Examining Implicit Price Variation for Lake Water Quality

Kristen Marie Swedberg

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Agricultural and Applied Economics

Kevin Boyle
Kelly Cobourn
Wei Zhang

December 16, 2020
Blacksburg, VA

Keywords: water quality, hedonic model, lake housing market

Examining Implicit Price Variation for Lake Water Quality

Kristen Marie Swedberg

ABSTRACT

Hedonic models are used to estimate implicit prices for water quality in housing markets. Recent studies aggregate sales across large spatial areas in scaled-up models leading to a concern that these models may overlook regional heterogeneity in water-quality preferences. We estimate scaled-up hedonic models comprised of multiple states and individual states and investigate how observations from subregions can differ. We find that the scaled-up model results are driven by select subregions. The results of this study call into question hedonic models using data for large geographic regions where substantial differences may arise across housing markets.

Examining Implicit Price Variation for Lake Water Quality

Kristen Marie Swedberg

GENERAL AUDIENCE ABSTRACT

Water quality in lakes impacts the prices people pay for lakefront properties. However, these effects can vary across different housing markets. We study whether owners of lakefront properties throughout the northeast and upper Midwest are willing to pay the same amount for water in lakes. We find that in multi-state housing markets, the effects from one state can dominate the overall results and there is likely heterogeneity in preferences across housing markets.

Table of Contents

1 Introduction

Lakes are an important resource providing drinking water, aesthetics, and recreational opportunities. However, nutrient pollution in U.S. lakes is widespread, with 40% of lakes sampled in 2012 exceeding recommended levels of phosphorous and 35% exceeding recommended levels of nitrogen (U.S. Environmental Protection Agency, 2016). Nitrogen and phosphorous are natural components of lake ecosystems, but in high levels impact water quality through increased algae growth and reduced water clarity. An important consideration for policy analyses is the value people place on the effects of nitrogen and phosphorous on lake water quality (Griffiths et al., 2012; Keiser, 2019). One approach is to study the benefits of water quality capitalized in housing markets using hedonic models of property sale prices (Rosen, 1974; Taylor, 2017), and a number of studies have estimated the implicit prices of water quality embedded in sale prices of lake homes (Gibbs et al., 2002; Michael et al., 1996; Moore et al., 2020; Walsh & Milon, 2016).

Hedonic price models require variation in the water quality measurement to estimate implicit price. As lake water quality changes slowly over time, researchers often use cross-sectional data incorporating measurements and property sales from multiple lakes where there is spatial variation in lake water quality. The standard approach, therefore, is to identify housing markets within a state that contain multiple lakes with differing levels of water quality (Gibbs et al., 2002; Kashian et al., 2006; Michael et al., 1996; Walsh et al., 2011; Wolf & Klaiber, 2017). A few studies have used an alternative approach exploiting temporal variation within a single lake (Carey & Leftwich, 2007; Weng et al., 2020). Recently Moore et al. (2020) estimated implicit prices for water quality at a national scale, incorporating data from 113 lakes across 32 states in a single hedonic model.[1] The goal of this paper is to is to understand the implications of applying cross-sectional analysis on the national scale, *a la* Moore et al., as compared to

[1] Other national scale studies are present in air quality literature (Bayer et al., 2009; Chay & Greenstone, 2005).

the more standard approach of exploiting spatial variation in water quality in lakes within a common real estate market.

Cross-sectional hedonic analyses reveal the impacts of water quality within assumed market areas, relying on specific information related to property sales and water quality. Yet there is a tension between identifying spatial variation in lake water quality to support hedonic-model estimation and risking going to such a large area that markets and preferences may differ, potentially obscuring important differences in the implicit prices of water quality across markets. For example, prior research suggests implicit prices are heterogenous across distinct regions in New Hampshire and Maine (Gibbs et al., 2002; Poor et al., 2001). In contrast, Moore et al. (2020) conclude preferences for water quality are consistent across all 32 states in their study area. To expand upon this line of investigation we estimate implicit prices across multiple states and examine how the use of cross-sectional data over a large study area affects implicit-price estimates.

We do this because estimating water-quality benefits at broad geographic scales is essential to federal regulatory impact analyses under the Clean Water Act and state-level actions to implement national policy guidance. The open question is whether a single model over a large area can support decision making or is it important to estimate models specific to markets in smaller regions at the state or substate levels. Using sales across broad areas is appealing because one or a small number of models need to be estimated and there is not a need to reconcile differences in estimation results across models. However, it may be important to consider whether estimation results across large geographic areas obscure important subregion differences. By estimating the implicit price for each subregion, we investigate whether the scaled-up model results truly reflect the preferences for the entire geographic area.

In this study, we analyze implicit prices for water quality across six states throughout the northeast and upper Midwest with hedonic models at different geographic scales. We consider both traditional hedonic models comprised of sales from a specific region within a state and scaled-up hedonic models with sales aggregated at the state level and across multiple states. We find water quality preferences are heterogenous at both substate and state levels, which is not accounted for when using a scaled-up hedonic model. The key insight from our research is that using a single hedonic model over a large geographic area may not be appropriate when preferences for environmental amenities vary across markets, and the scaled-up models results in our study were driven by one subregion. We recommend using regional scale models consistent with traditional hedonic literature to support decision making and using these models to scale up to national welfare estimates rather than using a single scaled-up national model.

2 Modelling Water Quality Preferences

An overview of prior lake water quality research reveals the prevalence of regional and state-level variation in implicit prices. [2] Studies across Maine repeatedly find differences in water quality preferences between different regions ranging from $2,000 to as much as $13,000 for a 1 m increase in water clarity (Boyle et al., 1999; Boyle & Taylor, 2001; Michael et al., 2000; Poor et al., 2001). Research in Florida reveals even larger implicit prices, with Walsh et al. (2011) reporting $5,595 for only a 1 ft increase clarity. Given these results, one would expect a national scale study to find heterogenous water quality preferences. However, Moore et al. (2020) report consistent marginal effects for water quality, even after intentionally testing for state level variation. We must therefore consider how their modelling approach differs from previous studies and how these differences affect implicit price estimates.

[2] Nicholls & Crompton (2018) provide an exhaustive literature review of hedonic studies prior to 2018 for different water quality measures and aquatic features.

To analyze these differences we compare Moore et al.'s (2020) work to similar hedonic studies for lake water quality in Table 1. The key modification lies in the geographic scale of the model. Prior studies identify property sales nearby lakes concentrated in a single region, whereas Moore et al. (2020) use a broad sample of lakes identified in the 2007 and 2012 National Lakes Assessment (NLA).[3] This scaled-up modelling approach assumes consistent preferences across all lake housing markets in the study. While NLA lakes are representative of all U.S. lakes, one sample was collected for each lake during the summer of the assessment year (U.S. Environmental Protection Agency, 2016). Moore et al. (2020) then combine the individual measurement for each lake with sales from 2010 – 2013. This method differs from prior studies which select a water quality measurement relative to the year the property was sold, commonly the year prior to sale. Moore et al. (2020) address both spatial and temporal concerns surrounding the NLA data. They include state fixed effects and interaction terms between Secchi depth and the states with the largest number of transactions to control for state level heterogeneity, and they address potential measurement error using an instrumental variables approach.

Other modifications to the prior modelling structure by Moore et al. (2020) relate to the functional form. Moore et al. (2020) use a log-linear model that allows for implicit price variation relative to sales price but implies preferences for water quality have no relation to existing water quality on the lake. This specification is not only a departure from the established literature but does not account for a homeowner's inability to recognize changes in Secchi depth in lakes with high levels of water clarity (Smeltzer & Heiskary, 1990). Linear-log and double-log specifications are more common as they allow for diminishing implicit prices at high levels of water clarity. Prior studies also included interactions between water quality and lake area to allow for another dimension of implicit price variation representing the tradeoffs that homeowners make when selecting a lake, like choosing a small lake with less boat traffic

[3] More information on the National Lakes Assessment data can be found at https://www.epa.gov/national-aquatic-resource-surveys/nla

over a large pristine lake (Boyle et al., 1999). Significant lake area and water quality interaction terms (Gibbs et al., 2002; Walsh et al., 2011) indicate implicit prices are not consistent for different types of lakes. Intuitively, this implicit price variation relates to different lake uses. Clear lakes may be good for swimming, but do not provide the same fishing opportunities as lakes with more algae content. As types of lakes vary regionally within states and across states, it is important to account for heterogenous water quality preferences in hedonic models across large geographic areas.

Table 1 Hedonic Models for Secchi Depth[4] and Selected Implicit Prices

Reference	Study Area	Housing Data	Secchi Depth Measurements	Model Specification	Implicit Prices (elasticities)
(Michael et al., 1996)	Maine 4 Markets (4 lakes, 6 lakes, 7 lakes, 5 lakes)	Lakefront Single Family Homes & Unimproved Land	Minimum Sample Year Prior to Sale Trend (Sample – 10 Year Average) Additional lake area interactions (different for each market)	Semi-log in WQ N=90, N=84, N=14, N=155	\$11 - \$200 per foot frontage for 1 m increase in Secchi depth (evaluated at mean for each lake)
(Boyle et al., 1999)	Maine 4 Markets (25 total lakes)	Lakefront Houses	Secchi Depth * Lake Area Minimum Sample Year of Sale	Semi-log in WQ N=48, N=112, N=68, N=21	\$2,337 - \$12,938 for 1 m increase in Secchi depth (evaluated at mean for each market)
(Michael et al., 2000)	Maine 3 Markets (4 lakes, 13 Lakes, 5 lakes)	Lakefront Single Family Homes & Unimproved Land	Minimum Sample Year Varies (Current, Historical, Interaction, Trends, Seasonal Changes)	Semi-log in WQ N=89, N=295, N=147	\$2,040 - \$8,084 for 1 m increase in Secchi depth (year prior to sale sample evaluated at mean for each market)
(Boyle & Taylor, 2001)	Maine 4 Markets (34 total lakes)	Lakefront Properties	Secchi Depth * Lake Area Minimum Sample Year Prior to Sale	Semi-log in WQ N=55, N=158, N=74, N=31	\$2,000 - \$8,000 for 1 m increase in Secchi depth (evaluated at mean for each market)
(Poor et al., 2001)	Maine 4 Markets (4 – 13 lakes per market)	Lakefront Properties	Secchi Depth * Lake Area Minimum Sample Year of Sale	Semi-log in WQ N=56, N=174, N=52, N=66	\$2,600 - \$6,279 for 1 m increase in Secchi depth (evaluated at mean for each market)
(Gibbs et al., 2002)	New Hampshire 4 Markets (69 total lakes)	Lakefront Properties	Secchi Depth * Lake Area Minimum Sample Year of Sale	Semi-log in WQ N=115, N=178, N=80, N=74	\$1,135 - \$9,756 for 1 m increase in Secchi depth (evaluated at mean for each market)
(Walsh et al., 2011)	Orange County, Florida 1 Market (146 lakes)	Single Family Homes within 1 km of lake	Mean Sample Year of Sale Additional interactions with lakefront, distance to lake, and lake area	Double-log in WQ, Distance, & Area N=54,712	\$18,356 for 1 m increase in Secchi depth (on lakefront edge effect model 3 evaluated at mean) 0.133 elasticity (on lakefront edge effect model 1)
(Weng et al., 2020)	Lake Mendota, Wisconsin 1 Lake	Single Family Homes within 5 km of lake	Secchi Depth * Lakefront Indicator Mean Sample Summer Prior to Sale	Double-log in WQ N=13,169	0.353 elasticity
(Moore et al., 2020)	Nationwide (32 states) 1 Market (113 lakes)	Houses within .1 mile of lake Sales from 2010-2013	Sample Year 2007, 2012 One Sample per Lake Instrument using nitrogen, phosphorous and lake temperature	Log-linear N=1,462, N=600	\$39,711 for 1 m change in Secchi depth 0.20 elasticity (authors' preferred model evaluated at mean sales price)

[4] Secchi depth represents water clarity and is the most common indicator for water quality in the hedonic literature. It measures the depth at which a Secchi disk remains visible beneath the surface of the water.

3 Study Area and Data

3.1 Study Area

Our study area consists of six states throughout the northeast and upper Midwest: Maine, Michigan, Minnesota, New York, Vermont, and Wisconsin. Our sample consists of lakefront properties on lakes where we could obtain reliable water quality and property sales data. For our regional level analysis, we identified four distinct regions in Minnesota and New York that had enough lakefront transactions to support hedonic modelling. The two Minnesota regions include Otter Tail County and Twin Cities metro area (excluding Hennepin County). In New York we define the Adirondacks region using six counties south of the Adirondacks mountains. The Finger Lakes region consists of seven counties in the central New York. Selected regions and study lakes are shown in figure 1.

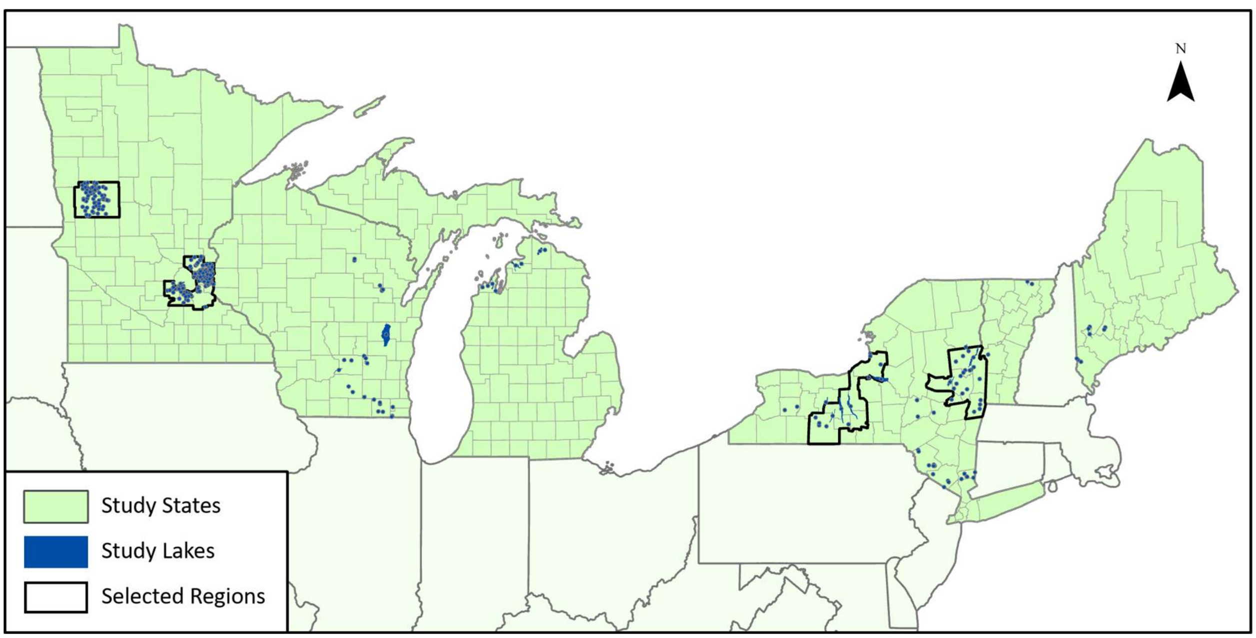

Figure 1 Map of Study Region with Selected Regions and Lakes

3.2 Lake and Water Quality Data

Lake and water quality data come from the Lake Multi-Scaled Geospatial and Temporal (LAGOS) Database, which combines detailed lake characteristics of more than 51,000 lakes and reservoirs with a broad range of water quality measurements and geospatial data (Soranno et al., 2015, 2017). We use the LAGOSNE R package developed by Stachelek et al. (2019) to access lake characteristics and other lake identifying information from the LOCUS module (Patricia A Soranno & Cheruvelil, 2017a) and water quality data from the LIMNO module (Patricia A Soranno et al., 2019). We downloaded geospatial data from the GIS module (Patricia A Soranno & Cheruvelil, 2017b).

Our water quality variables are Secchi depth and chlorophyll-a (chl-a) concentration as they are most common in LAGOS and can easily be observed prior property sales. While water clarity, measured by Secchi depth, is the primary way homeowners observe lake water quality, they can also observe excessive algae, which relates to chl-a concentration in the water. The impacts of algae are becoming more prominent in the hedonic water quality literature with studies reporting adverse effects of chl-a concentration (Liu et al., 2017; Walsh & Milon, 2016; Weng et al., 2020) and harmful algal blooms (Wolf & Klaiber, 2017) on local housing markets.

We identified lakes in our study area larger than 4 ha with Secchi or chl-a samples taken between June and September, when water quality is poorest, from 2000 to 2013. Some lakes had multiple water quality measurements in a given summer, so we used the mean summer sample for those lakes.[5] Homeowners observe water quality in the lake prior to purchase, but there are no conclusive results as to which year of sample or summary measure is best (Michael et al., 2000). We decided to use the measurement in the year prior to the sale similar to (Boyle & Taylor, 2001; Weng et al., 2020).

[5] We also ran each model using minimum and maximum summaries of these measures with similar results.

3.3 Housing Data

Sales data and tax parcel data were collected from a wide range of sources, most of which are publicly available and published online by state, county, and local governments. We accessed the data through government websites, parcel database downloads, and computer assisted mass appraisal systems (CAMA) managed by private companies. Parcel data was used to identify properties with lake frontage for our study lakes. We then searched for available sales data and tax records by parcel number. In the Twin Cities there was historical parcel data available that contained the most recent tax records and sales for each year. In Michigan parcel data was less available, so we obtained lakefront property sales and tax records directly from local assessors.[6,7] A full list of housing data sources is in Appendix A.

Data availability varied widely across and within the different states. Often state entities provide parcel data throughout the state, and local municipalities have discretion about providing public access to tax records and sales data. As a result, the municipalities we include in this study are constrained by those which provide public access to tax and sales records or in the case of Michigan had tax assessors willing to assist us. Minnesota is the richest both in terms of lakes, over 200 in our sample, as well as sales data, accounting for more than half of all transactions, over 2,500. Sample sizes for the remaining states range from as low 30 sales in Maine and Vermont to over 1,000 in New York. The richness of the Minnesota data allows us to test whether a single subregion can drive model results. Another benefit is all transactions lie within Otter Tail or the Twin Cities regions, providing insight into the behavior of state level hedonic models in the presence of regional heterogeneity.

[6] We'd like to thank Robert Englebrecht, Andrew Giguere, and Clayton McGovern for their assistance with this project.

[7] Data for Evangeline Michigan included all sales records, so we geocoded the addresses to identify lakefront properties.

3.4 Other Data Sources

Neighborhood attributes including census tract, block group, and median income for block group from the U.S. 2010 census were obtained from TIGER/Line shapefiles. We find the distance in miles from each property address to the nearest Walmart using Google Maps. Weather data is from the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010).[8] This dataset includes daily temperature summaries from 2001-2014 weather stations throughout our study area, which we averaged over the summer months in a given sale year for each weather station.

We report the descriptive statistics for all states combined in Table 2, and descriptive statistics for each state and selected region in Appendix B. All sales prices are adjusted to 2018 $ using regional yearly consumer price index data from FRED (https://fred.stlouisfed.org/).

Table 2 Descriptive Statistics by Water Quality Measure

	Secchi (N = 5,546)				Chl-a (N = 4,667)			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Sales Price (2018 $)[9]	367,966	278,081	12,780	1,802,000	372,363	285,543	12,780	1,745,250
Secchi Depth (m)	2.80	1.86	0.13	10.00				
Chl-a Concentration (ug/L)					20.02	29.24	0.20	438.00
Lake Area (ha)	3,145	7,411	4	53,417	3,546	7,993	4	53,417
Temperature (C)	20.80	1.61	14.45	23.98	20.81	1.59	14.71	23.98
Lot Size (acres)	0.69	0.94	0.05	8.32	0.67	0.91	0.05	8.32
Living Area (sqft)	1,690	890	272	12,348	1,695	894	280	12,348
House Age (years)	42.39	26.42	1	214	42.88	26.45	1	214
Walmart Distance (miles)	11.09	7.57	0.50	36.80	11.17	7.75	0.60	36.80
Median Income (2010 $)	71,245	23,881	27,045	195,313	70,532	24,044	27,045	195,313

[8] NCEP data obtained from https://globalweather.tamu.edu/.
[9] 2018 was the most recent year when compiled the data.

4 Methods and Model Specification

In this section we will first introduce the two hedonic price model equations that we estimate throughout our analysis. Then we discuss the different types of models and methods that we use to examine implicit price variation. Last, we present the specific hypotheses that we test for each model.

4.1 Model Specification

While there is no preferred functional form to estimate the hedonic price function for environmental amenities (Kuminoff et al., 2010), Secchi depth and chl-a concentration are commonly expressed with natural logs (Gibbs et al., 2002; Walsh & Milon, 2016; Weng et al., 2020), and recent studies frequently include the natural log for sales price (Bin et al., 2017; Liu et al., 2017; Moore et al., 2020; Wolf & Klaiber, 2017). Double-log functional forms are useful as they allow implicit price variation in relation to water quality and local housing prices, and the marginal effects can also be interpreted as elasticities. We estimate separate hedonic price models for Secchi depth and chl-a concentration with two different equations using the double-log functional form for all non-binary variables: [10,11]

$$\ln(Price) = \alpha + \beta_1 \ln(WQ) + \beta_2 \ln(Area) + \beta_3 \ln(Temp) + \beta_{\boldsymbol{S_0}} ln(WQ) \times \boldsymbol{S} \\ + \beta_{\boldsymbol{S_1}} \boldsymbol{S} + \beta_{\boldsymbol{P}} \boldsymbol{P} + \beta_{\boldsymbol{Y}} \boldsymbol{Y} + \beta_{\boldsymbol{Q}} \boldsymbol{Q} + \mu \tag{1}$$

$$\ln(Price) = \alpha + \beta_1 \ln(WQ) + \beta_2 \ln(Area) + \beta_3 \ln(Temp) + \beta_{\boldsymbol{P}} \boldsymbol{P} + \beta_{\boldsymbol{Y}} \boldsymbol{Y} + \beta_{\boldsymbol{Q}} \boldsymbol{Q} + \mu \tag{2}$$

where *Price* is the sales price, *WQ* is our water quality measure of interest, *Area* is the area of the lake, *Temp* is the mean summer ambient air temperature, ***S*** is vector of binary state variables, ***P*** is a vector of property attributes, ***Y*** indicates sales year by census tract fixed effects, and ***Q*** indicates sales quarter fixed effects.

[10] We removed all observations with 0 Secchi depth or chl-a concentration

[11] We also ran our models with a linear-log functional form and found similar results.

We introduce lake area and temperature into the model specification as different characteristics that homeowners may consider when purchasing a lake house. Although lake area is frequently interacted with water quality, we estimate the two variables separately as in (Moore et al., 2020).[12] We include average daily temperature, as it relates to consumer preferences with homeowners in the U.S. paying more to avoid high heat and cold (Albouy et al., 2016). Since the focus of our study is lakefront homes, we specify average daily temperature in the summer months only.

The remaining variables are common amongst all hedonic price models to control for physical attributes of the property, its location, and sale date. Physical attributes include lot size, square footage of living area, age of home, and median income of the block group. We include the distance from the nearest Walmart as these stores are often located nearby popular shopping areas. Other studies consider a range of other neighborhood characteristics (Moore et al., 2020) or spatial autocorrelation (Walsh et al., 2011) to control for unknown location specific differences. However, Kuminoff et al. (2010) show spatial fixed effects minimize bias from omitted neighborhood variables when compared to spatial autocorrelation methods.[13] We therefore use sale year by census tract fixed effects to control for both unknown neighborhood variables as well differences in sales prices across time and sales quarter fixed effects to control for seasonal variation in sales prices.

The difference between the equations is that equation (1) includes additional water quality and state interactions as well as state fixed effects, both using Minnesota for the base state. The interaction terms allow the marginal effect of water quality to vary across each state relative to the base state. Similar interaction terms are used by (Moore et al., 2020) to detect regional variation in water quality preferences. The state fixed effects pick up mean heterogeneity in sales prices between states.

[12] We initially ran models with a full set of interactions terms for lake area and temperature and state fixed effects but later chose to exclude them to focus the analysis on spatial definition of hedonic models.
[13] However spatial fixed effect cannot control for omitted variables within the reference area.

4.2 Model Types

The three model types we use in our analysis are: [14]

- Multi-State Models
- State Level Models
- Regional Models

The multi-state models combine lakefront property sales from lakes across multiple states in a single scaled-up hedonic model. We estimate two sets of multi-state models. The first set includes two models with observations from all six states in our study region, which we specify using both equations (1) and (2). The first specification assumes that all preferences are consistent across states except for water quality, allowing us to examine potential implicit price variation across states using a single scaled-up model. Significant interaction terms, state binary variables multiplied by water quality, are indicative of heterogeneity in implicit prices at the state level. Equation (2) contains the implicit assumption that preferences are consistent across states and allows us to test the restriction the coefficients on the state and water quality interaction variables are all insignificant. Insignificance suggest that there is not state-level heterogeneity in preferences.

The second set of multi-state models includes six models with observations from five states. In each model observations from a single state are dropped, and we estimate the effects of water clarity for the five remaining states using equation (2). These models allow us to investigate how sensitive the scaled-up hedonic model is to observations from a single subregion, in this case an individual state. This leave-one-out method is similar to an approach used by Boyle et al. (2013) in assessing the robustness of a meta-analysis by dropping study observations from the meta-equation. If the marginal effects of water quality

[14] We also modelled individual lakes in the sample with more than 100 observations but did not find many with significant effects possibly due to little variation in Secchi depth and chl-a.

are constant across models, then the assumption that water quality preferences are consistent across states holds. Alternatively, if the elasticity estimate changes upon dropping a state, that state could be driving the model results when all six states are included.

The state level models are comprised of lakefront property transactions aggregated throughout a state in a single scaled-up hedonic model, whereas the regional models are made up of sales from a well-defined region within an individual state. We specify six state level models and four regional models with equation (2). Estimating these models separately allows for preferences to vary across the different states and regions, meaning the elasticity estimates are not influenced by observations from other states or regions. We compare the results at the state level to those in the multi-state model to determine whether the multi-state model accurately reflects water quality preferences for each state's sample. If preferences differ across states, we can assess how this heterogeneity affects the elasticity estimates in the multi-state model. Similarly, we compare the regional models to the state level models to determine whether water quality preferences vary within individual states.

4.3 Hypothesis Testing

In the multi-state model, we test several hypotheses. For equation (1), we test whether all the state and water quality interactions terms are jointly significant using an F test.

$$H_0: \beta_{ME_0} = \beta_{MI_0} = \beta_{NY_0} = \beta_{VT_0} = \beta_{WI_0} = 0$$

Next, we test whether all the state and water quality interactions terms and state fixed effects are jointly significant using an F test.

$$H_0: \beta_{ME_0} = \beta_{MI_0} = \beta_{NY_0} = \beta_{VT_0} = \beta_{WI_0} = \beta_{ME_1} = \beta_{MI_1} = \beta_{NY_1} = \beta_{VT_1} = \beta_{WI_1} = 0$$

We then consider the multi-state model specified by equation (2). We test whether our water quality term has a significant effect on sales prices with a t test using the null hypothesis H_0: $\beta_1 = 0$. We perform the same hypothesis test each time we drop observations for each state from the model.

For each state level and regional model, we perform a single test. We specify each model using equation (2) and the relevant observations within each geographic area. We test the effect of our water quality variable with a t test using the null hypothesis H_0: $\beta_1 = 0$. For each model type, we expect a positive coefficient for Secchi depth and a negative coefficient for chl-a.

5 Results and Discussion

5.1 Multi-State Models

We estimate equations (1) and (2) for the first set of multi-state models using ordinary-least-squares (OLS) regression. Coefficient estimates for each water quality measure and state interactions are reported in Table 3, and we report results for the remaining continuous variables in Appendix C for all three model types.[15] While the interaction between Secchi and Maine is significant, we fail to reject the null hypothesis in joint significance test for all state and water quality interaction terms at a 5% significance level ($F = 1.94$, $F_{CRIT} = 2.21$). Similarly, the interaction between chl-a and Vermont is significant, but jointly the interaction terms for all state are not significant at the 5% level ($F = 0.72$, $F_{CRIT} = 2.21$). Following Moore et al. (2020), we could conclude from these results that water quality preferences do not vary across states even though the observations for our base case Minnesota account for more than half the sample size.

Next, we perform joint hypotheses tests involving state interactions and state fixed effects. We fail to reject the null hypotheses at a 5% significance level ($F_{CRIT} = 1.83$) for both Secchi ($F = 1.47$) and

[15] Coefficients for fixed effects are available from the authors by request.

chl-a (F = 0.86). These results indicate we can use equation (2) instead of (1) to estimate water quality elasticities. We find the coefficients for Secchi and chl-a are both highly significant with the expected signs, positive for Secchi and negative for chl-a. Even though Wisconsin is included in our study area, our Secchi coefficient 0.17 much smaller than the 0.35 coefficient reported by Weng et al. (2020) for an individual Wisconsin lake. However, we find similar effects for chl-a, -0.10 compared to -0.085 in the same study.

Table 3 Effects of Water Quality Measures and State Interactions on ln(Price) in Multi-State Model for All States

	Secchi			Chl-a	
	(1)	(2)		(1)	(2)
ln(Secchi)	0.17***	0.17***	ln(Chl-a)	-0.10***	-0.10***
	(0.03)	(0.03)		(0.03)	(0.03)
ln(Secchi) * ME	6.80**		ln(Chl-a) * ME	0.56	
	(2.4)			(1.26)	
ln(Secchi) * MI	-1.04		ln(Chl-a) * MI	0.21	
	(2.1)			(0.13)	
ln(Secchi) * NY	0.08		ln(Chl-a) * NY	-0.11	
	(0.18)			(0.1)	
ln(Secchi) * VT	0.14		ln(Chl-a) * VT	-0.30*	
	(0.08)			(0.12)	
ln(Secchi) * WI	-0.11		ln(Chl-a) * WI	0.27	
	(0.21)			(0.82)	
R^2	0.79	0.79	R^2	0.80	0.80
N	5,546		N	4,667	

Standard errors in parentheses. *$p < .05$, **$p < .01$, ***$p < .001$

In table 4 we report the water quality coefficients after dropping observations by state. For both water quality measures we find the results are highly significant and consistent with the original model except when dropping Minnesota where neither water quality coefficient is significant. Since the results in the multi-state model for all six states depend on the inclusion of Minnesota observations, it appears a subregion can drive the behavior of a single scaled-up hedonic model. Moreover, the elasticity estimates in the remaining multi-state models may not accurately represent the preferences for each state contradicting the results of the joint significance tests for the state interaction terms. This leads us to conclude that interaction terms cannot effectively control for heterogenous preferences in a scaled-up hedonic model.

Table 4 Effects of Water Quality Measures on ln(Price) after Dropping Observations by State

	All States	Drop ME	Drop MI	Drop MN	Drop NY	Drop VT	Drop WI
Secchi							
ln(Secchi)	0.17***	0.17***	0.17***	0.13	0.16***	0.17***	0.17***
	(0.03)	(0.03)	(0.03)	(0.12)	(0.03)	(0.03)	(0.03)
R^2	0.79	0.79	0.79	0.82	0.74	0.79	0.80
N	5,546	5,488	5,387	2,388	4,392	5,433	4,642
Chl-a							
ln(Chl-a)	-0.10***	-0.10***	-0.11***	-0.17	-0.09***	-0.10***	-0.10***
	(0.03)	(0.03)	(0.03)	(0.08)	(0.03)	(0.03)	(0.03)
R^2	0.80	0.80	0.80	0.84	0.75	0.80	0.81
N	4,667	4,635	4,561	2,086	3,553	4,629	3,871

Using equation (2). Standard errors in parentheses. *p < .05, **p < .01, ***p < .001

5.2 State Level Models

The state level model results in Table 5 indicate the presence of heterogenous water quality preferences. The coefficients vary across each state, and significant results are only found in Michigan, Minnesota, and New York. For Minnesota the results are identical to those in the multi-state model providing further evidence that the Minnesota observations are driving the multi-state model results. Accepting the results from the first two multi-state models without additional testing would lead us to erroneously conclude the estimated elasticities for Minnesota were consistent with all states in our sample. Since the multi-state model cannot effectively control for heterogenous preferences even with the inclusion of interaction terms, concerns surrounding the validity of scaled-up model estimates is warranted.

Table 5 Effects of Water Quality Measures on ln(Price) for Individual States

	All States	ME[16]	MI[17]	MN	NY	VT[16]	WI
Secchi							
ln(Secchi)	0.17***	-2.71	9.20*	0.17***	0.13	-0.32	0.10
	(0.03)	(2.72)	(4.4)	(0.03)	(0.18)	(0.56)	(0.18)
R^2	0.79	0.96	0.79	0.75	0.90	0.55	0.64
N	5,546	58	159	3,158	1,154	113	904
Chl-a							
ln(Chl-a)	-0.10***	1.95	-1.53*	-0.10**	-0.19*	0.20	0.38
	(0.03)	(1.06)	(0.59)	(0.03)	(0.08)	(0.26)	(0.74)
R^2	0.80	0.99	0.81	0.76	0.90	0.78	0.63
N	4,667	32	106	2,581	1,114	38	796

Using equation (2). Standard errors in parentheses. *p < .05, **p < .01, ***p < .001

5.3 Regional Models

Regional models also reflect heterogenous water quality preferences consistent with prior literature. In Table 6 we compare the marginal effects for the selected regions with the state level results for Minnesota and New York. The effects of Secchi are highly significant in Otter Tail and the Twin cities, but the coefficient is five times larger in Otter Tail, 0.50 compared to 0.10. The elasticities also differ from 0.353 for Lake Mendota, Wisconsin (Weng et al., 2020) and 0.133 in Orange County, Florida (Walsh et al., 2011). For chl-a we find a highly significant negative effect -0.24 in Otter Tail, while we find no effect in the Twin Cities. This elasticity estimate for Otter Tail is much larger in magnitude than those reported within in Florida or Wisconsin, approximately -0.06 (Walsh & Milon, 2016; Weng et al., 2020).

Within the selected New York regions, we find no effect for either water quality measure in the Adirondacks and Finger Lakes regions despite finding a significant effect for chl-a across the entire state, calling into question the state level results. The stark differences between state and regional coefficients in Minnesota further demonstrate how incorporating regions with distinct preferences in a scaled-up

[16] Given the sample size and census tract by sales year fixed effects, the results of these models are inconclusive.
[17] It is important to consider while both Michigan results are statistically significant at .05, we consider the large effect sizes indicative of potential omitted variable bias within the census tracts or other misspecification. When we remove census tract fixed effects, while keeping the sale year fixed effects, the results are no longer significant.

model impacts the marginal effects. Both regional Secchi coefficients differ from 0.17 reported for Minnesota. The state level coefficient closely resembles the simple weighted average between the two regions 0.18. Similarly, for chl-the state level result -0.10 is slightly larger than the weighted average of both coefficients -0.08, revealing how scaled-up hedonic models can obscure important differences in price effects for environmental amenities.

Looking at implicit prices tells another story. For a 1 m increase in Secchi depth calculated at the mean sales price and Secchi depth, we find similar results across Minnesota, $26,136 for Otter Tail, $24,818 for the Twin Cities, and $31,646 at the state level. However, if we use the only state level results to compute the implicit price for Secchi depth in Otter Tail, the estimate is approximately a third of the county level effect, $8,886 compared to $26,136. A difference this large could potentially change the outcome of a cost benefit analysis for water quality. Despite a functional form allowing for multiple levels of implicit price variation, the scaled-up model cannot adequately represent heterogenous preferences between two distinct markets.

Table 6 Effects of Water Quality Measures on ln(Price) for Selected Regions within MN and NY

	MN	Otter Tail	Twin Cities	NY	Adirondacks	Finger Lakes
Secchi						
ln(Secchi)	0.17***	0.50***	0.10**	0.13	-0.21	0.41
	(0.03)	(0.12)	(0.03)	(0.18)	(0.27)	(0.24)
R^2	0.75	0.60	0.68	0.90	0.92	0.84
N	3,158	629	2,529	1,154	424	572
Chl-a						
ln(Chl-a)	-0.10**	-0.24**	-0.04	-0.19*	-0.04	-0.24
	(0.03)	(0.08)	(0.03)	(0.08)	(0.11)	(0.12)
R^2	0.76	0.60	0.68	0.90	0.92	0.84
N	2,581	576	2,005	1,114	425	536

Using equation (2). Standard errors in parentheses. *p < .05, **p < .01, ***p < .001

6 Conclusions

From our results, we conclude preferences for water quality are not consistent across a large geographic region and reject the scaled-up model. Not only do we report different estimates for Secchi depth and chl-a concentration at both state and regional levels, but also find that scaled-up model estimates are driven by subregions. The scaled-up model results for all six states are driven by data from Minnesota, and the scaled-up model results for Minnesota results differ from the results for both subregions. Thus, future studies must consider the potential for regional heterogeneity when estimating benefits for water quality (or other environmental variables) over large geographic areas to provide credible implicit price estimates. Using regional hedonic models for specific housing markets within a state can capture this heterogeneity as longs as there is sufficient, systematic variation in water quality across water bodies in the region.

While it is tempting to incorporate different markets into a scaled-up hedonic model, our results indicate that researchers should proceed with caution when using this approach. In the presence of heterogenous preferences, scaled-up model results can be sensitive to observations from a single subregion and adding interaction terms between the variable of interest and each subregion may not fully control for regional heterogeneity. In our estimation the effect of Minnesota was not observed directly in the scaled-up model estimation results until we dropped the relevant observations from the model. Researchers using the scaled-up approach may consider similar sensitivity analyses to investigate the robustness of model results.

Most importantly we shade concern that a single scaled-up hedonic model cannot support regulatory decision making on large geographic scales let alone a national scale. Our results also demonstrate that even at the state level, heterogenous preferences cannot be represented by a single scaled-up hedonic model. Regional heterogeneity is not a water-quality specific phenomenon, and the results from this study are relevant to a broad range of environmental amenities. To support policy analyses over a large geographic area, we recommend using multiple regional models to capture heterogenous preferences with appropriate robustness checks.

Works Cited

Albouy, D., Graf, W., Kellogg, R., & Wolff, H. (2016). Climate Amenities, Climate Change, and American Quality of Life. *Journal of the Association of Environmental and Resource Economists*, *3*(1), 205–246. https://doi.org/10.1086/684573

Bayer, P., Keohane, N., & Timmins, C. (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, *58*(1), 1–14. https://doi.org/10.1016/j.jeem.2008.08.004

Bin, O., Czajkowski, J., Li, J., & Villarini, G. (2017). Housing Market Fluctuations and the Implicit Price of Water Quality: Empirical Evidence from a South Florida Housing Market. *Environmental and Resource Economics*, *68*(2), 319–341. https://doi.org/10.1007/s10640-016-0020-8

Boyle, K. J., Parmeter, C. F., Boehlert, B. B., & Paterson, R. W. (2013). Due Diligence in Meta-analyses to Support Benefit Transfers. *Environmental and Resource Economics*, *55*(3), 357–386. https://doi.org/10.1007/s10640-012-9630-y

Boyle, K. J., Poor, P. J., & Taylor, L. O. (1999). Estimating the Demand for Protecting Freshwater Lakes from Eutrophication. *American Journal of Agricultural Economics*, *81*(5), 1118–1122. JSTOR. https://doi.org/10.2307/1244094

Boyle, K. J., & Taylor, L. O. (2001). Does the Measurement of Property and Structural Characteristics Affect Estimated Implicit Prices for Environmental Amenities in a Hedonic Model? *The Journal of Real Estate Finance and Economics*, *22*(2), 303–318. https://doi.org/10.1023/A:1007855901029

Carey, R. T., & Leftwich, R. (2007). *Water Quality and Housing Value on Lake Greenwood: A Hedonic Study on Chlorophyll-a Levels and the 1999 Algal Bloom*. Strom Thurmond Institute, Clemson University.

Chay, K. Y., & Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy*, *113*(2), 376–424. JSTOR. https://doi.org/10.1086/427462

Gibbs, J. P., Halstead, J. M., Boyle, K. J., & Huang, J.-C. (2002). An Hedonic Analysis of the Effects of Lake Water Clarity on New Hampshire Lakefront Properties. *Agricultural and Resource Economics Review*, *31*(1), 39–46. https://doi.org/10.1017/S1068280500003464

Griffiths, C., Klemick, H., Massey, M., Moore, C., Newbold, S., Simpson, D., Walsh, P., & Wheeler, W. (2012). U.S. Environmental Protection Agency Valuation of Surface Water Quality Improvements. *Review of Environmental Economics and Policy*, *6*(1), 130–146. https://doi.org/10.1093/reep/rer025

Kashian, R., Eiswerth, M. E., & Skidmore, M. (2006). Lake Rehabilitation and the Value of Shoreline Real Estate: Evidence from Delavan, Wisconsin. *Review of Regional Studies*, *36*(2), 221–238.

Keiser, D. A. (2019). The Missing Benefits of Clean Water and the Role of Mismeasured Pollution. *Journal of the Association of Environmental and Resource Economists*, *6*(4), 669–707. https://doi.org/10.1086/703254

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, *60*(3), 145–160. https://doi.org/10.1016/j.jeem.2010.06.001

Liu, T., Opaluch, J. J., & Uchida, E. (2017). The impact of water quality in Narragansett Bay on housing prices. *Water Resources Research*, *53*(8), 6454–6471. https://doi.org/10.1002/2016WR019606

Michael, H. J., Boyle, K. J., & Bouchard, R. (1996). *Water Quality Affects Property Prices: A Case Study of Selected Maine Lakes* (No. 398). Maine Agricultural and Forest Experiment Station Miscellaneous Report 398. https://digitalcommons.library.umaine.edu/aes_miscreports/3/

Michael, H. J., Boyle, K. J., & Bouchard, R. (2000). Does the Measurement of Environmental Quality Affect Implicit Prices Estimated from Hedonic Models? *Land Economics*, *76*(2), 283–298. JSTOR. https://doi.org/10.2307/3147229

Moore, M. R., Doubek, J. P., Xu, H., & Cardinale, B. J. (2020). Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits. *Ecological Economics*, *176*, 106692. https://doi.org/10.1016/j.ecolecon.2020.106692

Nicholls, S., & Crompton, J. (2018). A Comprehensive Review of the Evidence of the Impact of Surface Water Quality on Property Values. *Sustainability*, *10*, 500.

Patricia A Soranno, & Cheruvelil, K. S. (2017a). LAGOS-NE-LOCUS v1.01: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013 ver 4. *Environmental Data Initiative.* https://doi.org/10.6073/pasta/0c23a789232ab4f92107e26f70a7d8ef

Patricia A Soranno, & Cheruvelil, K. S. (2017b). LAGOS-NE-GIS v1.0: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 2013-1925 ver 4. *Environmental Data Initiative.* https://doi.org/10.6073/pasta/4b04db93c53be4b65937d0fedd0fcf4b

Patricia A Soranno, Lottig, N. R., Delany, A. D., & Cheruvelil, K. S. (2019). LAGOS-NE-LIMNO v1.087.3: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013 ver 3. *Environmental Data Initiative.* https://doi.org/10.6073/pasta/08c6f9311929f4874b01bcc64eb3b2d7

Poor, P. J., Boyle, K. J., Taylor, L. O., & Bouchard, R. (2001). Objective versus Subjective Measures of Water Clarity in Hedonic Property Value Models. *Land Economics*, *77*(4), 482–493. JSTOR. https://doi.org/10.2307/3146935

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H., Juang, H.-M. H., Sela, J., … Goldberg, M. (2010). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Smeltzer, E., & Heiskary, S. A. (1990). Analysis and Applications of Lake User Survey Data. *Lake and Reservoir Management*, *6*, 109–118.

Soranno, P. A., Bacon, L. C., Beauchene, M., Bednar, K. E., Bissell, E. G., Boudreau, C. K., Boyer, M. G., Bremigan, M. T., Carpenter, S. R., Carr, J. W., Cheruvelil, K. S., Christel, S. T., Claucherty, M., Collins, S. M., Conroy, J. D., Downing, J. A., Dukett, J., Fergus, C. E., Filstrup, C. T., … Yuan, S. (2017). LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. *GigaScience*, *6*(gix101). https://doi.org/10.1093/gigascience/gix101

Soranno, P. A., Bissell, E. G., Cheruvelil, K. S., Christel, S. T., Collins, S. M., Fergus, C. E., Filstrup, C. T., Lapierre, J.-F., Lottig, N. R., Oliver, S. K., Scott, C. E., Smith, N. J., Stopyak, S., Yuan, S., Bremigan, M. T., Downing, J. A., Gries, C., Henry, E. N., Skaff, N. K., … Webster, K. E. (2015). Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science and data reuse. *GigaScience*, *4*(s13742-015-0067–4). https://doi.org/10.1186/s13742-015-0067-4

Stachelek, J., Oliver, S. K., & Farzan Masrour. (2019). *LAGOSNE: Interface to the Lake Multi-scaled Geospatial and Temporal Database. R package version 2.0.2.* https://CRAN.R-project.org/package=LAGOSNE

Taylor, L. O. (2017). Hedonics. In P. A. Champ, K. J. Boyle, & T. C. Brown (Eds.), *A Primer on Nonmarket Valuation* (pp. 235–292). Springer Netherlands. https://doi.org/10.1007/978-94-007-7104-8_7

U.S. Environmental Protection Agency. (2016). *National Lakes Assessment 2012: A Collaborative Survey of Lakes in the United States*. EPA 841-R-16-113, Washington, DC.

Walsh, P. J., & Milon, J. W. (2016). Nutrient Standards, Water Quality Indicators, and Economic Benefits from Water Quality Regulations. *Environmental and Resource Economics*, *64*(4), 643–661. https://doi.org/10.1007/s10640-015-9892-2

Walsh, P. J., Milon, J. W., & Scrogin, D. O. (2011). The Spatial Extent of Water Quality Benefits in Urban Housing Markets. *Land Economics*, *87*(4), 628–644. JSTOR. https://doi.org/10.3368/le.87.4.628

Weng, W., Boyle, K. J., Farrell, K. J., Carey, C. C., Cobourn, K. M., Dugan, H. A., Hanson, P. C., Ward, N. K., & Weathers, K. C. (2020). Coupling Natural and Human Models in the Context of a Lake Ecosystem: Lake Mendota, Wisconsin, USA. *Ecological Economics*, *169*, 106556. https://doi.org/10.1016/j.ecolecon.2019.106556

Wolf, D., & Klaiber, H. A. (2017). Bloom and bust: Toxic algae's impact on nearby property values. *Ecological Economics*, *135*, 209–221. https://doi.org/10.1016/j.ecolecon.2016.12.007

Appendix A Housing Data Sources by Municipality

State	Municipalities	Tax & Sales	Parcels
ME	Acton, Shapleigh	John E. O'Donnell & Associates https://jeodonnell.com/	Maine Geolibrary https://www.maine.gov/geolib/catalog.html
	Harrison	Harrison Property Cards https://www.harrisonmaine.org/assessor	
	Auburn	Patriot Properties https://www.patriotproperties.com/	
MI	Leelanau County	Leelanau Tax Assessor	-
	Evangeline	Evangeline Tax Assessor	
	Aloha, Benton, Inverness, Mullet	Cheboygan Tax Assessor	
MN	Otter Tail County	Otter Tail Property Search http://www.ottertailcounty.us/ez/publicsearch.php	Otter Tail GIS Web App https://ottertailcountymn.us/content-page/gis-maps/
	Anoka County, Carver County, Dakota County, Ramsey County, Scott County, Washington County	Minnesota Geospatial Commons https://gisdata.mn.gov/	
NY	Orange County, Oswego County, Otsego County, Patterson, Putnam County, Rensselaer County, Saratoga County, Schenectady County, Schoharie County, Schuyler County, Seneca County, Seneca Falls, Southeast, Steuben County, Sullivan County, Tompkins County, Warren County, Washington County, Wyoming County, Yates County	Systems Development Group: Image Mate Online https://sdgnys.com/#ImoShortcut	New York State GIS Services http://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1300
VT	Derby, Castleton	Patriot Properties https://www.patriotproperties.com/	Vermont Open Geodata Portal https://geodata.vermont.gov/pages/parcels
WI	Albion, Beaver Dam City, Burlington City, Caledonia, Dekorra, Delavan, Fox Lake, Geneva, Harrison, Lodi, Menominee, Monona City, Pardeeville, Pleasant Springs, Portage City, Sugar Creek, Twin Lakes, Upham, Wescott, Whitewater City	Accurate Assessor http://accurateassessor.com/municipalities/	Wisconsin Statewide Parcel Map Initiative https://www.sco.wisc.edu/parcels/data/

Appendix B Descriptive Statistics for States and Selected Regions

Descriptive Statistics for ME, MI, VT, and WI by Water Quality Measure

	Secchi				Chl-a			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
ME								
Sales Price (2018 $)	282,747	186,477	55,080	805,800	299,119	230,226	69,177	805,800
Secchi Depth (m)	5.56	1.06	4.25	9.01				
Chl-a Concentration (ug/L)					3.54	0.67	2.27	4.95
Lake Area (ha)	572	727	47	2,443	724	927	47	2,443
Temperature (C)	18.50	1.06	16.26	19.80	18.61	1.23	16.26	19.80
Lot Size (acres)	1.21	1.85	0.12	7.38	0.88	1.28	0.12	7.20
Living Area (sqft)	1,589	633	680	3,432	1,466	640	680	3,432
House Age (years)	52	31	2	99	43	28	2	87
Walmart Distance (miles)	7.96	6.23	3.00	23.20	9.56	6.50	3.30	22.00
Median Income (2010 $)	61,296	10,696	39,457	70,104	57,287	12,424	39,457	67,292
N		58				32		
MI								
Sales Price (2018 $)	593,165	374,483	135,783	1,802,000	582,325	380,426	135,783	1,664,174
Secchi Depth (m)	4.48	0.91	2.55	6.87				
Chl-a Concentration (ug/L)					1.48	0.48	0.55	2.62
Lake Area (ha)	3,709	1,696	100	7,016	3,201	794	153	7,016
Temperature (C)	17.94	1.69	14.45	20.98	18.19	1.68	14.71	20.98
Lot Size (acres)	0.97	1.08	0.12	8.32	0.99	1.19	0.12	8.32
Living Area (sqft)	1,839	904	480	6,106	1,782	758	480	4,414
House Age (years)	29	22	1	106	26	22	1	106
Walmart Distance (miles)	19.15	7.10	4.30	30.50	21.19	5.83	7.80	30.50
Median Income (2010 $)	54,481	8,465	37,621	104,906	55,020	6,797	37,621	64,643
N		159				106		
VT								
Sales Price (2018 $)	241,016	122,110	32,235	647,274	180,896	94,079	32,235	494,910
Secchi Depth (m)	6.76	1.76	2.24	8.78				
Chl-a Concentration (ug/L)					2.79	1.31	1.20	5.66
Lake Area (ha)	952	554	314	2,691	919	958	314	2,691
Temperature (C)	18.68	1.17	15.98	20.73	17.57	0.95	15.98	19.29
Lot Size (acres)	0.52	0.58	0.06	3.81	0.49	0.51	0.09	2.48
Living Area (sqft)	1,330	693	440	4,274	1,280	787	553	4,274
House Age (years)	49	26	1	111	40	23	1	74
Walmart Distance (miles)	13.32	5.92	2.70	21.30	6.31	4.53	2.70	18.40
Median Income (2010 $)	49,158	9,845	32,353	58,482	39,815	10,014	32,353	55,417
N		113				38		
WI								
Sales Price (2018 $)	278,502	156,645	18,380	998,675	277,999	159,499	18,380	998,675
Secchi Depth (m)	1.91	1.11	0.19	4.35				
Chl-a Concentration (ug/L)					35.75	41.28	0.44	307.50
Lake Area (ha)	5,032	13,285	9	53,417	5,666	14,280	9	53,417
Temperature (C)	20.60	1.62	16.34	23.98	20.63	1.60	16.34	23.98
Lot Size (acres)	0.47	0.59	0.06	7.63	0.48	0.62	0.06	7.63
Living Area (sqft)	1,783	864	272	5,409	1,790	889	384	6,197
House Age (years)	47	29	1	167	48	29	1	167
Walmart Distance (miles)	9.18	6.95	1.20	36.80	9.45	7.21	1.20	36.80
Median Income (2010 $)	56,549	16,821	27,045	94,375	56,343	16,899	27,045	94,375
N		904				796		

Descriptive Statistics for MN, Otter Tail, and Twin Cities by Water Quality Measure

	Secchi				Chl-a			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
MN								
Sales Price (2018 $)	407,677	276,214	12,780	1,728,060	421,816	284,071	12,780	1,728,060
Secchi Depth (m)	2.19	1.21	0.13	9.47				
Chl-a Concentration (ug/L)					22.34	28.22	1.23	438.00
Lake Area (ha)	541	861	4	5,676	611	944	4	5,676
Temperature (C)	21.46	1.24	17.98	23.49	21.46	1.26	17.98	23.49
Lot Size (acres)	0.83	1.09	0.05	7.99	0.82	1.07	0.05	7.99
Living Area (sqft)	1,758	934	320	12,348	1,778	939	325	12,348
House Age (years)	38	24	1	146	38	24	1	146
Walmart Distance (miles)	10.50	7.84	0.50	35.50	10.65	8.21	0.60	35.50
Median Income (2010 $)	80,794	24,256	28,538	195,313	80,086	24,835	28,538	195,313
N		3,158				2,581		
Otter Tail								
Sales Price (2018 $)	180,861	148,147	12,780	870,300	184,092	150,682	12,780	870,300
Secchi Depth (m)	3.46	1.18	0.19	8.77				
Chl-a Concentration (ug/L)					6.91	4.60	1.50	43.33
Lake Area (ha)	1,252	1,323	15	5,676	1,335	1,349	56	5,676
Temperature (C)	21.02	1.36	17.98	23.23	21.02	1.36	17.98	23.23
Lot Size (acres)	0.92	1.30	0.05	7.81	0.90	1.25	0.05	7.81
Living Area (sqft)	1,239	622	325	6,356	1,250	638	325	6,356
House Age (years)	41	23	1	135	40	23	1	135
Walmart Distance (miles)	22.83	5.96	3.00	35.50	23.26	5.64	8.90	35.50
Median Income (2010 $)	49,059	6,954	39,643	65,714	48,982	6,950	40,298	65,714
N		629				576		
Twin Cities								
Sales Price (2018 $)	464,090	271,737	15,615	1,728,060	490,110	276,526	20,140	1,728,060
Secchi Depth (m)	1.87	0.99	0.13	9.47				
Chl-a Concentration (ug/L)					26.77	30.52	1.23	438.00
Lake Area (ha)	364	577	4	5,573	403	656	4	5,573
Temperature (C)	21.56	1.18	18.13	23.49	21.59	1.21	18.13	23.49
Lot Size (acres)	0.81	1.03	0.05	7.99	0.79	1.01	0.05	7.99
Living Area (sqft)	1,887	954	320	12,348	1,929	957	352	12,348
House Age (years)	37	24	1	146	38	24	1	146
Walmart Distance (miles)	7.43	4.54	0.50	28.50	7.03	4.36	0.60	28.50
Median Income (2010 $)	88,687	20,244	28,538	195,313	89,022	20,548	28,538	195,313
N		2,529				2,005		

Descriptive Statistics for NY, Adirondacks, and Finger Lakes by Water Quality Measure

Variable	Secchi Mean	Secchi SD	Secchi Min	Secchi Max	Chl-a Mean	Chl-a SD	Chl-a Min	Chl-a Max
NY								
Sales Price (2018 $)	315,063	314,729	21,925	1,745,250	313,869	316,374	24,118	1,745,250
Secchi Depth (m)	4.39	2.16	0.79	10.00				
Chl-a Concentration (ug/L)					6.23	9.52	0.20	63.40
Lake Area (ha)	9,062	8,135	10	20,770	9,033	8,135	10	20,770
Temperature (C)	19.88	1.29	16.66	23.56	19.86	1.30	16.66	23.56
Lot Size (acres)	0.45	0.45	0.05	4.26	0.45	0.45	0.05	4.26
Living Area (sqft)	1,451	751	280	5,452	1,446	753	280	5,452
House Age (years)	52	28	1	214	52	27	1	214
Walmart Distance (miles)	13.04	6.33	1.70	36.60	12.86	6.08	1.70	36.00
Median Income (2010 $)	61,597	16,776	35,500	124,107	61,440	16,714	35,500	124,107
N		1,154				1,114		
Adirondacks								
Sales Price (2018 $)	501,107	420,571	37,128	1,745,250	500,234	420,462	37,128	1,745,250
Secchi Depth (m)	5.40	2.67	0.83	10.00				
Chl-a Concentration (ug/L)					4.45	6.58	0.20	43.51
Lake Area (ha)	5,097	5,442	10	11,559	5,086	5,441	10	11,559
Temperature (C)	19.46	1.25	16.66	22.15	19.46	1.25	16.66	22.15
Lot Size (acres)	0.49	0.47	0.05	3.46	0.49	0.48	0.05	3.46
Living Area (sqft)	1,621	844	280	5,000	1,620	846	280	5,000
House Age (years)	52	29	1	151	52	29	1	151
Walmart Distance (miles)	12.83	6.30	2.10	36.00	12.83	6.29	2.10	36.00
Median Income (2010 $)	69,921	19,754	35,500	117,196	69,869	19,750	35,500	117,196
N		424				425		
Finger Lakes								
Sales Price (2018 $)	197,671	144,950	21,925	1,482,130	189,727	135,913	24,881	1,482,130
Secchi Depth (m)	4.21	1.36	1.03	7.41				
Chl-a Concentration (ug/L)					4.58	4.92	0.34	41.41
Lake Area (ha)	14,450	7,013	13	20,770	14,686	6,889	13	20,770
Temperature (C)	20.03	1.09	17.80	22.81	20.01	1.12	17.80	22.81
Lot Size (acres)	0.45	0.47	0.05	4.26	0.45	0.47	0.05	4.26
Living Area (sqft)	1,350	682	300	5,452	1,340	683	300	5,452
House Age (years)	51	28	1	214	50	26	1	214
Walmart Distance (miles)	13.34	6.40	1.70	36.60	12.96	5.90	1.70	26.10
Median Income (2010 $)	54,597	9,202	36,578	99,167	54,080	8,530	36,578	77,273
N		572				536		

Appendix C Estimates for Continuous Variables on ln(Price) in Multi-State, Individual State, and Selected Region Models

	All States (1)	All States (2)	ME	MI	MN	NY	VT	WI	Otter Tail	Twin Cities	Adirondacks	Finger Lakes
Secchi												
ln(Area)	0.10***	0.11***	0.30**	0.26	0.10***	0.16**	0.04	0.06	0.06	0.11***	0.27***	-0.31*
	(0.01)	(0.01)	(0.10)	(0.17)	(0.01)	(0.05)	(0.30)	(0.03)	(0.03)	(0.01)	(0.07)	(0.12)
ln(Temp)	-2.04*	-1.86*	2.70	0.55	-1.53	0.21	-1.62	-4.40	-4.30	1.87	2.46	-0.45
	(0.90)	(0.90)	(2.51)	(1.41)	(1.91)	(1.97)	(1.47)	(5.69)	(4.38)	(2.13)	(3.28)	(2.49)
ln(Lot)	0.11***	0.11***	-0.04	0.02	0.10***	0.14***	0.00	0.09***	0.07	0.11***	0.14***	0.12***
	(0.01)	(0.01)	(0.08)	(0.05)	(0.01)	(0.02)	(0.07)	(0.03)	(0.04)	(0.01)	(0.03)	(0.02)
ln(Sqft)	0.56***	0.57***	0.68***	0.13	0.59***	0.56***	0.43***	0.56***	0.69***	0.54***	0.63***	0.51***
	(0.02)	(0.02)	(0.18)	(0.10)	(0.03)	(0.03)	(0.12)	(0.03)	(0.07)	(0.03)	(0.06)	(0.04)
ln(Age)	-0.10***	-0.10***	-0.10	0.06	-0.12***	-0.08***	-0.13	-0.08***	-0.17***	-0.11***	-0.07**	-0.10***
	(0.01)	(0.01)	(0.08)	(0.05)	(0.01)	(0.02)	(0.07)	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)
ln(Walmart)	-0.04	-0.03	0.74**	0.27	-0.11*	0.02	0.30	-0.02	0.34	-0.20***	0.11	-0.03
	(0.03)	(0.03)	(0.25)	(0.23)	(0.05)	(0.05)	(0.42)	(0.05)	(0.19)	(0.05)	(0.10)	(0.07)
ln(Income)	0.10	0.11	1.88	3.87***	0.07	-0.10	1.35*	0.05	-0.85	0.20*	-0.30	-0.33
	(0.07)	(0.07)	(5.15)	(0.61)	(0.10)	(0.12)	(0.58)	(0.12)	(0.57)	(0.09)	(0.22)	(0.22)
Chl-a												
ln(Area)	0.12***	0.12***	0.19**	-2.72***	0.11***	0.15***	0.22	0.01	0.08*	0.12***	0.23***	-0.21
	(0.01)	(0.01)	(0.06)	(0.76)	(0.01)	(0.04)	(0.14)	(0.08)	(0.04)	(0.02)	(0.05)	(0.12)
ln(Temp)	-1.79*	-1.77	-0.37	0.56	-1.90	0.19	0.88***	-1.07	-8.08	2.43	2.07	-0.76
	(0.91)	(0.91)	(2.03)	(1.48)	(2.01)	(1.93)	(0.18)	(6.22)	(4.71)	(2.16)	(3.21)	(2.46)
ln(Lot)	0.11***	0.11***	0.29	0.02	0.10***	0.14***	0.10	0.11***	0.04	0.12***	0.14***	0.13***
	(0.01)	(0.01)	(0.13)	(0.06)	(0.02)	(0.02)	(0.13)	(0.03)	(0.04)	(0.02)	(0.03)	(0.02)
ln(Sqft)	0.56***	0.56	-0.40	0.03	0.60***	0.54***	-0.07	0.56***	0.73***	0.53***	0.62***	0.49***
	(0.02)	(0.02)	(0.32)	(0.13)	(0.03)	(0.03)	(0.23)	(0.04)	(0.08)	(0.03)	(0.06)	(0.04)
ln(Age)	-0.10***	-0.10***	-0.06	0.06	-0.12***	-0.08***	-0.26*	-0.08***	-0.19***	-0.11***	-0.07*	-0.09***
	(0.01)	(0.01)	(0.08)	(0.05)	(0.01)	(0.02)	(0.10)	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)
ln(Walmart)	-0.06	-0.06	0.76**	0.35	-0.12*	-0.01	-0.78	-0.04	0.24	-0.22***	0.12	-0.08
	(0.03)	(0.03)	(0.21)	(0.27)	(0.06)	(0.06)	(0.47)	(0.05)	(0.21)	(0.06)	(0.10)	(0.07)
ln(Income)	0.09	0.09	9.62*	4.35***	0.07	-0.04	0.93***	0.04	-1.04	0.22*	-0.31	-0.17
	(0.07)	(0.07)	(3.92)	(0.70)	(0.11)	(0.12)	(0.23)	(0.14)	(0.66)	(0.10)	(0.22)	(0.24)

Standard errors in parentheses. *p < .05, **p < .01, ***p < .001