Modeling the Influence of Drinking Water Quality on Residential Property Value

Ghazal Pashanasangi

A Thesis in

Concordia Institute for

Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montreal, Quebec, Canada

August 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Ghazal Pashanasangi

Entitled: Modeling the Influence of Drinking Water Quality on Residential Property Value

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Dr. D. Dysart-Gale __________ Chair

Dr. K. Schmitt __________ Supervisor

Dr. C. Wang __________ CIISE Examiner

Dr. S. Li __________ External Examiner (BCEE)

Approved by ______________________________

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date September 2013____________________

ABSTRACT

Modeling the Influence of Drinking Water Quality on Residential Property Value

Ghazal Pashanasangi

This thesis is an endeavor to study the influence of drinking water quality on residential property value in the United States. Bayesian inference method is used to model the influence of water pollution on property value. Hierarchical Bayesian inference method is used when different levels of pollution are considered. Number of types of chemicals detected in water utilities of each state, amount of total trihalomethanes (TTHMs) and total haloacetic acids (HAAs) in big cities drinking water, average property value in each state and big cities, and population are the data used in analyses. Results show that water pollution has a negative impact on the average property price in state level. Both TTHMs and HAAs, which are the most important pollutants in big cities, also have negative impact on residential property price in big cities. The impact of TTHMs pollution is more than HAAs pollution. Results of the hierarchical analyses show the state that each big city belongs to, determines the effect of TTHMs and HAAs pollution on residential property price.

Acknowledgments

First, I would like to express my sincere gratitude and appreciation to my supervisor, Dr. Ketra Schmitt, for her motivation and generous guidance. I am grateful to her for giving me this opportunity and making this research possible.

I consider it an honor to have many friends who provided me a supportive and enjoyable environment in the past two years.

Finally, I would like to thank my family for their unconditional support and continuous encouragement. Special thanks to my husband, Mahdi Milani Fard, for his love, support and priceless advices on statistics. I owe my deepest gratitude to my parents, who always kept my happiness, success and education in their top priorities. This work is dedicated to my beloved family.

Table of Contents

List of Figures

List of Tables

1. Introduction

The goal of this thesis is to study the relationship between drinking water quality and residential property value in the United States. Different analyses in city and state level based on population have been done to achieve this goal.

The motivation for this work is answering questions about the relationship between water quality and property value in state level and city level, and based on different pollutions. Knowing the possible correlation can be helpful to find out hidden causal variables.

Water quality has a significant role in human health, ecosystem functionality, and economy. It is a global concern which has an undeniable influence on human life. Both infrastructure and technology improvements and regulations and policies improvements are needed to solve this problem.

States should comply with different federal and state level regulations, and health guidelines. However, policy enforcement is an important factor for state compliance to regulations. Policies make it possible to implement the practical solutions to water quality problem. Having stronger policies can make a path to improved water quality.

Economic analyses are needed to predict the economic value of environmental quality. These analyses provide practical information to build a decision support system for environmental policy makers. This thesis model the influence of drinking water quality on property value, which is a measure of water quality benefits.

Most of the studies about the relationship between water quality and property values focused on surface water quality. Studies on the influence of drinking water quality are really inadequate. In addition most of the studies used data of specific neighborhoods or cities and the results are dependent to that neighborhoods features and structures.

This thesis tries to find out the relationship between water pollution and property price in different states, the relationship between water pollution and property price in big cities, and the influence of water pollution on both city and state levels on property price in big cities. In addition to general water pollution, some analyses fuscous on the pollutants which are the top concerns in big cities.

Our focus is on the mean behavior. As we are using country scale data, the exact location and specific features are overshadowed by diversity. In addition to local studies, which can be helpful for local policy making, large scale studies are also needed.

Using drinking water and property value data of the United States gives the opportunity to conduct all the analyses in a large scale. Therefore the results are not exclusive to one small region and its structure and features.

In addition new methods are applied to analyze data in this thesis. Bayesian modeling method is used to model the average residential property value as a function of population, number of chemicals found in the water utilities, and the percentage of population affected by those chemicals. Markov chain Monte Carlo (MCMC) algorithms are used to make inferences about the model. Furthermore hierarchical Bayesian Modeling is used to model the effect of water pollution in both city and state levels.

Although Bayesian modeling is used in water quality prediction, water quality model uncertainty, and water management decision making, to our knowledge it has not been used in modeling the influence of water quality on property value [1] [2] [3].

These analyses model the relationship between water quality and property value, however water quality is not necessarily the direct reason of increase in property value. This correlation should not be interpreted as causation. The correlation between water quality and property value is not the same in different states or big cities. Environmental and economical situation of the region can shape different circumstances which can lead to different correlations between water quality and property value in long term.

2. Literature Review

The objective of this chapter is to review related literature in order to understand the relationship between water quality and property value thoroughly. The first part is a brief discussion on different aspects of water quality. The second part is a general overview to determine the scope of the research. And the last part summarizes previous works.

2.1. Water Quality Background

This part provides the background required to understand different aspects of water quality.

2.1.1. Definition

Water quality is a general term that refers to myriad chemicals and biological parameters. These inform the level of water quality. Although regulatory standards help to define acceptable quality, these vary by geographic region and designated use. The tolerated amount of different substances or combination of them is relative to the particular usage. In other words water which has a good quality for a specific purpose can be unacceptable for another. The physical, chemical, and biological condition of water should be known to determine the quality [4] [5] [6].

2.1.2. Importance

Water quality affects every aspect of life on earth from the basic ability of ecosystems to function to human health and the economy.

All living organisms need clean and sufficient water in order to survive. Clean and safe freshwater is necessary for all ecosystems to function properly. Degraded water quality cause serious damages to rivers, streams, lakes, groundwater, coastal zones, and vegetated wetlands [7]. This threatens ecosystems and biodiversity.

Human health is dependent on healthy ecosystems for adequate fresh water and food production and is negatively affected by poor water quality. Water-related diseases and health effects of high concentrations of nutrients are two important examples of these health threats. The effect on children is more dangerous. Children under 5 years old comprise most of deaths that are a consequence of unsafe water [8] [9] [10].

In addition polluted water has a large number of economical disadvantages. It has a negative impact on agriculture, industry, mining, tourism and recreation activities. It also leads to degradation of ecosystem services, increased water treatment costs, increased health-related costs, and reduced property values. These costs vary in different regions. These impacts seem to be the most pronounced in areas where water resource management is weak and ineffective. For example, in Middle East and North Africa these costs are between 0.5 to 2.5 percent of GDP each year [11].

2.1.3. Threats

Generally human settlements are considered as a threat to quality and quantity of freshwater resources. There are two main reasons. The first reason consists of pollution caused by agricultural activities, industrial production, power generation, mining, and untreated sewage discharge. The second reason has to do with changing the water balance, consists of climate changes caused by human activities and over extraction of ground waters [12] [6] [13].

In addition unsafe solid waste discharge, unhygienic disposal, and deficient treatment of industrial residue lead to poor water quality. For instance, in developing countries more than 80 percent of sewage is directly discharged to water bodies [14] [15].

Increased acidity and higher levels of nutrients, salts, metals, chemicals, and pathogenic organism are the results of polluted and contaminated water [16].

2.1.4. Standards

Drinking water quality standards determine the permitted concentration of different substances in drinking water. Nowadays most of the countries have some standards to control the water quality. In addition World Health Organization has established guidelines on the standards that should be achieved. However, the United States and European countries are the only countries in which these standards have legal bases because of the United States Environmental Protection Agency and the European Drinking Water Directive. Further, the numerical levels of standards differ in different jurisdictions [17] [18] [19] [20].

Surface water quality standards, which determine the permitted level of contaminants in wastewater, effluents, or ambient water, have also been established in some countries. However this type of standards is less common [7].

2.1.5. Solutions

Regulatory standards are mostly guideline values for treated water, and in some cases surface water or effluents [17] [20] [21]. Even when legally enforceable, water standards are not sufficient to protect water resources, as standards do not prevent pollution from happening or automatically detect and treat pollution when it does occur. According to United Nations Environment Programme (UNEP) the solution for water quality problem has four parts [7]:

- Prevention of pollution

This strategy focuses on the source of contaminants and makes an effort to reduce or eliminate them before entering the water cycle. This is the cheapest and most effective solution for achieving better water quality. However, in areas where environmental laws and enforcement are weak, this may not be successful.

- Treatment of polluted water

As efforts to prevent water from being polluted are not always sufficient or effective, and in many cases there is no effort, polluted water should be treated to achieve the acceptable quality level for drinking or other purposes. Contaminated water treatment has different approaches which vary from high technology methods to biological and ecological methods. Again, this is less likely with low levels of environmental enforcements.

- Treatment of wastewater

Wastewater from domestic, agricultural, and industrial activities must be treated before entering water bodies. In some cases wastewater can be recycled and reused safely. Use of wastewater reduces the threats from human activities for freshwater resources. This approach is very helpful in semi-arid regions with growing populations [14] [22].

- Protection of ecosystems

Healthy ecosystems filter and clean water naturally. Therefore protecting ecosystems causes improved water quality. Damaged ecosystems can be assisted in the process of ecological restoration. This approach costs lower than technical efforts for filtering the polluted water [23] [24]. This is extremely related to prevents of pollution and treatment of polluted water.

2.1.6. Policies

Policies make it possible to implement the practical solutions to water quality problem. The desirable targets are determined and the solutions to achieve them are enumerated. Environmental policies are structured to construct a path for moving toward these targets. To construct this path, policies in three fields are required:

- Knowledge and education

Understanding the importance of water quality and its role in human life is the first step. Public awareness and academic education and research on both technological and economical aspects are vital to make this possible.

- Technology and methods

Water treatment, sewage treatment, and ecosystem restoration methods and equipments should be developed in all regions. As most of these are complex and expensive procedures, improving the existing methods and technologies are required too.

- Law and management

Strong leadership is needed to make changes possible in all levels. Creating legal bases for employing standards, improving management, and allocating more budgets to solve water problem are important steps which need strong policies. Funding is required to make all of these policy goals possible. When conflicts funding needs arise, water treatment can be seen as a luxury or unnecessary expense.

On one hand water quality monitoring and water treatment are expensive acts. On the other hand funding to enforce water quality standards is inadequate in most countries. Therefore, the first step to make water quality regulation successful is finding appropriate sources of financing. This is only possible by investing in the expertise in order to define appropriate methods for analyzing and determining water quality improvement costs and benefits [7].

The cost of obtaining a specific level of quality can be predicted by engineers, nevertheless as environmental goods are non-marketed, there are different analysis needed to predict the benefits. This thesis model the influence of water quality on property value, which is a measure of water quality benefits. In other words this thesis provides economical analyses to support decision making procedure.

2.2. General overview

In the beginning the literature review was started by searching the web for sources that studied the influence of water quality on property value. Searching in sites like Google Scholar and Engineering Village for keywords resulted in a very limited number of papers. Therefore, we decided to perform a more general literature review which considers the influence of all environmental quality factors on property value.

The following is a review of studies about the influence of environmental quality factors on property value. The first part reviews existing studies and discusses which factors were considered and similar or contradictory results. The second part summarizes methodological approaches. The third part focuses on deficiencies in literature and explains requirements.

2.2.1. Influence of Environmental Quality Factors on Property Value

Concerns about the places we live, work, and play are a normal part of human life; which lead us to search for places which satisfy our expectations for safety and enjoyment. As a result of these concerns about daily residential area and surroundings, environmental quality factors have the ability to affect amenities and the value of residential properties. In terms of property value, the most effective environmental quality factors are greenery, water quality, air pollution, noise level, and landscape view. Water quality is one of the most important factors as it has a significant role in human health, ecosystem functionality, and economy.

Numerous studies have been done about the relationship between environmental quality factors and property value. These studies have mostly been conducted in specific neighborhoods. Therefore they have focused on different factors according to features of that neighborhood. These factors can vary from 'having a garden facing water' in the Netherlands [25] to 'Traveling time from the apartment to the central business district' in Hong Kong [26]. However, there are some common factors in most of the studies. Knowing considered factors, applied methods, and results will be helpful in order to find out the data gaps and advance the study by using new methods and approaches.

Environmental quality factors can be classified in different ways. However, the most practical approach is to categorize them into natural or manmade.

The first category consists of natural features such as greenery, water quality, air pollution, and landscape view. Green belts refer to green space (in the form of park land, natural spaces or agricultural space) surrounding urban spaces. According to previous studies, the presence of greenbelt, tree cover, and wetlands leads to higher property values in most neighborhoods [25] [27] [28] [29]. All the reviewed studies show that higher water quality has a direct relationship with prices [30] [31] [32] [33] [34]. Landscape view refers to visible features of an area, often considered in terms of their aesthetic appeal. Landscape view also increases property values. The most discussed scenery is sea view [26] [27] [35] [36]. Air pollution also decreases property prices [26] [37] [38] [39] [40].

Although most of these studies focused on suburbs and less crowded locations, the studies on high-rise and densely populated environments give interesting results too. This type of locations has different effective factors or in some cases, the same factor has a different influence. For instance a study conducted in Hong Kong suggests that nearby green belt area is not a significant factor [2].

The second category consists of manmade landscapes or equipment such as water parks, high voltage electric transmission lines, and noise level. These factors can affect the property value in both positive and negative ways. For example, water parks have a positive effect [41], while high voltage electric transmission lines have a negative effect [42] [43].

For some factors contrasting results exist in different locations. Noise impact is a good example. Different studies find different results such as statistically significant negative relationship, no relationship, and statistically significant positive relationship between noise and residential property values [26] [44] [45] [46]. This difference seems to be related to population density. In higher population densities property value seems to increase with increased noise pollution. This is likely the confounded variable for the locations of amenities, proximity to work and general benefits of city living and demand for city houses.

In addition to all quality factors risk perception can also affect property value. Normally the public belief about environmental risk, which can be effected by media, is different from scientifically determined risk. However, perceived risk has an important economic impact. For example based on previous studies perceived risk from being near a hazardous waste site has a negative impact on property value [47] [48] [49] [50].

2.2.2. Methodological Approaches

Most of the studies used Hedonic pricing method, which is a method to estimate the value of non-marketed goods that affects prices of marketed goods. This approach is extensively used to estimate the costs associated with environmental factors that affect the price of residential properties. Hedonic pricing method is based on the principle that the value of a marketed good is related to its features. Thus, the value of each feature can be estimated based on the willingness to pay for changes in that feature.

Hedonic pricing method assumes that price of a property consists of the prices of different factors such as size, building structure, location, site, neighborhood features, and environmental characteristics. The first step to apply this method is to define price of the property as a function of these parameters. Regression methods are employed to estimate the price function. The second step is to combine this model with actual quantities and do the regression analysis again to estimate the value of each parameter [51] [52] [53].

For example to find out the value of clean air using Hedonic pricing method, price function should be estimated based on all effective parameters. Next the value is estimated by using regression techniques on the price model and actual quantities. The estimated value reveals the difference between values of two completely similar properties, one in a place with polluted air and one in a place with clean air.

However, this method has some limitations. A lot of specific parameters for each property must be gathered in order to apply this method. Hedonic pricing method tends to focus on small very complete datasets. When larger datasets are used, these factors will be overshadowed by diversity. In addition using this method leads to the results which are heavily dependent on the model specification. This model has three assumptions which are rarely true in reality. The first assumption is that everyone is aware of the potential effects of changes in environmental factors. The second one is that a variety of houses with any combination of features is available in the market. And the third one is property prices will immediately change after any change in environmental factors [51] [52] [4] [54] [55].

2.2.3. Deficiencies and Requirements

2.2.3.1. Deficiencies

All the studies that are mentioned above used data of specific neighborhoods or cities. Therefore the results are completely dependent to that neighborhoods features and structures. There is no study in bigger scales for instance a country.

In addition most of the studies about the relationship between water quality and property values focused on surface water quality, not drinking water quality. Therefore studies on the influence of drinking water quality are really inadequate.

Moreover, some potentially effective factors such as population are not considered in previous studies.

2.2.3.2. Requirements

The importance of drinking water quality is undeniable. Pollution in drinking water has a direct impact on human health [56] [8] [57] [58] [59]. The number of people who die from a shortage of fresh and clean water each year is more than the number of people who die in violence and war [7]. Children are the most vulnerable group [9] [10].

In addition to basic infrastructure improvements, water regulation, both in terms of sound policies and enforcement, are key to improving human health at the global level.

While water availability and quality needs are most apparent in the developing world, quality issues and regulatory gaps exist in the developed world in rural areas and on native lands. Moreover, although high quality water is taken for granted in urban areas in the developed world, mishaps do happen. And when they do they impact large numbers of people.

An understanding of drinking water quality influence is required to take steps toward more strong policies. Similarly understanding the economic impacts of sewage outflows is also important for developing water policy. These economic analyses suggest a decision support system for environmental policy makers. That is a helpful approach toward improvement in sustainability, equity and economic efficiency of water resource policies.

2.3. Influence of water Quality on Property Value

The literature review is performed by searching the web for studies on the influence of water quality on property value. However, it resulted in a very limited number of papers. In addition to consulting librarians more books, journals, papers, and theses are checked for similar studies. Again the result was not adequate. To our knowledge the literature on this subject is really limited and more studies are required. All of these make conducting this research more important and unique.

This part is a summary of previous studies on the influence of water quality on property values. As this review demonstrates, to our knowledge all existing works on the influence of water quality on property value has explored surface water characteristics.

Leggett and Bockstael have done a study on this subject in Chesapeake Bay. The important feature of this area is the variation of water quality in different parts of Chesapeake Bay. This study discusses the effect of water quality on property value by measuring the benefits of future improvements in water quality for property owners. The results show that the owners of houses in front of the water have a positive willingness to pay for improving the water quality. It should be noted that the owners have concerns about fecal coliform bacteria counts, which is concluded to have a significant negative effect on property values. According to the results a change of 100 fecal coliform counts per 100 mL leads to around 1.5% change in property values. Based on predictions this reduction in fecal coliform can cause an approximate increase of $230,000 in the total property value of 41 residential properties in front of the water. This study use Hedonic techniques to measure the benefits [30].

Another study from 1968 used the same method, investigating the value of properties around 60 artificial lakes in Wisconsin. Results showed that water quality of lakes, which were classified to good, moderate, and poor, was a significant factor of property values. In other words the value of properties near more polluted lakes was less than adjacent properties near less polluted lakes. This study used tax data to find property values [31].

A study by Epp and Al-Ani supports the influence of water quality on property value, focused on rural nonfarm properties around small rivers and streams in Pennsylvania. Analyzing data of single-family owner-occupied houses that are located within 700 feet of the stream shows the significant effect of water quality on property value. This result is true both when the water quality is determined by an index of measured water quality characteristics or by the owners' perceptions of water quality [32].

In addition to the general analysis, Epp and Al-Ani also analyzed the clean streams and polluted streams separately. Streams which have all components within the normal range are determined as clean and streams which have one or more components outside the normal range are determined as polluted. According to analyses the pH level has influence on value of properties which are close to clean streams. However, it does not have influence on value of properties which are close to polluted streams. In other words in clean streams the increase of pH level in normal range, which is 6.5 to 8.5, leads to increase in property value. This increase is about $653.96 for 1 point increase in pH level in normal range. The reason could be the increasing opportunity of trout fishing. Data sampling was in a way that minimizes the variation of other effective factors between properties [32].

Young discusses the effect of perceived water quality on property values. This study, which is conducted in St. Albans Bay on Lake Champlain in Vermont, compares the value of properties located adjacent to St. Albans Bay with similar properties located adjacent to the shoreline of the main lake near the bay. There was a malfunctioning municipal waste treatment plant near the bay that caused some pollution problems. The results show that the properties adjacent to the bay have 20 percent lower values than the similar properties that are adjacent to the shoreline of the main lake [33].

A different approach is employed in a study by Mendelsohn et al. in the New Bedford, Massachusetts harbor. This study uses panel modeling, which is suitable for local pollution problems, in order to measure the effect of PCB pollution by using residential property values. The results show affected properties have a significant lower value. This price reduction, which is associated with timing and location of the waste site area, ranges from $7000 to $10,000 (1989 dollars) [34].

3. Methodology

The objective of this chapter is to discuss the employed methods in this thesis. First the required background is provided and Bayesian modeling and MCMC algorithms are introduced. This will be followed by model specification and details.

3.1. Background

3.1.1. Frequentist versus Bayesian

Frequentist inference and Bayesian inference are two different approaches to statistics. Frequentist approach focuses on the frequency of an event (a set of outcomes). The goal is to find out the conditional probability of observing specific data given the parameters, which is shown by $P(D|\theta)$, where D represents the observed data and θ represents the true parameter value. In this approach data are random and parameters are fixed or known [60].

On the other hand in Bayesian approach the goal is to find out the conditional probability of parameters given the specific data, which is shown by $P(\theta|D)$. In other words, the goal is the probability of the parameter value given the data. In this approach data are fixed (limited number) and the probability of parameters is unknown. Bayesian statistics normally leads to probability intervals of the parameters. Bayesian Inference updates probability interval estimations, using new data [61] [62] [63].

3.1.2. Bayes theorem and Bayesian Inference

Bayesian Inference uses Bayes' theorem to update the probability estimate for unknown parameter values, as additional information is learned from data [64] [65].

If θ is a set of parameters and (x, y) is a set of observations, Bayes' theorem says:

$$P(\theta|x, y) = \frac{P(y|x, \theta)P(\theta)}{P(y|x)}$$

Equation 3.1

Where $P(\theta)$ is the prior probability, $P(\theta|x, y)$ is the posterior probability, $P(y|x, \theta)$ is the likelihood function, and $P(y|x)$ is the marginal likelihood.

The prior probability of an unknown parameter is a distribution which replaces randomness with uncertainty, before observing data. The posterior probability of an unknown parameter is a conditional probability distribution estimate, after observing the data [65]. The likelihood function of a set of parameter values is the probability of a set of observations given those parameter values. Marginal likelihood is the probability of a set of observations regardless of parameter values.

Bayesian inference is a method to update the prior probability distribution to posterior probability distribution based on the observed data. Bayesian inference is closer to real world thinking; we enter each new situation with our opinion (prior) and update our view of reality based on our experience (new data) to develop a current point of view (a posterior).

3.1.3. Conjugate and Non-Conjugate Prior

When the combination of prior distribution and likelihood distribution results in a posterior which is in the same family of the prior, the prior is conjugate to the likelihood. A common example of conjugate prior is Gaussian distribution, which is a conjugate prior for a Gaussian likelihood. In other words having a Gaussian prior leads to a Gaussian Posterior if the likelihood function is Gaussian. This means that the distribution of the likelihood and the prior can be dealt with in a closed-form solution, so that the posterior can be determined analytically. These prior and posterior distributions are called conjugate [66]. In this case the distribution of posterior is known and it can be easily simulated.

However in most of the cases the prior is not conjugate and the posterior does not have a standard form. As no closed form solution exists for non-conjugate priors, posterior approximation must be done numerically. In case of non-conjugate prior, some simulation techniques are required to approximate the posterior distribution [67]

3.1.4. Markov chain

Markov Chain is defined as the random transition of a system between different states. The system is in a certain state in each step. This process is a memoryless process. It means that the next state only depends on the current state, not the whole sequence. Assuming $X_1, X_2, \ldots, X_n$ as possible states of the system:

$$X_{n+1} = f\ (X_n)$$

Equation 3.2

This means that X_{n+1} is only a function of X_n, and does not depend on any earlier states of the system.

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n)$$

Equation 3.3

Since the distribution of X_{n+1} depends only on X_n, we do not need to retain the previous system states. In other words, the probability of the next step is conditionally independent of previous time steps except for the immediate predecessor.

The probabilities of each possible next state based on the current state, are called transition probabilities. To describe a Markov chain a set of all possible states and transition matrix, which is consisted from transition probabilities, should be defined. It is not possible to certainly predict the next state of the system and the transition process can go on forever [68] [69] [70].

3.1.5. Monte Carlo

Monte Carlo methods are a group of numerical algorithms that use repeated random sampling to determine the properties of a phenomenon [71]. For example assume X as a random variable with the expected value of A. By generating n independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution, A can be approximated by:

$$A \approx A_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

Equation 3.4

A_n gets closer to A as the number of samples increases.

One of the main applications of Monte Carlo methods is to generate samples from a probability distribution. This is a helpful approach when it is impossible to obtain a closed form expression for the desired distribution [72] [73].

3.1.6. Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a class of computing methods. At first it was developed by physicists to compute complex integrals by using random number generation.

The main application of these numerical algorithms is to simulate distributions. By generating samples the target distribution can be simulated numerically. It is a useful approach to use MCMC when it is not possible to derive a closed form of desired distribution.

Markov chain Monte Carlo methods refer to a group of algorithms which sample from probability distributions based on constructing a Markov chain. After a large number of transition steps, the state of the chain is used as a sample of the desired probability distribution. Normally Markov chains in MCMC have discrete steps and continuous state space [74] [75].

In Bayesian inference MCMC can be used as a simulation technique for sampling from posterior. This method is a so practical when the posterior does not have a standard form.

MCMC algorithms generate samples from the posterior density to approximate the required distribution by constructing Markov chains. These samples only depend on the previous one, and improve as the number of steps increase. Each Markov chain starts with an initial value and has posterior as a target distribution [76] [77].

3.2. Model Specification

The Average residential property value in each state is modeled as a function of population and water pollution in each state. The first idea of our model comes from studying sewage outflows data and trying to find a model for the influence of water quality on property value. We observe that property value has a linear relationship with water quality. We tried some potential models such as linear, exponential, and etc. We found that having a linear model with variable coefficients for pollution and population is the best choice for our study.

Model likelihood, prior probability distributions, and posterior probability distributions should be defined, in order to determine a Bayesian model.

In this model $price_i$, pop_i, and $poll_i$ are respectively the average residential property price in the ith state ($100,000), the population of the ith state(1000 persons), and the water pollution in the ith state. Water pollution in each state is defined as the number of chemicals detected in the water utilities of the state times percentage of population served by drinking water contained those chemicals.

Where $price_i$ is a Normal distribution with mean μ_i and variance τ, the model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 3.5a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i$$

Equation 3.6b

This definition of $price_i$ is equivalent to have a noise function with a Normal distribution and define $price_i$ as:

$$price_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i + n_i$$

Equation 3.6a

$$n \sim N(0, \tau)$$

Equation 3.7b

The first definition of $price_i$ is used as the model likelihood function. In this function β_0, β_1, and β_2 are coefficients. β_0 shows the effect of all unconsidered factors. β_1 shows the effect of population. β_2 shows the effect of pollution. Bayesian inference is used to improve our knowledge about the probability distributions of these coefficients. In other words data is used to upgrade prior distribution of these coefficients to posterior distributions.

The uninformative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \ \beta_1 \sim N(0, 10^4)\ ,\ \beta_2 \sim N(0, 10^4)\ ,\ \ \tau \sim \Gamma(0.01, 0.01)$$

Equation 3.8

This means β_0, β_1, and β_2 are Normal distributions with mean 0 and variance 10^4 and τ is a Gamma distribution with shape 0.01 and rate 0.01.

Gamma distribution is a continuous distribution which has two positive real parameters. The probability density function of Gamma distribution is defined as below:

$$X \sim \Gamma(\alpha, \beta)$$

Equation 3.9a

$$g(x; \alpha, \beta) = \beta^{\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} x^{-\beta x}$$

$$for\ x \geq 0\ and\ \alpha, \beta \geq 0$$

Equation 3.10b

The prior probability distributions of β_0, β_1, and β_2 are selected to be normal distributions with zero mean and very large variance in order make it as uninformative as possible. This is also known as a diffuse prior. The prior probability distribution for τ is defined to be gamma distribution since it is the conjugate prior to normal distribution with known mean.

In addition to uninformative prior probability distributions a set of informative prior probability distributions are defined based on general knowledge of the process. Informative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0.1, 10^2)\ ,\ \beta_2 \sim N(-0.1, 10^2)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 3.11

We chose to leave a diffuse prior for β_0 because it shows the effect of all not considered factors and we didn't want to limit it. On the other hand we chose informative priors for β_1 and β_2. The informative prior probability distributions imply that the effect of population on property value is positive and the effect of water pollution on property value is negative.

The purpose is to estimate the posterior probability distributions of β_0, β_1, and β_2 by Bayesian inference. This will be done by upgrading the prior probability distributions based on data. Posterior distributions are in fact probability distributions of coefficients of μ_i function. As mentioned above likelihood function is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 3.10a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i$$

Equation 3.12b

It is important to understand the difference between μ_i and $Typical\, price$. $Typical\, price$ is defined as the value of price when all covariates are equal to their sample means.

$$Typical\, price = \beta_0 + \beta_1 * mean(pop_i) + \beta_2 * mean(poll_i)$$

Equation 3.13

$Typical\,price$ is an instant value for each dataset. However, μ_i has a separate distribution for each sample set $(pop_i, poll_i)$ from the dataset. It should be noted that typical price is equal to the expected value of μ_i in case of having a linear likelihood function.

4. Data

Two groups of data are used for analysis and modeling. These two are water quality data and residential data. The novelty of our final dataset is in being a combination of these two groups of data. In addition our final dataset consists of all the U.S. states data and has detailed data for big cities and specific pollutants. Data characteristics and sources are discussed in this section.

4.1. Water Quality Data

In the process of data gathering many resources were consulted to find water quality data. Most of them only consist of the result of specific tests for one water utility on specific dates. Altogether, to our knowledge the National Drinking Water database is the biggest and completest existing database. This database, which is a group of separate state reports, is used to gather and make our required datasets for modeling.

The National Drinking Water database is gathered by Environmental Working Group, which is an American environmental organization. "The mission of the Environmental Working Group (EWG) is to use the power of public information to protect public health and the environment." [78]

Tap water contaminant data of 43 U.S. states over a five year period from 2004 to 2009 is recorded. Other states failed to provide water quality data. A large section of the data is from analyses conducted by water utilities. Individual chemicals are tested as state or federal requirements. A small part of the data is from the tests done by state agencies in short-term monitoring programs. The whole dataset includes nearly 20 million test results [79]. This dataset consists of separate reports from states that show contaminants which exceeding limits. Each state report has a wide variety of sources from different areas in that state. Consisting all the state, federal and agencies test results has made this dataset large and complete. Number of chemicals detected in water utilities of each state and the amount of top concern pollutants in big cities are used in analyses.

Water quality data is used in different analyses in this thesis. The first analysis used number of chemicals detected in water utilities of each state and influenced population. The second and third analyses used total trihalomethanes (TTHMs) and total haloacetic acids (HAAs) amount in big cities water utilities. Later analyses used a combination of all the data. Following is a description of each group of data features and properties.

4.1.1. Number of Chemicals Detected in Water Utilities of Each State

The dataset used for the first model consists of five parameters:

- Number of water utilities in each state
- The total number of detected chemicals in water utilities of each state
- Population served by water utilities in each state
- Population served by chemicals detected in water utilities in each state
- Percentage of population served by chemicals detected in water utilities in each state

Number of detected chemicals means number of types of chemical which are detected in water utilities. This dataset is gathered for the modeling from separate state reports of the database over a five year period from 2004 to 2009. Table 1 shows our dataset.

State	***Water utilities***	***Detected chemicals***	***Population***	***Population served with chemicals***	***% Population served with chemicals***
Alabama	*362*	*72*	*4,883,170*	*4,565,163*	*93.49*
Arizona	*753*	*68*	*4,995,854*	*4,959,760*	*99.28*
Arkansas	*713*	*94*	*2,929,506*	*2,929,506*	*100*
California	*2655*	*182*	*54,146,711*	*52,625,466*	*97.19*
Connecticut	*607*	*67*	*2,696,783*	*2,695,754*	*99.96*
Delaware	*212*	*170*	*889,111*	*889,111*	*100*
Florida	*1743*	*103*	*18,662,845*	*18,337,674*	*98.26*
Hawaii	*106*	*37*	*1,442,030*	*1,421,834*	*98.6*
Idaho	*726*	*61*	*1,063,278*	*1,057,640*	*99.47*
Illinois	*1,765*	*101*	*13,084,369*	*13,083,669*	*99.99*
Indiana	*778*	*81*	*4,630,253*	*4,545,413*	*98.17*
Iowa	*1135*	*83*	*2,661,554*	*2,660,281*	*99.95*
Kentucky	*386*	*63*	*4,908,697*	*4,804,170*	*97.87*
Maine	*388*	*74*	*656,078*	*656,078*	*100*
Maryland	*478*	*93*	*5,156,372*	*5,150,059*	*99.88*
Massachusetts	*494*	*94*	*7,695,428*	*7,252,397*	*94.24*
Michigan	*926*	*54*	*6,710,306*	*6,074,253*	*90.52*
Minnesota	*937*	*94*	*4,150,414*	*4,144,080*	*99.85*
Missouri	*1506*	*109*	*5,184,785*	*5,184,005*	*99.98*
Montana	*684*	*67*	*712,991*	*712,611*	*99.95*
Nebraska	*605*	*90*	*1,412,179*	*1,411,786*	*99.97*
Nevada	*222*	*73*	*2,414,998*	*2,395,224*	*99.18*
New Hampshire	*1140*	*91*	*925,608*	*897,987*	*97.02*
New Jersey	*627*	*119*	*8,619,862*	*8,612,455*	*99.91*
New Mexico	*616*	*96*	*1,657,756*	*1,633,045*	*98.51*
New York	*2,317*	*181*	*17,300,092*	*17,072,434*	*98.68*
North Carolina	*2146*	*112*	*7,352,522*	*7,348,186*	*99.94*
North Dakota	*327*	*29*	*562,310*	*561,620*	*99.88*

Ohio	*1297*	*80*	*10,103,190*	*10,088,792*	*99.86*
Oklahoma	*847*	*39*	*3,320,844*	*3,097,608*	*93.28*
Oregon	*845*	*57*	*3,191,117*	*3,171,607*	*99.39*
Pennsylvania	*2060*	*59*	*10,834,741*	*10,827,007*	*99.93*
Rhode Island	*77*	*80*	*983,259*	*958,426*	*97.47*
South Carolina	*103*	*21*	*2,236,545*	*729,333*	*32.61*
South Dakota	*250*	*26*	*785,973*	*658,214*	*83.75*
Texas	*4641*	*122*	*20,389,435*	*20,352,009*	*99.82*
Utah	*434*	*50*	*3,802,780*	*3,731,947*	*98.14*
Vermont	*436*	*71*	*410,630*	*386,801*	*94.2*
Virginia	*1171*	*93*	*7,456,511*	*7,437,750*	*99.75*
Washington	*2182*	*91*	*5,288,422*	*4,980,486*	*94.18*
West Virginia	*32*	*19*	*1,285,582*	*234,926*	*18.27*
Wisconsin	*4278*	*110*	*4,694,237*	*4,461,879*	*95.05*
Wyoming	*212*	*41*	*400,966*	*399,277*	*99.58*

Table 1. Number of Chemicals Detected in Water Utilities of Each State from 2004 to 2009

In this dataset number of detected chemicals includes detected chemicals which exceed health guidelines, detected chemicals which exceed legal limits, and also unregulated detected chemicals.

Water utilities nationwide detected 316 different kinds of contaminants in water supplied to 256 million Americans. 202 of the total 316 detected chemicals are not subject to any government regulation or safety standards for drinking water [79]. This means these chemicals are legal in any amount. In other words only 114 contaminants from 316 detected contaminants were subject of U.S. Environmental Protection Agency (EPA) enforceable drinking water safety standard.

According to EWG's drinking water quality analyses water utilities show 92 percent compliance with the United States Environmental Protection Agency for 114 regulated chemicals [79]. This means the regulations are effective. However, there is no regulation for many contaminants.

Most of the data comes from regularly conducted analyses by water utilities. All these test results are recorded. Water utilities are identified as out of compliance only if the annual average of the test results is above the maximum contaminant level (MCL). Since MCLs are based on annual averages for most of the contaminants, water utilities are not necessarily out of compliance for exceeding the MCL in one test [79].

4.1.2. TTHMs and HAAs Amounts in Big Cities Water Utilities

In addition to state reports, an analysis of big cities drinking water is available in National Drinking Water database. In this analysis drinking water quality in 100 big cities (cities with population over 250,000) of United States are ranked.

The top concern for each city is defined as the chemical with the highest average level relative to the legal limit for regulated contaminates, or to the national average concentration for unregulated contaminates. In most of these cities the top concern is TTHMs or HAAs.

As these two pollutants are the top concern in most cities, we decide to discuss the amount of TTHMs and HAAs in big city drinking water in the second and third modeling.

Table 2 shows a ranking (from best to worst) of 100 big city's drinking water quality, the top concern in each city, and the amount of TTHMs and HAAs.

City	State	Top concern	THMs (ppb)	HAAs (ppb)
Arlington	*TX*	*TTHMs*	*6.7*	*2.3*
Providence	*RI*	*TTHMs*	*22.5*	*16.4*
Fort Worth	*TX*	*TTHMs*	*22.6*	*11.9*
Charlestone	*SC*	*HAAs*	*23.7*	*21.3*
Boston	*MA*	*HAAs*	*3.7*	*3.7*
Honolulu	*HI*	*Dieldrin*	*0.8*	*0.1*
Austin	*TX*	*TTHMs*	*30.6*	*14.8*
Fairfax County	*VA*	*TTHMs*	*23*	*16*
St. Louis	*MO*	*HAAs*	*15.4*	*19.2*
Minneapolis	*MN*	*HAAs*	*29.6*	*27.3*
Richmond	*VA*	*HAAs*	*15*	*21.5*
Dallas	*TX*	*TTHMs*	*32.1*	*24*
New York	*NY*	*HAAs*	*30.8*	*32*
Oklahoma City	*OK*	*TTHMs*	*40*	*23.5*
Buffalo	*NY*	*TTHMs*	*32.3*	*18.9*
Cincinnati	*OH*	*TTHMs*	*33*	*8.1*
St. Paul	*MN*	*TTHMs*	*31.1*	*18.2*
Sacramento	*CA*	*HAAs*	*1.4*	*17.7*
Milwaukee	*WI*	*Bromate*	*5.4*	*1.1*
Odessa	*FL*	*HAAs*	*16.4*	*14*
Bridgeport	*CT*	*HAAs*	*40*	*31*
Louisville	*KY*	*TTHMs*	*22.8*	*13.9*
Hartford	*CT*	*TTHMs*	*33.2*	*23.5*
Springfield	*MA*	*HAAs*	*50.8*	*38.4*
Seattle	*WA*	*HAAs*	*31.5*	*25.2*
Greensboro	*NC*	*HAAs*	*42.6*	*36.6*
Newport News	*VA*	*HAAs*	*24.6*	*23.5*
Winston-Salem	*NC*	*HAAs*	*29.4*	*27.9*
San Antonio	*TX*	*HAAs*	*15.4*	*11.6*
Kent	*OH*	*TTHMs*	*45.8*	*34.2*
Birmingham	*AL*	*TTHMs*	*30.5*	*23.2*
Pittsburgh	*PA*	*TTHMs*	*54*	*13.1*
San Francisco	*CA*	*Nitrate*	*8.6*	*24.9*
Chicago	*IL*	*Combined Radium (-226*	*16.1*	*8.2*

		& -228)		
Toledo	*OH*	*TTHMs*	*35.7*	*11.9*
Oakland	*CA*	*TTHMs*	*28*	*16*
New Haven	*CT*	*TTHMs*	*37*	*26*
Manatee County	*FL*	*TTHMs*	*22.8*	*14.1*
Monroe County	*NY*	*TTHMs*	*53.8*	*6*
Long Beach	*CA*	*Arsenic (total)*	*2.6*	*10.5*
New York	*NY*	*HAAs*	*37.3*	*42.6*
Layton	*UT*	*Thallium (total)*	*18.2*	*15.6*
Fort Wayne	*IN*	*TTHMs*	*22.8*	*14.4*
Corpus Christi	*TX*	*TTHMs*	*38*	*26.5*
Tacoma	*WA*	*HAAs*	*6.5*	*20.6*
Miami	*FL*	*HAAs*	*18.7*	*22.4*
Palm Beach County	*CA*	*HAAs*	*24.7*	*17.8*
Bryn Mawr	*PA*	*TTHMs*	*31.7*	*22.3*
Erie County	*NY*	*TTHMs*	*53.7*	*21.7*
Plano	*TX*	*TTHMs*	*58.3*	*22*
Albuquerque	*NM*	*Arsenic (total)*	*6.2*	*2.3*
St. Louis County	*MO*	*HAAs*	*12*	*24.8*
Phoenix	*AZ*	*TTHMs*	*38.5*	*16.3*
Stockton	*CA*	*Arsenic (total)*	*27.1*	*8.9*
New York	*NY*	*HAAs*	*46.5*	*47.9*
Mobile	*AL*	*TTHMs*	*72.2*	*26.2*
Pinellas County	*FL*	*TTHMs*	*30.1*	*16.1*
Los Angeles Suburbs	*CA*	*TTHMs*	*30.8*	*14*
Portland	*OR*	*HAAs*	*20.2*	*26.1*
Philadelphia	*PA*	*TTHMs*	*44.6*	*31.4*
Santa Ana	*CA*	*Nitrate*	*15*	*4.2*
Totowa	*NJ*	*TTHMs*	*38.1*	*26.5*
Mahwah	*NJ*	*HAAs*	*34.6*	*26.9*
Norfolk	*VA*	*TTHMs*	*49.5*	*30*
Hillsborough County	*FL*	*TTHMs*	*27.7*	*13.4*
Tucson	*AZ*	*Arsenic (total)*	*5.7*	*0.2*
Lexington	*KY*	*TTHMs*	*48.9*	*26.8*
Tampa	*FL*	*TTHMs*	*26.4*	*16.2*
Baltimore	*MD*	*HAAs*	*44.5*	*42.1*

Alameda County	*CA*	*HAAs*	*14.2*	*20.9*
Tulsa	*OK*	*TTHMs*	*48.4*	*18.9*
Cleveland	*OH*	*HAAs*	*22.5*	*22.3*
Pittsburgh City	*PA*	*TTHMs*	*57.5*	*15.4*
Raleigh	*NC*	*TTHMs*	*42.2*	*31.4*
Columbus	*OH*	*Phosphorus*	*32.5*	*27.3*
Mesa	*AZ*	*TTHMs*	*33.3*	*13.7*
Charlotte	*NC*	*TTHMs*	*43.1*	*17.1*
Little Rock	*AR*	*Dibenz[a,h]anthracene*	*49.8*	*29.7*
Anaheim	*CA*	*Alpha particle activity*	*3.2*	*18*
Salt Lake City	*UT*	*HAAs*	*12.2*	*26.7*
Orlando	*FL*	*TTHMs*	*50.4*	*20.1*
Montgnomery & Prince George's Counties	*MD*	*HAAs*	*31.8*	*34.6*
Los Angeles	*CA*	*HAAs*	*34.4*	*29*
San Jose	*CA*	*Nitrate*	*1.8*	*10*
Haworth	*NJ*	*Dieldrin*	*13.3*	*3.4*
Cocoa	*FL*	*TTHMs*	*49.8*	*25.4*
West Milford	*NJ*	*HAAs*	*48.4*	*51.4*
Chino Hills	*CA*	*Nitrate*	*0.4*	*7.9*
Fresno	*CA*	*Nitrate*	*0.1*	*1.7*
Indianapolis	*IN*	*HAAs*	*33.6*	*36.6*
Jacksonville	*FL*	*TTHMs*	*48*	*13.3*
San Diego	*CA*	*Gross beta particle activity*	*41*	*17.8*
North Las Vegas	*NV*	*TTHMs*	*50*	*18*
Omaha	*NE*	*Arsenic (total)*	*16.6*	*17*
Houston	*TX*	*Alpha particle activity*	*20.1*	*22.2*
Reno	*NV*	*Arsenic (total)*	*39.6*	*20.4*
Riverside County	*CA*	*Perchlorate*	*8.8*	*14.5*
Las Vegas	*NV*	*TTHMs*	*62*	*27*
Riverside	*CA*	*Perchlorate*	*0.9*	*0.9*
Pensacola	*FL*	*MTBE*	*0.2*	*0.6*

Table 2. Big Cities Ranking, Top Concern and Amount of TTHMs and HAAs (ppb) from 2004 to 2009

Here is a review of ranking factors and methods.

Three factors are considered by EWG to rank drinking water quality. These factors are [79]:

- Number of Chemicals found
- Percentage of detected chemicals in tested chemicals
- Average level of each pollutant comparative to either legal limits or national average

The first factor is the total number of detected chemicals which is consists of chemicals which exceed health guidelines, chemicals which exceed legal limits, and also unregulated chemicals.

The second factor is the percentage of detected chemicals in all tested chemicals. Some states did not provide data for all the chemicals which are tested but not detected. Therefore, the number of reported tests is fewer than mandatory number of tests. In these cases EWG assumed that these water utilities tested for all 80 contaminants which are mandatory to test [79].

The third factor is the average level of each pollutant comparative to either legal limits or national average. The amount of regulated contaminates are compared to legal limits. However the amount of unregulated contaminates are compared to national average concentrations. National average concentrations are computed based on the data from utilities that reported detecting that contaminant [79].

The calculated value for each factor is scaled from 0 to 100 for all utilities. In addition weights are assigned to each of three factors according to their importance. A weight of

0.3 is assigned to the number of chemicals found (first factor), a weight of 0.2 is assigned to the percentage of detected chemicals in tested chemicals (second factor), and a weight of 0.5 is assigned to the average level of each pollutant comparative to either legal limits or national average. The final ranking is calculated by summing weighted ranks for three factors [79].

4.2. Residential Data

In addition to drinking water quality data, residential data is also used in analysis and modeling. Average property value in all the states and in big cities is obtained by average listing price. In addition population of all the states and big cities are also used in modeling.

4.2.1. Residential property price

Residential property price in states and in big cities are collected from a wide range of real estate information services. These data can change over time. The data collection process has done in September 2012. Table 3 shows the average residential property price in each state. The average price was available for all the states. However, to our knowledge there is no complete source including standard deviation and other statistical parameters.

State	Average Property Price (1000$)	State	Average Property Price (1000$)
Alabama	*625*	***New Hampshire***	*313*
Arizona	*288*	***New Jersey***	*131*

Arkansas	*217*	***New Mexico***	*120*
California	*405*	***New York***	*707*
Connecticut	*566*	***North Carolina***	*27*
Delaware	*333*	***North Dakota***	*316*
Florida	*612*	***Ohio***	*177*
Hawaii	*912*	***Oklahoma***	*602*
Idaho	*335*	***Oregon***	*642*
Illinois	*470*	***Pennsylvania***	*746*
Indiana	*178*	***Rhode Island***	*450*
Iowa	*174*	***South Carolina***	*488*
Kentucky	*607*	***South Dakota***	*319*
Maine	*498*	***Texas***	*279*
Maryland	*402*	***Utah***	*407*
Massachusetts	*535*	***Vermont***	*348*
Michigan	*594*	***Virginia***	*352*
Minnesota	*255*	***Washington***	*347*
Missouri	*203*	***West Virginia***	*194*
Montana	*493*	***Wisconsin***	*227*
Nebraska	*191*	***Wyoming***	*523*
Nevada	*398*		

Table 3. Average Residential Property Value in Different States

The average price in each state is calculated by averaging the price listings.

4.2.2. Population

Population of all the states and big cities are also used in modeling. These data are from 2010 United States Census Bureau [80].

5. Analysis and Modeling

Related bud distinct analyses have been conducted in order to study the influence of drinking water quality on residential property value. Section 5.1 presents Bayesian inference of residential property value in different states based on population and drinking water quality. Sections 5.2 and 5.3 are Bayesian inference of residential property value in big cities based on population and different factors of drinking water quality. Sections 5.4 to 5.7 are hierarchical Bayesian inference of residential property value in big cities based on population and different factors of drinking water quality in city and state level. Figure 1 shows how each section and analysis is related.

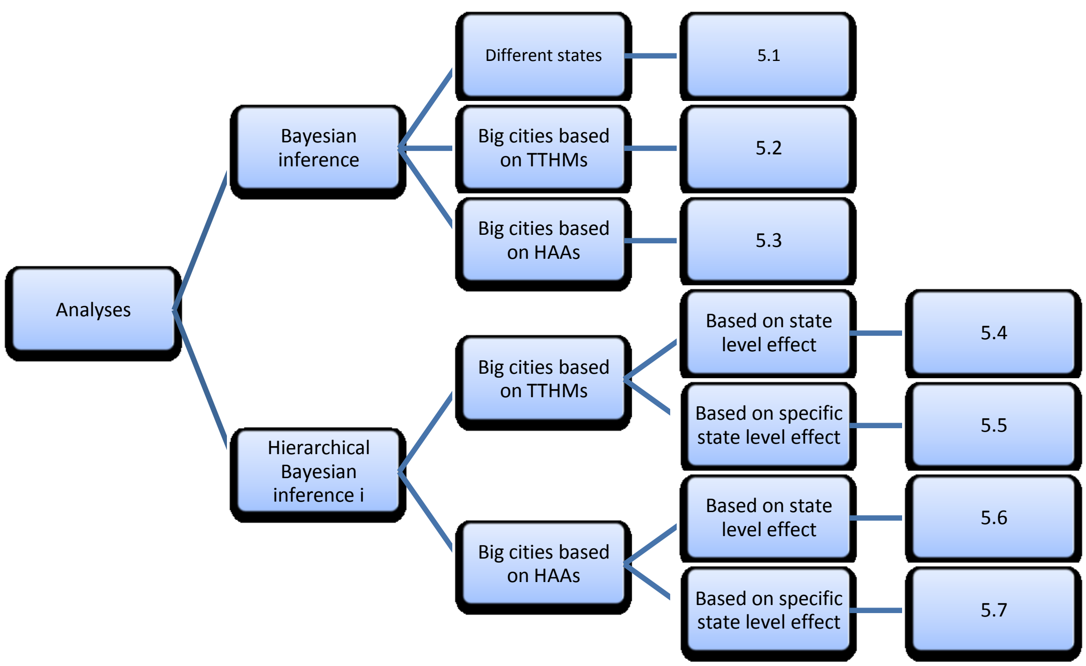

Figure 1. Connection between Sections and Analyses

5.1. Bayesian Inference of Property Value in Different States

In this analysis a model of residential property value in different states is developed based on population, number of detected chemicals in water utilities, and percentage of population served by drinking water containing those chemicals. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.1.1. Model

The Average residential property value in each state is modeled as a function of population and water pollution in each state. In this model $price_i$, pop_i, and $poll_i$ are respectively the average residential property price in the i^{th} state (\$100,000), the population of the i^{th} state(1000 persons), and the water pollution in the i^{th} state.

Where $price_i$ is a Normal distribution with mean μ_i and variance τ, the model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.1a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i$$

Equation 5.2b

The uninformative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ \beta_2 \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.3

Informative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0.1, 10^2)\ ,\ \beta_2 \sim N(-0.1, 10^2)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.4

The purpose is to estimate the posterior probability distributions of $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_2}$ by Bayesian inference. This will be done by upgrading the prior probability distributions based on data.

5.1.2. Implementation

All details of the model are explained above. The purpose of this model is to find posterior distributions which lead to an upgraded function for average residential property price in each state based on water population and water pollution in that state. Since the combination of prior distributions and likelihood distribution does not result in a posterior from the same family of the prior, the prior and posterior distributions are non-conjugate. Therefore MCMC algorithms are used to make inference about the model.

The model is developed using WinBUGS software. WinBUGS is software for Bayesian inference. This software is a version of BUGS (Bayesian inference Using Gibbs Sampling) project [81]. WinBUGS is a practical tool for using MCMC algorithms in Bayesian inference.

Two experiments have been done, one with uninformative prior distribution and the other one with informative prior distributions. In each experiment these steps are followed:

- Defining model specifications
- Loading data
- Compiling the model
- Setting required nodes
- Updating model for 10000 iterations
- Recording the results

5.1.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, β_2, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 4 shows the parameter values properties of posterior distributions using uninformative priors. Table 5 shows the parameter values properties of posterior distributions using informative priors.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.0011*	*0.1523*	*-0.3*	*-0.0026*	*0.2972*
Beta1	*0.2943*	*0.1877*	*-0.0769*	*0.2933*	*0.6687*
Beta2	*-0.3559*	*0.189*	*-0.7261*	*-0.7044*	*0.0168*
Tau	*1.031*	*0.1893*	*0.6961*	*1.018*	*1.432*

Table 4. State Based Properties of posterior based on uninformative prior

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.0001*	*0.1571*	*-0.3087*	*-0.0026*	*0.3067*
Beta1	*0.4996*	*0.1969*	*0.124*	*0.4944*	*0.9042*
Beta2	*-0.7106*	*0.2046*	*-1.132*	*-0.7044*	*-0.3331*
Tau	*0.9742*	*0.1894*	*0.639*	*0.9609*	*1.378*

Table 5. State Based Properties of posterior based on informative prior

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i$$

Equation 5.5

Based on the above results the expected value of μ_i for each sample set $(pop_i, poll_i)$ using uninformative priors is equal to:

$$E(\mu_i) = -0.0011 + 0.2943 * pop_i - 0.3559 * poll_i$$

Equation 5.6

The expected value of μ_i for each sample set $(pop_i, poll_i)$ using informative priors is equal to:

$$E(\mu_i) = -0.0001 + 0.4996 * pop_i - 0.7106 * poll_i$$

Equation 5.7

These functions for expected value of μ_i for each sample set is written using mean of β_0, β_1, and β_2. Results of experiments with informative prior and uninformative prior may seem similar. However, the small changes in model's coefficients can lead to a significant change in price estimation. Based on the equation with uninformative prior (equation 5.5) the average value of a property will be $29,430 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be $35,590 lower with each extra chemical detected in the water utilities (consuming that the chemical effect all the state population). However, based on the equation with informative prior (equation 5.6) the average value of a property will be $49,960 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be $71,060 lower with each extra chemical detected in the water utilities (consuming that the chemical effect all the state population).

The relationship between residential property value, drinking water pollution, and population in the form of the 50^{th}, $2.5^{th,}$ and 97.5^{th} confidence surfaces is shown in following 3D graphs. Figure 2, 3, 4 and 5 show different views of the 3D graph.

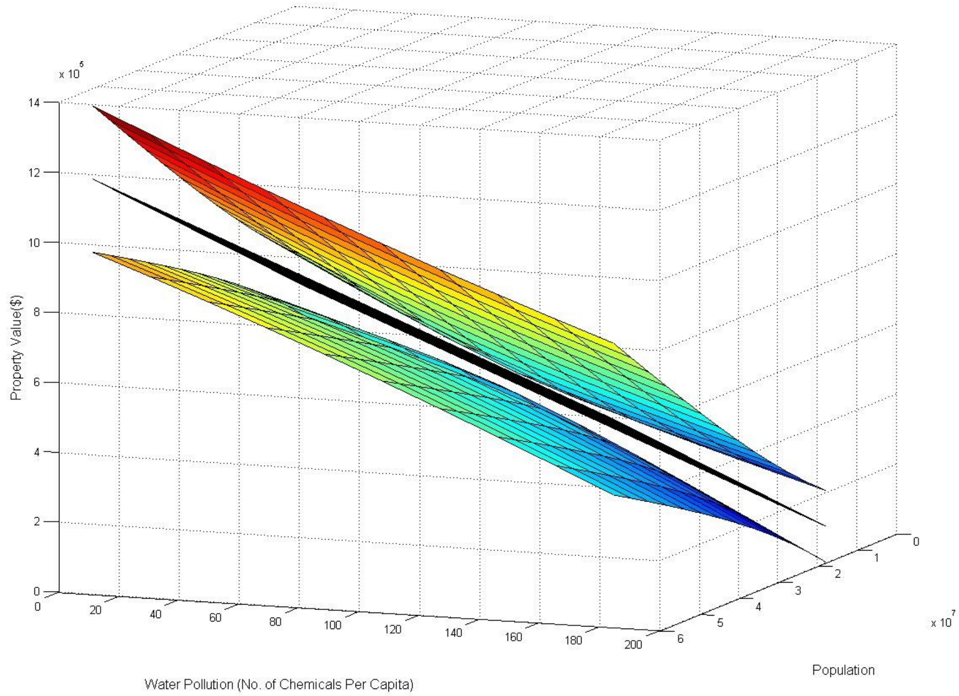

Figure 2. First view of 3D graph including median and confidence surfaces

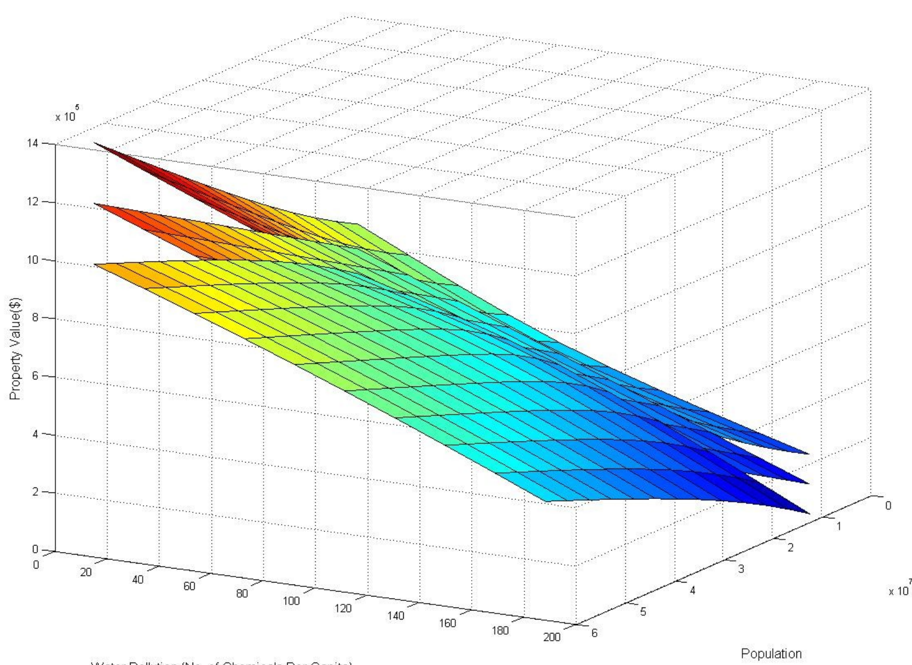

Figure 3. Second view of 3D graph including median and confidence surfaces

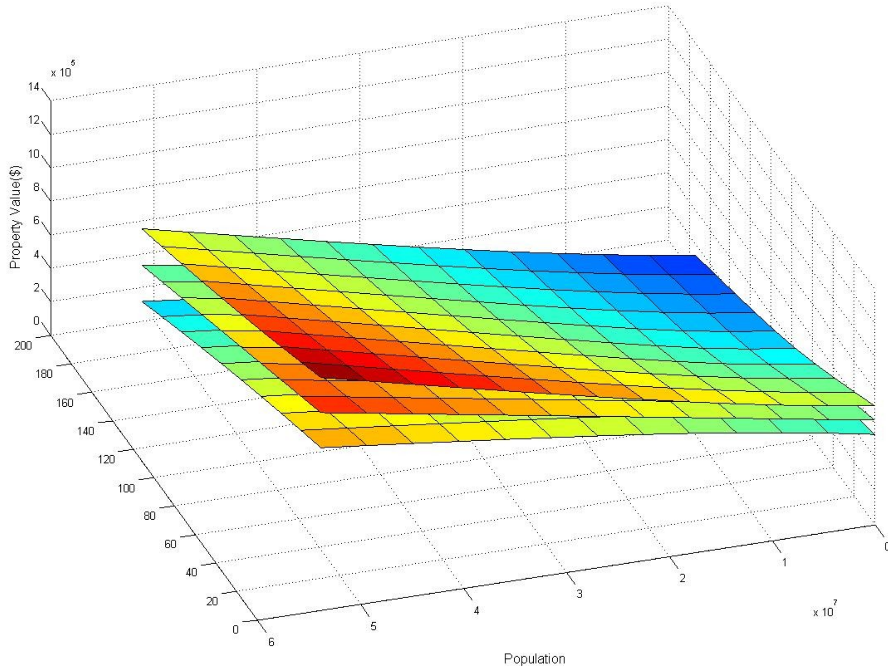

Figure 4. Third view of 3D graph including median and confidence surfaces

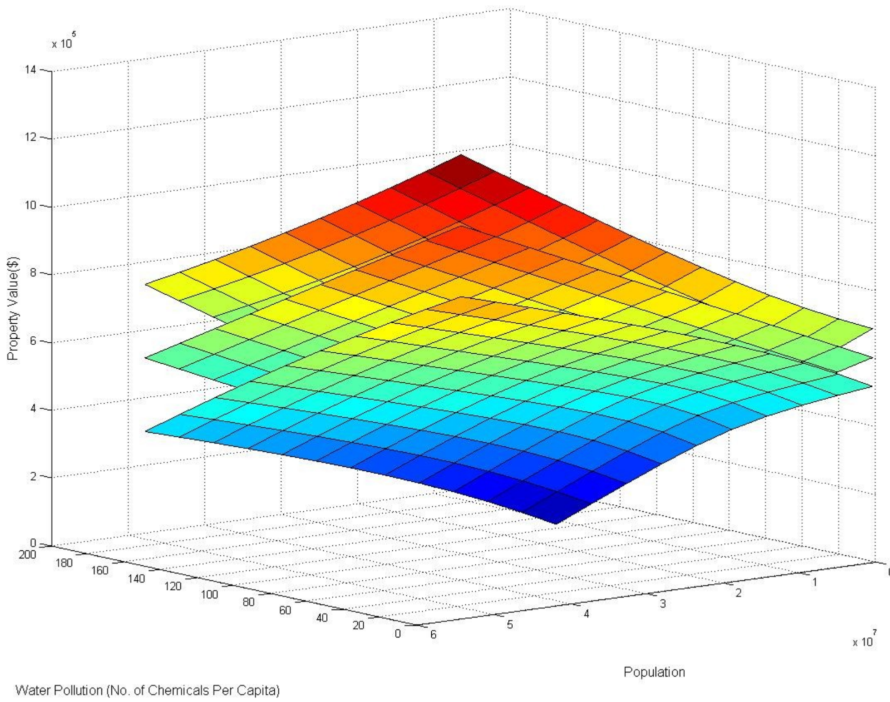

Figure 5. Fourth view of 3D graph including median and confidence surfaces

These surfaces provide estimation for residential property value of the state based on different levels of population and water pollution in the state.

5.1.4. Discussions

In this analysis the Average residential property value in each state is modeled as a function of population and water pollution in each state.

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.8a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * poll_i$$

Equation 5.9b

The final equation of μ_i (the mean value of property price distribution) gives us a model to estimate property price based on population and water pollution. The expected value of μ_i for each sample set $(pop_i, poll_i)$ using uninformative priors is equal to:

$$E(\mu_i) = -0.0011 + \ 0.2943 * pop_i - 0.3559 \ * poll_i$$

Equation 5.10

And with informative priors is equal to:

$$E(\mu_i) = -0.0001 + \ 0.4996 * pop_i - 0.7106 \ * poll_i$$

Equation 5.11

This model shows the strong direct effect of population and strong indirect effect of water pollution on the residential property price in different states.

β_0, β_1, and β_2 are coefficients. β_0 shows the effect of all unconsidered factors.

β_1 shows the effect of population. Positive amounts of β_1 show a direct relationship between property price and population and negative amounts show an indirect relationship between them. According to the results β_1 has positive means of 0.2943 and 0.4996 in inferences with uninformative prior and informative prior. This means that the relationship between property price and population is direct.

β_2 shows the effect of water pollution. Positive amounts of β_2 show a direct relationship between property price and water pollution and negative amounts show an indirect relationship between them. According to the results β_2 has negative means of -0.3559 and -0.7106 in inferences with both uninformative prior and informative prior. This means that the relationship between property price and water pollution is indirect.

Informative prior probability distributions are defined in a way that implies a direct relationship between property price and population and an indirect relationship between property price and water pollution.

The mean of β_1, which is 0.4996, is a larger positive number when using informative prior in compare of using uninformative prior, which is 0.2943. This means informative prior leads to a stronger direct relationship between property value and population.

The mean of β_2 , which is 0.7106, has higher negativity when using informative prior in compare of using uninformative prior, which is -0.3559. This means informative prior leads to a stronger indirect relationship between property value and water pollution. These correlations don't necessarily imply causation.

Relationship between water quality and property value is different in each state. This difference between states is a result of various geographical, environmental, and economical situations. States should comply with different federal and state level regulations. However enforcement policies are also an important factor in the state compliance. Stronger relationship can be interpreted as the important role of environmental factors in that state which can lead to more value in long term.

The functions for expected value of μ_i for each sample set is written using mean of of β_0, β_1, and β_2. However each one of β_0, β_1, and β_2 is a probability distribution. Having sample sets of $(price_i, pop_i, poll_i)$ for all the states and using the mean value of β_0 and β_1, gives β_2 for each state. These numbers belong to a normal distribution with a negative mean. They can be positive or negative; however they have a negative average. Table 6 shows β_2 for each state.

State	Beta	State	Beta
Alabama	*0.08*	***New Hampshire***	*0.14*
Arizona	*-0.16*	***New Jersey***	*0.20*
Arkansas	*0.11*	***New Mexico***	*0.07*
California	*-36.29*	***New York***	*-0.46*
Connecticut	*0.11*	***North Carolina***	*0.37*
Delaware	*1.86*	***North Dakota***	*-0.84*
Florida	*-0.14*	***Ohio***	*0.00*
Hawaii	*0.12*	***Oklahoma***	*0.31*
Idaho	*-0.41*	***Oregon***	*0.14*
Illinois	*-0.44*	***Pennsylvania***	*0.09*
Indiana	*0.00*	***Rhode Island***	*0.05*
Iowa	*0.02*	***South Carolina***	*1.19*
Kentucky	*0.13*	***South Dakota***	*-1.00*
Maine	*0.09*	***Texas***	*0.49*
Maryland	*-36.08*	***Utah***	*7.99*
Massachusetts	*-0.09*	***Vermont***	*-0.34*
Michigan	*0.23*	***Virginia***	*0.37*
Minnesota	*0.14*	***Washington***	*0.15*
Missouri	*0.21*	***West Virginia***	*-0.52*
Montana	*0.20*	***Wisconsin***	*0.20*
Nebraska	*0.07*	***Wyoming***	*0.45*
Nevada	*-3.01*		

Table 6. Coefficient of the effect of water pollution for each state

These coefficients are used in hierarchical Bayesian inferences in sections 5.5 and 5.7.

5.2. Bayesian Inference of Property Value in Big Cities Based on TTHMs

In this analysis a model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total trihalomethanes (TTHMs) water pollution. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.2.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population and TTHMs water pollution in each city.

In this model $price_i$, pop_i, and $tthms_i$ are respectively the average residential property price in the ith big city($100,000), the population of the ith big city(1000 person), and the TTHMs water pollution in the ith big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.12a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.13b

The uninformative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \ \beta_1 \sim N(0, 10^4)\ ,\ \ \beta_2 \sim N(0, 10^4)\ ,\ \ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.14

Informative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(1, 10^2)\ ,\ \beta_2 \sim N(-1, 10^2)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.15

We chose to leave a diffuse prior for β_0 because it shows the effect of all not considered factors and we didn't want to limit it. On the other hand we chose informative priors for β_1 and β_2. The informative prior probability distributions imply that the effect of population on property value is positive and the effect of TTHMs water pollution on property value is negative.

The purpose is to estimate the posterior probability distributions of β_0, β_1, and β_2 by Bayesian inference. This will be done by updating the prior probability distributions based on data.

5.2.2. Implementation

The purpose of this model is to find posterior distributions which lead to an updated function for average residential property price in each city based on population and TTHMs water pollution in that city. Since the combination of prior distributions and likelihood distribution does not result in a posterior from the same family of the prior, the prior and posterior distributions are non-conjugate. Therefore MCMC algorithms are used to make inference about the model.

Again the model is developed using WinBUGS software. Two experiments have been done, one with uninformative prior distribution and the other one with informative prior distributions.

5.2.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, β_2, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 7 shows the parameter values properties of posterior distributions using uninformative priors. Table 8 shows the parameter values properties of posterior distributions using informative priors.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.0005*	*0.105*	*-0.2079*	*-0.0018*	*0.2057*
Beta1	*0.1595*	*0.1052*	*-0.0470*	*0.1589*	*0.3653*
Beta2	*-0.313*	*0.1057*	*-0.5235*	*-0.3179*	*-0.111*
Tau	*1.117*	*0.1769*	*0.7997*	*1.106*	*1.49*

Table 7. TTHMs Based Properties of posterior based on uninformative prior

Node	Mean	SD	2.5%	Median	97.5%

Beta0	*-0.0005*	*0.105*	*-0.2079*	*-0.0018*	*0.2057*
Beta1	*0.1606*	*0.1052*	*-0.0459*	*0.16*	*0.3666*
Beta2	*-0.319*	*0.1057*	*-0.5246*	*-0.3189*	*-0.1119*
Tau	*1.117*	*0.1769*	*0.7993*	*1.106*	*1.49*

Table 8. TTHMs Based Properties of posterior based on informative prior

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.16

Based on the above results the expected value of μ_i for each sample set $(pop_i, tthms_i)$ using uninformative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1595 * pop_i - 0.313 * tthms_i$$

Equation 5.17

The expected value of μ_i for each sample set $(pop_i, tthms_i)$ using informative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1606 * pop_i - 0.319 * tthms_i$$

Equation 5.18

These functions for expected value of μ_i for each sample set is written using mean of of β_0, β_1, and β_2. Based on the with uninformative prior (equation 5.14) the average value of a property will be $15,950 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be $31,300 lower with each extra unit of

TTHMs pollution. However, based on the equation with informative prior (equation 5.15) the average value of a property will be \$16,060 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be \$31,900 lower with each extra unit of TTHMs pollution. Using informative priors change the results a little. However these changes were more significant in state level. An explanation is that in big cities more diverse groups of factors affect the property price. Therefore, no matter how limited the priors are, the influence of water quality is smaller than state level and we cannot increase it a lot by using informative priors.

5.2.4. Discussion

In this analysis the Average residential property value in 100 big cities of the United States is modeled as a function of population and TTHMs water pollution in each city.

The final equation of μ_i (the mean value of property price distribution) gives us a model to estimate property price based on population and TTHMs water pollution. The expected value of μ_i for each sample set $(pop_i, tthms_i)$ using uninformative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1595 * pop_i - 0.313 * tthms_i$$

Equation 5.19

And for informative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1606 * pop_i - 0.319 * tthms_i$$

Equation 5.20

This model shows the direct effect of population and indirect effect of TTHMs water pollution on the residential property price in big cities.

Similar to the first study results, β_1 has a positive mean in inferences with both uninformative prior and informative prior. It has a mean of 0.1595 for uninformative prior and 0.1606 for informative prior. This means that the relationship between property price and population is direct. However β_1 has smaller means compare to the state study which has a mean of 0.4996. This can imply that the effect of population is stronger in state level. In addition only cities with population over 250,000 are considered in this study and they are not selected from cities with diverse range of populations. This can explain the less strong effect of population in the model.

β_2 has a negative mean in inferences with both uninformative prior and informative prior. It has a mean of -0.313 for uninformative prior and -0.319 for informative prior. This means that the relationship between property price and TTHMs water pollution in big cities is indirect. These correlations don't necessarily imply causation.

It seems that residential property price is more related to water pollution in state level compare to big cities. An explanation is that in big cities more diverse groups of factors affect the property price.

5.3. Bayesian Inference of Property Value in Big Cities Based on HAAs

In this analysis a model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total haloacetic acids (HAAs) water pollution. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.3.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population sand HAAs water pollution in each city.

In this model $price_i$, pop_i, and $haas_i$ are respectively the average residential property price in the i^{th} big city ($100,000), the population of the i^{th} big city (1000 persons), and the HAAs water pollution in the i^{th} big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.21a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * haas_i$$

Equation 5.22b

The uninformative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ \beta_2 \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.23

Informative prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(1, 10^2)\ ,\ \beta_2 \sim N(-1, 10^2)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.24

We chose to leave a diffuse prior for β_0 because it shows the effect of all not considered factors and we didn't want to limit it. On the other hand we chose informative priors for β_1 and β_2. The informative prior probability distributions imply that the effect of population on property value is positive and the effect of HAAs water pollution on property value is negative.

The purpose is to estimate the posterior probability distributions of β_0, β_1, and β_2 by Bayesian inference. This will be done by upgrading the prior probability distributions based on data.

5.3.2. Implementation

The purpose of this model is to find posterior distributions which lead to an upgraded function for average residential property price in each city based on population and HAAs water pollution in that city. Since the combination of prior distributions and likelihood distribution does not result in a posterior from the same family of the prior, the prior and posterior distributions are non-conjugate. Therefore MCMC algorithms are used to make inference about the model.

Again the model is developed using WinBUGS software. Two experiments have been done, one with uninformative prior distribution and the other one with informative prior distributions.

5.3.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, β_2, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 9 shows the parameter values properties of posterior distributions using uninformative priors. Table 10 shows the parameter values properties of posterior distributions using informative priors.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.0005*	*0.1099*	*-0.2176*	*-0.0019*	*0.2153*
Beta1	*0.1617*	*0.1101*	*-0.0545*	*0.1611*	*0.3771*
Beta2	*-0.132*	*0.1107*	*-0.3471*	*-0.132*	*0.0846*
Tau	*1.02*	*0.1615*	*0.7301*	*1.01*	*1.36*

Table 9. HAAs Based Properties of posterior based on uninformative prior

Node	Mean	SD	2.5%	Median	97.5%

Beta0	*-0.0005*	*0.1099*	*-0.2176*	*-0.0019*	*0.2152*
Beta1	*0.1629*	*0.1101*	*-0.0533*	*0.1623*	*0.3785*
Beta2	*-0.1332*	*0.1107*	*-0.3484*	*-0.1331*	*0.0835*
Tau	*1.02*	*0.1615*	*0.7297*	*1.009*	*1.36*

Table 10. HAAs Based Properties of posterior based on informative prior

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * haas_i$$

Equation 5.25

Based on the above results the expected value of μ_i for each sample set $(pop_i, haas_i)$ using uninformative priors is equal to:

$$E(\mu_i) = -0.0005 + \ 0.1617 * pop_i - 0.132 \ * haas_i$$

Equation 5.26

The expected value of μ_i for each sample set $(pop_i, haas_i)$ using informative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1629 * pop_i - 0.1332 \ * haas_i$$

Equation 5.27

These functions for expected value of μ_i for each sample set is written using mean of of β_0, β_1, and β_2. Based on the equation with uninformative prior (equation5.22) the average value of a property will be $16,170 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be $13,200 lower with each extra unit of HAAs pollution. However, based on the equation with informative prior (equation 5.23) the average value of a property will be \$16,290 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be \$13,320 lower with each extra unit of HAAs pollution. Using informative priors change the results a little. However these changes were more significant in state level. An explanation is that in big cities more diverse groups of factors affect the property price. Therefore, no matter how limited the priors are, the influence of water quality is smaller than state level and we cannot increase it a lot by using informative priors.

5.3.4. Discussion

In this analysis the Average residential property value in 100 big cities of the United States is modeled as a function of population and HAAs water pollution in each city.

The final equation of μ_i (the mean value of property price distribution) gives us a model to estimate property price based on population and TTHMs water pollution. The expected value of μ_i for each sample set (($pop_i, haas_i$) using uninformative priors is equal to:

$$E(\mu_i) = -0.0005 + 0.1617 * pop_i - 0.132 \ * haas_i$$

Equation 5.28

And with informative priors is equal to:

$$E(\mu_i) = -0.0005 + \ 0.1629 * pop_i - 0.1332 \ * haas_i$$

Equation 5.29

This model shows the direct effect of population and indirect effect of HAAs water pollution on the residential property price in big cities.

Similar to the first study results, β_1 has a positive mean in inferences with both uninformative prior and informative prior. It has a mean of 0.1617 for uninformative prior and 0.1629 for informative prior. This means that the relationship between property price and population is direct. However again β_1 has smaller means compare to the state study, which is 0.4996. This can imply that the effect of population is stronger in state level.

β_2 has a negative mean in inferences with both uninformative prior and informative prior. It has a mean of -0.132 for uninformative prior and -0.1332 for informative prior. This means that the relationship between property price and HAAs water pollution in big cities is indirect. These correlations don't necessarily imply causation.

Comparing the results of the experiments based on TTHMs and HAAs water pollutions show that in big cities TTHMs water pollution has a greater correlation on property price than HAAs water pollution.

5.4. Hierarchical Bayesian Inference of Property Value in Big Cities Based on TTHMs and General State Level Effect

In this analysis a hierarchical model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total trihalomethanes (TTHMs) water pollution. In addition the general state level effect of water pollution on property price is used. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.4.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population, TTHMs water pollution in each city, and general state level effect of water pollution on property price.

In this model $price_i$, pop_i, and $tthms_i$ are respectively the average residential property price in the i^{th} big city, the population of the i^{th} big city, and the TTHMs water pollution in the i^{th} big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.30a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.31b

$$\beta_2 = a * \beta + b$$

Equation 5.32

β_2 , the coefficient of $tthms_i$, is a function of β. β is the expected value for coefficient of water pollution in estimating the property price in state level. According to the first analysis β is equal to -0.3559.

The prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ a \sim N(0, 10^4)\ ,\ b \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.33

5.4.2. Implementation

The purpose of this model is to find posterior distributions which lead to an updated function for average residential property price in each city based on population, TTHMs water pollution in that city and general state level effect of water pollution on property price. MCMC algorithms are used to make inference about the model and the model is developed using WinBUGS software.

5.4.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, a, b, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 11 shows the parameter values properties of posterior distributions.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*0.0004*	*0.1047*	*-0.2058*	*-0.0005*	*0.2061*
Beta1	*0.1599*	*0.1048*	*-0.0471*	*0.1592*	*0.3659*
A	*0.2745*	*0.9343*	*-1.844*	*0.1798*	*1.868*

B	*-0.2176*	*3.325*	*-6.602*	*-0.1885*	*6.614*
Tau	*1.119*	*1.1782*	*1.7984*	*1.109*	*1.487*

Table 11. TTHMs Based General Hierarchical Properties of posterior

The mean of β_2 can be calculated based on the results for a and b.

$$\beta_2 = (0.2745 * -0.3559) - 0.2176 = -0.3153$$

Equation 5.34

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.35

$$\beta_2 = a * \beta + b$$

Equation 5.36a

$$\beta = -0.3559$$

Equation 5.37b

Based on the above results the expected value of μ_i for each sample set $(pop_i, tthms_i)$ is equal to:

$$E(\mu_i) = 0.0004 + 0.1599 * pop_i - 0.3153 * tthms_i$$

Equation 5.38

This function for expected value of μ_i for each sample set is written using mean of of β_0, β_1, and β_2. Based on this equation (equation 5.32) the average value of a property will be \$15,990 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be \$31,530 lower with each extra unit of TTHMs pollution.

5.5. Hierarchical Bayesian Inference of Property Value in Big Cities Based on TTHMs and Specific State Level Effect

In this analysis a hierarchical model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total trihalomethanes (TTHMs) water pollution. In addition the specific state level effect of water pollution on property price is used. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.5.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population, TTHMs water pollution in each city, and specific state level effect of water pollution on property price.

In this model $price_i$, pop_i, and $tthms_i$ are respectively the average residential property price in the i^{th} big city, the population of the i^{th} big city, and the TTHMs water pollution in the i^{th} big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.39a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.40b

$$\beta_2 = a * \beta_i + b$$

Equation 5.41

β_2 , the coefficient of $tthms_i$, is a function of β_i. β_i is the coefficient of water pollution for the state that the ith city belongs to, in estimating the property price in state level. Table 5 in the first analysis shows the amount of β for different states.

The prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ a \sim N(0, 10^4)\ ,\ b \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.42

5.5.2. Implementation

The purpose of this model is to find posterior distributions which lead to an updated function for average residential property price in each city based on population, TTHMs water pollution in that city, and specific state level effect of water pollution on property price. MCMC algorithms are used to make inference about the model and the model is developed using WinBUGS software.

5.5.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, a, b, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 12 shows the parameter values properties of posterior distributions.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.061*	*0.1074*	*-0.2756*	*-0.0615*	*0.1497*
Beta1	*0.1989*	*0.1047*	*-0.01*	*0.1985*	*0.4034*
A	*0.0137*	*0.007*	*0.0002*	*0.0137*	*0.0274*
B	*-0.1881*	*0.1215*	*-0.4292*	*-0.1881*	*0.0539*
Tau	*1.161*	*0.1862*	*0.8253*	*1.15*	*1.456*

Table 12. TTHMs Based Specific Hierarchical Properties of posterior

The mean of β_2 for each state can be calculated based on the results for a and b.

$$\beta_2 = (0.0137 * \ \beta_i) - 0.1881$$

Equation 5.43

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.44

$$\beta_2 = a * \beta_i + b$$

Equation 5.45

Based on the above results the expected value of μ_i for each sample set $(pop_i, tthms_i)$ is equal to:

$$E(\mu_i) = -0.061 + 0.1989 * pop_i + \big((0.0137 * \beta_i) - 0.1881\big) * tthms_i$$

Equation 5.46

This function for expected value of μ_i for each sample set is written using mean of of β_0, and β_1. Based on this equation (equation 5.39) the average value of a property will be \$19,890 higher with each additional 1000 inhabitants. However changes in the average value of a property is related to the coefficient of water pollution for the state that the city belongs to. For example for a city in California, which has a coefficient equal to -36.2877, the average value of a property will be \$68,524 lower with each extra unit of TTHMs pollution. However for a city in New York, which has a coefficient equal to -0.4571, the average value of a property will be \$19,436 lower with each extra unit of TTHMs pollution.

5.5.4. Discussion and Comparison

In hierarchical analyses based on TTHMs, average residential property value in 100 big cities of the United States is modeled as a function of population, TTHMs water pollution in each city, and state level effect of water pollution on property price. The general analysis used the general state level effect of water pollution on property price and the specific analysis used the specific state level effect of water pollution on property price.

The general analysis provides the expected value of μ_i for each sample set $(pop_i, tthms_i)$ as:

$$E(\mu_i) = 0.0004 + 0.1599 * pop_i - 0.3153 \ * tthms_i$$

Equation 5.47

However the specific analysis provides the expected value of μ_i for each sample set $(pop_i, tthms_i)$ as:

$$E(\mu_i) = -0.061 + 0.1989 * pop_i + \big((0.0137 * \ \beta_i) - 0.1881\big) * tthms_i$$

Equation 5.48

The general model uses the average data of states, however, the specific model uses the data from the state that each city belongs to.

β_1 is the coefficient of population and β_2 is the coefficient of TTHMs. Since β_2 is defined as $\beta_2 = a * \ \beta + b$, both a and b are indirect coefficients of TTHMs. However a can be interpreted as the magnitude of the coefficient of state level water population because it is the direct coefficient of β. And b can be interpreted as the coefficient of city TTHMs water pollution.

In general analyses the effect of TTHMs and population on property price is similar to the experiment that only considers the city effect. This means that in big cities the effect of water pollution on property price is not related to the average effect of state level water pollution on property price. The explanation is that big cities have more complex structure and they are not necessarily similar to other parts.

On the other hand in specific analyses the effect of TTHMs and population on property price is not similar to the experiment that only considers the city effect. This means that the state that each city belongs to, determines the effect of TTHMs pollution on residential property price.

5.6. Hierarchical Bayesian Inference of Property Value in Big Cities Based on HAAs and General State Level Effect

In this analysis a hierarchical model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total haloacetic acids (HAAs) water pollution. In addition the general state level effect of water pollution on property price is used. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.6.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population, HAAs water pollution in each city, and general state level effect of water pollution on property price.

In this model $price_i$, pop_i, and $haas_i$ are respectively the average residential property price in the i^th^ big city, the population of the i^th^ big city, and the HAAs water pollution in the i^th^ big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.49a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * haas_i$$

Equation 5.50b

$$\beta_2 = a * \beta + b$$

Equation 5.51

β_2 , the coefficient of $haas_i$, is a function of β. β is expected value for coefficient of water pollution in estimating the property price in state level. According to the first analysis β is equal to -0.3559.

The prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ a \sim N(0, 10^4)\ ,\ b \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.52

5.6.2. Implementation

The purpose of this model is to find posterior distributions which lead to an updated function for average residential property price in each city based on population, HAAs water pollution in that city, and general state level effect of water pollution on property price. MCMC algorithms are used to make inference about the model and the model is developed using WinBUGS software.

5.6.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, a, b, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 13 shows the parameter values properties of posterior distributions.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*0.0004*	*0.1095*	*-0.2154*	*-0.0005*	*0.2157*
Beta1	*0.1621*	*0.1097*	*-0.0545*	*0.1614*	*0.3777*
A	*0.1928*	*0.9831*	*-1.939*	*0.0552*	*1.963*
B	*-0.1038*	*1.298*	*-2.579*	*-0.0925*	*2.58*
Tau	*1.021*	*0.1627*	*0.7289*	*1.013*	*1.357*

Table 13. HAAs Based General Hierarchical Properties of posterior

The mean of β_2 can be calculated based on the results for a and b.

$$\beta_2 = (0.1928 * \ -0.132) - 0.1038 = -0.1292$$

Equation 5.53

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * haas_i$$

Equation 5.54

$$\beta_2 = a * \ \beta + b$$

Equation 5.55a

$$\beta = \text{-0.132}$$

Equation 5.56b

Based on the above results the expected value of μ_i for each sample set $(pop_i, haas_i)$ is equal to:

$$E(\mu_i) = 0.0004 + 0.1621 * pop_i - 0.1292 * haas_i$$

Equation 5.57

This function for expected value of μ_i for each sample set is written using mean of of β_0, β_1, and β_2. . Based on this equation (equation 5.48) the average value of a property will be $16,210 higher with each additional 1000 inhabitants. On the other hand the average value of a property will be $12,920 lower with each extra unit of HAAs pollution.

5.7. Hierarchical Bayesian Inference of Property Value in Big Cities Based on HAAs and Specific State Level Effect

In this analysis a hierarchical model of residential property value in 100 big cities (with population over 250,000) of the United States is developed based on population and total haloacetic acids (HAAs) water pollution. In addition the specific state level effect of water pollution on property price is used. The following sections explain model specification, implementation and results and continue with a brief discussion.

5.7.1. Model

The Average residential property value in 100 big cities of the United States is modeled as a function of population, HAAs water pollution in each city, and specific state level effect of water pollution on property price.

In this model $price_i$, pop_i, and $haas_i$ are respectively the average residential property price in the ith big city, the population of the ith big city, and the HAAs water pollution in the ith big city. Other model parameters are the same as the first analysis.

The model likelihood is defined as:

$$price_i \sim N(\mu_i, \tau)$$

Equation 5.49a

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * haas_i$$

Equation 5.58b

$$\beta_2 = a * \beta_i + b$$

Equation 5.59

β_2 , the coefficient of $haas_i$, is a function of β_i. β_i is the coefficient of water pollution for the state that the ith city belongs to, in estimating the property price in state level.

The prior probability distributions are defined as:

$$\beta_0 \sim N(0, 10^4)\ ,\ \beta_1 \sim N(0, 10^4)\ ,\ a \sim N(0, 10^4)\ ,\ b \sim N(0, 10^4)\ ,\ \tau \sim \Gamma(0.01, 0.01)$$

Equation 5.60

5.7.2. Implementation

The purpose of this model is to find posterior distributions which lead to an updated function for average residential property price in each city based on population, HAAs water pollution in that city, and specific state level effect of water pollution on property price. MCMC algorithms are used to make inference about the model and the model is developed using WinBUGS software.

5.7.3. Results

Results of Bayesian inference consist of parameter values' properties of posterior probability distributions. Having these properties, the posterior probability distributions of β_0, β_1, a, b, and τ can be estimated. These properties are mean, standard deviation, 2.5th percentile, median, and 97.5th percentile.

The results of 10,000 iterations for both experiments, discarding n iteration for burn-in, are shown. Table 14 shows the parameter values properties of posterior distributions.

Node	Mean	SD	2.5%	Median	97.5%
Beta0	*-0.0109*	*0.1119*	*-0.2359*	*-0.01168*	*0.2073*
Beta1	*0.1785*	*0.114*	*-0.0494*	*0.1778*	*0.4015*
A	*0.0042*	*0.0078*	*-0.0109*	*0.0043*	*0.0197*
B	*-0.0963*	*0.1259*	*-0.3462*	*-0.09634*	*0.1545*
Tau	*1.013*	*0.1624*	*0.7199*	*1.003*	*1.348*

Table 14. HAAs Based Specific Hierarchical Properties of posterior

The mean of β_2 for each state can be calculated based on the results for a and b.

$$\beta_2 = (0.0042 * \beta_i) - 0.0963$$

Equation 5.61

As mentioned in model specification model likelihood is defined as:

$$\mu_i = \beta_0 + \beta_1 * pop_i + \beta_2 * tthms_i$$

Equation 5.62

$$\beta_2 = a * \beta_i + b$$

Equation 5.63

Based on the above results the expected value of μ_i for each sample set $(pop_i, tthms_i)$ is equal to:

$$E(\mu_i) = -0.0109 + 0.1785 * pop_i + (0.0042 * \beta_i) - 0.0963) * haas_i$$

Equation 5.64

This function for expected value of μ_i for each sample set is written using mean of of β_0, and β_1. Based on this equation (equation 5.55) the average value of a property will be \$17,850 higher with each additional 1000 inhabitants. However changes in the average value of a property is related to the coefficient of water pollution for the state that the city belongs to. For example for a city in California, which has a coefficient equal to -36.2877, the average value of a property will be \$24,870 lower with each extra unit of TTHMs pollution. However for a city in New York, which has a coefficient equal to -0.4571, the average value of a property will be \$9,821 lower with each extra unit of TTHMs pollution.

5.7.4. Discussion and Comparison

In hierarchical analyses based on HAAs, average residential property value in 100 big cities of the United States is modeled as a function of population, HAAs water pollution in each city, and state level effect of water pollution on property price. The general analysis used the general state level effect of water pollution on property price and the specific analysis used the specific state level effect of water pollution on property price.

The general analysis provides the expected value of μ_i for each sample set $(pop_i, haas_i)$ as:

$$E(\mu_i) = 0.0004 + 0.1621 * pop_i - 0.1292 * haas_i$$

Equation 5.65

However the specific analysis provides the expected value of μ_i for each sample set $(pop_i, haas_i)$ as:

$$E(\mu_i) = -0.0109 + 0.1785 * pop_i + (0.0042 * \beta_i) - 0.0963) * haas_i$$

Equation 5.66

The general model uses the average data of states, however, the specific model uses the data from the state that each belongs to.

β_1 is the coefficient of population and β_2 is the coefficient of HAAs. Since β_2 is defined as $\beta_2 = a * \beta + b$, both a and b are indirect coefficients of HAAs. However a can be interpreted as the magnitude of the coefficient of general state level water population because it is the direct coefficient of β. And b can be interpreted as the coefficient of city HAAs water pollution.

In general analyses the effect of HAAs and population on property price is similar to the experiment that only considers the city effect. This means that in big cities the effect of water pollution on property price is not related to the average effect of state level water pollution on property price. The explanation is that big cities have more complex structure and they are not necessarily similar to other parts.

On the other hand in specific analyses the effect of HAAs and population on property price is not similar to the experiment that only considers the city effect. This means that the state that each city belongs to, determines the effect of HAAs pollution on residential property price.

6. Conclusion

Drinking water quality is an important concern all over the world as poor water quality can easily threaten human, environmental and economic health. In terms of improving water quality, strong policies and enforcement are as important as technology and infrastructure improvements. To take steps toward stronger policies, an understanding of economic impact of water quality is required.

In this thesis the relationship between drinking water quality and residential property value based on population is modeled in different levels. This modeling leads to estimate the economic impact of drinking water quality, which is an important factor in environmental policy making process.

To select the proper model we study our data. We observe that property value has a linear relationship with water quality. We checked other factors such as population and area; and also tried some potential models such as linear and exponential. We found that having a linear model with variable coefficients for pollution and population is the best choice for our study. We also used hierarchical models to consider water pollution in different levels such as city and state.

This thesis features three inter-related statistical analyses. The influence of water quality on property price is modeled separately in different states and in big cities. The results of the state-level analysis are used as inputs into a hierarchical model of city-level data. This model studies the influence of general water quality and the most important pollutants in big cities (TTHMs and HAAs) in state and city levels. Population is considered in all the analyses.

Working on data from an entire country gives us the opportunity to develop a general model that is not strongly dependent on local features, which is a novel approach in relating property values to local factors. Further, merging data from disparate sources has created a novel dataset focusing on urban areas, mixing water quality data and property price data in a large scale, and considering population as a factor in all analyses. In addition using Bayesian and hierarchical Bayesian methods allow us to study both state and city influence at the same time.

Results of the analyses show that water pollution impact on property price in states has a negative mean. In other words, pollution reduces property value. Both TTHMs and HAAs, which are generally the most important pollutants in big cities, also have negative impact on residential property price in big cities. The impact of TTHMs pollution is more than HAAs pollution.

Results of general hierarchical analyses show that in big cities the effect of TTHMs and HAAs pollution in the city is more important than general effect of water pollution. As big cities have more complex structure and there are a lot of factors influencing residential property price, the effect of general water pollution is not that important. This is because big cities have different features which are not necessarily similar to the rest of the state they belong to.

On the other hand, results of specific hierarchical analyses show that the state that each city belongs to, determines the effect of TTHMs and HAAs pollution on residential property price.

Based on the results reducing water pollution in state level can lead to higher property price which is a factor that shows people willingness to pay for better water quality. In other words this can be counted as one of the economic benefits of improving water quality. This could be useful in policy making cost-benefit analyses.

It is important to notice that what we have found is the relationship between water quality and property price and it does not mean that water quality is the direct reason of these price changes. In other words this correlation does not imply causation.

Property price is a function of different parameters such as property features, neighborhood features (including environmental quality factors) and economy. We have shown that improvement in water quality is correlated to increase in property value. Better water quality is not necessarily the direct reason of increase in property value.

Different states and cities have different level of correlation between property value and water quality. Stronger relationship can be interpreted as the important role of environmental factors in that states or cities which can lead to more value in long term. However water quality is not necessarily the direct reason of increase in property value.

The same thing is true for big cities as well. However in big cities the importance of top concern pollutant is much higher. As an example, to gain more economic benefits from water quality improvement in big cities, it's better to focus on top concern pollutants (which are normally TTHMs and HAAs).

Moreover, we need to consider the state that each big city belongs to, in order to estimate the economic benefits from decreasing TTHMs and HAAs pollution in that big city. In other words the effect of water pollution on property value in that state level can change the effect of TTHMs pollution on property value in city level.

Analyzing data from an entire country lets us to focus on the mean behavior. This means that the exact location and structural features are overshadowed by diversity. These findings can be used as a decision support system for environmental policy making. In addition to general modeling in state level, city level modeling gives us the opportunity to consider different and more complex structures of big cities.

These models are helpful to predict the economical impact of improvement in water quality in different scales. Getting more economic benefits can be a motivation to invest more budget in water quality improvement.

Solving water pollution problems has different aspects including knowledge and education, technology and methods, and law and management. Knowledge and education is the first step. Public awareness and academic education and research on both technological and economical aspects are necessary. The second step is developing water treatment, sewage treatment, and ecosystem restoration methods and equipments. At last creating legal bases for employing standards, improving management, and allocating more budgets to solve water problem are necessary actions to solve the problem.

All of these aspects need strong policies to be practical and effective. These analyses show the economic benefits of water quality improvement in different scales. Knowing the economic benefits of this improvement is helpful for making related decisions. They can be used as decision support data to motivate more investments and stronger policies.

7. Future Works

There are many opportunities for future research. Here are some examples of how this study can be extended in order to further explore the topic.

Conducting a survey on willingness to pay for better water quality will be helpful. Using the results of modeling and survey together will lead to stronger and more reliable predictions.

What this research offers is the economic benefit of drinking water quality improvement. By finding the cost of this improvement in different scales, the cost-benefit analysis can be done. The cost-benefit analysis can be used to provide a basis for comparing different possible projects. In addition the probable results of the cost-benefit analysis can be an economic motivation for improving policies.

Another possible research is to extend the present models. This could be done by using more water quality factors or by using more residential factors. Using more factors will lead to have a more accurate model and better predictions of economic benefits of drinking water quality improvement. In addition the same models can be used to predict the economic impact of sewage outflows, which is an important threat to water quality.

And at the end there are still a lot of data gaps in big scales. There is no complete database of mishaps in water utilities. Having access to these data can resolve some encountered limitations during doing this research. Data of water quality in rural areas and in big scales, data of outbreaks in regulated chemicals, data of unregulated chemicals, and also policy enforcement data are still not available. Data gathering can be a step towards more reliable studies.

8. Bibliography

[1] D.W. Dilks, R.P. Canale, P.G. Meier, "Development of Bayesian Monte Carlo techniques for water quality model uncertainty," *Ecological Modelling,* vol. 62.1, pp. 149-162, 1992.

[2] K.H. Reckhow, "Water quality prediction and probability network models," *Canadian Journal of Fisheries and Aquatic Sciences,* vol. 56.7, pp. 1150-1158, 1999.

[3] C.K. Uejio, T.W. Peters, J.A. Patz, "Inland lake indicator bacteria: Long-term impervious surface and weather influences and a predictive Bayesian model," *Lake and Reservoir Management,* vol. 28.3, 2012.

[4] F.A. Ward, "Decision support for water policy: a review of economic concepts and tools," *Water policy ,* vol. 9.1, pp. 1-31, 2007.

[5] D.L. Johnson, et al., "Meanings of environmental terms," *Journal of Environmental Quality,* vol. 26.3, pp. 581-589, 1997.

[6] G.M. Carr, J.P. Neary, Water Quality for Ecosystem and Human Health, 2nd Edition, United Nations Environment Programme, Global Environment Monitoring

System, 2008.

[7] M. Palaniappan, "Clearing the Waters: A focus on water quality solutions," United Nations Environment Programme, Division of Environmental Policy Implementation, 2010.

[8] P.H. Gleick , Dirty-water: Estimated Deaths from Water-related Diseases 2000-2020, Pacific Institute for Studies in Development, Environment, and Security, 2002.

[9] R.E. Black, S.S. Morris, J. Bryce, "Where and why are 10 million children dying every year?," *The Lancet ,* vol. 361.9376, pp. 2226-2234, 2003.

[10] J. Pronczuk de Garbino, Children’s health and the environment - A global perspective, Geneva, 2004.

[11] J. Bucknall, Making the most of scarcity: accountability for better water management results in the Middle East and North Africa, World Bank Publications, 2007.

[12] N.E. Peters, M. Meybeck, D.V. Chapman, "Effects of Human Activities on Water Quality," *Encyclopedia of Hydrological Sciences,* 2005.

[13] N.E. Peters, M. Meybeck., "Water quality degradation effects on freshwater availability: Impacts of human activities," *Water International,* vol. 25.2, pp. 185-193, 2000.

[14] UN-Water, "Water Quality: Policy Brief," http://www.unwater.org/downloads/waterquality_policybrief.pdf, 2011.

[15] World Health Organization, "Progress on drinking water and sanitation: Special focus on sanitation.," UNICEF, New York, and World Health Organization, Geneva, 2008.

[16] United Nations Word Water Assessment Programme, "The United Nations World Water Development Report 3: Water in Changing Word," UNESCO, Paris and Earthscan, London, 2009.

[17] World Health Organization, Guidelines for drinking-water quality, World Health Organization, 2011.

[18] European Commission, "Drinking Water Directives," http://ec.europa.eu/environment/water/water-drink/index_en.html.

[19] United States Environmental Protection Agency, "Safe Drinking Water Act,"

http://water.epa.gov/lawsregs/rulesregs/sdwa/index.cfm.

[20] Safe Drinking Water Foundation, "Purpose of Drinking Water Quality Guidelines/ Regulations," http://www.safewater.org/PDFS/PurposeofDrinkingWaterQualityGuidelinesRegulations.pdf.

[21] L. Fewtrell, J. Bartram., Water Quality: Guidelines, Standards and Health, Assessment of Risk and Risk Management for Water-Related Infectious Disease, International Water Assn, 2001.

[22] R. Choukr-Allah, A. Hamdy, "Wastewater recycling and reuse in Mediterranean region as a potential resource for water saving," *Options,* vol. 89, 2003.

[23] JA Harris, RJ Hobbs, E Higgs, J Aronson, "Ecological restoration and global climate change," *Restoration Ecology,* vol. 14.2, pp. 170-176, 2006.

[24] "The value of New Jersey's ecosystem services and natural capital," The Gund Institute of Ecological Economics, Burlington, VT and The New Jersey Department of Environmental Protection, Trenton, New Jersey, 2007.

[25] J. Luttik, "The value of trees, water and open space as reflected by house prices in

the Netherlands," *Landscape and Urban Planning,* vol. 48.3, pp. 161-167, 2000.

[26] E. Hui, C.K. Chau, L. Pun, M.Y. Law, "Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix," *Building and Environment,* vol. 42.6, pp. 2333-2343, 2007.

[27] R. Cellmer, A. Senetra, A. Szczepanska, "The effect of environmental factors on property value," in *FIG Working Week 2012*, Rome, 2012.

[28] D.J. Morales, "The contribution of trees to residential property value," *Journal of Arboriculture,* vol. 6.11, pp. 305-308, 1980.

[29] B.L. Mahan, S. Polasky, and R.M. Adams, "Valuing urban wetlands: a property price approach," *Land Economics,* pp. 100-113, 2000.

[30] C.G. Leggett, N.E. Bockstael, "Evidence of the effects of water quality on residential land prices," *Journal of Environmental Economics and Management,* vol. 39.2, pp. 121-144, 2000.

[31] E.L. David, "Lakeshore property values: a guide to public investment in recreation," *Water Resources Research,* vol. 4.4 , pp. 697-707, 1968.

[32] D.J. Epp, K.S. Al-Ani, "The effect of water quality on rural nonfarm residential property values," *American Journal of Agricultural Economics,* vol. 61.3, pp. 529-534, 1979.

[33] C.E. Young, "Perceived Water Quality and the Value of Seasonal Homes.," *JAWRA Journal of the American Water Resources Association,* vol. 20.2, pp. 163-166, 2007.

[34] R. Mendelsohn, D. Hellerstein, M. Huguenin, R. Unsworth, R. Brazee, "Measuring hazardous waste damages with panel models," *Journal of Environmental Economics and Management,* vol. 22.3, pp. 259-271, 1992.

[35] G.M. Brown, H.O. Pollakowski, "Economic valuation of shoreline," *The Review of Economics and Statistics,* vol. 59.3, pp. 272-278, 1977.

[36] B.T. Michael, S.L. Vicky, S.J. Michael, "Residential real estate prices: a room with a view," *Journal of Real Estate Research,* vol. 23.1, pp. 129-138, 2002.

[37] S. Chattopadhyay, "Estimating the demand for air quality: new evidence based on the Chicago housing market," *Land Economics,* pp. 22-38, 1999.

[38] R.G. Ridker, J. A. Henning, " "The determinants of residential property values with

special reference to air pollution," *The Review of Economics and Statistics,* vol. 49.2, pp. 246-257, 1967.

[39] V.K. Smith, J.C. Huang, "Can markets value air quality? A meta-analysis of hedonic property value models," *Journal of Political Economy,* pp. 209-227, 1995.

[40] R.J. Anderson, T.D. Crocker, "Air pollution and residential property values," *Urban Studies,* vol. 8.3, pp. 171-180, 1971.

[41] A.H. Darling, "Measuring benefits generated by urban water parks," *Land Economics,* vol. 49.1, pp. 22-34, 1973.

[42] S.W. Hamilton, G.M. Schwann, "Do high voltage electric transmission lines affect property value?," *Land Economics,* vol. 71.4, pp. 436-444, 1995.

[43] G. Blomquist, "The effect of electric utility power plant location on area property value," *Land Economics,* vol. 50.1, pp. 97-100, 1974.

[44] M. Espey, H. Lopez, "The impact of airport noise and proximity on residential property values," *Growth and Change,* vol. 31.3, pp. 408-419, 2002.

[45] G. Pennington, N. Topham, R. Ward, "Aircraft noise and residential property

values adjacent to Manchester International Airport," *Journal of Transport Economics and Policy ,* pp. 49-59, 1990.

[46] I.R. Lake, I.J. Bateman, B.H. Day, A.A. Lovett, "Improving land compensation procedures via GIS and hedonic pricing," *Environment and Planning C,* vol. 18.6, pp. 681-696, 2000.

[47] J.J. McCluskey, G.C. Rausser, "Estimation of perceived risk and its effect on property values," *Land Economics,* vol. 77, no. 1, pp. 42-55, 2001.

[48] G.H. McClelland, W.D. Schulze, B. Hurd, "The Effect of Risk Beliefs on Property Values: A Case Study of a Hazardous Waste Site," *Risk Analysis,* vol. 10, no. 4, pp. 485-497, 1990.

[49] H. Jenkins-Smith, G.W. Bassett, "Perceived Risk and Uncertainty of Nuclear Waste: Differences Among Science, Business, and Environmental Group Members," *Risk Analysis,* vol. 14, no. 5, pp. 851-856, 1994.

[50] T. Gayer, J.T. Hamilton, W.K. Viscusi, "Private values of risk tradeoffs at superfund sites: housing market evidence on learning about risk," *Review of Economics and Statistics,* vol. 82, no. 3, pp. 439-451, 2000.

[51] A.M. Freeman III, "Hedonic prices, property values and measuring environmental benefits: a survey of the issues," *The Scandinavian Journal of Economics,* pp. 154-173, 1979.

[52] K.G. Mäler, "A note on the use of property values in estimating marginal willingness to pay for environmental quality," *Journal of environmental Economics and Management,* vol. 4.4, pp. 355-369, 1977.

[53] G. Garrod, K.G. Willis, Economic valuation of the environment: methods and case studies, Cheltenham: Edward Elgar, 1999.

[54] K.G. Mäler, J.R. Vincent, Handbook of Environmental Economics: Valuing environmental changes. Vol. 2, North Holland, 2005.

[55] E. Cassel, R. Mendelsohn, "The choice of functional forms for hedonic price equations: comment," *Journal of Urban Economics,* vol. 18.2, pp. 135-142, 1985.

[56] J.I. Barzilay, W.G. Weinberg, J.W. Eley, The water we drink: Water quality and its effects on health, Rutgers University Press, 1999.

[57] C. Wu, C. Maurer, Y. Wang, S. Xue, D.L. Davis, "Water pollution and human health in China," *Environmental Health Perspectives,* vol. 107.4, pp. 251-256,

1999.

[58] A.H. Smith, E.O. Lingas, M. Rahman, "Contamination of drinking-water by arsenic in Bangladesh: a public health emergency," *Bulletin of the World Health Organization,* vol. 78.9, pp. 1093-1103, 2000.

[59] T. Yoshida, H. Yamauchi, G.F. Sun, "Chronic health effects in people exposed to arsenic via the drinking water: dose–response relationships in review," *Toxicology and applied pharmacology,* vol. 198.3, pp. 243-252, 2004.

[60] H.B. Stauffer, Contemporary Bayesian and frequentist statistical research methods for natural resource scientists, Wiley-Interscience, 2007.

[61] P.M. Lee, Bayesian statistics: an introduction, Wiley, 2012.

[62] W.M. Bolstad, Introduction to Bayesian statistics, Wiley-Interscience, 2007.

[63] D.V. Lindley, Introduction to probability and statistics from bayesian viewpoint. part 2 inference, CUP Archive, 1965.

[64] G.E.P. Box, G.C. Tiao, Bayesian inference in statistical analysis, Wiley-Interscience, 2011.

[65] B. Hall, "Bayesian Inference," Cran.r-project, LaplacesDemon Package, http://cran.r-project.org/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf.

[66] H. Raiffa, R. Schlaifer, Applied Statistical Decision Theory, Harvard Business School Publications, 1961.

[67] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian data analysis, Chapman & Hall/CRC, 2004.

[68] J.R. Norris, Markov chains, Cambridge series in statistical and probabilistic Mathematics, Cambridge university press, 1998.

[69] S.P. Meyn, R.L. Tweedie, P.W. Glynn, Markov chains and stochastic stability. Vol. 2, Cambridge University Press, 2009.

[70] C.M. Grinstead, J.L. Snell, "Markov Chains," in *Introduction to probability*, American Mathematical Soc., 1998.

[71] M.H. Kalos, P.A. Whitlock, Monte carlo methods, Wiley-VCH, 2008.

[72] D.J.C. MacKay, "Introduction to monte carlo methods," in *Learning in graphical*

models, Springer Netherlands, 1998, pp. 175-204.

[73] P. Glasserman, Monte Carlo methods in financial engineering, Springer, 2003.

[74] R.M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," University of Toronto, 1993.

[75] C.J. Geyer, "Practical markov chain monte carlo," *Statistical Science,* vol. 7.4, pp. 473-483, 1992.

[76] D. Gamerman, H.F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference, Vol. 68, Chapman & Hall/CRC, 2006.

[77] C. Andrieu, N. De Freitas, A. Doucet, M.I. Jordan, "An introduction to MCMC for machine learning," *Machine learning,* Vols. 50.1-2, pp. 5-43, 2003.

[78] Environmental Working Group, "About the Environmental Working Group," http://www.ewg.org/about.

[79] Environmental Working Group, "National Drinking Water Database," http://www.ewg.org/tap-water/, 2009.

[80] U.S. Census Bureau, U.S. Census Bureau Data, Retrieved 2011-07-13, 2010.

[81] MRC Biostatistics Unit, Cambridge, "The BUGS project, WinBUGS," http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.