Optimal weight settings in locally weighted regression: A guidance through cross-validation approach

Roshan Puri

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 Arts
in
Data Analysis and Applied Statistics

Jennifer Van Mullekom, Chair
Klaus Moeltner
Anne Driscoll

September 5, 2023
Blacksburg, Virginia

Keywords: locally weighted regression, weight variables, cross-validation.

Optimal weight settings in locally weighted regression: A guidance through cross-validation approach

Roshan Puri

ABSTRACT

Locally weighted regression is a powerful tool that allows the estimation of different sets of coefficients for each location in the underlying data, challenging the assumption of stationary regression coefficients across a study region. The accuracy of LWR largely depends on how a researcher establishes the relationship across locations, which is often constructed using a weight matrix or function. This paper explores the different kernel functions used to assign weights to observations, including Gaussian, bi-square, and tri-cubic, and how the choice of weight variables and window size affects the accuracy of the estimates. We guide this choice through the cross-validation approach and show that the bi-square function outperforms the choice of other kernel functions. Our findings demonstrate that an optimal window size for LWR models depends on the cross-validation (CV) approach employed. In our empirical application, the full-sample CV guides the choice of a higher window-size case, and CV *by proxy* guides the choice of a lower window size. Since the CV *by Proxy* approach focuses on the predictive ability of the model in the vicinity of one specific point (usually a policy point/site), we note that guiding a model choice through this approach makes more intuitive sense when the aim of the researcher is to predict the outcome in one specific site (policy or target point). To identify the optimal weight variables, while we suggest exploring various combinations of weight variables, we argue that an efficient alternative is to merge all continuous variables in the dataset into a single weight variable.

Optimal weight settings in locally weighted regression: A guidance through cross-validation approach

Roshan Puri

GENERAL AUDIENCE ABSTRACT

Locally weighted regression (LWR) is a statistical technique that establishes a relationship between dependent and explanatory variables, focusing primarily on data points in proximity to a specific point of interest/target point. This technique assigns varying degrees of importance to the observations that are in proximity to the target point, thereby allowing for the modeling of relationships that may exhibit spatial variability within the dataset.

The accuracy of LWR largely depends on how researchers define relationships across different locations/studies, which is often done using a "weight setting". We define weight setting as a combination of weight functions (determines how the observations around a point of interest are weighted before they enter the model), weight variables (determines proximity between the point of interest and all other observations), and window sizes (determines the number of observations that can be allowed in the local regression). To find which weight setting is an optimal one or which combination of weight functions, weight variables, and window sizes generates the lowest predictive error, researchers often employ a cross-validation (CV) approach. Cross-validation is a statistical method used to assess and validate the performance of a predictive model. It entails removing a host observation (a point of interest), predicting that point, and evaluating the accuracy of such predicted point by comparing it with its actual value.

In our study, we employ two CV approaches. The first one is a full-sample CV approach, where we remove a host observation, and predict it using the full set of observations used in the given local regression. The second one is the CV *by proxy* approach, which uses a similar mechanism as full-sample CV to check the accuracy of the prediction, however, by focusing only on the vicinity points that shares similar characteristics as a target point.

We find that the bi-square function consistently outperforms the choice of Gaussian and tri-cubic weight functions, regardless of the CV approaches. However, the choice of an optimal window size in LWR models depends on the CV approach that we employ. While the full-sample CV method guides us toward the

selection of a larger window size, the CV *by proxy* directs us toward a smaller window size. In the context of identifying the optimal weight variables, we recommend exploring various combinations of weight variables. However, we also propose an efficient alternative, which involves using all continuous variables within the dataset into a single-weight variable instead of striving to identify the best of thousands of different weight variable settings.

Acknowledgments

I would like to express my heartfelt gratitude to the members of my thesis committee, without whom this research would not have been possible.

First and foremost, I extend my sincere appreciation to Dr. Jennifer Van Mullekom, who served as the chair of my committee. Her guidance and unwavering support throughout my entire master's program, as well as for this research, were invaluable.

I owe special thanks to Dr. Klaus Moeltner for his pivotal role as a committee member during my master's program and also as the chair of my PhD program. His continued support and guidance throughout the research development have been instrumental. This research is an integral part of my dissertation, and I am genuinely grateful for his mentorship and expertise.

I also wish to extend my gratitude to Dr. Anne Driscoll for her valuable feedback and her investment in my academic journey.

Moreover, I want to acknowledge the unwavering support of my family, friends, and colleagues who have been a constant source of encouragement throughout my academic pursuit.

Contents

List of Tables **vii**

1 Introduction 1

2 Empirical Framework 4

3 Data 8

4 Results and Discussion 10

4.1 Full-sample CV results 11

4.2 CV *by proxy* 13

5 Conclusion 14

Bibliography 15

List of Tables

1 Source Studies and Number of Observations . . . 22

2 Descriptive statistics of variables used in our analysis . . . 23

3 External variables that serves in different weight variables combination . . . 24

4 Weight variables combination . . . 25

5 Full-sample CV results using `wvar 1` as weight . . . 26

6 Full-sample CV results using `wvar 2` as weight . . . 27

7 Full-sample CV results using `wvar 3` as weight . . . 28

8 Full-sample CV results using `wvar 4` as weight . . . 29

9 CV *by proxy* results using `wvar 1` as weight . . . 30

10 CV *by proxy* results using `wvar 2` as weight . . . 31

11 CV *by proxy* results using `wvar 3` as weight . . . 32

12 CV *by proxy* results using `wvar 4` as weight . . . 33

1 Introduction

Locally weighted regression (LWR), also known as geographically weighted regression, is a widespread tool in a variety of disciplines such as environmental science and health (Gilbert and Chakraborty, 2011; Harris et al., 2010), and housing markets (Brunsdon et al., 1998; Cao et al., 2019; Manganelli et al., 2014), geology (Atkinson et al., 2003; Pasculli et al., 2014), landscape ecology (Gao and Li, 2011; Zhang et al., 2019), remote sensing (Kamarianakis et al., 2008; Propastin, 2012), geography and urban planning studies (Cleveland, 1979; Cleveland and Devlin, 1988; Li et al., 2019; Loader, 1999; McMillen and Redfearn, 2010; Wang et al., 2020). This estimation approach challenges the assumption made by ordinary least squares that the regression coefficients remain stationary across the study region. When location—often in a geospatial space which could usually be a specific study site or a set of study sites with shared characteristics—characterizes a data-generating process, then one set of coefficients does not reflect the true relationship between the response and the predictor variables for the entire data, making an estimation from ordinary least squares approach questionable (Fotheringham and Brunsdon, 1999). In such a case, it is reasonable to consider that there could be inherent disparities in relationships across different locations or that there could be some issue with the model specification used to measure the relationships, resulting in varying parameter estimates across space (Brunsdon et al., 1998).

LWR allows us to examine such relationships across locations, by estimating a different set of coefficients for each location in the underlying data. However, the accuracy of the predicted outcome variable in this estimation framework largely depends on how a researcher establishes a relationship across locations. A local relationship is often constructed using the weight matrix or function, which is then used to assign weights to each observation in the dataset based on its distance (geographic or Euclidean distance, for instance) from the location being predicted (Fotheringham et al., 2015). The weights are higher for observations that are closer or more similar to the location being predicted and lower for observations that are farther away.[1]

Weights are typically defined using a kernel function such as Gaussian, bi-square, or tri-cubic, which assigns a weight to each observation based on its distance from the point of interest (Fotheringham et al., 2002, 2009). Gaussian kernel assigns a higher weight to observations that are closer to the point of interest and gradually reduces the weight as the distance increases. This makes it a suitable choice for models that require smoothness and continuity. The bi-square kernel assigns a higher weight to observations that are

[1]The closeness or the proximity can be determined just by the geographic distance, or the distance among the few or all explanatory variables (see section 2 for more details).

closer to the point of interest but reduces the weight abruptly as the distance increases beyond a certain threshold. This makes it a good choice for models that need to be robust to outliers and have a sharper cut-off. Similarly, the tri-cubic kernel assigns an even higher weight to observations that are closer to the point of interest and reduces the weight more slowly as the distance increases. This makes it a good choice for models that require a higher degree of smoothing than the Gaussian kernel.[2]

Each of the kernel functions mentioned above itself is the function of weight variables and window size. McMillen and Redfearn (2010) discuss that the weight-determining variables could be all of the predictor variables, a subset of those (e.g. only continuous variables), a simple geographic distance measure, or variables that are external to the set of predictor variables. Including too many variables into the weight function may degrade the precision of estimates, especially when the model (baseline meta-regression model (MRM) in our case) consists of many explanatory variables (McMillen and Redfearn, 2010). Including only one variable as weight into the weight function, however, may not sufficiently capture the features of a source study's underlying population or spatial characteristics. Hence, it is necessary to find an optimal combination of weight variables that increases the precision of estimates while adequately capturing the source study's features.

The choice of window size in the above-mentioned kernel function is as important as the choice of weight variables since it determines the number of observations that can be allowed in the local regression (McMillen and Redfearn, 2010). This window size approach sets the same specific value for all locations. This allows the number of observations receiving some positive weights to be fixed for each target, producing the same number of observations for each local regression.[3] A smaller value of window size should be chosen if the goal is to predict the dependent variable (Pagan and Ullah, 1999) as it usually reduces predictive bias. However, a very small value of window size relative to the number of explanatory variables may produce highly variable results (Pagan and Ullah, 1999). Choosing a very large value of window size, on the other hand, can yield locally biased estimates as observations far away from the target point are used in regression (Fotheringham et al., 2009). Therefore, it is important to work with an optimal value of window size.

In summary, the choice of kernel function and its components (weight variables and window size) is an important consideration when defining weights in the statistical modeling of LWR. Understanding the properties

[2] More details on the choice of weight function is provided in section 2.

[3] Alternatively, a fixed bandwidth that allows the number of observations receiving some weights to vary on the availability of the observations around the target point is possible. However, such fixed bandwidth will either lead to excessive smoothing (case, when there are too many observations near the target point) or to highly variable results (case, where data/observations are sparsely populated around target point) (Fotheringham et al., 2009; McMillen and Redfearn, 2010).

of different kernels can help researchers select the appropriate weighting scheme for their specific needs and ensure that their models accurately reflect the underlying data. However, since a researcher often doesn't know the true data-generating process, assuming a weight setting (a combination of the weight function, weight variables, and window size) to fit the local regression could deteriorate the predictive accuracy of the model. To circumvent this problem, we use an experimental set-up that guides toward the optimal weight settings through a cross-validation (CV) approach.

A CV approach allows us to evaluate the within-predictive accuracy of our different model choices, given by different weight settings. In this application, we use two different CV approaches. The first one is a full-sample CV, an exercise similar to the standard leave-one-out exercise, where a home observation is first removed and then predicted using output from local regression as discussed in Brunsdon et al. (1996), Farber and Páez (2007), and Fotheringham et al. (2009). Intuitively, we can interpret this as an out-of-sample prediction, but within the *sample-at-large*. We can then use the output obtained from this CV exercise to calculate summary statistics for assessing the predictive fitness of the model.

The above procedure allows us to determine the ability of a particular weight setting that generate accurate within-sample predictions for the *sample-at-large*. However, when deciding on the optimal weight settings if the researcher believes that the target point (where an outcome variable needs to be predicted) shares similar underlying characteristics with only the sub-sample of the data, then it may be more relevant to focus on the predictive ability of the model in the vicinity of such point rather than the full sample. Hence, we also consider an alternative CV approach which we labeled as CV *by proxy* (see next section for details).

In this paper, we provide guidance on the choice of optimal weight settings in LWR by systematically varying the specification of the weight functions, weight variables, and window sizes across a suite of models. We use two CV approaches, i.e., full-sample CV and CV *by proxy* to compute the predictive fitness of our different model choices. We then compare the predictive performance of those model choices with that of the generic regression model (referred to as the baseline model). We consider the weight setting that produces the lowest predictive error as the optimal one. We carry out the entire analysis in the Bayesian estimation framework as it provides the advantage of non-reliance on asymptomatic theory, given the small sample size of the data as in our case, and allows us to obtain a full posterior distribution of estimates, which can be conveniently used for generating predictive distributions for secondary outputs if the researcher aims to do so (Moeltner et al., 2019). We find that LWR has the potential to improve the predictive accuracy of the model as this consistently generates lower prediction errors than the baseline model. We then show that such gain in accuracy is largely driven by the choice of window size and weight functions.

The remainder of this paper is comprised of the following sections: Section 2 presents the empirical framework required for LWR and presents its experimental setup in a Bayesian framework; Section 3 presents the data used for the meta-regression purpose along with the spatially explicit variables that serve as our external weight variables; Section 4 summarizes and discusses the results; and Section 5 concludes.

2 Empirical Framework

We frame our empirical setup to fit the meta-data on willingness-to-pay (WTP) for water quality improvement.[4] In an environmental valuation context, such meta-data is used to predict the WTP in policy sites (typically not investigated) through the transfer of estimates, functions, or models. We start by adopting a "best-practiced" MRM, referred as MRM2 on Moeltner (2019) and Moeltner et al. (2019), and consider it as our baseline model. For any observation i in location/study t, MRM2 can be written as:

$$log\left(\frac{y_{it}}{q_{1,it} - q_{0,it}}\right) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{m}'_{it}\boldsymbol{\gamma} + \delta\left(\frac{q_{1,it} + q_{0,it}}{2}\right) + \epsilon_{it}, \tag{1}$$

$$\epsilon_{it} \sim n(0, \sigma^2)$$

where y_{it} is the WTP for a water quality change from $q_{0,it}$ to $q_{1,it}$, $\mathbf{x}_{it}$ is a vector of context-specific regressors, i.e., variables that are usually available for the policy site, and $\mathbf{m}_{it}$ is a vector of methodological indicators which needs to be controlled to avoid any potential omitted variables problem, ϵ_{it} is the idiosyncratic error term distributed normally with mean 0 and variance σ^2.

Equation (1) produces a single set of estimates of parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and δ across all locations t. However, as discussed earlier, these estimates of parameters or coefficients may vary across locations, requiring us to fit the regression locally to produce local estimates at each target point.[5] Location t in our meta-analysis is a set of study sites that share common characteristics; however, this can also be a single study site. This non-parametric approach of estimating the local estimates is known as the LWR method. Simply put, LWR provides estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and δ at the target location t, i.e., $\hat{\boldsymbol{\beta}}_t$, $\hat{\boldsymbol{\gamma}}_t$, and $\hat{\delta}_t$ for a given value of $\mathbf{x}$, $\mathbf{m}$ and q-average, i.e., $\frac{q_1+q_0}{2}$. This necessitates an assignment of the weight to each observation in equation (1). We

[4] See next section for details on meta-data.
[5] We use "target point" to specify the location where parameters are being estimated.

use the weight function ψ_t and estimate the following equation:

$$\sqrt{\psi_t}\Bigg(log\Bigg(\frac{y_{it}}{q_{1,it}-q_{0,it}}\Bigg)\Bigg) = \sqrt{\psi_t}\Bigg(\mathbf{x'}_{it}\boldsymbol{\beta}_t + \mathbf{m'}_{it}\boldsymbol{\gamma}_t + \delta_t\Bigg(\frac{q_{1,it}+q_{0,it}}{2}\Bigg) + \epsilon_{it}\Bigg), \tag{2}$$

$$\psi_t = \mathcal{K}(\mathbf{z}_{it}, \mathbf{z}_t, w)$$

$$\epsilon_{it} \sim n(0, \sigma_t^2)$$

where ψ_t is a non-negative valued weights determined by a kernel function $\mathcal{K}(.)$. $\mathcal{K}(.)$ itself is characterized by $\mathbf{z}_{it}$—a vector of variables that determine observation i's proximity or similarity with the target point, $\mathbf{z_t}$ and a window size w. An error term is still normally distributed, however, with location-specific variance σ_t^2.

Weight function ψ_t determines how the observations around target point t are weighted before they enter the model. We pick the three most commonly used weight functions. i.e., Gaussian, bi-square, and tri-cubed from geography literature (Farber and Páez, 2007; Fotheringham et al., 2002; Redfearn, 2009). All of these functions produce weights that monotonically decrease with the distance[6], i.e., observations closer to the home observation/s receive higher weights than others.

The first, the Gaussian function, is given by:

$$\psi_t = exp\Bigg(-\frac{1}{2}\Bigg(\frac{d_{it}}{d^*_{t,w}}\Bigg)^2\Bigg) \tag{3}$$

where $d_{it} = \sqrt{\mathbf{z_t'z_i}}$ is a Euclidean distance between observation(s) at location t and any other observation i, $d^*_{t,w}$ is the maximum distance in the set of observations at location t, as (implicitly) imposed by window size w. Since weight must decay as the distance between observation i and location t increases, it is necessary to sort the entire sample in increasing order of distance. This produces distance d_{it} of 0 for all the observations associated with home location t, thereby providing unit weight to all the observations in the same location. The weight will decrease in conformity with the Gaussian curve as we move away from the observations at location t and eventually reach close to 0.6065 $\Big(= \exp(-\frac{1}{2})\Big)$ once we are at the furthest point (or cut-off point as determined by window size w) from observation i.

The second and third weight functions, i.e., bi-square and tri-cubed kernel functions differ from the Gaussian

[6]By saying "distance", we refer to the Euclidean distance between two locations based on the different choice of weight variables. Hence, it should not be confused with the geospatial distance between two locations.

weight function in weight received by the observation/s at the cut-off point. In these cases, weights decrease more rapidly as the distance between home observation i and the observations at location t decreases, approaching zero at the cut-off point.

Equations (4) and (5) provide the functional form of the bi-square and tri-cubed weight functions:

$$\psi_t = \left(1 - \left(\frac{d_{it}}{d^*_{t,w}}\right)^2\right)^2 \tag{4}$$

$$\psi_t = \left(1 - \left(\frac{d_{it}}{d^*_{t,w}}\right)^3\right)^3 \tag{5}$$

where ψ_t in equations (4) and (5) takes a value of 0 if $d_{it} < d^*_{t,w}$.

As evidenced in equation (2) and discussed in section 1, the variables comprised in $\mathbf{z}$, and window size w ultimately affect the estimates of all parameters. We consider a different combination of weight variables and window sizes (discussed in sections 3 and 4 respectively). To find an optimal combination of weight functions, weight variables, and window size, we use the CV method. We use two CV approaches, i.e., full-sample CV and CV *by proxy*. In the full-sample CV approach, we first consider observation j at location t as our target point and predict the WTP of water quality change y_{jt} by removing the target point—a leave-one-out exercise. In cases where a location t has multiple observations, we repeat this procedure locally for each observation within t. Then, we average the predicted y_{js} from this CV method across locations to calculate the Mean Absolute Percentage Error (MAPE), a measure of prediction accuracy.

CV *by proxy* builds on similar concepts. However, there is a subtle difference. Unlike a full-sample CV, this approach allows us to find an optimal weight setting that generates more accurate predictions for locations in the vicinity of the target or a policy point. When the aim of the researcher is to predict the outcome in one specific site, which would often be the case in an environmental valuation context such as predicting WTP in a policy site, then it makes an intuitive sense to make the model choice based on the sub-sample of the data or nearby location as the policy point might share a same data-generating process as its vicinity points. To account for this possibility, we first define a hypothetical policy point, and select the 10 closest sample locations based on the Euclidean distance (using weight variables) between the hypothetical policy point and location in the meta-data. We then predict those 10 sample locations based on the number of observations determined by the window size and average them to generate the MAPE.

Mathematically, the MAPE associated with a given weight setting can be written as:

$$MAPE = \frac{100}{n}\sum_{t=1}^{T}\sum_{j=1}^{n\in t}\left|\frac{y_{it}-\bar{\hat{y}}_{jt,j\neq i}(w)}{y_{it}}\right|, \tag{6}$$

$$\bar{\hat{y}}_{jt} = exp\Bigg(\mathbf{x'}_{jt}\boldsymbol{\beta}_t + \mathbf{m'}_{jt}\boldsymbol{\gamma}_t + \delta_t\Big(\frac{q_{1,jt}+q_{0,jt}}{2}\Big) + log(q_{1,jt}-q_{0,jt}) + \epsilon_t\Bigg)$$

where n is the total number of predicted $\hat{y}_{js}$ from CV approach across all locations, and $\hat{y}_{jt}$ is the predicted value of y_{jt} after removing target point from the location t, $\bar{\hat{y}}_{jt}$ is the mean of predicted values of y_{jt}.

We use a Bayesian approach to carry out an entire estimation. For this, we assume the terms in equation (2) as $\sqrt{\psi_t}log\Big(\frac{y_{it}}{q_{1,it}-q_{0,it}}\Big)= y^*_{it}$, $\sqrt{\psi_t}\mathbf{x}_{it}{=}\mathbf{x}^*_{it}$, $\sqrt{\psi_t}\mathbf{m}_{it}{=}\mathbf{m}^*_{it}$, and $\sqrt{\psi_t}\Big(\frac{q_{1,it}+q_{0,it}}{2}\Big){=}q^*_{it}$ and estimate the following likelihood function for location t:

$$p(\mathbf{y}^*|\theta_t,\sigma^2_t,\tilde{\mathbf{X}}^*) = (2\pi)^{-\frac{n_w}{2}}(\sigma^2_t)^{-\frac{n_w}{2}}exp\Bigg(-\frac{1}{2\sigma^2_t}(\mathbf{y}^*-\tilde{\mathbf{X}}^*\theta_t)'(\mathbf{y}^*-\tilde{\mathbf{X}}^*\theta_t)\Bigg) \tag{7}$$

where n_w is the number of observations that receive weight greater than zero [7], $\mathbf{y}^*$ is the vector of y^*_i within the location t, $\tilde{\mathbf{X}}^*$ is the weighted matrix of our explanatory data set that collects $\mathbf{X}^*$, $\mathbf{M}^*$, and $\mathbf{q}^*$ which themselves are the vectors of $\mathbf{x}^*_i$, $\mathbf{m}^*_i$ and q^*_i at location t, respectively. Coefficients $\boldsymbol{\beta_t}$, $\boldsymbol{\gamma_t}$, and δ_t, as provided in equation (2), are collected in $\boldsymbol{\theta}_t$, i.e., $\boldsymbol{\theta}_t{=}\begin{bmatrix}\boldsymbol{\beta}'_t & \boldsymbol{\gamma}'_t & \delta_t\end{bmatrix}'$.

Equation (7) produces a posterior distribution of $\boldsymbol{\theta}_t$ and σ^2_t, which we obtain by employing a Gibbs Sampling approach. We assign the same prior distribution as in Moeltner et al. (2019) and Moeltner (2019), i.e, a standard multivariate normal prior for $\boldsymbol{\theta}_t$ and inverse-gamma priors for σ^2_t. We repeat the analysis for all available locations, as determined by our weight variables.

We summarize our entire analysis in the following fashion:

1. First, we select a set of variables—both intraneous and extraneous to our data set that serve as our weight vector $\mathbf{z}$. These weight variables allow us to split the dataset into small groups that share the same underlying population or spatial characteristics, provided by the same value of the weight variables in each group. These small groups represent our locations. Note that the observations in these locations may not necessarily overlap with the observations of the individual source studies.

2. We then compute the Euclidean distance between the local observation(s) and all other observa-

[7] The number of observations that receives positive weight are implicitly determined by the window-size w as discussed above.

tions based on $\mathbf{z}$. We sort the data in ascending order of calculated distance and retain the first w observations as local sample for estimation purposes.

3. Next, we assign a weight based on a specific weight function ψ_t to each observation in the local sample and estimate the LW-MRM model using the Bayesian approach.

4. We repeat this analysis for all the locations as determined in step 1.

5. We then repeat steps 1-4 for each weight setting, i.e., all possible combinations of weight variables, weight functions and window sizes.

6. Finally, we obtain a set of posterior distributions of parameter estimates for each weight setting and determine the predictive performance following the full-sample CV and CV *by proxy* exercise as discussed above, for which we again use the Bayesian estimation approach similar to step 3. This process generates r draws of predicted WTP for each observation within each model, of which we calculate the mean and compare it to the actual value of WTP to then compute the MAPE statistics. A lower value of MAPE means higher predictive accuracy.

3 Data

We use the same water quality meta-data as in Moeltner et al. (2022). The core of this data set was first used by the U.S. EPA in designing regulations on steam electric power plant discharges (U.S. Environmental Protection Agency, 2015).[8] This meta-data originally consisted of 140 observations drawn from 51 different primary research efforts, which we refer as source studies. We revise the original sample by adding observations from other studies (e.g., more recent contributions such as Moore et al. (2018), Choi and Ready (2021), see Table 1) that are found appropriate in generating a benefit transfer prediction in an environmental valuation context. Specifically, we add a total of 73 observations, from 14 source studies. Some of these studies overlap with existing ones in terms of populations and study sites. Therefore, seven studies contributing to a total of 25 observations were dropped from the original meta-data. The final meta-data, hence, consists of 58 studies, providing a total of 188 observations.

[8]This data set is then adopted by Newbold et al. (2018) and Moeltner (2019) to examine and achieve the consistency of the theoretical properties, which is then followed by Johnston et al. (2019) to systematically model the distance-decay mechanism while using meta-regression in BT context.

While Table 1 presents the final meta-data with a brief overview of all the studies that were added to the original data set, Table 2 presents the descriptive statistics and the description of the variables used in our analysis. Our meta-data consists of 22 explanatory variables, 16 of which are context-specific and six are moderators/methodological indicators. Context-specific variables (i.e., $\mathbf{x}$) consist `lnyear`, `noann`, `nonusers`, `sub-proportion`, `northeast`, `central`, `south`, `swim`, `fish`, `lnag`, `lnincome`, `paytax`, `payuse`, `ln_ar_ratio`, $\mathtt{q}_0$, and $\mathtt{q}_1$ and are known for policy sites while the methodological variables (i.e., $\mathbf{m}$) consist of `thesis`, `volunt`, `nonrev`, `oneshotval`, `rum`, and `ibi` which are more of a nuisance for the actual transfer step but need to be controlled to avoid any potential omitted variable problems. Out of these 22 regressors, seven variables (`lnyear`, `sub-proportion`, `lnag`, `lnincome`, `ln_ar_ratio`, $\mathtt{q}_0$, and $\mathtt{q}_1$) are continuous; the rest enter the regression model as binary indicators with values of 0 and 1.

Table 1 and Table 2

The average willingness to pay to obtain the change in water quality in our sample is $140.85 (adjusted to 2019 dollar value). The baseline water quality in our samples is 46.90 (100-point scale) while the endpoint water quality is 59.99 (100-point scale), on average. This provides us with an average point difference of 13.09 for water quality from the baseline through the endpoint. While around 60% of the observations come from studies conducted in the central and southern regions of the U.S., only 13% come from studies conducted in the northeast region of the country. Our sample not only consists of observations from published studies in a peer-reviewed (PR) journal, but also from theses (8%) and other non-PR sources (16%).

Table 3 presents the descriptive statistics of external variables (not included in MRM) that enters as a weight variable for a different weight combination used in our analysis. The first two variables, i.e., `ln_pop` (the log of the population within the affected resource area) and `ln_sz_ratio` (logged ratio of the size of the affected resource area over the population-averaged distance from the resource) had already been included in previous MRM installments, while the remaining four, i.e., `pctdev` (percentage of the catchment area that is developed land), `pctopen` (percentage of catchment area that is open and ag. land), `pctfor` (percentage of the catchment that is forest land), and `pctwet` (percentage of the catchment area that is wetland) were derived by merging the spatial information from our source studies with the EPA's database on land features of lakes and stream catchment regions, as described in Hill et al. (2016) and Hill et al. (2018).

Table 3

We use the context-specific variables and external variables from Tables 2 and 3, respectively and define four different combinations of weight variables as `wvar 1`, `wvar 2`, `wvar 3` and `wvar 4` (see Table 4). The first weight variable combination `wvar 1` includes the `pctdev`, `pctopen`, `pctfor`, and `pctwet`. All of these variables are extraneous to our baseline model. We add 5 more variables, i.e., `sub-proportion`, `lnag`, `lnincome`, `ln_pop`, `ln_sz_ratio` to `wvar 1` (the first three of which are internal to our baseline model) and define our `wvar 2`. While our third weight variable combination `wvar 3` consists of all the continuous variables presented in panels 1 and 2, the third weight variable combination `wvar 4` goes beyond and includes all the variables in panels 1 and 2 (i.e., all context-specific and external variables).

Table 4

4 Results and Discussion

With this data in hand, we estimate the models discussed in section 2. First, we use the estimates from equation (1) to calculate the mean APE for the baseline MRM. Then, we estimate equation (2) with the use of different weight functions and window sizes as described in section 2. We use a full sample (188 observations) as a starting point for the window size, indicating the full use of observations in the regression process. We then decrease our window size until we reach a window size of 100. This process yields seven window sizes (188, 170, 150, 130, 110, 105, 100), the combination of which with three functions (Gaussian, bi-square and tri-cubed) provides us with a total of 21 models for each weight version, providing a total of 84 models for four weight versions (`wvar 1`, `wvar 2`, `wvar 3`, and `wvar 4`). We estimate each model and perform a CV exercise to calculate MAPE as in equation (6). Within each model, we estimate WTP locally, i.e., for each identified location (see details below) using a Bayesian estimation framework. For each coefficient that $\boldsymbol{\theta}$ takes in equation (7), we specify their priors as a mean vector of zeros and a variance of ten. For the error variance σ^2, we specify inverse-gamma priors with the value of 1/2 for both shape and scale. We remove the first 10,000 draws as "burn-ins" and keep the following 10,000 draws for the purpose of inference. These specifications of priors, burn-ins, and "keepers" are exactly the same as in Moeltner et al. (2019).

4.1 Full-sample CV results

The results from the full-sample CV approach for four weight versions, i.e., `wvar 1`, `wvar 2`, `wvar 3`, and `wvar 4` are presented in Table 5 through Table 8, respectively. Across all tables, we present two effective window sizes: minimum and maximum. While the minimum window size represents the window sizes that we pre-specify in our model, the maximum window size goes beyond and captures all observations in the "boundary or cut-off" point (as dictated by the minimum window size) sharing the same distance. For any given model, this number would be either equal to or slightly higher than its corresponding minimum window size as evidenced in Tables 5 through 8 (second row vs third row, for example). The fourth column "identified local regression" presents the number of locations that we identify in our regression. These identified locations vary and range from 37 to 65 for `wvar 1`, 44 to 82 for `wvar 2`, 98 to 143 for `wvar 3`, 106 to 155 for `wavr 4`, and depending on the window sizes. For larger window sizes, we retain all possible locations in the model. However, this number decreases as we decrease the window size because some of the observations—hence their corresponding locations—fall outside of the specified window size. Even for the included locations, when we decrease the window size, there is a possibility that all "0" or "1" of one of the binary indicators may drop out or create perfect collinearities, thus causing a rank violation in the underlying local regression. For the same reason, we may lose an additional number of observations while dropping the host observation during the leave-one-out exercise. This further reduces the number of identified locations in our model. The fifth column captures only those observations that are retained for the APE calculation while running local regressions within each model. The last four columns report the mean, median, minimum, and maximum absolute percentage error.

Table 5, Table 6, Table 7 and Table 8

The first row of Tables 5 through 8 reports the statistics for the baseline MRM. This baseline MRM is free of any weight functions and generates a MAPE of 112% (approx). All other rows present MAPEs computed with different settings of window sizes and weight functions. We consistently find an accuracy gain in each of these settings relative to the baseline MRM. These gains range from 6% to 55% and are consistent across the choice of all weight variables. We notice a decrease in the value of MAPE (hence, improvement in accuracy) as the window size decreases when the Gaussian weight function is used for LW-MRM. The overall trend is consistent for the use of the bi-square and tri-cubed weight functions. The only exceptions are mid-range window sizes where we find the larger values of MAPEs relative to higher window-sized cases (170 vs 130 for the choice of `wvar 4` in Table 8, for example). This can occur when the home observation is effectively an outlier for its own local regression model, which is possible when the home observation takes different explanatory variables than the majority of the observations. The MAPE values, then, decrease for the lower window size cases and become the smallest (largest accuracy gain) for the window size range of 110-105 since such troublemaker locations get no longer identified in lower window size cases. However, these gains occur with the loss of usable observations for APE. Although we succeed to retain a relatively larger number of observations in our model even for the use of lower window size, the corresponding MAPEs still can't be interpreted and compared directly with the baseline MRM as they don't truly represent the entire data. These results are consistent across the choice of all weight variables.

Next, we just look at those window-size cases where we were able to retain full observations back in the model within the choice of each weight function and weight variable. We notice that window size 170 is out-performing window size 188 across all cases. We then examine the performance of the weight function when the window size is 170 and find that choice of the bi-square function outperforms the choice of Gaussian and tri-cubic function, the latter one being close. We then proceed toward the choice of weight variables. Specifically, we evaluate the performance of the weight variables in the afore-selected weight function and the window size configuration. While the choice of `wvar 1` and `wvar 2` yield similar results, they slightly outperform the choice of `wvar 3` and `wvar 4`. This result holds not only in the chosen weight function and the window size configuration (i.e., bi-square, 170), but also applies in general across all choices. It is worth mentioning that we perform a lot of trials before we reach the choice of `wvar 1` and `wvar 2` So, this paper doesn't provide any guidance on how to make these choices and hence requires an experiment with different combination of weight variables. However, when selecting `wvar 3` and `wvar 4`, we exclusively employ all continuous variables and all context-specific and external variables, respectively. As can be seen in Tables 5 through 8, the performance of `wvar 3` and `wvar 4` is not significantly inferior to that of `wvar 1` and `wvar 2`. So, a researcher might save some computational time and can avoid the cumbersome exercise of varying the analysis with the different choices of weight combinations by picking all continuous variables (for say) as a weight.

4.2 CV *by proxy*

As discussed in section 2, CV *by proxy* requires us to create a hypothetical scenario for the policy point based on which we choose 10 sample locations in the vicinity. For this, we create a stylized scenario where a researcher wants to predict the WTP for the change in water quality from $q_0 = 65$ to $q_1 = 70$. We pre-define context-specific variables as follows: "`study year`" = log(2023-1980), "`non-annual payments`" = 0, "`tax payment`" = 1, "`non-users`" = 1, "`central`" = 1 (all the other regions get the value 0), "`swimming`" = 1, "`fishing`" = 1, "`user fees`" = 0, and all other continuous variables that are both internal and external to context-specific variables are set to their sample mean. We use these values to define the weight variables related to this scenario, which is then used to find the 10 closest sample points based on the Euclidean distance. Each of these points is first removed and then predicted using the given weight setting and then averaged to calculate the MAPE—a similar exercise as we carried out before.

Table 9, Table 10, Table 11, and Table 12

Tables 9 through 12 present the results from CV *by proxy* for the choice of `wvar 1`, `wvar 2`, `wvar 3`, and `wvar 4`, respectively. Similar to the full-sample case, the first row in each of these tables presents the results for the baseline MRM. The baseline MRM is free of weight function and the MAPE is solely dictated by the weight variables. Hence, unlike in full-sample, the MAPE for baseline is different (i.e., 77.812%, 85.656%, 94.561%, and 509.976% respectively for `wvar 1 - wvar 4` for the different choice of weight variables. All other rows present the MAPE statistics for the different weight functions and window-size configurations.

A quick overview of these statistics across all tables suggests that LWR can bring a substantial gain in predictive accuracy compared to the baseline model. These gains are consistent across the tables and are up to 40%, 50%, 60%, and 90% for the choice of `wvar 1`, `wvar 2`, `wvar 3`, and `wvar 4`, respectively. A close look at these tables suggests that `wvar 4` even though it generates up to 90% accuracy gain, has some exceptions where the MAPE generated by LWR is higher than that of baseline (mid-window size ranges, for example). This can occur because of the similar reason we discussed above, i.e., when the window size is reduced the home observation can effectively be an outlier for its own local regression model. There is a similar exception for the choice of `wvar 3` when the window size is 170-150 and the weight function is Gaussian. Besides this, our result is stable across all weight settings.

Next, we look at the performance of the window size and weight function for the choice of each weight variable setting (i.e., in each table). We notice that within each weight function, lower window sizes (105 and 100) are outperforming all other choices and for these window sizes, the bi-square weight function consistently brings more accuracy gain into the model.[9] While `wvar 2` clearly outperforms other weight variables for the afore-selected configuration, it is important to note that `wvar 3`—the weight variable that includes all continuous variables, is not significantly inferior. Hence, for the same reason discussed in section 4.1, a researcher may want to consider all continuous variables available in the dataset as one combination of weight.

5 Conclusion

We use the Bayesian Locally-Weighted MRM and provide guidance toward the choice of optimal weight settings. We first show that LWR has the potential to improve the predictive accuracy of the models and highlight the importance of exploring different weight functions and window sizes to find the most appropriate settings. We examine the importance of the weight function in providing guidance on model choice and suggest the use of the bi-square function as it outperforms the choice of other weight functions, regardless of the CV approach used. Additionally, we show that the optimal window size for LWR models depends on the type of cross-validation method employed. While the full-sample CV guides the use of a higher window-size case, CV *by proxy* guides the choice of a lower window size.

While searching for optimal weight variables is recommended, combining all available continuous variables within the data and using them as one single weight can still yield good results and can serve as a practical and efficient alternative. Our findings have important implications for researchers and practitioners who wish to employ LWR models in their work as they can optimize the performance of their models and improve decision-making by considering our configuration. Further, if a researcher is interested in policy prediction, then she could use the top ten (for say) weight settings from each CV approach and evaluate the performance of each model (given by weight settings) by directly comparing the predictive efficiency of these models with that of baseline MRM. Further research can continue to explore the use of different weight functions and window sizes for LWR models in various applications and the development of new weight variable selection methods.

[9] The performance of tri-cubed is similar and very close to that of bi-square.

Bibliography

Aiken, R., 1985. Public benefits of environmental protection in Colorado. Ph.D. thesis. Colorado State University.

Atkinson, P.M., German, S.E., Sear, D.A., Clark, M.J., 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. Geographical Analysis 35, 58–82.

Banzhaf, H.S., Burtraw, D., Criscimagna, S., Cosby, B.J., Evans, D.A., Krupnick, A., Siikamäki, J., et al., 2014. Valuation of ecosystem services in the southern appalachian mountains. Discussion Paper-Resources for the Future (RFF) .

Banzhaf, H.S., Burtraw, D., Evans, D., Krupnick, A., 2006. Valuation of natural resource improvements in the adirondacks. Land Economics 82, 445–464.

Bockstael, N., McConnel, K., Strand, I., 1989. Measuring the benefits of improvements in water quality: The Chesapeake Bay. Marine Resource Economics 6, 1–18.

Borisova, T., Collins, A., D'Souza, G., Benson, M., Leigh Wolfe, M., Benham, B., 2008. A benefit-cost analysis of total maximum daily load implementation 1. JAWRA Journal of the American Water Resources Association 44, 1009–1023.

Brunsdon, C., Fotheringham, A., Charlton, M., 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. Geographical Analysis 28, 281–298.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician) 47, 431–443.

Cameron, T.A., Huppert, D.D., 1989. Ols versus ml estimation of non-market resource values with payment card interval data. Journal of environmental economics and management 17, 230–246.

Cao, K., Diao, M., Wu, B., 2019. A big data–based geographically weighted regression model for public housing prices: A case study in singapore. Annals of the American Association of Geographers 109, 173–186.

Carson, R.T., 1994. Prospective interim lost use value due to ddt and pcb contamination in the southern california bight: Volumes 1&2. report to the National Oceanic and Atmospheric Administration .

Choi, D., Ready, R., 2021. Measuring benefits from spatially-explicit surface water quality improvements: The roles of distance, scope, scale, and size. Resource and Energy Economics 63, 1101108.

Cleveland, W., 1979. Robust locally-weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74, 829–836.

Cleveland, W., Devlin, S., 1988. Locally-weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association 83, 596–610.

Clonts, H.A., Malone, J.W., 1990. Preservation attitudes and consumer surplus in free-flowing rivers, in: Social science and natural resource recreation management. Routledge, pp. 301–317.

Collins, A., Rosenberger, R., Fletcher, J., 2009. Valuing the restoration of acidic streams in the Appalachian Region, in: Thurston, H., Heberling, M., Schrecongost, A. (Eds.), Environmental Economics for Watershed Restoration, pp. 29–52.

Collins, A.R., Rosenberger, R.S., 2007. Protest adjustments in the valuation of watershed restoration using payment card data. Agricultural and Resource Economics Review 36, 321–335.

Corrigan, J.R., Kling, C.L., Zhao, J., 2008. Willingness to pay and the cost of commitment: an empirical specification and test. Environmental and Resource Economics 40, 285–298.

Croke, K., Fabian, R., Brenninam, G., 1987. Estimating the value of improved water quality in an urban river system. Journal of Environmental Systems 16, 13–24.

Desvousges, W.H., Smith, V.K., Fisher, A., 1987. Option price estimates for water quality improvements: a contingent valuation study for the monongahela river. Journal of Environmental Economics and Management 14, 248–267.

Farber, S., Griner, B., 2000. Using conjoint analysis to value ecosystem change.

Farber, S., Páez, A., 2007. A systematic investigation of cross-validation in gwr model estimation: Empirical analysis and Monte Carlo simulations. Journal of Geographical Systems 9, 371–396.

Fotheringham, A., Brunsdon, C., Charlton, M., 2002. Geographically weighted regression: The analysis of spatially varying relationships. Wiley.

Fotheringham, A.S., Brunsdon, C., 1999. Local forms of spatial analysis. Geographical analysis 31, 340–358.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2009. Geographically weighted regression. The Sage handbook of spatial analysis 1, 243–254.

Fotheringham, A.S., Crespo, R., Yao, J., 2015. Geographical and temporal weighted regression (gtwr). Geographical Analysis 47, 431–452.

Gao, J., Li, S., 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Applied Geography 31, 292–302.

Gilbert, A., Chakraborty, J., 2011. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in florida. Social Science Research 40, 273–286.

Harris, P., Fotheringham, A.S., Juggins, S., 2010. Robust geographically weighted regression: A technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. Annals of the Association of American Geographers 100, 286–306.

Hayes, K.M., Tyrrell, T.J., Tyrell, T.J., ANDERSON, G., 1992. Estimating the benefits of water quality improvements in the upper narragansett bay. Marine Resource Economics 7, 75–85.

Herriges, J.A., Shogren, J.F., 1996. Starting point bias in dichotomous choice valuation with follow-up questioning. Journal of environmental economics and management 30, 112–131.

Hill, R., Weber, M., Debbout, R., Leibowitz, S., Olsen, S., 2018. The lake-catchment (LakeCat dataset: Characterizing landscape features for lake basins within the conterminous USA. Freshwater Science 37, 208–221.

Hill, R., Weber, M., Leibowitz, S., Olsen, S., Thronbrugh, D., 2016. The stream-catchment (StreamCat dataset: A database of watershed metrics for the conterminous United States. Journal of the American Water Resources Association 52, 120–128.

Hite, D., Hudson, D., Intarapapong, W., 2002. Willingness to pay for water quality improvements: The case of precision application technology. Journal of Agricultural and resource Economics , 433–449.

Holland, B., Johnston, R., 2017. Optimized quantity-within-distance models of spatial welfare heterogeneity. Journal of Environmental Economics and Management 85, 110–129.

Huang, J.C., Haab, T.C., Whitehead, J.C., 1997. Willingness to pay for quality improvements: should revealed and stated preference data be combined? Journal of environmental economics and management 34, 240–255.

Initiative, W.B.S.R., Hansen, E., Collins, A., Svetlik, J., McClurg, S., 2008. An economic benefit analysis for abandoned mine drainage remediation in the west branch susquehanna river watershed, pennsylvania .

Interis, M., Petrolia, D., 2016. Loaction, location, habitat: How the value of ecosystem services varies across location and by habitat. Land Economics 92, 292–307.

Irvin, S., Haab, T., Hitzhusen, F.J., 2007. Estimating willingness to pay for additional protection of ohio surface waters: contingent valuation of water quality. Economic valuation of river systems. Edward Elgar, Cheltenham , 35–51.

Johnston, R., Besedin, E., Stapler, R., 2017. Enhanced geospatial validity for meta-analysis and environmental benefit transfer: An application to water quality improvements. Environmental and Resource Economics 90, 773–795.

Johnston, R., Ramachandran, M., 2014. Modeling spatial patchiness and hot spots in stated preference willingness to pay. Environmental and Resource Economics 59, 363–387.

Johnston, R., Swallow, S., Allen, C., Smith, L., 2002. Designing multidimensional environmental programs: Assessing tradeoffs and substitution in watershed management plans. Water Resources Research 38, 479–510.

Johnston, R.J., Besedin, E.Y., Holland, B.M., 2019. Modeling distance decay within valuation meta-analysis. Environmental and Resource Economics 72, 657–690.

Kamarianakis, Y., Feidas, H., Kokolatos, G., Chrysoulakis, N., Karatzias, V., 2008. Evaluating remotely sensed rainfall estimates using nonlinear mixed models and geographically weighted regression. Environmental Modelling & Software 23, 1438–1447.

Kaoru, Y., 1993. Differentiating use and nonuse values for coastal pond water quality improvements. Environmental and Resource Economics 3, 487–494.

Lant, C.L., Roberts, R.S., 1990. Greenbelts in the cornbelt: riparian wetlands, intrinsic values, and market failure. Environment and Planning A 22, 1375–1388.

Lant, C.L., Tobin, G.A., 1989. The economic value of riparian corridors in cornbelt floodplains: a research framework. The Professional Geographer 41, 337–349.

Li, B., Cai, Z., Jiang, L., Su, S., Huang, X., 2019. Exploring urban taxi ridership and local associated factors using gps data and geographically weighted regression. Cities 87, 68–86.

Lichtkoppler, F.R., Blaine, T.W., 1999. Environmental awareness and attitudes of ashtabula county voters concerning the ashtabula river area of concern: 1996–1997. Journal of Great Lakes Research 25, 500–514.

Lindsey, G., 1994. Market models, protest bids, and outliers in contingent valuation. Journal of Water Resources Planning and Management 120, 121–129.

Lipton, D., 2004. The value of improved water quality to chesapeake bay boaters. Marine Resource Economics 19, 265–270.

Loader, C., 1999. Local regression and likelihood. New York: Springer.

Londoño Cadavid, C., Ando, A.W., 2013. Valuing preferences over stormwater management outcomes including improved hydrologic function. Water Resources Research 49, 4114–4125.

Lyke, A.J., 1993. Discrete choice models to value changes in environmental quality: a Great Lakes case study. The University of Wisconsin-Madison.

Manganelli, B., Pontrandolfi, P., Azzato, A., Murgante, B., 2014. Using geographically weighted regression for housing market segmentation. International Journal of Business Intelligence and Data Mining 13 9, 161–177.

Mathews, L.G., Homans, F.R., Easter, K.W., 1999. Reducing Phosphorus Pollution in the Minnesota River: How Much Is It Worth? Technical Report.

McMillen, D., Redfearn, C., 2010. Estimation and hypothesis testing for nonparametric hedonic house price functions. Journal of Regional Science 50, 712–733.

Moeltner, K., 2019. Bayesian nonlinear meta regression for benefit transfer. Journal of Environmental Economics and Management 93, 44–62.

Moeltner, K., Balukas, J., Besedin, E., Holland, B., 2019. Waters of the United States: Upgrading wetland valuation via benefit transfer. Ecological Economics 164, 106336.

Moeltner, K., Puri, R., Johnston, R., Besedin, E., Balukas, J., Le, A., 2022. Locally-weighted meta-regression and benefit transfer. Paper presented at the annual W-4133 regional project meeting, Hood River, OR, Apr. 26-27, 2022 .

Moore, C., Guignet, D., Maguire, K., Dockins, C., Simon, N., 2018. Valuing ecological improvements in the Chesapeake Bay and the importance of ancillary benefits. Journal of Benefit Cost Analysis 9, 1–26.

Nelson, N., Loomis, J., Jakus, P., Kealy, M., von Stackelburg, N., Ostermiller, J., 2015. Linking ecological data and economics to estimate the total economic value of improving water quality by reducing nutrients. Ecologica Economics 118, 1–9.

Newbold, S., Massey, D., Walsh, P., Hewitt, J., 2018. Using structural restrictions to achieve theoretical consistency in benefit transfer. Environmental and Resource Economics 69, 529–553.

Opaluch, J.J., Grigalunas, T., Diamantides, J., Mazzotta, M., Johnston, R., 1999. Recreational and resource economic values for the peconic estuary system. Economic Analysis Inc. Final Report .

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Themes in Modern Econometrics, Cambridge University Press.

Pasculli, A., Palermi, S., Sarra, A., Piacentini, T., Miccadei, E., 2014. A modelling methodology for the analysis of radon potential based on environmental geology and geographically weighted regression. Environmental Modelling & Software 54, 165–181.

Propastin, P., 2012. Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. International Journal of Applied Earth Observation and Geoinformation 18, 82–90.

Redfearn, C., 2009. How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. Regional Science and Urban Economics 39, 297–306.

Roberts, L.A., 1997. Economic valuation of some wetland outputs of Mud Lake, Minnesota-South Dakota. North Dakota State University.

Rowe, R.D., Schulze, W.D., Hurd, B., Orr, D., 1985. Economic assessment of damage related to the eagle mine facility. energy and resource consultants. Inc., Boulder, CO .

Sanders, L.D., Walsh, R.G., Loomis, J.B., 1990. Toward empirical estimation of the total value of protecting rivers. Water Resources Research 26, 1345–1357.

Schulze, W., Boyce, R., McClelland, G., Robert, D., Breffle, W., 1995. Contingent valuation of natural resource damages due to injuries to the Upper Clark Fork River Basin. Creative Media Partners, LLC.

Shrestha, R.K., Alavalapati, J.R., 2004. Valuing environmental benefits of silvopasture practice: a case study of the lake okeechobee watershed in florida. Ecological Economics 49, 349–359.

Stumborg, B.E., Baerenklau, K.A., Bishop, R.C., 2001. Nonpoint source pollution and present values: A contingent valuation study of lake mendota. Applied Economic Perspectives and Policy 23, 120–132.

Sutherland, R.J., Walsh, R.G., 1985. Effect of distance on the preservation value of water quality. Land Economics 61, 281–291.

Takatsuka, Y., 2004. Comparison of the contingent valuation method and the stated choice model for measuring benefits of ecosystem management: a case study of the Clinch River Valley, Tennessee. The University of Tennessee.

U.S. Environmental Protection Agency, 2015. Benefit and cost analysis for the effluent limitations guidelines and standards for the steam electric power generating point source category. Technical Report EPA-821-R-15-005. United States Environmental Protection Agency, Office of Water.

van Houtven, G., Mansfield, C., Phaneuf, D., von Haefen, R., Milstead, B., Kenney, M., Reckhow, K.H., 2014. Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. Ecological Economics 99, 40–52.

Wang, Z., Fan, C., Zhao, Q., Myint, S.W., 2020. A geographically weighted regression approach to understanding urbanization impacts on urban warming and cooling: A case study of las vegas. Remote Sensing 12, 222.

Wattage, P.M., 1992. Measuring the benefits of water resource protection from agricultural contamination: results from a contingent valuation study. Iowa State University.

Welle, P.G., 1986. Potential Economic Impacts of Acid Deposition: A Contingent Valuation Study of Minnesota (Rain). The University of Wisconsin-Madison.

Welle, P.G., Hodgson, J.B., 2011. Property owners' willingness to pay for water quality improvements: contingent valuation estimates in two central minnesota watersheds. Journal of Applied Business and economics 12, 81–94.

Wey, K.D., 1990. A social welfare analysis of congestion and water quality of great salt pond, black island, rhode island .

Whitehead, J., Blomquist, G., Hoban, T., Clifford, W., 1995. Assessing the validity and reliability of contingent valuation: A comparison of on-site users, off-site users, and non-users. Journal of Environmental Economics and Management 29, 238–251.

Whitehead, J.C., 2006. Improving willingness to pay estimates for quality improvements through joint estimation with quality perceptions. Southern Economic Journal 73, 100–111.

Whitehead, J.C., Groothuis, P.A., et al., 1992. Economic benefits of improved water quality: a case study of north carolina's tar-pamlico river. Rivers 3, 170–178.

Whittington, D., Cassidy, G., Amaral, D., McClelland, E., Wang, H., Poulos, C., 1994. The economic value of improving the environmental quality of galveston bay. Publications. Galveston Bay National Estuary Program 38, 292.

Zhang, Y., Middel, A., Turner, B., 2019. Evaluating the effect of 3d urban form on neighborhood land surface temperature using google street view and geographically weighted regression. Landscape Ecology 34, 681–697.

Zhao, M., Johnston, R., Schultz, E., 2013. What to value? Ecological indicator choices in stated preference valuation. Environmental and Resource Economics 56, 3–25.

de Zoysa, A.D.N., 1995. A benefit evaluation of programs to enhance groundwater quality, surface water quality and wetland habitat in Northwest Ohio. The Ohio State University.

Tables

Table 1: Source Studies and Number of Observations

Source Studies	No. of observations.	Remarks
Aiken (1985)	1	-
Banzhaf et al. (2006)	2	-
Banzhaf et al. (2014)	1	-
Bockstael et al. (1989)	2	new addition (1 out of 2 obs.)
Borisova et al. (2008)	2	-
Cameron and Huppert (1989)	1	-
Carson (1994)	2	-
Choi and Ready (2021)	6	new addition
Clonts and Malone (1990)	2	-
Collins and Rosenberger (2007)	1	-
Collins et al. (2009)	1	new addition
Corrigan et al. (2008)	1	-
Croke et al. (1987)	6	-
de Zoysa (1995)	1	-
Desvousges et al. (1987)	12	-
Initiative et al. (2008)	2	-
Farber and Griner (2000)	6	-
Hayes et al. (1992)	2	-
Herriges and Shogren (1996)	1	-
Hite et al. (2002)	2	-
Holland and Johnston (2017)	6	new addition
Huang et al. (1997)	2	-
Interis and Petrolia (2016)	10	new addition
Irvin et al. (2007)	4	-
Johnston and Ramachandran (2014)	3	new addition
Johnston et al. (2002)	1	new addition
Johnston et al. (2017)	3	new addition
Kaoru (1993)	1	-
Lant and Roberts (1990)	3	-
Lant and Tobin (1989)	9	-
Lichtkoppler and Blaine (1999)	1	-
Lindsey (1994)	8	-
Lipton (2004)	1	-
Londoño Cadavid and Ando (2013)	2	-
Lyke (1993)	2	-
Mathews et al. (1999)	1	-
Moore et al. (2018)	2	new addition
Nelson et al. (2015)	2	new addition
Opaluch et al. (1999)	1	-
Roberts (1997)	1	-
Rowe et al. (1985)	1	-
Sanders et al. (1990)	4	-
Schulze et al. (1995)	4	new addition (2 out of 4 obs.)
Shrestha and Alavalapati (2004)	2	-
Stumborg et al. (2001)	2	-
Sutherland and Walsh (1985)	1	-
Takatsuka (2004)	4	-
van Houtven et al. (2014)	32	new addition
Wattage (1992)	2	-
Welle and Hodgson (2011)	3	-
Welle (1986)	4	-
Wey (1990)	1	-
Whitehead et al. (1992)	2	-
Whitehead et al. (1995)	1	new addition
Whitehead (2006)	3	-
Whittington et al. (1994)	1	-
Zhao et al. (2013)	3	new addition
Total no. of studies= 58	Total no. of obs. = 188	Total addition = 73 obs

Table 2: Descriptive statistics of variables used in our analysis

Variables	Description	Mean	Std. Dev.	Min	Max
wtp	estimated WTP for specified water quality change, 2019$	140.85	122.91	1.73	498.38
	context-specific variables (known for policy context)				
lnyear	log (years since the earliest study in the sample (1980))	2.64	0.98	0.00	3.61
nonann	1= payments are not annual (e.g. lump sum)	0.18	0.39	0.00	1.00
nonusers	1= survey population only includes non-users	0.06	0.24	0.00	1.00
sub-proportion	proportion of water bodies of same type in the state or region	0.35	0.40	0.00	1.00
northeast	1= study conducted in the northeast region of the U.S.	0.13	0.34	0.00	1.00
central	1= study conducted in the central (midwest or mountain plains) region of the U.S.	0.30	0.46	0.00	1.00
south	1= study conducted in the southeast and southwest region of the U.S.	0.30	0.46	0.00	1.00
swim	1 = changes in swimming use emphasized in survey	0.22	0.41	0.00	1.00
fish	1= changes in game fishing emphasized in survey	0.19	0.39	0.00	1.00
lnag	log of affected resource areas that is agricultural	-1.64	0.90	-4.26	-0.08
lnincome	log of median U.S. income in the year of data collection, 2019$	10.95	0.16	10.65	11.48
paytax	1= payment mech = increased taxes	0.40	0.49	0.00	1.00
payuse	1 = payment mechanism = increased user cost	0.02	0.14	0.00	1.00
ln_ar_ratio	log of sampled area ("market area") divided by the affected resource area	-0.59	2.41	-8.48	6.65
q_0 (baseline)	baseline water quality, 100-point WQI scale	46.90	15.16	10.00	85.00
q_1	target water quality, 10-point WQI scale	59.99	14.07	12.50	95.00
	moderators/methodological indicators				
thesis	1= study was a PhD or Master's thesis	0.08	0.27	0.00	1.00
volunt	1=payment vehicle described as voluntary	0.05	0.23	0.00	1.00
nonrev	1=study was not published in a peer-reviewed journal	0.16	0.37	0.00	1.00
oneshotval	1=only 1 valuation question given	0.53	0.50	0.00	1.00
rum	1=RUM model used	0.56	0.50	0.00	1.00
ibi	1= water quality derived from a biological index	0.08	0.27	0.00	1.00

Std. Dev. = standard deviation

Table 3: External variables that serves in different weight variables combination

Label	Description	Mean	Std. Dev.	Min	Max
ln__pop	log of population within the affected resource area	14.04	1.64	9.34	16.87
ln__sz__ratio	log of size of affected resource area/population-weighted average distance from resource	7.80	2.52	3.48	14.13
pctdev	percentage of catchment area that is developed land	9.35	15.52	0.17	77.06
pctopen	percentage of the catchment area that is open and ag. Land	29.22	21.26	0.26	91.40
pctfor	percentage of catchment area that is forest land	31.95	22.62	0.06	84.47
pctwet	percentage of catchment area that is wetland	11.10	12.51	0.01	64.30

Std. Dev. = standard deviation
Iden.Loc. = number of identified locations

Table 4: Weight variables combination

Variables	wvar 1	wvar 2	wvar 3	wvar 4
Context Specific				
lnyear			✓	✓
noann				✓
nonusers				✓
sub_proportion		✓	✓	✓
northeast				✓
central				✓
south				✓
swim				✓
fish				✓
lnag		✓	✓	✓
lnincome		✓	✓	✓
paytax				✓
payuse				✓
ln_ar_ratio			✓	✓
q0			✓	✓
q1			✓	✓
external				
ln_sz_ratio		✓	✓	✓
lnpop		✓	✓	✓
pctdev	✓	✓	✓	✓
pctopen	✓	✓	✓	✓
pctfor	✓	✓	✓	✓
pctwet	✓	✓	✓	✓

Table 5: Full-sample CV results using `wvar 1` as weight

weight function	effective window size		identified local regression	usable obs. for APE	absolute percentage error (APE)			
	min	max			mean	median	min	max
baseline MRM	188	188	-	188	112.378	57.834	0.410	4337.237
	188	188	65	188	104.023	54.101	0.191	3946.724
	170	177	65	188	101.476	50.324	0.342	4160.760
	150	159	62	176	97.391	48.848	0.211	3379.196
Gaussian	130	139	50	152	84.383	46.010	0.115	3455.496
	110	119	47	140	58.532	46.028	0.206	342.849
	105	114	43	127	55.709	39.976	0.586	353.240
	100	110	37	110	52.565	40.920	0.058	265.115
	188	188	65	188	83.233	44.674	1.346	2738.958
	170	177	65	188	75.485	43.989	0.189	1824.906
	150	159	62	176	71.175	42.369	0.146	1820.820
Bi-square	130	139	50	152	58.852	39.019	0.194	1764.262
	110	119	47	140	49.206	36.764	1.259	472.841
	105	114	43	127	49.455	34.482	2.343	611.397
	100	110	37	110	54.391	34.045	0.872	943.952
	188	188	65	188	86.004	46.279	0.654	2929.213
	170	177	65	188	77.486	44.560	0.023	1901.706
	150	159	62	176	72.494	42.636	0.173	1874.652
Tri-cubed	130	139	50	152	58.302	39.295	0.266	1580.559
	110	119	47	140	50.014	35.764	0.106	613.270
	105	114	43	127	53.031	34.604	1.814	921.374
	100	110	37	110	61.809	34.847	0.020	1655.268

Table 6: Full-sample CV results using `wvar 2` as weight

weight function	effective window size		identified local regression	usable obs. for APE	absolute percentage error (APE)			
	min	max			mean	median	min	max
baseline MRM	188	188	-	188	112.378	57.834	0.410	4337.237
Gaussian	188	188	82	188	103.837	54.183	0.458	3903.880
	170	174	82	188	97.094	50.277	1.548	3345.113
	150	155	76	175	96.680	48.611	0.468	3248.141
	130	136	62	152	82.131	45.619	0.089	3330.325
	110	119	58	140	58.120	45.683	0.055	307.322
	105	110	50	123	54.646	41.247	0.423	316.664
	100	110	44	109	52.871	41.128	0.022	262.198
Bi-square	188	188	82	188	82.942	44.810	0.404	2699.419
	170	174	82	188	74.541	43.697	0.045	1732.68
	150	155	76	175	70.148	41.754	0.310	1765.746
	130	136	62	152	57.896	39.070	0.200	1625.299
	110	119	58	140	49.155	36.824	0.278	413.611
	105	110	50	123	48.163	34.245	1.442	475.693
	100	110	44	109	52.309	34.001	0.567	740.225
Tri-cubed	188	188	82	188	85.686	46.536	0.655	2880.105
	170	174	82	188	76.399	44.610	0.994	1791.780
	150	155	76	175	71.281	42.229	0.014	1795.492
	130	136	62	152	57.250	38.711	0.966	1439.636
	110	119	58	140	49.820	36.155	0.137	515.645
	105	110	50	123	50.889	34.665	0.126	643.628
	100	110	44	109	57.968	36.884	0.550	1238.147

Table 7: Full-sample CV results using `wvar 3` as weight

weight function	effective window size		identified local regression	usable obs. for APE	absolute percentage error (APE)			
	min	max			mean	median	min	max
baseline MRM	188	188	-	188	112.378	57.834	0.410	4337.237
Gaussian	188	188	143	188	105.756	54.150	0.170	4201.613
	170	172	143	188	100.698	51.265	0.229	3883.361
	150	153	130	175	92.580	50.082	0.193	3630.094
	130	137	110	154	85.844	42.222	0.230	4300.907
	110	115	107	149	65.921	44.385	0.964	1074.069
	105	112	104	145	64.890	42.796	1.273	1167.375
	100	106	97	129	63.134	43.856	0.076	1194.019
Bi-square	188	188	143	188	90.331	47.201	1.173	3910.452
	170	172	143	188	85.760	42.414	0.039	4529.575
	150	153	130	175	85.713	40.970	0.017	5083.155
	130	137	110	154	89.803	34.698	0.997	6688.175
	110	115	107	149	49.655	31.602	0.017	722.273
	105	112	104	145	51.843	30.767	0.152	800.306
	100	106	97	129	55.447	31.209	0.804	1082.147
Tri-cubed	188	188	143	188	92.632	47.856	0.551	3990.099
	170	172	143	188	90.440	44.369	0.260	5229.870
	150	153	130	175	92.557	39.477	0.267	6073.337
	130	137	110	154	103.626	35.450	0.759	8687.406
	110	115	107	149	52.081	31.734	0.146	852.394
	105	112	104	145	54.459	31.036	0.012	1086.150
	100	106	97	129	59.447	31.137	0.476	1355.856

Table 8: Full-sample CV results using `wvar 4` as weight

weight function	effective window size		identified local regression	usable obs. for APE	absolute percentage error (APE)			
	min	max			mean	median	min	max
baseline MRM	188	188	-	188	112.378	57.834	0.410	4337.237
Gaussian	188	188	155	188	105.766	54.636	0.168	4201.357
	170	172	155	188	100.698	50.966	0.227	3882.789
	150	153	142	175	92.558	50.082	0.196	3629.623
	130	137	122	154	85.985	42.216	0.226	4300.526
	110	115	119	149	65.868	45.085	0.958	1074.316
	105	112	115	145	65.596	42.795	0.585	1167.649
	100	107	106	129	63.114	43.899	0.094	1194.383
Bi-square	188	188	155	188	90.315	47.269	1.628	3908.951
	170	172	155	188	85.712	42.404	0.033	4526.251
	150	153	142	175	85.693	40.959	0.135	5079.262
	130	137	122	154	89.710	35.165	0.924	6693.306
	110	115	119	149	50.579	31.613	0.157	724.337
	105	112	115	145	51.707	30.479	0.124	795.612
	100	107	106	129	55.284	31.147	0.694	1068.234
Tri-cubed	188	188	155	188	92.617	48.143	0.555	3988.627
	170	172	155	188	90.399	44.419	0.318	5229.790
	150	153	142	175	92.581	39.471	0.138	6071.216
	130	137	122	154	103.586	35.514	0.849	8702.623
	110	115	119	149	52.944	31.656	0.108	991.518
	105	112	115	145	54.408	31.257	0.092	1079.080
	100	107	106	129	59.314	31.141	0.399	1333.163

Table 9: CV *by proxy* results using `wvar 1` as weight

weight function	window size	usable obs. for APE	absolute percentage error (APE)			
			mean	median	min	max
baseline MRM	188	10	77.812	84.928	23.637	139.960
Gaussian	188	10	74.274	78.498	25.659	133.564
	170	10	74.263	77.516	25.887	132.930
	150	10	72.998	74.820	26.676	132.398
	130	10	70.737	70.922	28.037	125.600
	110	10	62.39	59.838	31.473	108.548
	105	10	62.653	58.926	31.458	107.566
	100	10	56.867	54.345	31.602	93.773
Bi-square	188	10	66.977	65.919	30.557	118.919
	170	10	64.175	61.720	32.205	113.462
	150	10	61.654	54.117	33.523	109.362
	130	10	50.684	47.329	32.384	82.527
	110	10	44.401	45.576	29.653	64.699
	105	10	44.202	46.159	25.928	64.294
	100	10	39.006	39.777	21.084	54.183
Tri-cubed	188	10	69.168	69.447	29.060	123.513
	170	10	66.202	65.046	30.804	117.711
	150	10	63.637	56.432	31.920	113.897
	130	10	50.250	47.051	32.804	80.953
	110	10	43.651	44.802	29.588	62.495
	105	10	43.431	45.542	25.238	62.081
	100	10	39.678	39.830	26.245	53.448

Table 10: CV *by proxy* results using `wvar 2` as weight

weight function	window size	usable obs. for APE	absolute percentage error (APE)			
			mean	median	min	max
baseline MRM	188	10	85.656	89.722	9.972	180.788
Gaussian	188	10	80.450	83.781	17.351	153.065
	170	10	72.939	71.167	26.198	134.592
	150	10	66.700	67.383	5.104	131.505
	130	10	65.114	64.386	15.387	124.422
	110	10	64.404	56.108	27.193	122.688
	105	10	44.207	41.512	7.151	82.382
	100	10	43.318	40.946	4.263	81.065
Bi-square	188	10	65.747	60.304	30.677	120.363
	170	10	59.326	54.565	12.011	113.476
	150	10	57.789	51.306	14.723	107.952
	130	10	47.070	46.399	2.795	83.078
	110	10	43.952	43.106	21.763	68.849
	105	10	34.316	33.492	10.468	53.990
	100	10	34.285	33.318	10.908	53.184
Tri-cubed	188	10	67.656	62.273	29.196	125.089
	170	10	61.301	57.554	13.010	117.788
	150	10	59.987	54.275	16.888	112.531
	130	10	47.498	46.299	9.861	81.661
	110	10	44.521	42.639	28.576	65.661
	105	10	35.712	39.572	8.980	53.216
	100	10	35.678	38.829	9.425	52.417

Table 11: CV *by proxy* results using `wvar 3` as weight

weight function	window size	usable obs. for APE	absolute percentage error (APE)			
			mean	median	min	max
baseline MRM	188	10	94.561	63.380	16.909	351.030
Gaussian	188	10	87.884	59.478	20.850	344.462
	170	10	97.302	59.612	22.342	471.427
	150	10	102.126	67.653	17.458	423.170
	130	10	87.342	57.811	13.894	380.129
	110	10	59.205	58.222	12.278	130.275
	105	10	57.031	54.727	11.813	105.957
	100	10	49.104	51.854	14.324	105.498
Bi-square	188	10	75.124	48.878	6.733	355.157
	170	10	68.791	43.349	1.123	304.370
	150	10	49.972	27.623	3.983	200.786
	130	10	37.305	30.538	2.470	105.177
	110	10	39.577	36.826	9.492	71.894
	105	10	41.213	37.407	10.949	70.504
	100	10	41.333	38.219	12.100	69.988
Tri-cubed	188	10	77.732	50.401	9.219	359.430
	170	10	68.653	45.755	2.031	290.942
	150	10	46.829	25.925	5.292	167.209
	130	10	33.577	29.967	1.158	79.507
	110	10	39.498	35.880	11.170	70.749
	105	10	41.750	39.934	13.254	71.334
	100	10	41.425	40.325	14.829	72.168

Table 12: CV *by proxy* results using `wvar 4` as weight

weight function	window size	usable obs. for APE	absolute percentage error (APE)			
			mean	median	min	max
baseline MRM	188	10	509.980	63.380	16.909	4280.200
Gaussian	188	10	493.718	59.475	14.945	4165.462
	170	10	476.762	59.607	12.870	3882.093
	150	10	450.596	83.548	17.459	3537.070
	130	10	514.393	69.277	14.243	4303.417
	110	10	52.870	43.190	12.274	130.279
	105	10	47.086	33.747	11.811	105.980
	100	10	40.093	33.376	14.316	105.485
Bi-square	188	10	459.329	49.761	5.446	3884.917
	170	10	513.733	54.410	1.098	4479.821
	150	10	547.285	40.982	7.902	4977.284
	130	10	709.371	37.165	2.956	6727.388
	110	10	39.146	41.009	9.505	71.878
	105	10	42.841	42.826	10.952	70.500
	100	10	43.281	43.312	12.129	70.004
Tricubed	188	10	469.565	53.138	8.634	3963.471
	170	10	582.416	57.038	2.566	5164.475
	150	10	641.320	40.096	7.940	5950.650
	130	10	908.797	36.853	1.218	8759.781
	110	10	39.111	38.022	11.188	70.749
	105	10	43.240	43.288	13.258	71.334
	100	10	42.962	44.014	14.862	72.172