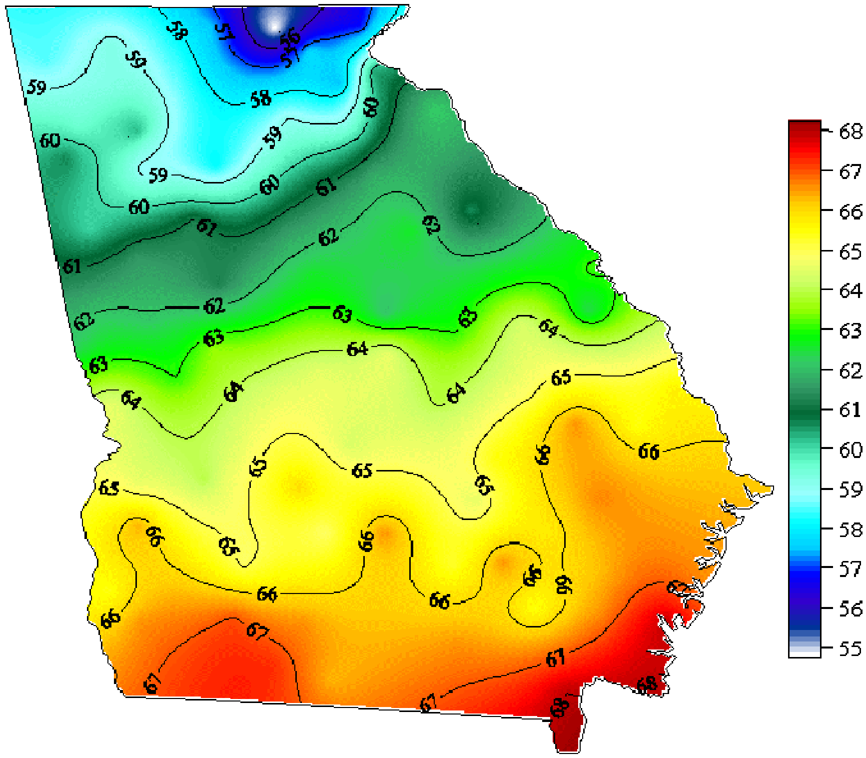
Regarding *temperature*, it was assumed that hotter climates would have less water available, so the cost of service and residential bill would rise. But, as the next figure shows, the direct relationship between temperature and the bill is clearly negative.

Figure 18: Water Bill versus Temperature



Even when controlling for source water, rainfall, population density (urban versus rural areas), and population, as well as other factors, the temperature variable still holds this negative relationship with the water bill. One explanation is that the variable may be highly correlated to another meaningful variable that is omitted from the water model. This omitted variable might be regional as the temperature increases fairly regularly from the northwest to the southeast, as the following figure shows.

Figure 19: Average Annual Temperature in Georgia (1961-1990)



Source: Georgia Automated Environmental Monitoring Network

The different regions of temperature nearly match Georgia's physiographic regions (Figure 6). Ostensibly, though, source water is associated with the physiographic area (Figure 7), so this regional information is already included in the model.

Another explanation could be that the method for assigning temperature data to utilities is incorrect. If the assignments were based on isograms and not on the closest weather station, the resulting sign in the water model could be different. At any rate, it is uncertain if this is the cause, the variable is correlated to an omitted variable, or if another explanation exists. Further research would be necessary.

Moderately Significant Variables in the Water Model

The following table shows the actual percent change in the bill for variables that are statistically significant at the 10% level at one or more consumption.

Table 29: Marginal Effects of Moderately Significant Variables – Water Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
UO: Authority?	11.4%	13.3%*	14.1%**	15.4%**
Population growth (% per year)	-2.1%***	-1.4%**	-1.1%*	-0.8%
RS: Flat fee?	17.7%	-15.1%*	-34.1%***	-45.9%***
Interbasin transfer?	11.4%	12.1%*	11.9%*	12%*
Loans?	6.5%	7.0%*	7.4%*	7.6%*
Outside rates?	-6.2%*	-5.3%	-4.7%	-4.4%

Notes: 1) * significant at 10%; ** significant at 5%; *** significant at 1%
2) Dependent variables are logarithmically transformed

All significant coefficients have the predicted sign discussed in Table 4. There is a moderate difference between bills from *authorities* and bills from *municipalities*, especially at high consumption levels. The board members of authorities are appointed and can charge high rates without fear of being voted out of office. In addition, municipally-owned utilities can be subsidized by other government departments. The *population growth* variable matters only for low levels of consumption. Utilities with high growth rates may be passing on the savings due to economies of scale to impoverished and/or elderly customers who generally consume less water. Utilities with *flat fees* have much lower bills than utilities with *uniform blocks*. This makes sense because bills which are simply flat fees are the same at any consumption, whereas bills with uniform rate structures continue to increase with consumption. The *interbasin transfers* variable is weakly significant for 6000-12000 GM, and the variable is price inelastic. The *loans* variable is also weakly significant at those levels. Residential water bills for inside customers have little association with higher bills for outside users. In other words, it does not seem that utilities are charging inside residential customers any differently, even though the utilities are receiving increased revenue from outside customers.

Non-significant Variables in the Water Model

Some coefficients were not significant at any consumption level. The following table lists these variables by group.

Table 30: Variables with Non-Significant Coefficients at All Consumption Levels – Water Model

Variable Group	Independent Variables
Cost Factors	Median year homes built UO: County? Commercial tap rates? RS: Decreasing block?
Control Factors	Rainfall (inches/year) Median HH income

The *median year homes built* variable obviously is not the best proxy for the age of a water and sewer system in Georgia. However, Figure 11 shows the variable has an even distribution, which could mean that old and new systems counteract the effect on the bill. As far as *commercial rates*, residential water bills for inside customers are not associated. The model shows that water bills with *decreasing blocks* are no different than bills with *uniform blocks*. Even though the mean water bills are different, especially at 12000 GM ($36 for uniform blocks versus $29 for decreasing blocks), the models show these values are not statistically different from each other. Despite the variability of *rainfall*, this has no impact on the water bill. Finally, areas with high *median household incomes* do not have different water bills than areas with lower incomes.

Examination of Categorical Variables in the Water Model

The categorical variables in the water model – utility ownership, primary source, and rate structures – have municipalities, GW (south of the Fall Line), and uniform block rates, respectively, as their reference variables. It would be interesting, though, to know the following comparisons: authorities vs. counties, SW vs. SWP, and decreasing block rates and flat rates vs. increasing block rates. Since authorities are the semi-autonomous counterpart to counties, we should know if the bills are different. Also, it would be good to know if the bills are different between the two types of surface water. Finally, the conventional wisdom says increasing block rates promote conservation by increasing the marginal price to consumers, and decreasing blocks and flat fees have the opposite effect. The following table shows the regression results from switching the reference variables. Only the categorical variables are shown because the coefficients and standard errors for the other variables in the regression are exactly the same.

Table 31: Alternative Regression Results for the Categorical Variables in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
UO: Authority?	-0.004	0.034	0.047	0.057
	[0.086]	*[0.078]*	*[0.078]*	*[0.080]*
UO: Municipality?	-0.112	-0.09	-0.084	-0.086
	[0.083]	*[0.075]*	*[0.076]*	*[0.077]*
UO: Utility commission?	-0.308**	-0.302**	-0.314***	-0.329***
	[0.131]	*[0.119]*	*[0.119]*	*[0.122]*
PS: GW (N. of Fall Line)?	-0.111*	-0.122**	-0.126**	-0.122**
	[0.065]	*[0.059]*	*[0.059]*	*[0.061]*
PS: GW (S. of Fall Line)?	-0.149*	-0.273***	-0.331***	-0.364***
	[0.076]	*[0.069]*	*[0.069]*	*[0.070]*
PS: SW?	-0.089	-0.092*	-0.096*	-0.097*
	[0.057]	*[0.052]*	*[0.052]*	*[0.053]*
RS: Decreasing block?	-0.06	-0.021	-0.056	-0.112
	[0.079]	*[0.071]*	*[0.072]*	*[0.073]*
RS: Flat fee?	0.076	-0.245**	-0.521***	-0.748***
	[0.116]	*[0.105]*	*[0.106]*	*[0.108]*
RS: Uniform block?	-0.087**	-0.081**	-0.104**	-0.134***
	[0.044]	*[0.040]*	*[0.040]*	*[0.041]*

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Notice there is little difference between the bills from authorities and counties. Despite the fact that authorities have appointed board members, everything else being equal, the bills are not different from the bills of counties. And like municipalities, counties have bills that are higher than bills from utility commissions (see the subsection "*Highly Significant Variables in the Water Model*" for an explanation). The difference between bills for utilities with SW and SWP is weak at 6000-12000 GM. Surprisingly, there is no difference between bills for utilities with decreasing blocks and those with increasing blocks. Yet there is a difference between bills for utilities with uniform blocks and those with increasing blocks. It was assumed that the difference between increasing blocks and decreasing blocks would be greater than the difference between increasing blocks and uniform blocks. Finally, as expected, there is a major difference between bills for utilities with flat fees versus those with increasing blocks.

Highly Significant Variables in the W&S Model

The following table shows the actual percent change in the bill for variables that are statistically significant at the 5% level for three or four consumption levels. Unlike the previous model, the dependent variable for the W&S model is not transformed, so the coefficients are interpreted as dollars per unit (or percent) change in the variable. The first seven variables have statistically significant coefficients at the 5% level for four consumption levels, and the rest have three or four coefficients at the 10% level.

Table 32: Marginal Effects of Highly Significant Variables in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
PS: SWP?	28.7%**	45.2%***	53.4%***	58.2%***
Grants ($)	18.1%***	14.6%***	12.6%**	11.9%**
OR: Water only?	-14.2%**	-13.8%**	-13.6%**	-13.7%**
Connection fees (log $)	0.04%***	0.048%***	0.051%***	0.052%***
Year of last rate change	2.4%***	2.2%***	2.1%***	2.2%***
PS: SW?	19%*	30.7%***	35.9%***	39.1%***
OR: Water & sewer?	-8.2%*	-8.2%**	-8.5%**	-8.5%**
PS: GW (N. of Fall Line)?	9.9%	21.8%**	27.8%***	32.3%***
RS: W-Uniform,S-Flat?	-17.6%	-27.9%***	-33.8%***	-37.7%***

Notes: 1) * significant at 10%; ** significant at 5%; *** significant at 1%
2) Dependent variables are logarithmically transformed

The signs for most coefficients meet the predictions in Tables 4 and 5, except *grants* and *connection fees*, which was the same result for the water model. The explanations for that model equally fit the W&S model. The highly significant variables *SWP*, *SW*, *grants*, *connection fees*, *year of last rate change*, and *GW (north of the Fall Line)* in the above table are also highly significant in the water model (Table 28). *Outside rates* for water and sewer customers as well as for water customers only are highly significant, which is the opposite from the water model. Oddly, utilities with uniform water rates and flat sewer fees are more significant than utilities with flat W&S fees (see next table).

Moderately Significant Variables in the W&S Model

The following table summarizes the change in the bill for statistically significant variables at the 10% level.

Table 33: Marginal Effects of Moderately Significant Variables in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
RS: W&S Increasing?	6%	10.4%*	15.5%***	19.8%***
RS: W&S-Flat?	37.7%*	-3.7%	-25.8%*	-39.3%***
Population growth	-2%***	-1%*	-0.7%	-0.5%
RS: W-Increasing,S-Uniform?	17.5%**	12.4%	10.8%	11.5%

Notes: 1) * significant at 10%; ** significant at 5%; *** significant at 1%
2) Dependent variables are logarithmically transformed

And, *increasing W&S block rates* are significant at the higher consumption levels when compared to uniform W&S block rates. *W&S-Flat* fees are weakly significant at the consumption extremes and have opposite signs. The sign is positive at 3000 GM, which makes sense because flat fees are probably based on the average consumption of 6000 GM. So, it is expected that flat fees for lower consumption levels would be higher than the bills for block rates. The significance of the *population growth* factor is similar to the water model (Table 29). The last variable in the table has little association with the W&S bill.

Non-significant Variables in the W&S Model

Some coefficients were not significant at any consumption level. The following table lists these variables by group.

Table 34: Variables with Non-Significant Coefficients at All Consumption Levels – W&S Model

Variable Group	Independent Variables
Cost Factors	Interbasin transfer? Median year homes built UO: Authority? UO: County? UO: Utility commission? Population density Population with water (log) Loans ($) CR: Sewer only? CR: Water & sewer? CR: Water only?
Control Factors	Rainfall (inches/year) Temperature (°F) Median HH income ($/year) RS: W&S-Decreasing? RS: W-Decreasing,S-Uniform? RS: W-Uniform,S-Decreasing? RS: W-Uniform,S-Cap?

The result for *median year homes built* reinforces the fact that this is not an adequate proxy for the age of the system, or that the even mix of old and new systems counteracts the effect on the bill. The bills for *authorities*, *counties*, and *utility commissions* had no difference with the bills for *municipalities*. Surprisingly, *population* and *population density* have no significant coefficients at any consumption level; a stark contrast to the water model where the variables were significant for all consumption levels. *Loans* were weakly significant in the water model and not significant at all in the W&S model. As in the water model, *commercial rates*, *rainfall*, and *median household income* were not associated with the bill. Contrary to the water model, *temperature* was not associated with the bill. This suggests that future research should focus on an omitted variable correlated to temperature which is relevant for water bills only. The rate structure variables in the table have at least one service with decreasing blocks. This includes utilities with sewer caps where all have decreasing sewer block rates capped at an average of 9000 GM. Even when these variables are combined into one, it is not significant when compared to uniform W&S blocks.

Examination of Categorical Variables in the W&S Model

Similar to the water model, this section makes the following comparisons: authorities vs. counties, SW vs. SWP, and decreasing W&S block rates and flat W&S fees vs. increasing W&S block rates. The following table shows the results from switching the reference variables. Only the categorical variables are shown because the coefficients and standard errors for the other variables in the regression are exactly the same.

Table 35: Alternative Regression Results for the Categorical Variables in the W&S Model

Independent Variables	3000 GM	6000 GM	9000 GM	12000 GM
UO: Authority?	1.527	6.138	10.028*	15.125**
	[3.195]	[4.299]	[5.816]	[7.470]
UO: Municipality?	-0.63	-0.718	-1.517	-1.927
	[3.343]	[4.498]	[6.086]	[7.816]
UO: Utility commission?	-2.223	-4.716	-8.306	-11.519
	[4.233]	[5.695]	[7.705]	[9.896]
PS: GW (N. of Fall Line)?	-4.291*	-8.516***	-12.606***	-15.972***
	[2.384]	[3.208]	[4.340]	[5.574]
PS: GW (S. of Fall Line)?	-7.234***	-16.122***	-24.811***	-33.242***
	[2.696]	[3.627]	[4.907]	[6.302]
PS: SW?	-1.572	-4.179*	-6.855**	-9.298**
	[1.856]	[2.497]	[3.379]	[4.340]
OR: Water & sewer?	1.383	1.637	1.568	1.673
	[1.718]	[2.311]	[3.127]	[4.017]
OR: None	4.103**	5.757**	7.390**	9.037**
	[1.760]	[2.368]	[3.204]	[4.115]
CT: Sewer only?	-0.434	-2.956	-4.931	-6.077
	[5.204]	[7.001]	[9.472]	[12.166]
CT: Water & sewer?	1.092	-0.723	-2.355	-3.841
	[2.431]	[3.271]	[4.426]	[5.684]
CT: None	-0.5092	-0.9321	-1.222	-1.208
	[2.397]	[3.224]	[4.363]	[5.603]
RS: W&S-Decreasing?	1.101	0.573	-2.54	-6.952
	[3.353]	[4.510]	[6.102]	[7.838]
RS: W-Decreasing,S-Uniform?	0.942	-0.047	-3.942	-7.943
	[4.259]	[5.729]	[7.751]	[9.956]
RS: W&S-Uniform?	-1.554	-4.080*	-8.297***	-13.391***
	[1.612]	[2.169]	[2.935]	[3.770]
RS: W-Increasing,S-Uniform?	3.009	0.595	-3.101	-6.444
	[2.542]	[3.419]	[4.626]	[5.942]
RS: W&S-Flat?	8.338	-4.303	-18.871**	-34.122***
	[5.237]	[7.045]	[9.532]	[12.243]
RS: W-Uniform,S-Decreasing?	-6.488	-7.476	-8.928	-15.147
	[5.440]	[7.318]	[9.901]	[12.717]
RS: W-Uniform,S-Flat?	-5.768	-13.972***	-24.035***	-35.143***
	[3.769]	[5.070]	[6.860]	[8.810]
RS: W-Uniform,S-Cap?	0.918	-0.097	-4.171	-12.67
	[3.719]	[5.003]	[6.769]	[8.694]

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Contrary to the water model, there is a slight difference between the bills from authorities and counties at high consumption levels and a moderate difference between bills for utilities with *SW* and those with *SWP*. A similarity is that there is no difference between bills for utilities with decreasing W&S blocks and those with increasing W&S blocks. Even when the variables with decreasing water and/or sewer block rates (including utilities with sewer caps) are combined into one variable, there is no significance. Also, the *flat fee* variables are significant compared to increasing W&S blocks, especially at high consumption levels.

Summary of the Statistically Significant Variables

Table 36 summarizes the statistically significant factors in both models at three or more consumption levels. These are the most robust associations with the bills. Any comparison of Georgia utilities' rates should have, at the least, these factors included. The table lists only the coefficients for 6000 GM to illustrate the relative magnitude of the relationship with the bills in both models.

Table 36: Highly Significant Variables in Both Models (6000 GM)

Independent Variables	Water Model	W&S Model
PS: GW (N. of Fall Line)?[4]	-11.5%**	-16.1%**
PS: GW (N. of Fall Line)?[3]	16.3%**	21.8%**
PS: SW?[3]	20%***	30.7%***
PS: SWP?[3]	31.4%***	45.2%***
Grants?	10.5%**	14.6%***
Connection fees (log $)	0.026%**	0.048%***
Year of last rate change	2.2%***	2.2%***

Notes: 1) ** significant at 5%; *** significant at 1%
2) Dependent variables are logarithmically transformed
3) Relative to GW (S. of Fall Line)
4) Relative to SWP

The following variables are non-significant in both models at all consumption levels: *median year homes built*, *commercial rates*, *rainfall*, and *median household income*. Simple comparisons of two utilities' rates probably should not consider these factors. In addition, *median year homes built* is probably not an adequate proxy for the age of a water and sewer system. Also, commercial rates do not appear to be subsidizing the bills for residents inside the utilities' jurisdictions.

The categorical variables *utility ownership* and *rate structures* also have a few categories that are not significant in both models at all levels: *authority* vs. *county*, *county* vs. *municipality*, *decreasing block* vs. *increasing block*, and *decreasing block* vs. *uniform block*. However, these categorical variables should be included in the simple comparisons because other categories (i.e., *authority* vs. *municipality* and *increasing block* vs. *uniform block*) are moderately significant.

Comparing the actual results with the predictions from Tables 4 and 5, the signs on *grants* (+), *connection fees* (+), *utility commissions* (–), and *temperature* (–) were the opposite of expected. The fact that *grants* and *connection fees* are associated with an increase in the bill is probably a sign of progressive financial practices. At the least, it does not appear the utilities use these revenue streams to subsidize the bill for residents inside their jurisdiction. Hollman and Boyet (1975) found *grants and subsidies* (–) and *connection fees* (+) to be significant; however, the former variable includes government subsidization, so it is not comparable to *grants* in this report. In addition, the donor requirements have changed since the 1970's when grants were more likely to subsidize the residential bill.

Moving on, *utility commissions* and *temperature* are significant in the water model only. *Utility commissions* probably have lower water bills because they provide other utility services which could subsidize the bill. It is possible that, instead, municipally-owned utilities have higher bills because they subsidize other government departments, but this is unlikely because operating expenses exceed operating revenue for many utilities. The *temperature* variable is probably correlated to another variable omitted from the model. However, it is possible that the method for assigning temperature data to each utility was incorrect.

Other surprises include the fact that decreasing blocks compared to both increasing and uniform blocks are not significant, yet increasing blocks compared to uniform blocks are significant. The assumption is that bills with increasing blocks are the most expensive followed by bills with uniform blocks and then decreasing blocks. So, a difference exists between the most expensive block (increasing block) and the midrange block (uniform block), but not between the most expensive block and the least expensive block (decreasing block). It is unclear the mechanism behind this, and further research would be necessary.

Another interesting result is that bills from *authorities* are not significantly different than bills from *counties* in both models. This contradicts the hypothesis that county commissioners may feel compelled to keep rate increases to a minimum in order not to upset voters. Also, *outside rates* are not significant in the water model, but they are highly significant in the W&S model. So, higher outside rates for water and W&S may be subsidizing inside W&S bills. This effect was also found in Thorsten et al. (2007).

CHAPTER 5

DISCUSSION

This chapter summarizes the entire report, findings, and limitations of the regression models. The first section summarizes each chapter in one paragraph. The second section discusses the general implications of the model estimates for each utility. It also summarizes three important findings: authorities and counties do not have significantly different bills, bills with decreasing blocks are not significantly different from bills with uniform or increasing blocks, and grants are actually associated with higher rates. The third section summarizes the limitations of the analysis. The final section discusses the use of significant variables in this report to augment future rate surveys.

Report Summary

Currently, utilities compare their rates directly to other utilities' rates and/or to the unconditional regional average. This helps them gauge a fair price increase and measure their performance, but these comparisons do not account for factors that influence the bill. As such, the current method is like comparing apples to oranges; each utility has different reasons for the value of their bill. This report demonstrates the use of multiple regression models to incorporate the influential factors and provide estimates of average bills for each utility's unique class. Then, each utility can compare more accurately their actual bill against this average measure. The models also show the factors associated with the bill, which reveal a few policy implications.

To make comparisons, utilities use rate surveys, which are regional compilations of utilities' rates. One of the first rate surveys in the US was produced in the early 1970's, and the number of rate surveys has increased exponentially since then. As rate setting techniques and rate surveys have advanced, a conventional wisdom has developed around a group of factors that influence the rates (Raftelis 2005). This group was the basis of the independent variables in the models. Through multiple regression, researchers (Hollman and Boyet 1975; Mann 1972; Thorsten, Eskaf and Hughes 2007) have verified the predicted effect of some of these variables, but they did not estimate each utility's bill. Finally, the full sample (n = 415) in this report consists of Georgia utilities defined as CWS's owned by local governments. These utilities serve 93% of the population (6.9 million people) using water from CWS's.

In this report, two types of bills – water only and combined water and sewer – are modeled at four consumption levels: 3000, 6000, 9000, and 12000 GM. The first model provides estimates for utilities that serve water only, and the second model provides estimates to utilities that serve at least water. The different consumption levels mainly test the sensitivity of the estimates and statistically significant factors. The primary data comes from the GEFA-EFC survey conducted in late 2006, and the secondary data comes from Ga EPD, GEFA, USDA, NCDC, SDWIS, and US Census. To be included in the sample for either model, utilities needed 1) at least water service, 2) enough observations ($n \geq 2$) for each dummy variable, 3) data for all variables in the model, and 4) rate structures consistent throughout the modeled consumption levels and throughout the year.

The models were developed and tested by 1) transforming variables, 2) running regressions with different combinations of raw and transformed variables, 3) selecting models that passed specification tests, 4) selecting models with VIF less than 10, 5) checking homoscedasticity and normality of the residuals, and 6) selecting the model with the highest adjusted R^2.

The resulting models produced estimates of the average bill for each utility along with a 95% confidence interval. A table of the actual bills, estimates, and confidence intervals for each model and consumption level can be found online (see Appendix B). For most utilities in each model, the actual bill in relation to the confidence interval was consistent across three or four consumption levels (below CI, within CI, or above CI). The following variables in both models were statistically significant at the 5% level across three or four consumption levels: all of the source water variables in relation to *GW (south of the Fall Line)* and *SWP* (except *SW* for the latter); *year of last rate change* (+); *grants* (+); and *connection fees* (+). All source waters were positively associated with the bill when compared to *GW (south of the Fall Line)* and both GW's were negatively associated with the bill when compared to *SWP*. Compared to *SWP*, *SW* was significant at the 10% level for 6000, 9000, &12000 GM in both models. *Median year homes built*, *commercial rates*, *rainfall*, and *median household income* had no significance in both models at any consumption level.

Implications of the Findings

For many utilities, the model estimates of average bills are closer to the actual bill than the state average. Utilities with relatively high bills could then feel more open to increase rates if necessary. Likewise, utilities with relatively low bills may feel less inclined to raise rates and/or to advertise themselves as low-cost service providers. Also, for a number of utilities, their bills may actually be "below average" (below the confidence interval) even if they are greater than the state average, and vice-versa. In addition, bills for some utilities may be considered "average" (within the confidence interval) even if they are considerably higher/lower than the state average. The regression models out perform rules-of-thumb comparisons and show that several more factors should be considered when contrasting utility rates.

Interesting policy implications include the fact that authorities' bills are not significantly different than counties' bills. From the perspective of management and efficiency, authorities could have better operating ratios than their counterparts (counties), but this does not necessarily mean they charge higher bills, everything else being equal. Perhaps controlling for operating ratio might show a difference between authorities and counties.

In addition, bills with increasing block rates are significant and higher than bills with uniform block rates. Paradoxically, though, the bills with decreasing block rates are no different than bills with increasing or uniform block rates. While the first finding is good news for organizations and utilities that want to promote increasing block rates as a conservation measure (the overall bill is higher), the second finding does not seem sensible. If there is a difference between increasing block rates and the next lowest rates (uniform block), then there also should be a difference with the lowest rates (decreasing block). In any case, the fact that bills for increasing block rates are high does not mean that utilities switching to this rate structure will automatically have higher bills. The model only confirms that utilities currently with increasing blocks have higher overall bills, which is the desired result for many proponents of this rate structure.

Grants have the opposite effect predicted in Raftelis (2005). It was assumed that utilities would use grants to subsidize the residential bill, and this was, to some extent, confirmed by Hollman and Boyet. However, since their study was conducted during the 1970's when government grants for W&S systems were more prevalent, the result is not surprising (CBO, 2002). Since then, federal and state governments have steadily reduced grants and required progressive financial practices. The models show a strong, positive correlation with both types of bills, and this may confirm the new trend in grants.

Limitations

While the regression models met the goals of this reports, there are a number of limitations to consider. These include the use of proxy variables instead of the actual variables listed in Raftelis (2005), manipulation of certain variables, omitted variables, and specification. Nearly all variables in both models were proxies, except *grants* and *rate structures*. The *loans*, *grants*, *temperature*, *rainfall*, *rate structures*, and *year of last rate change* data needed to be manipulated for use in the models. Omitted variables include geography, number of commercial customers, infiltration and inflow, and rate-setting approach from Raftelis. Other variables could be correlated to *temperature* and may explain its negative sign. As far as specification, both models had logged dependent variables with slightly lower goodness-of-fit measures than the models with raw variables. Also, spatial autocorrelation was not considered in this report.

Proxy variables, such as *median year homes built* for the age of the system, were needed because the actual data were inaccessible. Many of these variables proved to be significant, such as the *source water* variables, *grants*, and *connection fees*. However, *median year homes built* was not significant in either model, and *population* was only significant in the water model. Perhaps if the variables in Raftelis ("age of the system" and "peak demand") corresponding to the previously mentioned variables were available, then they would be significant.

As far as the manipulation of the variables, *loans* and *grants* are dichotomous variables based on USDA and GEFA money executed within an arbitrary timeframe. A better approach would be to use the amounts given on utilities' financial audits, but this information was not conveniently available. Utilities were assigned to the closest weather stations with *temperature* and *rainfall* data, but it might be better to use isograms. The closest station may not always be in the same microclimate as a utility. The variables for the rate structures were a mix between effective and universal rate structures where the former accounts for the consumption level of the model and the latter is based on infinite consumption. All rate structures above 12000 GM were considered uniform and the rest were designated according to their block structure, regardless of the consumption modeled. It may be more accurate to use effective rate structures for all models, but testing both methods was outside the scope of this report. The *year of last rate change* variable had a significant amount of missing data, and the simplest procedure was to impute the mean value. While this method has become obsolete in the face of more robust techniques, such as multiple imputation, the latter was beyond the scope of the report. An alternative was to remove the utilities with missing years. But, mean substitution seemed to be adequate since both models had practically the same results when the data was omitted. In the future, studies should consider the modern procedures if resources permit.

Data for other variables from Raftelis – geography, number of commercial customers, infiltration and inflow, and rate-setting approach – could not be collected. While it is uncertain if they would be significant, these variables could still be affect the significance of other variables and the estimated average bills. In addition, one of these variables could be correlated to the *temperature* variable, which has a sign (negative) opposite of expectations (see Table 4).

Both models contained logged dependent variables because slightly more specification tests were passed and the scatter and normality of the residuals appeared better than the models with untransformed dependent variables. However, the models with logged dependent variables had a slightly lower goodness-of-fit. This may cause the confidence intervals for the average bill to be larger, which makes this estimate more conservative. In other words, more utilities' actual bills will fall within the confidence interval. In addition, retransforming the prediction into meaningful units slightly biases the estimate. Finally, utilities' rates may be correlated with their neighbors' rates (spatial autocorrelation), but accounting for this was beyond the scope of this report.

Using Significant Variables for Rate Studies

As the report demonstrates, regression modeling can incorporate several factors into an estimated average bill for each utility. However, some organizations that conduct rate surveys may have limited resources to perform a full regression analysis. In this case, the organizations may want to include the following information for each utility in their rate surveys: source water type, grants (amount or yes/no), connection fees, and year of last rate change. Then utilities could search for other utilities with the closest characteristics and compare rates. However, this will still be a rough process because few utilities will have the same values for all four characteristics. In addition, the comparisons do not account for factors that may not be statistically significant, yet still influence the bill to some degree. The advantage of multiple regression is that a sample of utilities with the exact same characteristics is not needed for the analysis. The model can essentially interpolate the average bill for any utility.

APPENDIX A:

COVER LETTER TO UTILITIES

GEFA GEORGIA ENVIRONMENTAL FACILITIES AUTHORITY

PAUL R. BURKS
Executive Director

SONNY PERDUE
Governor

November 6, 2006

To Whom It May Concern:

The Georgia Environmental Facilities Authority (GEFA), in partnership with the Environmental Finance Center (EFC), is conducting a statewide survey of water and sewer rates and connection fees. The EFC has received a grant from the United States Environmental Protection Agency to collect this data from Southeastern states. The goal of this survey is to help Georgia's water and sewer systems by providing public officials with an accurate portrait of water and sewer rates across the state. Once the survey is concluded, all participating utilities will receive a copy of the results.

We would appreciate your assistance in this process by providing data on your water/sewer system to the EFC. Information should include, where appropriate:

- In-town and out-of-town rates
- Rates for all consumer classes (residential, commercial, industrial, etc.)
- Effective date (if known)

In addition, the rate schedule should include, where appropriate:

- Billing cycle (monthly, bi-monthly, etc.)
- Conservation-orientated rates (e.g. seasonal rates, peak rates or summer surcharges)
- Minimum charges, and if a minimum consumption amount is included

The connection fee schedule should contain one-time fees for new connections to water and/or sewer systems, which include, but are not limited to:

- Tap fees
- Meter fees
- Cost recovery charges
- Impact fees

If you do not have printed sheets, or the printed sheets are missing any items listed above, please provide this data on the attached "Supplemental Information" sheet. You can submit this information via:

1. Fax: (919) 843-2528
2. E-mail: efc@sog.unc.edu
3. Internet: http://www.efc.unc.edu/ga/rates.htm
4. U.S. Mail: Environmental Finance Center
   Post Office Box 671346
   Marietta, GA 30066

As you know, appropriately setting water and sewer rates is critical to the operation of your system. Rate schedules and fees determine whether a community will have sufficient funds to protect public health and the environment. By conducting this statewide survey, we hope to provide you with the information to better assist you in setting your rates. Your assistance with this survey is greatly appreciated. If you have questions, please contact Jason Bodwell at GEFA at (404) 584-1011 or the EFC - Andrew Westbrook (westbrok@sog.unc.edu) at (919) 966-4199 or Mark Horowitz (msh@email.unc.edu) at (919) 966-4199. We appreciate your assistance.

Sincerely,

Paul Burks

233 Peachtree Street, NE • Harris Towers, Suite 900 • Atlanta, Georgia 30303-1927 • 404/584-1000 • Fax 404/584-1069

APPENDIX B:

ESTIMATES OF THE AVERAGE BILLS

Please see http://msh345.googlepages.com/ModelEstimates.xls for a list of estimates from the water and W&S models for all utilities in the respective samples.

APPENDIX C:

ANALYSIS OF VARIANCE FOR THE OLS MODELS

WATER MODEL - 3000 GM (n=391)				F(24, 366)	=	9.07
Source	SS	df	MS	Prob > F	=	0
Model	22.2	24	0.93	R-squared	=	0.373
Residual	37.4	366	0.10	Adj R-squared	=	0.332
Total	59.6	390	0.15	Root MSE	=	0.319
WATER MODEL - 6000 GM (n=391)				F(24, 366)	=	13.81
Source	SS	df	MS	Prob > F	=	0
Model	28.0	24	1.17	R-squared	=	0.475
Residual	30.9	366	0.08	Adj R-squared	=	0.441
Total	59.0	390	0.15	Root MSE	=	0.291
WATER MODEL - 9000 GM (n=391)				F(24, 366)	=	17.33
Source	SS	df	MS	Prob > F	=	0
Model	35.3	24	1.47	R-squared	=	0.532
Residual	31.1	366	0.08	Adj R-squared	=	0.501
Total	66.4	390	0.17	Root MSE	=	0.291
WATER MODEL - 12000 GM (n=391)				F(24, 366)	=	20.24
Source	SS	df	MS	Prob > F	=	0
Model	42.8	24	1.78	R-squared	=	0.570
Residual	32.3	366	0.09	Adj R-squared	=	0.542
Total	75.1	390	0.19	Root MSE	=	0.297
W&S BILL MODEL - 3000 GM (n=269)				F(31, 237)	=	3.37
Source	SS	df	MS	Prob > F	=	0
Model	9.9	31	0.32	R-squared	=	0.306
Residual	22.6	237	0.10	Adj R-squared	=	0.215
Total	32.5	268	0.12	Root MSE	=	0.31
W&S MODEL - 6000 GM (n=269)				F(31, 237)	=	6.26
Source	SS	df	MS	Prob > F	=	0
Model	15.2	31	0.49	R-squared	=	0.450
Residual	18.5	237	0.08	Adj R-squared	=	0.378
Total	33.7	268	0.13	Root MSE	=	0.28
W&S MODEL - 9000 GM (n=269)				F(31, 237)	=	8.51
Source	SS	df	MS	Prob > F	=	0
Model	20.9	31	0.68	R-squared	=	0.527
Residual	18.8	237	0.08	Adj R-squared	=	0.465
Total	39.8	268	0.15	Root MSE	=	0.28
W&S MODEL - 12000 GM (n=269)				F(31, 237)	=	10.24
Source	SS	df	MS	Prob > F	=	0
Model	26.3	31	0.85	R-squared	=	0.572
Residual	19.6	237	0.08	Adj R-squared	=	0.517
Total	45.9	268	0.17	Root MSE	=	0.29

REFERENCES

Allison, Paul D. Missing Data. Vol. 07-136. Sage University Papers Series on Quantitative Applications in the Social Sciences, 2001.

AWWA. Principles of Water Rates, Fees, and Charges (M1). 5th edition ed., 2000.

Black & Veatch. California Water Rate Survey., 2006. <http://www.bv.com>.

Chen, X., P. Ender, M. Mitchell, and C. Wells. "Regression with Stata." 2003. <http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter1/statareg1.htm>.

City of Cartersville. "Water and Sewer Rates Very Low." The Cartersville Messenger Winter 2007.

Congressional Budget Office. Future Investment in Drinking Water and Wastewater Infrastructure. Washington, DC:, 2002. <http://www.cbo.gov>.

DAA. 18th Annual Virginia Water and Wastewater Rate Survey., 2006. <http://www.daa.com/news/surveys/2006-18thVA-Water-Wastewater-Rate-Report.pdf>.

Dallas Water Utilities Department. "Dallas Rate Survey." Journal of the American Water Works Association 67.5 (1975): 232-8.

Dalton Utilities. "Rates Lowest Statewide." 2007. <http://www.dutil.com/residential/water_rates.php>.

Donahue, E. J.,III. "Rate-Making: A Cooperative Process." Journal of the New England Water Works Association 110.3 (1996): 238-44.

EPA. Clean Watersheds Needs Survey 2000: Report to Congress., 2003.

---. "Public Drinking Water Systems: Facts and Figures." February 28, 2006 <http://www.epa.gov/safewater/pws/factoids.html>.

GEFA and EFC. Water and Wastewater Rates and Rate Structures in Georgia., 2007. <http://www.efc.unc.edu/ga/rates.htm>.

Georgia Municipal Association. 2005 Water and Wastewater Survey Findings., 2005. <http://www.gmanet.com/data/pdf/2005_watersewer.pdf>.

Hamilton, L. C. Statistics with Stata: Updated for Version 8. Belmont, CA: Brooks/Cole-Thomson Learning, 2004.

Hollman, K. W., and W. E. Boyet. "An Empirical Analysis of Water-Price Determinants in Small Municipalities." Journal of the American Water Works Association 67.5 (1975): 274-7.

Jordan, Jeffrey. 1994/95 Survey of Private Water Rates in Georgia. Vol. 96-18 (Faculty Series). Griffin, GA: University of Georgia, Department of Agriculture & Applied Economics, 1996.

---. 1998 Survey of Wastewater Rates and Facilities. Griffin, GA: University of Georgia, Department of Agriculture & Applied Economics, 1998.

Mann, Patrick C. "A New Focus in Water Supply Economics - Urban Water Pricing." Journal of the American Water Works Association 62.9 (1970): 534-7.

---. "Water Service Prices: A Principal Component and Regression Analysis of Determinants." Available from the National Technical Information Service as Pb-213 808 (1972) .

Manning, Paige S., Alan Barefield, and Jonathan Mays. Mississippi Survey of Drinking Water System Characteristics and Rate Structure. Department of Agricultural Economics, Mississippi State University Extension Service, 2005.

Meeus, Jean. "Astronomical Algorithms." Willmann-Bell, 1991. 84-85.

Metropolitan North Georgia Water Planning District. Water & Wastewater Rate Survey 2006. Atlanta Regional Commission - Environmental Planning Division, 2006. <http://www.northgeorgiawater.com/files/MNGWPD_2006_Rates_Survey.pdf>.

Metropolitan Water District of Orange County. Water Rates: Water System Operation and Financial Information., 2006. <http://www.mwdoc.com/documents/2006WaterRateSurvey.pdf>.

MWRA Advisory Board. Annual Water and Sewer Retail Rate Survey., 2006. <http://www.mwraadvisoryboard.com>.

National Association of Clean Water Agencies. 2005 Financial Survey - A National Survey of Municipal Wastewater Management Financing and Trends. Washington, DC: NACWA, 2006.

NCDC. Monthly Normals of Temperature, Precipitation, and Heating and Cooling Degree Days: 1971-2000 Georgia. Climatography of the United States No. 81., 2002.

Nelson, Charles R. "Applied Time Series Analysis for Managerial Forecasting." San Francisco: Holden-Day, 1973. 161-165.

New Hampshire Department of Environmental Services. 2004 Water Rate Survey - Larger Water Systems., 2005.

Ohio EPA. 2005 Sewer and Water Rate Survey., 2006. <http://www.epa.state.oh.us/ofa/sw05/Ohio_EPA_2005_Sewer_and_Water_Rate_Survey.pdf>.

Raftelis Financial Consultants, and American Water Works Association. 2006 Water and Wastewater Rate Survey. Charlotte, NC: Raftelis Financial Consulting, 2007.

Raftelis, GA. "1988 National Water Rate Survey." Journal of the American Water Works Association 80.9 (1988): 78-81.

---. Water and Wastewater Finance and Pricing: A Comprehensive Guide. Third Edition ed. Boca Raton, FL: CRC Press, 2005.

Randall, A., and J. Dewbre. Inventory of Water Diversions and Rate Structures for Cities, Towns, and Villages in New Mexico. Vol. 241 (Las Cruces). University Park, NM: New Mexico Agricultural Experiment Station, 1972.

San Diego County Water Authority. 2005 Member Agency Water Rate Survey., 2006. <http://www.sdcwa.org/about/member-agencies.phtml#survey>.

Southeast Michigan Consortium for Water Quality. Water and Sewer Rate Analysis., 2004. June 18, 2007 <http://www.semcog.org/products/pdfs/SEMCOG%202003-2004%20Rate%20Analysis.pdf>.

StataCorp. "Reference G-M." Stata Statistical Software: Release 8.0. College Station, TX: Stata Corporation, 2003a. 274-279.

---. "Reference N-R." Stata Statistical Software: Release 8.0. College Station, TX: Stata Corporation, 2003b. 363-365.

---. "Reference N-R." Stata Statistical Software: Release 8.0. College Station, TX: Stata Corporation, 2003c. 378-380.

---. "Reference S-Z." Stata Statistical Software: Release 8.0. College Station, TX: Stata Corporation, 2003d. 171-173.

Thorsten, R., S. Eskaf, and J. Hughes. Cost Plus – Estimating Real Determinants of Water and Sewer Bills. Working Paper, Chapel Hill, NC: Environmental Finance Center, 2007.

Tighe & Bond. Massachusetts Water & Sewer Rate Surveys., 2004. <http://rates.tighebond.com/>.

Triangle J Council of Governments. 2007 Triangle Area Water and Sewer Rates Survey., 2007.

United States Geological Survey. "Water Q&A: Water use at home." 2007. <http://ga.water.usgs.gov/edu/qahome.html#HDR3>.

Water Infrastructure Finance Authority of Arizona. 2006 Water and Wastewater Residential Rate Survey for the State of Arizona., 2006. <http://azwifa.gov/QuickLinks/2006RateStudy.pdf>.

Western Resource Advocates. Water Rate Structures in New Mexico: How New Mexico Cities Compare using this Important Water use Efficiency Tool., 2006. <http://www.westernresourceadvocates.org/>.

Wooldridge, Jeffrey. Introductory Econometrics: A Modern Approach. 2nd ed. South-Western, 2003. 199-210.

Zieburtz, William. 2004 Georgia Water/Wastewater Rate Survey. Alpharetta, Georgia: Black & Veatch, 2004.