The Cost of Algae Contamination in Fresh Water Lakes: Identification of Environmental Quality Marginal Bid Functions Using Hydrology-Based Instrument

David Wolf

The Ohio State University

Wolf.527@osu.edu

H. Allen Klaiber

The Ohio State University

Klaiber.16@osu.edu

Sathya Gopalakrishnan

The Ohio State University

Gopalakrishnan.27@osu.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

The Cost of Algae Contamination in Fresh Water Lakes: Identification of Environmental Quality Marginal Bid Functions Using Hydrology-Based Instrument

Abstract

Unbiased estimation of hedonic, marginal willingness to pay functions has proven to be a difficult task despite Rosen publishing his seminal paper over 40 years ago. Concerns over endogeneity in Rosen's second stage have led researchers to either assume an explicit utility functional form or use data from multiple markets to imperfectly identify marginal bid functions. The methodology proposed in this paper overcomes problems associated with both approaches by deriving marginal bid function estimates using data collected from multiple markets and exogenous instruments of environmental attributes developed from ecological and hydrological processes unique to our study setting. Unbiased parameter estimates recovered from this process are then converted into an inverse Hicksian demand function using the methodology proposed by Hausman (1981). We empirically demonstrate our model using remote-sensing water quality data, housing transactions and demographic data collected from 7 counties bordering Lake Erie. Using our estimated Cobb-Douglas demand function, we find annualized benefits of $3,215 per household when water quality conditions meet the standards discussed under the 2012 Great Lakes Water Quality Agreement. This large improvement in welfare has significantly different implications for policy compared to the naïve estimate of $1,465 derived from first-stage estimates of WTP. Our analysis provides new policy insights that highlight the potential welfare gain from early intervention to avoid large damages produced from environmental change.

Keywords: hedonic; demand estimation; instrumental variables; Lake Erie; harmful algal bloom

JEL Codes: Q25, Q51, Q53, Q57

The Cost of Algae Contamination in Fresh Water Lakes: Identification of Environmental Quality Marginal Bid Functions Using Hydrology-Based Instrument

Highlights

- Overcome endogeneity difficulties present within hedonic literature by developing instrumental variable unique to study setting
- Demonstrate significant difference in welfare estimates using first and second stage hedonic estimates
- Evaluate welfare implications associated with fulfilling the 2012 Great Lakes Water Quality Agreement
- Near lake home owners accrue $136 million in annualized benefits

The Cost of Algae Contamination in Fresh Water Lakes: Identification of Environmental Quality Marginal Bid Functions Using Hydrology-Based Instrument

I. Introduction

Recovering public good demand curves using hedonic theory has proven to be an elusive task. Rosen (1974)'s seminal paper initiated this undertaking by laying out a two-step procedure. Within the first stage a hedonic price function is recovered by regressing the market price of a differentiated product onto its bundle of underlying characteristics. The resulting price function can then decomposed into a set of attribute (implicit) prices which Rosen believed to be akin to observable market prices, if they existed. A number of studies have used these estimates to evaluate marginal changes in public goods (Black 1999; Chay and Greenstone 2005; Greenstone and Gallagher 2005) as they are often interpreted as the consumer's marginal willingness to pay (MWTP) for each attribute at their utility-maximizing decision. Rosen, however, intended to recover marginal bid function estimates using these attribute prices as the dependent variable within a second stage regression, allowing researchers to not only recover MWTP estimates but also entire demand functions for each characteristic of the differentiated product.

A number of identification issues have arisen within the literature, however, which have discouraged using Rosen's full model. Brown and Rosen (1982) argue no new information is added between the first and second stage, within a single market setting, unless additional a priori functional form restrictions are imposed on the hedonic price function.[1] The subsequent second stage estimation procedure will therefore reproduce information already recovered from the first stage, making it impossible to recover unbiased demand parameter estimates. Further

[1]Ekeland and et al. (2004) discovered that additional functional form restrictions are not needed if the hedonic price function is estimated non-parametrically.

identification issues arise when examining the consumer's simultaneous decision to choose both the quantity they consume and the price they pay for each attribute through their optimal consumption decision.[2] Consumers with strong tastes for an amenity will be incentivized to find a market and a product that allow them to consume as much of that good as desired at as low of a price as possible. Unobserved consumer preferences within the second stage, in other words, will be correlated with both the quantity and price variables, leading to inconsistent estimation of the marginal bid function (Bartik 1987).

Despite these problems there is still a need to recover entire demand functions rather than MWTP point estimates. First-stage MWTP estimates are formed and only valid at the point of tangency between the consumer's bid curve and the hedonic price function. Any policy that would non-marginally change the availability or quantity of a public good could therefore not be evaluated accurately using these first stage estimates as this would likely alter the equilibrium hedonic price schedule (Palmquist, 2005). Evaluation of non-marginal changes and the recovery of consumer preferences is crucial, however, for policy decision-making. In particular optimal policy intervention requires knowledge of where, when and to what extent changes in public goods are needed, and these decisions depend in part on the demand for the public good of interest.

Researchers have responded to this dilemma by either adding data from multiple markets (Bartik 1987; Epple 1987; Zabel and Kiel 2000; Zhang et al. 2016) or by assuming an explicit functional form for the utility function (Chattopadhyay 1999; Bajari and Benkard 2005; Bishop and Timmins 2008). Although the multi-market approach circumvents the problem of having to

[2]This is especially true within the property hedonics literature where consumers implicitly decide the availability and price they pay for public goods through their location decision (Epple and Sieg 1999).

specify a full utility functional form, it requires market segmentation exist either across time or space and the assumption that unobserved consumer preferences are not stratified across markets.[3] Empirical studies tend to support the former assumption but not the latter (see Zhang et al. (2016) for a complete discussion on this topic).

The methodology proposed within this paper overcomes problems associated with both approaches by deriving marginal bid function estimates using data collected from multiple markets and exogenous instruments of environmental attributes developed from ecological and hydrological processes unique to the study setting. Unbiased parameter estimates recovered from this process are then converted into an inverse Hicksian demand function using the methodology proposed by Hausman (1981). We empirically demonstrate our model using remote-sensing water quality data, housing transactions and demographic data collected from 7 counties bordering Lake Erie. Finally, welfare implications associated with fulfilling the 2012 Great Lakes Water Quality Agreement are calculated using our recovered Cobb-Douglas marginal bid function, with aggregate, annualized benefits estimated to be $136 million for near Lake Erie homeowners.

The rest of the paper is structured as follows. The next section will provide an overview of the literature and discuss the current limitations present within the multi-market, second-stage hedonic setting. Section 3 describes our study setting, while Section 4 introduces our model and econometric specification. Finally Section 5 presents our study's results and section 6 concludes.

---

[3] Advancements made by Zhang et al. (2016) have been able to relax the market stratification assumption, however, by partially identifying demand parameter slope estimates, even in the presence of within and across market taste-based sorting, using imperfect instruments (Nevo and Rosen 2012) rooted in the logic of Tiebout sorting behavior (Tiebout 1956).

II. Literature Review

Following the recovery of attribute prices from the first stage hedonic regression, Rosen (1974) proposed using a system of simultaneous equations to recover consumer marginal bid curves and supplier marginal offer curves for each attribute of the differentiated product:

(1) $\frac{\partial \mathrm{P(Z)}}{\partial \mathrm{Z_j}} = \theta'_j\left(\mathrm{Z_{i1}}, \dots, \mathrm{Z_{iK}}, \mathrm{X_i}^O, \mathrm{X_i}^U\right)$

(2) $\frac{\partial \mathrm{P(Z)}}{\partial \mathrm{Z_j}} = \phi'_j\left(\mathrm{Z_{i1}}, \dots, \mathrm{Z_{iK}}, \mathrm{Y_i}^O, \mathrm{Y_i}^U\right)$

Where j indexes product attributes 1, …, K and i indexes consumer-supplier pairs 1, …, N, $\theta'_j(.)$ denotes the marginal bid function for attribute j, while $\phi'_{\mathrm{j}}(.)$ represents the marginal offer function for attribute j. The marginal bids and offers made for attribute j will vary depending on the bundle of product characteristics transacted $\{\mathrm{Z_{i1}}, \dots, \mathrm{Z_{iK}}\}$ and by the observed and unobserved characteristics of the buyer, $\mathrm{X_i}^O$ and $\mathrm{X_i}^U$, and seller, $\mathrm{Y_i}^O$ and $\mathrm{Y_i}^U$.[4] Finally P(Z) represents the hedonic price equilibrium, which is a function of all of the product bundles purchased/sold within the market (i.e. $\mathrm{Z} = \{\mathrm{Z_{11}}, \dots, \mathrm{Z_{1K}}, \mathrm{Z_{21}}, \dots, \mathrm{Z_{2K}}, \mathrm{Z_{31}}, \dots, \mathrm{Z_{NK}}\}$), while $\frac{\partial \mathrm{P(Z)}}{\partial \mathrm{Z_j}}$ is the price for attribute j and is derived by taking the derivative of the hedonic price function with respect to characteristic j.

Rosen (1974)'s proposed second stage simply restates the outcomes produced under a hedonic equilibrium. At each point along the hedonic price equilibrium consumers and suppliers are paired such that neither party can improve their objective function by selecting a new bundle of

[4]The superscript on $\mathrm{X_i}$ and $\mathrm{Y_i}$ is used to distinguish observed consumer/supplier characteristics (O) from unobserved characteristics (U). In addition, for clarity purposes we denote the marginal bid made by consumer i for attribute j as θ'_{ij} and the marginal offer made by supplier i for attribute j as ϕ'_{ij}.

characteristics bought/sold at a different point on the hedonic price function. This optimal pairing of buyers and sellers only occurs when consumers are on their lowest bid curve tangent to the hedonic price function (θ_{ij}) and when suppliers are on their highest offer curve tangent to the hedonic price function (ϕ_{ij}). The occurrence of three such pairing (denoted as A, B and C) along the hedonic price equilibrium is graphically depicted in Figure 1, while the equalities presented in equations (1) and (2) capture these tangency conditions.[5]

For any vector of chosen characteristics $\{Z_{i1}, \ldots, Z_{iK}\}$ the derivative of the hedonic price function with respect to attribute j ($\frac{\partial P(Z)}{\partial Z_j}$) can therefore provide a point estimate of both consumer i's marginal bid function for attribute j and supplier i's marginal offer function for attribute j.[6] Figure 2 displays the intersection between the marginal bid curve, marginal offer curve and the derivative of the hedonic price function at three separate equilibrium outcomes. Notice how the hedonic price function intersects each curve only once, providing an estimate of a single point on both the marginal bid function and the marginal offer function. Rosen believed both the consumer's marginal bid function and supplier's marginal offer function could be fully recovered using the estimated hedonic price function, if the above system of equations was treated as a "garden variety identification problem". In particular he proposed using observed demand shifters, X_i^{O}, as instrumental variables for the endogenous quantity regressor (Z_j) within

[5]The hedonic price equilibrium can be interpreted as a product of the interactions between a distribution of buyers and sellers operating within a market and represents the set of prices needed for market-clearing conditions to hold for all variants of the differentiated product.
[6]Note: by further assuming a constant marginal utility of income, consumer i's marginal bid function for attribute j becomes equivalent to an inverse ordinary demand curve (see McConnell and Phipps 1987). Similarly, Rosen (1974) refers to supplier i's marginal offer curve for attribute j as a profit-compensated supply function for attribute j.

the marginal offer function and similarly using observed supply shifters, Y_i^O, as instrumental variables for the endogenous quantity regressor (Z_j) within the marginal bid function.[7]

This approach seems plausible under a standard system of supply and demand equations since exogenous shifts in the supply curve should trace out the demand curve. However as mentioned by Brown and Rosen (1982) and Bartik (1987), this system of equations is not equivalent to a standard set of supply and demand curves. In particular under a hedonic setting, shifts in the marginal offer function will also correspond to shifts in the marginal bid function. Consumer preference parameters derived from this process will be identified based on movements from one hedonic equilibrium to the next and not by shifts up and down the same curve. Figure 3 demonstrates the bias produced from this identification strategy. Suppose you are attempting to recover consumer B's marginal bid curve for attribute j ($\theta'_{B,j}$) by exogenously shifting the marginal offer curve from $\phi'_{B,j}$ to $\phi'_{A,j}$ using variation in Y_i^O. Rather than recovering points B and D from this process, which would allow for the identification of consumer B's marginal bid function, points A and B are recovered instead. Estimates of $\theta'_{B,j}$ (θ'_{Rosen}) will therefore be biased as variation in supplier characteristics will not be cleanly separated from variation in consumer characteristics. Econometrically this suggests that supplier characteristics are unable to meet the stringent exogeneity condition required for valid instrumental variables since the correlation between Y_i^O and X_i^U is nonzero.

One way to circumvent this problem is by adding additional information from multiple markets. The intuition behind this approach is based off of two assumptions: first the shape of the hedonic price function must vary across hedonic markets, and second preferentially-identical

[7] OLS estimation of equations (1) and (2) will produce bias parameter estimates for the variable of interest, Z_j, as $Corr(X^U, Z_j) \neq 0$ and $Corr(Y^U, Z_j) \neq 0$.

households are assumed to be randomly distributed. The first assumption appears to be an intuitive result of spatially and time-varying conditions that form the hedonic price equilibria.[8] The second assumption, on the other hand, is less accepted in the literature but allows preferentially-identical households, located across multiple markets, to select different optimal bundles. If both assumptions hold estimation of multiple points along a shared marginal bid function is possible.

Figure 4 demonstrates the intuition behind the multi-market approach. Consider three consumers (A, B and E) living in two separate markets characterized by the hedonic price functions $P'(Z_j|Z_{-j})^{1}$ and $P'(Z_j|Z_{-j})^{2}$, where the superscript on $P'(Z_j|Z_{-j})$ indexes the hedonic market. Note that consumers A and B live in market 1, while Consumer E lives in market 2. Further suppose that consumers B and E have identical observable characteristics and preferences (i.e. $X_B{}^{O} = X_E{}^{O}; X_B{}^{U} = X_E{}^{U}$). Under this setting consumers B and E will select different quantities of characteristic Z_j despite sharing a common marginal bid curve, θ'_j. This is due to the difference in attribute prices they face in their respective markets. Consumer B faces lower attribute prices for characteristic j than consumer E, resulting in a higher quantity of attribute j selected by consumer B. Differences in the hedonic price function across markets then allows variation in Z_j to be cleanly separated from variation in $X_i{}^{U}$ allowing for unbiased estimation of consumer B and E's shared marginal bid curve.

Although using data from multiple markets appears to be an intuitive work-around to Rosen's endogeneity problem, this approach is likely not applicable in most empirical settings since it depends upon the assumption that consumer preferences are not stratified across markets.

---

[8]Empirical evidence supports this hypothesis as well, with researchers finding variation in the price function across cities (Witte et al. 1979) and metro regions (Zabel and Kiel 2000).

Tiebout (1956)'s "voting with your feet" model provides theoretically rationale explaining why we should expect consumers with similar preferences to live near each other. In particular individuals who are able to sort themselves into separate but preferentially-distinct communities are often able to obtain higher levels of utility than those who are restricted in their location decision. This is most apparent when society uses majority rule to determine its allocation of public goods. When preferentially-distinct communities are formed through sorting the bundle of public goods that residents receive from living within these communities will closely resemble their own utility-maximizing choice. Residents will, in other words, move as a means to minimize the difference between their preference for public goods and the median voter's preference for public goods. In a world without sorting, however, each community will host a wide-range of individuals each with their own unique set of preferences. The median voter's decision in this case will not appease many from within his/her own community since the difference between their decision and the remaining residents' desired allocation of goods will be substantial. This will incentivize homeowners and communities to obtain higher utility levels by segregating into distinct markets, eventually causing the market to move back into a sorting equilibrium.

Consequently the marginal bid function estimates produced from a multi-market hedonic approach will still be biased due to the presence of sorting. Figure 5 further examines the identification issues arising from this approach. Following our previous example, suppose there are three consumers (A, B and E) living in two hedonic markets. Unlike the previous example, however, relax the assumption that consumer preferences are randomly distributed across markets and allow for the possibility of a sorting equilibrium. Under this scenario even if identification of the marginal bid function is drawn from observably similar consumers (i.e.

$X_B{}^O = X_E{}^O$) purchasing different bundles of attributes across markets, the estimated parameter coefficients will likely be biased due to systematic differences in unobserved preferences. Variation in the hedonic price equilibrium across markets will be correlated with unobserved consumer preferences and can therefore no longer be used to isolate variation in Z_j from variation in $X_i{}^U$. The vertical distance between consumer B and consumer E's marginal bid function in figure 5 depicts this bias. In this particular example the slope parameter estimate for attribute j will be positively biased due to positive correlation between unobserved preferences for attribute j and the quantity of attribute j consumed (i.e. $Corr(X^U, Z_j) > 0$) and the positive stratification of preferences across markets (i.e. $X_B{}^U > X_E{}^U$).

Subsequently, our paper extends the multi-market approach by providing unbiased estimates of marginal bid functions using data across multiple markets and exogenous instruments developed from processes unique to our public good of interest. Our approach avoids issues prevalent within the structural, single market literature (Chattopadhyay 1999; Bajari and Benkard 2005; Bishop and Timmins 2008) and the reduced-form, multi-market literature (Bartik 1987; Zhang et al. 2016) by not having to explicitly specify a full utility functional form for all attributes of the differentiated product while still being able to derive an unbiased estimate of the marginal bid function for a single public good. We demonstrate this advancement by recovering marginal bid functions for water quality using remote-sensing harmful algal bloom data and housing transactions located in multiple markets along the Lake Erie shoreline.

III. Application to Lake Erie

Lake Erie is one of the most valuable natural resources located within the United States. It provides drinking water for millions of people each year, supports a $10.7 billion annual tourist industry (Great Lakes Commission 2014) and is diverted daily to help generate power, grow crops and manufacture goods (NOAA 2016). However many of these services are under threat due to persistent harmful algal blooms (HABs) that impair water quality through their production of harmful toxins. In 2014, for example, over 500,000 Toledo, OH residents were unable to drink their tap water due to heightened levels of *microcystin*, a freshwater toxin, produced from a nearby HAB.

In addition the toxic algal blooms experienced during the “Toledo Water Crisis” and in Lake Erie’s western basin are predicted to worsen in duration and frequency due to rising summer temperatures and increased agricultural runoff (Robson and Hamilton 2003; Mooij et al. 2005). A number of management strategies have been proposed in response to these expected conditions which aim to limit the amount of agricultural loadings occurring within Lake Erie’s watershed. Some of the more favored policy suggestions include the use of buffer strips (Scavia et al. 2016), a tax on fertilizer (Sohngen et al. 2015) and shorter, more diverse crop rotation cycles (Smith et al. 2015). Little is known, however, in regards to the benefits accrued from taking such actions. Our study fills in this gap by empirically estimating a water quality marginal bid function for near Lake Erie homeowners, which allows for unbiased estimation of welfare changes resulting from non-marginal changes in water quality.

Housing, water quality, recreational and demographics data have been collected to empirically estimate a marginal bid function for water quality. Housing transactions data was assembled from county auditor and tax offices covering the 7 counties and subsequently augmented with lender information (i.e. loan amount, lender name, owner name, etc.) using data purchased from CoreLogic. Records from this combined dataset include historic sales information, a geolocation and select structural characteristics for each property sold between July 2002 and December

2015. We further cleaned our sample by removing houses that sold for less than $50,000, were sold more than once over a 12 month span, were delinquent or vacant, or had extreme structural characteristics[9]; this was done to remove potential sources of omitted variable bias that are associated with these non-standard housing transactions.

Additional spatial characteristics were added to each housing transaction using its geolocation provided by the county parcel shapefiles. Tract and blockgroup IDs were attached to each property using shapefiles collected from the Census, while a continuous measure of Lake Erie proximity was calculated using a hydrological shapefile provided by the USGS's National Hydrography Dataset. Two additional discrete, mutually-exclusive lake proximity measures were also constructed from our continuous distance measure: a dummy variable controlling for whether or not the property was adjacent to Lake Erie (i.e. within 20 meters of the lake) and an additional non-adjacent, near lake dummy variable was formed for houses located between 20 and 500 meters from the lake. Finally we merged individual attribute data (i.e. income, gender and race) obtained through the Home Mortgage Disclosure Act (HMDA) with our property transactions dataset using a set of variables common across both datasets. Approximately 64% of the individuals within the transactions dataset were matched with information from HMDA using the following set of the variables: transaction date, lender name, loan amount and the property's census tract, garnering a total matched sample of 140,708 housing transactions. A description of all the variables used in this study is provided in Table 1, while Table 2 displays summary statistics for the entire sample as well as at the commuting zone level for all of the property-related characteristics.

[9]Any observation with a covariate value in the 1st or 99th percentile is tagged as a potential outlier and removed from the sample.

Housing transactions were then categorized into distinct housing markets based on the time period and location it was sold in. Commuting zones were first used as a natural breaking point between markets given the distinct political, demographic and economic factors that form each region.[10] We further disaggregated our sample using two time periods identified by major shifts in the national and local housing market: a "boom" period between 2002 and 2007 and a "bust" period between 2008 and 2015.[11] A total of 8 housing markets are identified using this disaggregation strategy, with hedonic price functions expected to change not only across space but across time. Although housing markets are traditionally segmented based on spatial jurisdictions or political boundaries (Palmquist 1984; Goodman and Thibodeau 1998), a newer line of literature has recently defined preferentially-distinct communities across both time and space (Zabel and Kiel 2000; Banzhaf and Walsh 2008). This paper follows more closely with the latter by assuming differences across space and time shift the hedonic equilibrium.

In addition 10 day algal-composite data spanning 2002 - 2014 for all of Lake Erie have been acquired. These data were uniformly gridded into 1100 meter by 1100 meter squares using remote sensing data collected from the National Oceanic and Atmospheric Administration (NOAA 2015) and converted into a Cyanobacterial Index value by Wynne and Stumpf (2015). A snapshot of the time-varying algae data at two time periods (September 2011, 2012) highlights the heterogeneity in the water quality across space and time as shown in Figures 7 and 8. To attach these algal readings to housing transactions a number of temporal and spatial aggregates

[10]Commuting zone information was obtained from the USDA's Economic Research Service (ERS) division. According to the ERS, commuting zones were formed with the explicit intention to reflect local economies where people live and work (USDA ERS 2016)

[11]Figure 6 displays the national and Cleveland Case-Shiller Housing Price Index between July 2002 and December 2015. Changing the temporal breakpoint from Jan 2008 to Jan 2007 has no quantitative impact on our results.

were examined. The primary set of results discussed below, however, attach algal concentration levels at the household level using readings only taken from the three closest observed algal locations.[12] Given the measurements taken from these locations, mean annual algal reading were then estimated for every observation within the study. Only algal measurements taken within 12 months of the transaction sale month were used to create this estimate, with the current month of algal concentrations excluded from this process to remove the possibility that future algal readings were used to predict current housing values.

VI. Empirical Estimation

Estimation of the first-stage hedonic price function took the following form:

$$(3)\quad \ln \mathrm{P_{ilt}}^{\mathrm{m}} = \alpha_0{}^{m} + \alpha_1{}^{m}\mathrm{X_i} + \alpha_2{}^{m}\delta_{l} + \alpha_3{}^{m}\mathrm{M_t} + \alpha_4{}^{m}\mathrm{Y_t} + \alpha_5{}^{m}\mathrm{LakeAdj_i} +$$
$$\alpha_6{}^{m}\mathrm{NearLake_i} + \alpha_7{}^{m}\mathrm{Distancetolake_i} + \alpha_8{}^{m}\mathrm{NearLake_i} * \mathrm{Distancetolake_i} +$$
$$\alpha_9{}^{m}(\mathrm{NearLake_i} + \mathrm{LakeAdj_i}) * \log(\mathrm{Algae_{it}}) + \epsilon_{ilt}{}^{m}$$

Where the log of the price of house i sold in location l during time period t within market m is given by $\ln \mathrm{P_{ilt}}^{\mathrm{m}}$. Separate regressions were run for each housing market providing a total of 8 hedonic price function estimates. House-specific structural attributes $\mathrm{X_i}$ (i.e. square footage, number of bathrooms, lot size, etc.), spatial fixed effects δ_{l}, and month $\mathrm{M_t}$ and year $\mathrm{Y_t}$ dummies were all included within each of these regressions. Inclusion of both spatial and temporal fixed effects within equation (3) is an essential component of our analysis as it greatly reduces the potential for omitted variable bias by limiting identification of our key parameters of interest

[12]Our results are robust to using more spatially aggregated/disaggregated measures of algae.

($\alpha_0{}^m$, … , $\alpha_9{}^m$ for m =1, … , 8) to come from either variation across space, but within the spatial fixed effect, or across time.

Additional terms that control for proximity to lake and water quality were also added to (3). $\mathrm{LakeAdj_i}$, $\mathrm{NearLake_i}$, $\mathrm{Distancetolake_i}$ and $\mathrm{NearLake_i} * \mathrm{Distancetolake_i}$ control for any proximity effect that may exist for non-lake, near lake and lakeshore properties. Inclusion of the two discrete variables ($\mathrm{LakeAdj_i}; \mathrm{NearLake_i}$) controls for any premium a house gains from either being adjacent to the lake or within the lake community but not adjacent. Holding all else equal, we expect lake adjacent homes to be valued higher than near lake homes (i.e. $\alpha_5{}^m > \alpha_6{}^m$ for m = 1, … , 8) given previous findings within the literature which show significant variation in the capitalization of local amenities across space (Abbott and Klaiber 2013). We further include two lake proximity covariates which vary continuously across space. The first term, $\mathrm{NearLake_i} * \mathrm{Distancetolake_i}$, allows the lake proximity effect to vary within the NearLake distance band and is expected to negatively influence a house's value (i.e $\alpha_8{}^m < 0$ for m = 1, ... , 8). The premium associated with living within a lake community is lessened, in other words, as you move further from the lake yet remain within the lake community. The second continuous measure, $\mathrm{Distancetolake_i}$, captures any additional premium that is associated with living close to a lake, but not within the lake community.

The primary variable of interest, however, is the logged algae variable interacted with the sum of the mutually exclusive distance bands: $(\mathrm{NearLake_i} + \mathrm{LakeAdj}) * \log(\mathrm{Algae_{it}})$. By interacting logged algae with this term we limit algae's impact on property values to occur only within the pre-specified near lake distance band. Estimates from Wolf and Klaiber (2017) suggest this band should be set somewhere between 500 - 600 meters; we therefore follow this recommendation and set our distance dummy equal to 500 meters. Finally, a vector of idiosyncratic and independently distributed error terms ($\epsilon_{\mathrm{ilt}}{}^{m}$) were also included within (5) to capture any remaining unobserved variation in housing values.

Following the estimation of eight separate hedonic regressions, the implicit price for algae is then recovered by taking the derivative of the hedonic price equilibrium with respect to algae. Given the functional form assumed in (3), derivation of these prices takes the following form:

$$(4)\ \frac{\partial \widehat{\mathrm{P_{ilt}}}}{\partial \mathrm{algae}}^{m} = \widehat{\alpha_9}^{m} * \mathrm{P_{ilt}}\left(\frac{\mathrm{NearLake_i} + \mathrm{LakeAdj_i}}{\mathrm{Algae_{it}}}\right)$$

The attribute prices recovered from (4) are useful by themselves since they can be used to determine consumer's marginal willingness to pay for public goods such as water quality. However, as previously mentioned, these attribute prices are only valid near the consumer's optimal consumption decision. Large changes in public goods, therefore, cannot be accurately evaluated unless an entire marginal bid function is recovered. Consequently we estimate a marginal bid function for water quality by using the attribute prices recovered from (4) as the dependent variable within Rosen (1974)'s second stage regression. We assume the consumers' marginal bid function for water quality follows a Cobb-Douglas specification:

$$(5)\ \log\left(\frac{\partial \widehat{\mathrm{P_{ilt}}}}{\partial \mathrm{algae}}^{m}\right) = \gamma_0 + \gamma_1 \log(\mathrm{Algae}_{it}) + \gamma_2 \log(\mathrm{W_i}) + \gamma_3 \log(\mathrm{Demographics}_i) + \gamma_4 \log(\mathrm{CompositeGood}_i) + \upsilon_{ilt}$$

Where Algae_{it} is the same vector of algal measurements used in the first stage regression, W_i is a vector of product attributes other than algae (i.e. square footage, distance to Lake Erie, parcel lot size, etc.) ,[13] $\mathrm{Demographics}_i$ is a vector observable consumer characteristics that are expected to

[13]0-1 indicator variable quantities such as Fireplace, Garage, etc. are not log transformed within equation (5)'s specification.

influence demand for water quality, CompositeGood_i is the amount of numeraire consumed,[14] υ_{ijt} is an error term and γ are a vector of structural utility parameters. OLS estimation of (5) would result in biased estimation of our primary variable of interest, γ_1, since unobserved preferences for water quality will be correlated both with how much algae the consumer accepts at his residential location and the implicit price he pays for water quality. Consistent estimation of (5) therefore requires a valid and exogenous instrumental variable for Algae_{it}.

Exogenous Instrumental Variable Strategy

Our proposed instrumental variable is derived from hydrological and ecological processes that directly affect the density and frequency of harmful algal blooms (HABs) on Lake Erie. In particular we use a measure of discharge flow rates from the Maumee watershed, one of the prominent Lake Erie tributaries, as a proxy for algal concentrations. Discharge flow rates are predicted to be positively correlated with algae, but uncorrelated with unobserved consumer preferences for water quality. Water flowing out of the Maumee watershed and into Lake Erie's western basin often contains high concentrations of agricultural runoff collected from nearby fields.[15] Consequently when heavy rainfall events occur in the spring, recently applied fertilizer will be transported from the farmers' fields to Lake Erie's shallow western basin via the Maumee River. Faster and heavier outflow from the Maumee River will therefore increase the rate at which these nutrients are disseminated across the. Subsequently, as the water column becomes

[14] CompositeGood_i is formed by subtracting household i's income from their annualize home price. Annualized home values are estimated by taking its market price and multiplying it by 11% (Poterba 1984).

[15] Approximately 80% of the land within the Maumee watershed is used for agricultural purposes (Richards et al. 2002).

inundated with higher concentrations of Phosphorous and Nitrogen, co-limited[16] blooms of *Microcystis* form (Michalak et al. 2013; Zhou et al. 2013; Stumpf et al. 2012).

Given this finding uncovered from the hydrological and biological literature, we adjust our second stage hedonic approach by first instrumenting for the endogenous algae variable using an aggregate measure of discharge flow rates:

$$(6)\ \log\left(\widehat{\mathrm{Algae}}_{it}\right) = \xi_0 + \xi_1 \mathrm{Discharge_t} + \xi_2 \log(\mathrm{W_i}) + \xi_4 \log(\mathrm{Demographics_i}) + \xi_5 \log(\mathrm{CompositeGood}_i) + \mu_{\mathrm{ilt}}$$

$$(7)\ \log\left(\frac{\partial \widehat{\mathrm{P_{ilt}}}}{\partial \mathrm{algae}}^{m}\right) = \gamma_0 + \gamma_1 \log(\widehat{\mathrm{Algae}}_{it}) + \gamma_2 \log(\mathrm{W_i}) + \gamma_3 \log(\mathrm{Demographics}_i) + \gamma_4 \log(\mathrm{CompositeGood}_i) + \upsilon_{ilt}$$

To obtain a measure of discharge flow rates, we collected data from the United States Geologic Survey (USGS)'s monitoring station located on the Maumee River in Waterville, Ohio (USGS 2016). Water flow rates are recorded daily and then aggregated by the USGS into monthly averages. Using this monthly data, we construct an annual maximum spring time flow rate for every time period within our study sample.[17] Only discharge measures taken between March and June are taken into consideration, however, as they are known to have the strongest impact on HAB intensity.[18]

A pair of dummy variables indicating whether or not the homeowner owns a boat or fishing license are also included within the vector of demographics in equation (7). Boat license data for

[16]Most cyanobacteria are considered nitrogen-fixing and phosphorous-limited, which indicates that P is the key nutrient limiting HAB growth. The species of *Microcystis* found on Lake Erie, however, is believed to be co-limited by both P and N (Paerl et al. 2011).
[17]As indicated by the sole time subscript on $\mathrm{Discharge_t}$, this measure does not vary across space. All observations occurring within the same time period, in other words, have the same aggregate discharge value attached to it.
[18] This aggregate estimate follows closely with the measures used in Stumptf et al. (2012)'s HAB forecasting model. In particular Stumptf et al. (2012) create yearly discharge rates using only readings taken between March and June.

permits sold between 2009 – 2015 and fishing license data for permits sold between 2011 – 2015 were collected from the Ohio Department of Natural Resources. Both of these statewide datasets contain address information which allowed us to geographically reference the boating and fishing license datasets to the housing transactions data using Google Earth. Households that purchased a boat or fishing permit after their sale date, but before the property's next sale date, if there was one, were labeled as boat owners or fishermen. A description of all the variables used in equations (6) and (7) are included at the bottom of Table 1. In addition Table 3 presents a set of summary statistics for these variables.

Converting Inverse Ordinary Demand Curves into Inverse Compensated Demand Curves

The marginal bid function derived from equations (6) and (7) is equivalent to a consumer's inverse uncompensated (Marshallian) demand curve for water quality. Compensated (Hicksian) inverse demand functions are preferred, however, when evaluating the welfare implications of a quantity change as they hold consumer utility constant. Following the methods discussed in Hausman (1981) and Palmquist (2005), we convert our fitted inverse Marshallian demand curve into a Hicksian demand curve by substituting the expenditure function in for income. Within our study's current setting this substitution would take the following form:

$$(8)\ \log\left(\frac{\partial\hat{\theta}_{\mathrm{i}}}{\partial\mathrm{Algae}}\right) = \hat{\gamma}_0 + \hat{\gamma}_1\log(\mathrm{Algae}_{it}) + \hat{\gamma}_2\log(\mathrm{W_i}) + \hat{\gamma}_3\log(\mathrm{Demographics}_i) + \hat{\gamma}_4\log(\mathrm{m}_i - \hat{\theta}_i)$$

Where m_i is equal to annual income for household i and $\frac{\partial\hat{\theta}_{\mathrm{i}}}{\partial\mathrm{Algae}}$ is the derivative of the consumer i's bid function with respect to algae. The first substitution, $\frac{\partial\widehat{\theta_{\mathrm{i}}}}{\partial\mathrm{Algae}} = \frac{\partial\widehat{\mathrm{P_{ilt}}}}{\partial\mathrm{algae}}^{m}$, follows from the tangency condition outlined in Rosen (1974)'s first stage, while the second substitution, $\mathrm{m}_i -$

$\hat{\theta}_i = \text{CompositeGood}_i$, is simply the definition of our composite good. Taking the exponent of the left and right-hand side produces a standard ordinary differential equation:

$$(9)\ \frac{\partial \hat{\theta}_i}{\partial \text{Algae}} = e^{\hat{\gamma}_0} \text{Algae}_{it}{}^{\hat{\gamma}_1}\ W_i{}^{\hat{\gamma}_2} \text{Demographics}_i{}^{\hat{\gamma}_3} (m_i - \hat{\theta}_i)^{\hat{\gamma}_4}$$

Solving for $\hat{\theta}_i$ then yields equation (10),

$$(10)\ \hat{\theta}_i = m_i - \left[-(1-\hat{\gamma}_4)\frac{e^{\hat{\gamma}_0}\text{Algae}_{it}{}^{\hat{\gamma}_1+1}\ W_i{}^{\hat{\gamma}_2}\text{Demographics}_i{}^{\hat{\gamma}_3}}{\hat{\gamma}_1+1} + C\right]^{1/(1-\hat{\gamma}_4)}$$

Where C is the constant of integration. As suggested by Palmquist (2005), we invert (10), solving for C, and set it equal to our constant level of utility:

$$(11)\ C = u = \left(m_i - \hat{\theta}_i\right)^{1-\hat{\gamma}_4} + (1-\hat{\gamma}_4)\frac{e^{\hat{\gamma}_0}\text{Algae}_{it}{}^{\hat{\gamma}_1+1}\ W_i{}^{\hat{\gamma}_2}\text{Demographics}_i{}^{\hat{\gamma}_3}}{\hat{\gamma}_1+1}$$

The equation for compensating variation resulting from a quantity change in algae can then be calculated using the information provided in (10) and (11):

$$(12)\ CV_{Algae_i} = \hat{\theta}_i(\text{Algae}_N, u_0) - \hat{\theta}_i(\text{Algae}_0, u_0)$$

$$(13)\ CV_{Algae_i} = m_i - \left[-\left(\frac{1-\hat{\gamma}_4}{\hat{\gamma}_1+1}\right)\left(e^{\hat{\gamma}_0}\text{Algae}_N{}^{\hat{\gamma}_1+1}\ W_i{}^{\hat{\gamma}_2}\text{Demographics}_i{}^{\hat{\gamma}_3} - e^{\hat{\gamma}_0}\text{Algae}_0{}^{\hat{\gamma}_1+1}\ W_i{}^{\hat{\gamma}_2}\text{Demographics}_i{}^{\hat{\gamma}_3}\right) + (m_i - P_{ilt})^{1-\hat{\gamma}_4}\right]^{\frac{1}{1-\hat{\gamma}_4}} - P_{ilt}$$

Where Algae_0 indicates the original level of algae observed by consumer i during time period t (i.e. $\text{Algae}_0 = \text{Algae}_{it}$) and Algae_N is the new level of algae after an improvement/degradation in water quality conditions occurs. In addition we know the value of consumer i's bid function is equal to the price they initially paid for their house; this allows us to substitute P_{ilt} for $\hat{\theta}_{\mathrm{i}}(\mathrm{Algae}_0, \mathrm{u}_0)$ in equations (11) and (12).

IV. Results

Figure 9 displays the first-stage hedonic point estimates for $\alpha_9{}^m$ and their corresponding 95% confidence intervals. 7 of the 8 markets have negatively signed values for $\alpha_9{}^m$, which matches our expectations and previous findings within the literature which show a robust negative relationship between housing/land values and various measures of water impairment (Leggett and Bockstael 2000; Poor et al. 2001; Walsh et al. 2011). As indicated by the varying confidence intervals in Figure 9, the rate at which water conditions are capitalized into near lake housing values varies significantly across markets, with near lake homes located within the "Erie Bust" market, for example, losing approximately 7.3% of its value when algal concentrations increase by 1% as compared to the 1.5% drop experienced in the "Toledo Bust" market. The relationship between housing values and structural attributes[19] across all markets matches expectations: housing values increasing at a decreasing rate as square footage and parcel lot size increase. In addition property values increase when improvements are made to the house (i.e. adding a fireplace or a garage) but lose value as the building ages. Lakeshore and near lake homes are also valued more because of their proximity to Lake Erie. Lake adjacent homes are valued approximately 80%[20] more than non-adjacent homes, while homes within 500 meters, but non-adjacent, obtain a premium of around 35%. This proximity effect appears to be spatially limited

[19]Full first-stage hedonic results are available in Table A1.
[20] Dummy variable estimates presented in the text have been corrected using the technique suggested by (Halvorsen and Palmquist 1980).

to 500 meters, however, with non-lake houses gaining no additional value the closer they are located to the lake.

We further test the robustness of our first-stage results by changing the functional form assumption made on the hedonic price function. In our first approach we specify a log-log hedonic price function which allows changes in algae to non-linearly affect the market value of a home. However, due to the concerns presented in Kuminoff et al. (2010) and Cropper et al. (1988), we specify two additional models which vary how algae and housing values are specified within the first stage regression. Results from these modified regressions are displayed in Table A2 and are qualitatively similar to our initial findings. Omitted variable bias is also a major concern in the hedonics literature (Abbott and Klaiber 2011; Kuminoff and Pope 2014) given the number of unknown factors that can affect a home's value. One way to mitigate these concerns is through the use of spatial and temporal fixed effects. We test how various combinations of spatial and temporal fixed effects affect our results and present them in Table A3. The sign and magnitude on most of the results remain unchanged when moving from Table A1 to Table A3; however some significance is lost when weaker fixed effects are implemented, which further highlights the importance of using more stringent fixed effects.

Demand slope estimates along with important preference shifters recovered from the second-stage regression are displayed in Table 4. Only housing transactions located within 500 meters of Lake Erie that had a correctly signed implicit price estimate[21] were used to estimate the second-stage hedonic regression. Observations located further than 500 meters from Lake Erie were

[21]Given that our public good of interest is a bad rather than a good, we convert our original marginal damage estimates into implicit water quality prices by multiplying the vector of $\alpha_9{}^m$ estimates by -1. Observations with incorrectly signed implicit price estimates were dropped (i.e. observations with negative implicit price values) due to our Cobb-Douglas functional form specification.

always given an implicit price value of 0 due to the functional form restrictions implemented within the first-stage regression and therefore had to be dropped from the analysis. Using this subsample of 4,553 households three separate marginal bid functions were estimated using OLS and two-stage least squares.

The first column in Table 4 presents results from an OLS regression, while the final two columns use multi-market dummies and discharge flow rates respectively as instrumental variables for the endogenous algae covariate. Estimates on the slope parameter across all three columns have a magnitude greater than 1, indicating that water quality is an elastic good. Both the OLS and multi-market estimate are similar in magnitude and significance, however, which suggests the market dummies are not an acceptable exogenous instrumental variable within the given setting. Results from a Wu-Hausman test (Wu 1973; Hausman 1978) of endogeneity further collaborate this finding. In particular, we are unable to reject the null hypothesis that $\log(\text{Algae}_{it})$ is an exogenous variable ($F(1,95) = 0.0233$; $p = 0.879$) using the recovered parameter values from column 2. However, we are able to reject this hypothesis at the 10% level using values from our preferred model ($F(1,95) = 3.316$; $p = 0.0717$).

We further test the robustness of our finding by changing the functional form assumption made in (7). The second panel in Table 4 presents results using a semi-log functional specification. Similar to the top panel the instrumented slope estimate in column 3 is much larger in magnitude than those recovered from OLS and the multi-market approach. The market instrument appears to perform better under this scenario however, successfully rejecting the null hypothesis that Algae_{it} is an exogenous variable at the 1% level ($F(1,95) = 70.69$; $p<0.001$).[22] The tendency for

[22]The parameter value derived from the 3rd column is also able to reject the Wu-Hausman test at the 1% level ($F(1,95) = 17.01$; $p < 0.001$

both the OLS and market IV estimate to underestimate the true magnitude of the slope coefficient suggests that welfare estimates derived from these fitted marginal bid functions will be biased too. Under this particular setting, welfare predictions attributed to an improvement in water quality would underestimate homeowners' true gains, while predictions made under the opposite scenario would overestimate homeowners' losses.

In addition maximum, March – June discharge rates appear to be a relevant instrument for algae densities on Lake Erie obtaining a first-stage F-score of 48.81 within the Cobb-Douglas specification and 43 within the semi-log specification. As indicated by Table A4 discharge is also positively correlated with algae and the log transformation of algae, matching predictions uncovered from the biology and hydrology literature. Coefficients from the other covariates within the marginal bid function are for the most part insignificant. Boat owners and fishermen, for example, do not appear to have stronger preferences for water quality than non-boat owners and non-fishermen. This non-result is likely due to the lack of boating and fishing license data availability between 2002 and 2009 which may be biasing our estimates towards 0. Finally, we find a positive and robust relationship between the implicit price people pay for water quality and the amount of the composite good they consume. Parameter estimates for γ_4 are positive and significant in all but one of our specifications, indicating that water quality is a normal good.

Differences in Welfare Estimates

Although theoretically it's well understood why we can't use first-stage MWTP estimates to evaluate large changes in non-market goods, there are only a handful of studies that analyze the difference in welfare estimates using both MWTP point and marginal bid function estimates. Starting from the mean algae reading for the 4,453 households within 500 meters of Lake Erie

(9.48) we plot the difference in welfare estimates, in Figure 10, across these two approaches using the preferred Cobb-Douglas marginal bid function recovered in Table 4 and the average of the MWTP estimates recovered from (4). The solid black line in Figure 10 corresponds to the absolute difference in welfare estimates between those predicted under a second-stage setting and those derived using MWTP point estimates, while the dotted gray line shows the percentage difference in welfare estimates.[23] Welfare estimates derived from small changes in water quality are equivalent across both approaches; however the further we move from the hedonic equilibrium the more inaccurate the MWTP point estimates becomes. For example, a 40 reduction in mean algae readings would cause welfare estimates derived from first-stage hedonics to be off by more than 11%. Within the current setting this would equate to an absolute difference in welfare estimates of approximately $2,500 per household per year! Given these findings, we suggest to proceed with caution when attempting to provide non-marginal, policy-relevant estimates using first-stage hedonic estimates.

V. Discussion and Conclusion

A number of steps have been taken by policymakers to reduce the frequency and intensity of HABs on Lake Erie. During the 1960s and 1970s large algal blooms formed across the lake due to industrial, septic and agricultural runoff. The Great Lakes Water Quality Agreement (GLWQA) was formed jointly by the United States and Canada in response to these conditions in 1972 which called for the reduction of phosphorous runoff into the lake. Water conditions dramatically improved over the course of several years after the agreement had been reached and the ecosystem appeared to recover from its previous eutrophic state. However these improved

[23] $\frac{\Delta \text{Welfare}_{\text{Function}} - \Delta \text{Welfare}_{\text{Point}}}{\text{Welfare}_0}$, where Welfare_0 is equal to total consumer surplus at the initial algae level.

conditions did not last long; by the late 1990s annual harmful algal blooms had re-emerged due to the introduction of Zebra Mussels in the late 1980s (Vanderploeg et al. 2001) and the rise of soluble reactive phosphorous loadings in the 1990s (Joosse and Baker 2011).

Policymakers have attempted to curb these worsening conditions by passing amendments to the original GLWQA in 1983, 1987 and 2012. The most recent version calls for a 40% reduction in spring phosphorous loadings to the western basin from both the United States and Canada (GLWQA 2016). Despite this call for action little is still known about the economic benefits gained from a reduction in algae. Our paper attempts to partially fill in this gap by providing benefit estimates for near lake homeowners using our derived MWTP function. Estimates from Stumpf et al. (2012)'s HAB forecasting model[24] predict a 40% reduction in June total phosphorous loadings would correspond to approximately a 32.3% reduction in algae. Under these new conditions near lake homeowners are expected to gain, on average, $3,215 in annualized benefits per household. This estimate is more than double what would be predicted using MWTP estimates ($1465)! Aggregating this estimate across all Ohio, single family residents within 500 meters of Lake Erie produces an annual benefit of $136 million.[25] These

[24]Stumpf et al. (2012)'s forecasting model uses either June total phosphorous loadings or March – June Maumee watershed discharge rates to predict future HABs. Given the GLWQA's focus on reducing phosphorous runoff, we estimate future conditions using total phosphorous data provided by Stumpf et al. (2012).

[25]A total of 33,305 single family houses are located within 500 meters of Lake Erie in our 6 county study area. We include all of these homes within our aggregated benefit calculation due to the rationale provided by McCluskey and Rausser 2003. McCluskey and Rausser (2003) claim that homeowners incur additional losses/benefits when nearby environmental quality changes, even if the affected property is never sold on the market. Assuming properties are similar to capital assets, a change in liquidity, produced by a change in environmental quality, should be offset by higher or lower rates of return. However, under our study's setting unsold property owners do not experience a drop in return rates after water quality improves despite the fact that their property becomes more liquid. Property owners who hold onto their properties, even after Lake Erie water quality improves, therefore accrue additional benefits by not being penalized for having a more liquid asset.

large gains help justify the already stringent land management practices and fertilizer restrictions implemented by both the United States and Canadian government.

Our study's findings show that precise marginal bid function estimates can be recovered when exogenous, second-stage hedonic instruments are available. Our slope estimates are more negative than those predicted under an OLS and multi-market setting, suggesting that welfare estimates derived from these parameters would underestimate the true gains attributed to an improvement in water quality and would overestimate the losses associated with a degradation in water quality. In addition we find large differences in welfare estimates derived from first stage and second stage hedonic estimates; MWTP point estimates, for example, undervalue a 40% reduction in algae by approximately $2,500 per household per year. This large discrepancy further highlights the need to recover entire MWTP functions when evaluating large scale changes in non-market goods. Finally, estimates from our paper were used to predict future benefits recovered under the 2012 GLWQA. Near lake homeowners will receive, on average, an additional $3,215 in annualized benefits when total phosphorous loadings are reduced by 40%. Pairing this estimate with existing agricultural loss estimates caused by more stringent land management practices will allow policymakers to better understand the implicit tradeoffs that are made between higher agricultural yields and improved water conditions.

VI. Works Cited

Abbott, J. K., & Klaiber, H. A. (2011). An embarrassment of riches: Confronting omitted variable bias and multi-scale capitalization in hedonic price models. *Review of Economics and Statistics*, *93*(4), 1331-1342.

Bajari, P., & Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of political economy*, *113*(6), 1239-1276.

Banzhaf, S. H., & Walsh, R. P. (2008). Do people vote with their feet? An empirical test of Tiebout's mechanism. *The American Economic Review*, *98*(3), 843-863.

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models. *Journal of political Economy*, *95*(1), 81-88.

Bishop, K., & Timmins, C. (2008). Simple, consistent estimation of the marginal willingness to pay function: recovering Rosen's second stage without instrumental variables. *Duke University Department of Economics Working Paper*.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*(2), 577-599.

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.

Brown, J. N., & Rosen, H. S. (1982). On the Estimation of Structural Hedonic Price Models. *Econometrica (pre-1986)*, *50*(3), 765.

Chattopadhyay, S. (1999). Estimating the demand for air quality: new evidence based on the Chicago housing market. *Land Economics*, 22-38.

Chay, K. Y., & Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of political Economy*, *113*(2), 376-424.

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the choice of funtional form for hedonic price functions. *The Review of Economics and Statistics*, 668-675.

Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of political economy*, *112*(S1), S60-S109.

Epple, D. (1987). Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. *Journal of political economy*, *95*(1), 59-80.

Epple, D., & Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of political economy*, *107*(4), 645-681.

GLWQA 2016. The United States and Canada adopt phosphorus load reduction targets to combat Lake Erie algal blooms. Available at: https://www.epa.gov/glwqa/recommended-binational-phosphorus-targets.

Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of housing economics*, *7*(2), 121-143.

Greenstone, M., & Gallagher, J. (2005). *Does hazardous waste matter? Evidence from the housing market and the superfund program* (No. w11790). National Bureau of Economic Research.

Halvorsen, R., & Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American economic review*, *70*(3), 474-75.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251-1271.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *The American Economic Review*, *71*(4), 662-676.

Joosse, P. J., & Baker, D. B. (2011). Context for re-evaluating agricultural source phosphorus loadings to the Great Lakes. *Canadian Journal of Soil Science*, *91*(3), 317-327.

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.

Kuminoff, N. V., & Pope, J. C. (2014). Do “Capitalization Effects” for Public Goods Reveal the Public's Willingness to Pay?. *International Economic Review*, *55*(4), 1227-1250.

Leggett, C. G., & Bockstael, N. E. (2000). Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management*, *39*(2), 121-144.

McCluskey, J. J., & Rausser, G. C. (2003). Hazardous waste sites and housing appreciation rates. *Journal of Environmental Economics and Management*, *45*(2), 166-176.

McConnell, K. E., & Phipps, T. T. (1987). Identification of preference parameters in hedonic models: Consumer demands with nonlinear budgets. *Journal of Urban Economics*, *22*(1), 35-52.

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., ... & DePinto, J. V. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, *110*(16), 6448-6452.

Mooij, W. M., Hülsmann, S., Domis, L. N. D. S., Nolet, B. A., Bodelier, P. L., Boers, P. C., ... & Portielje, R. (2005). The impact of climate change on lakes in the Netherlands: a review. *Aquatic Ecology*, *39*(4), 381-400.

National Oceanic and Atmospheric Administration. (2015). *Experimental Lake Erie Harmful Algal Bloom Bulletin.* [Bulletin]. Ann Arbor: Rick Stumpf and Tim Wynne.

National Oceanic and Atmospheric Administration. (2016). About Our Great Lakes: Great Lakes Basin Facts. Retrieved from https://www.glerl.noaa.gov//pr/ourlakes/facts.html

Nevo, A., & Rosen, A. M. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, *94*(3), 659-671.

Paerl, H. W., Hall, N. S., & Calandrino, E. S. (2011). Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment*, *409*(10), 1739-1745.

Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. *The review of economics and statistics*, 394-404.

Palmquist, R. B. (2005). Property value models. *Handbook of environmental economics*, *2*, 763-819.

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.

Poterba, J. M. (1984). Tax subsidies to owner-occupied housing: an asset-market approach. *The quarterly journal of economics*, *99*(4), 729-752.

Reynolds, C. S. (2006). *The ecology of phytoplankton*. Cambridge University Press.

Richards, R. P., Calhoun, F. G., & Matisoff, G. (2002). The Lake Erie agricultural systems for environmental quality project. *Journal of Environmental Quality*, *31*(1), 6-16.

Robson, B. J., & Hamilton, D. P. (2003). Summer flow event induces a cyanobacterial bloom in a seasonal Western Australian estuary. *Marine and Freshwater Research*, *54*(2), 139-151.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, *82* (1), 34-55.

Scavia, D. O. N. A. L. D., Kalcic, M. A. R. G. A. R. E. T., Muenich, R. L., Aloysius, N., Arnold, J. E. F. F. R. E. Y., Boles, C. H. E. L. S. I. E., ... & Martin, J. (2016). Informing Lake Erie agriculture nutrient management via scenario evaluation. *University of Michigan: Ann Arbor, MI, USA.*

Smith, D. R., Francesconi, W., Livingston, S. J., & Huang, C. H. (2015). Phosphorus losses from monitored fields with conservation practices in the Lake Erie Basin, USA. *Ambio*, *44*(2), 319-331.

Sohngen, B., King, K. W., Howard, G., Newton, J., & Forster, D. L. (2015). Nutrient prices and concentrations in Midwestern agricultural watersheds. *Ecological Economics*, *112*, 141-149.

Stumpf, R. P., Wynne, T. T., Baker, D. B., & Fahnenstiel, G. L. (2012). Interannual variability of cyanobacterial blooms in Lake Erie. *PloS one*, *7*(8), e42444.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, 416-424.
Witte, A. D., Sumka, H. J., & Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets. *Econometrica: Journal of the Econometric Society*, 1151-1173.

USDA ERS. (2016). Commuting Zones and Labor Market Areas. Retrieved November 10, 2016, from https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/

USGS (2016). http://waterwatch.usgs.gov/. Accessed: 2016 Dec 13.

Vanderploeg, H. A., Liebig, J. R., Carmichael, W. W., Agy, M. A., Johengen, T. H., Fahnenstiel, G. L., & Nalepa, T. F. (2001). Zebra mussel (Dreissena polymorpha) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(6), 1208-1221.

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.

Wolf, D., & Klaiber, H. A. (2017). Bloom and bust: Toxic algae's impact on nearby property values. *Ecological Economics*, *135*, 209-221.

Wu, D. M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica: journal of the Econometric Society*, 733-750.

Wynne, T. T., & Stumpf, R. P. (2015). Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western lake erie from 2002–2014. *Toxins*, *7*(5), 1649-1663.

Zabel, J. E., & Kiel, K. A. (2000). Estimating the demand for air quality in four US cities. *Land Economics*, 174-194.

Zhang, C., Boyle, K. J., & Kuminoff, N. V. (2015). Partial identification of amenity demand functions. *Journal of Environmental Economics and Management*, *71*, 180-197.

Zhou, Y., Obenour, D. R., Scavia, D., Johengen, T. H., & Michalak, A. M. (2013). Spatial and temporal trends in Lake Erie hypoxia, 1987–2007. *Environmental science & technology*, *47*(2), 899-905.

Figure 1 – Hedonic Equilibrium

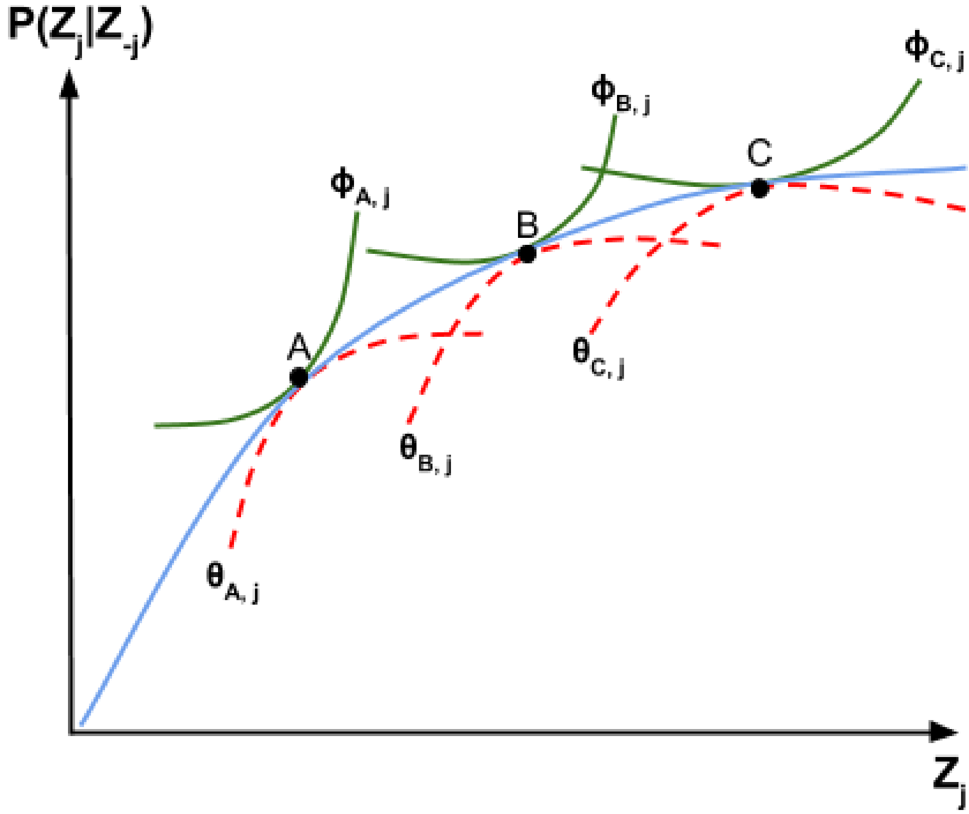


Figure 2 – Hedonic Equilibrium in Derivative Space

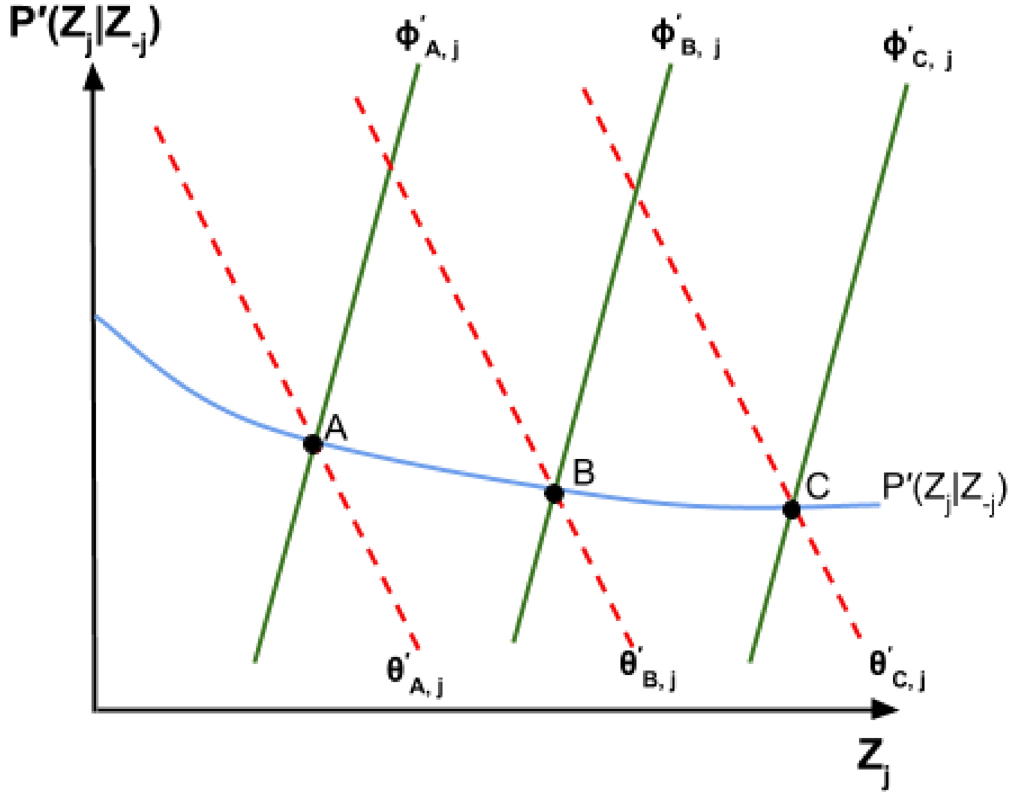


Figure 3 – Rosen's Second Stage Estimation

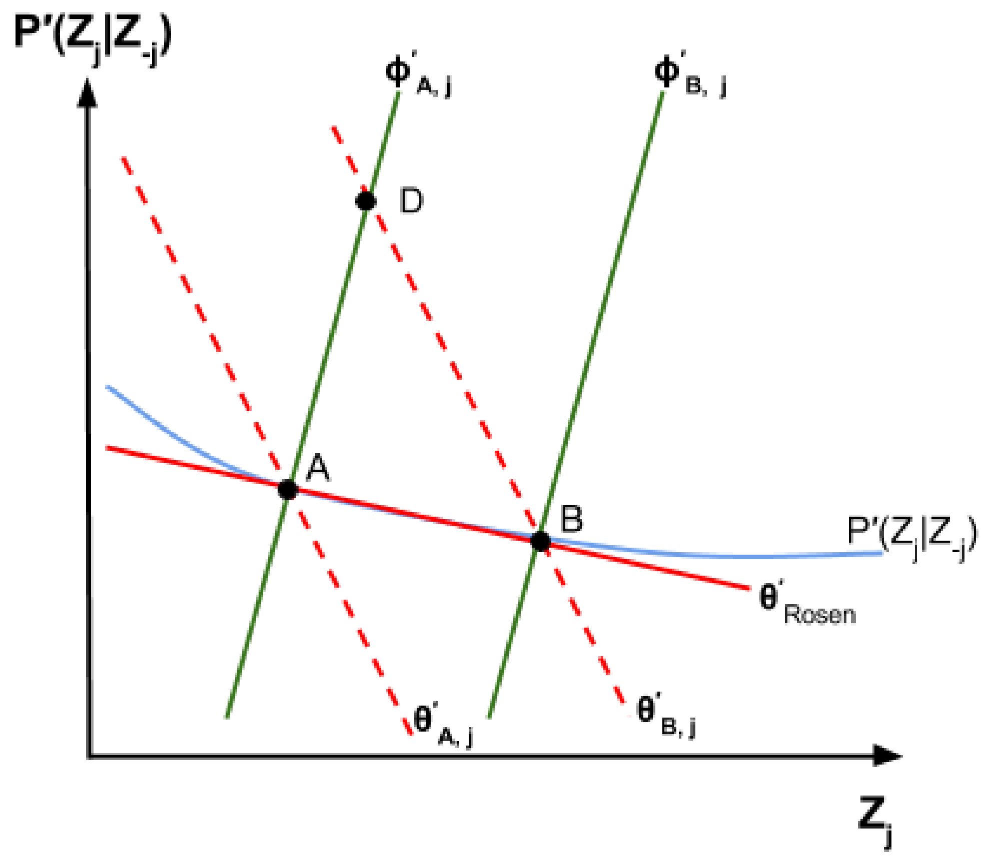


Figure 4 – Intuition Behind a Multi-market Approach

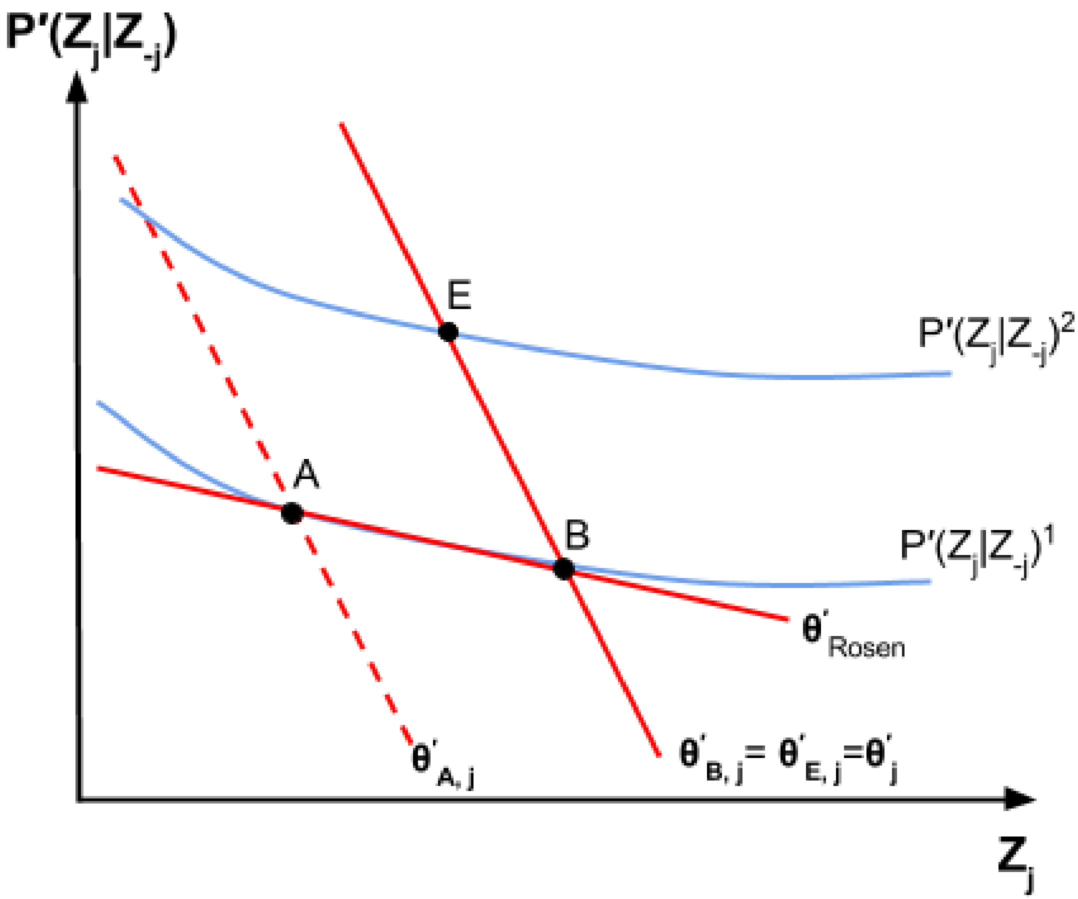


Figure 5-Bias Produced under a Multi-Market Hedonic Setting

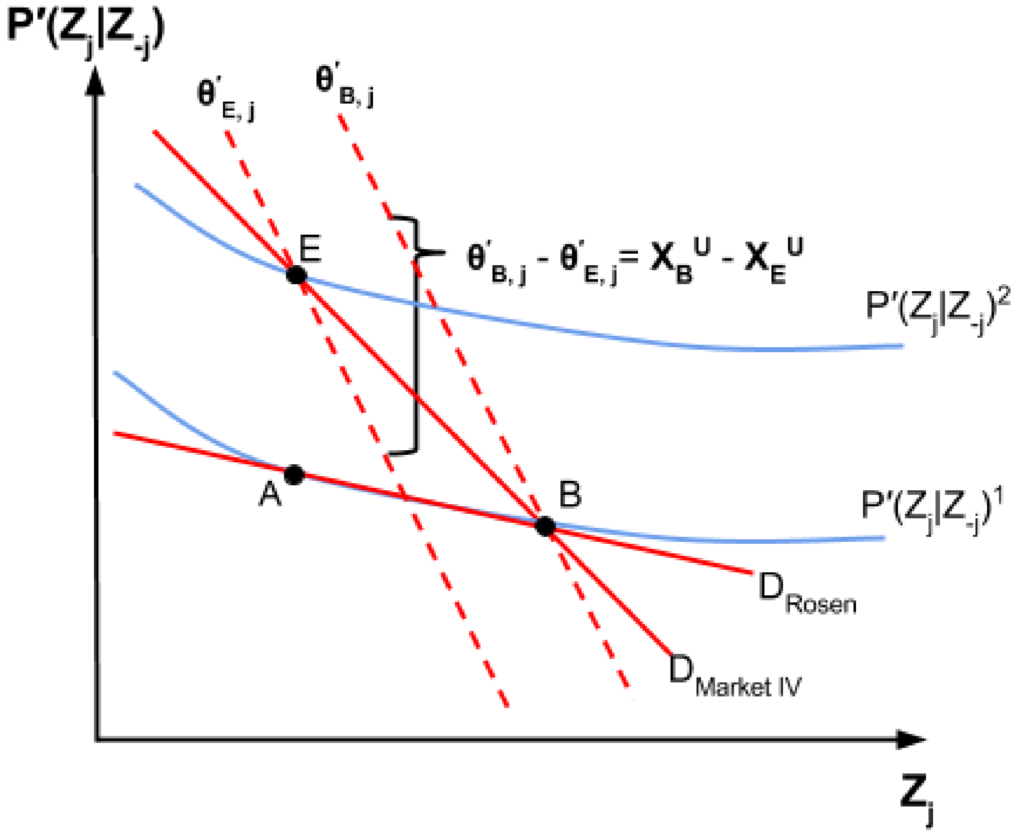
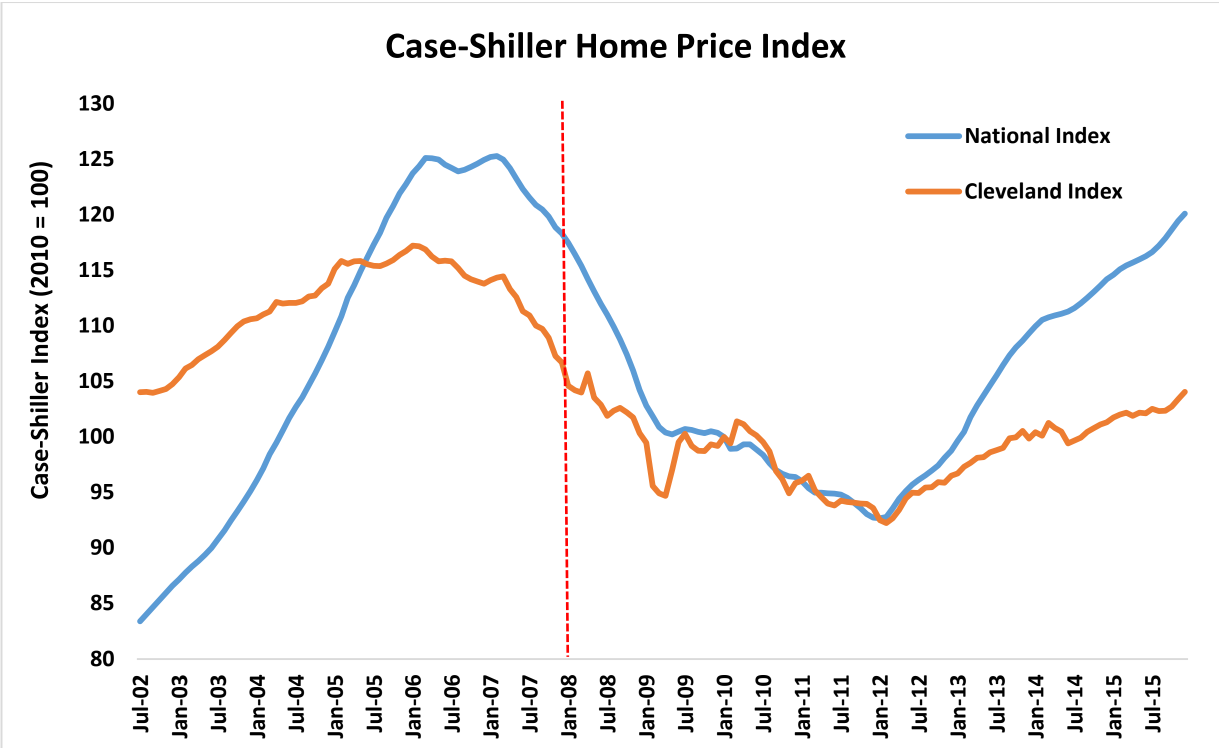
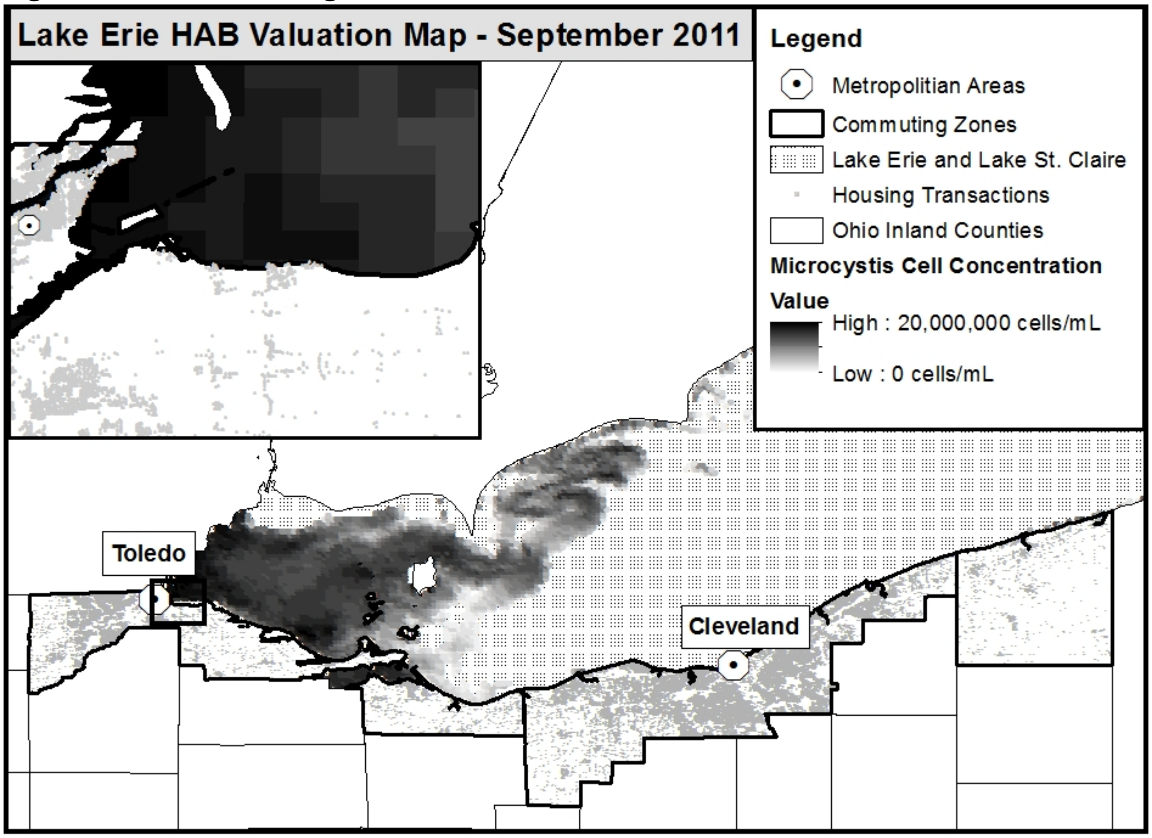


Figure 6 - Case-Shiller Home Price Index



Figures 7 and 8 – Housing Transactions and HAB Data



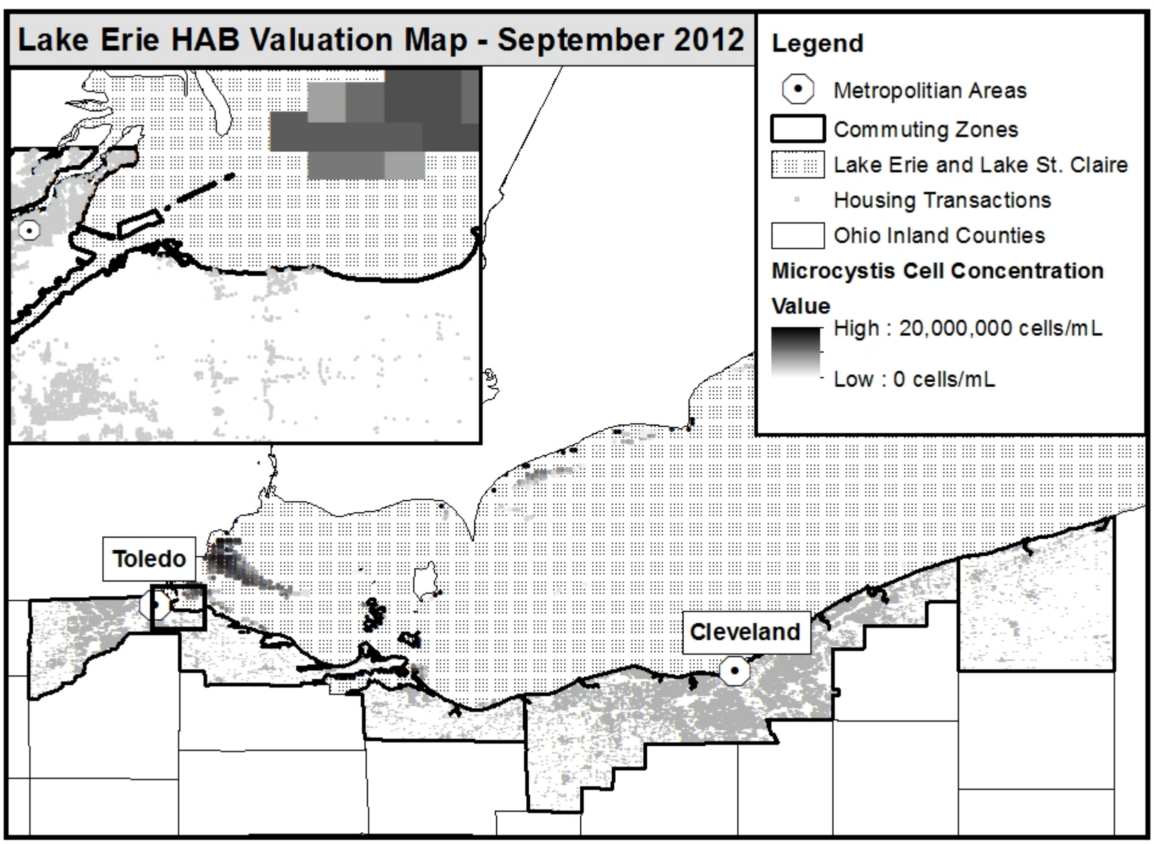


Figure 9: First-Stage Coefficients

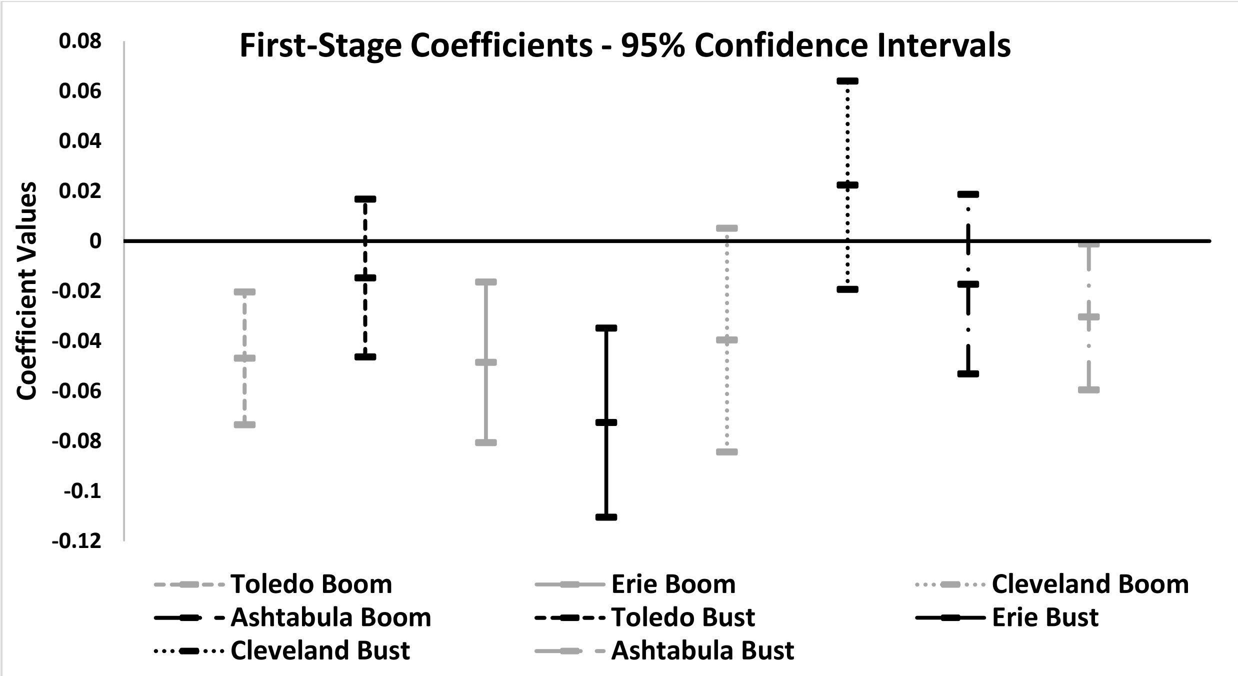


Figure 10 – Difference in Welfare Estimates

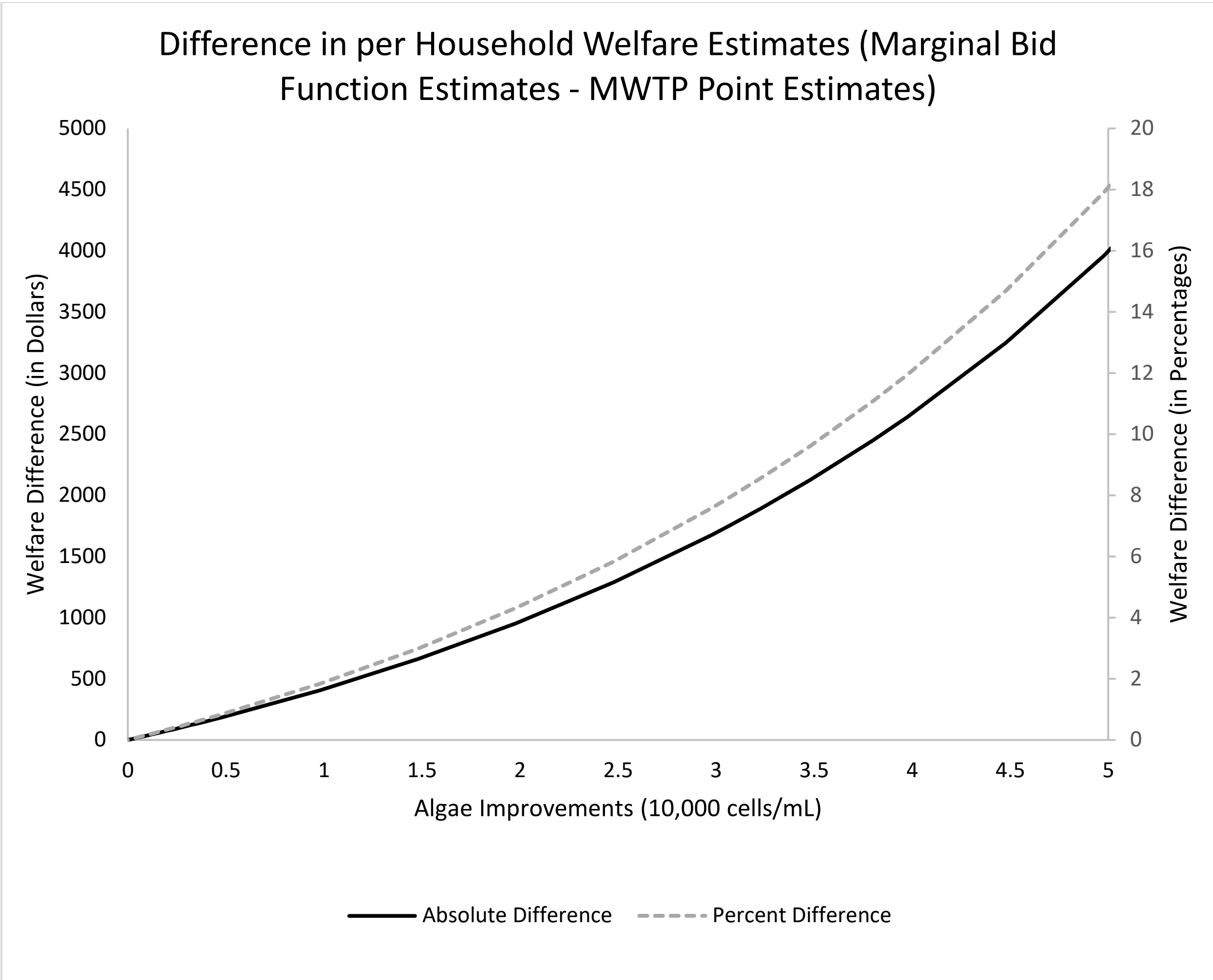


Table 1: Variable Descriptions

Variable	Description
Totalrooms	Number of rooms
Totalbaths	Number of baths
Sqft	Structural square footage in hundreds of feet
Acres	Parcel acreage
Age	Age of the house
Fireplace	Indicator variable for fireplace (includes both wood-burning and fabricated fireplaces)
Garage	Indicator variable for garage
Basement	Indicator variable for basement
Pool	Indicator variable for pool
Central AC	Indicator variable for central air conditioning
LakeAdj	Indicator variable for properties located within 20 meters of a lake
NearLake	Indicator variable for properties located between 20 and 500 meters of a lake
Distancetolake	Distance to closest lake in hundreds of meters
Algae	Annual mean Cyanobacterial Index value. One unit is equivalent to 10,000 *Microcystis* cells/mL
Discharge	March - June max water discharge from Maumee watershed. Measured in cubic feet per second
WaterTemperature	Annual maximum surface water temperature. Measured in degrees Celsius.
Fishing License	Indicator variable for whether or not the household owned a fishing license between 2011 and 2015
Boating License	Indicator variable for whether or not the household owned a boating license between 2009 and 2015
Composite Good	Annual income spent on non-housing related goods. Calculated by subtracting income from annualized housing costs

Table 2: Housing Summary Statistics

	All Counties (N=156112)				Ashtabula (N=5009)				Cuyahoga (N=73703)				Erie (N=7721)			
Variable Name	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Purchase Price	144,223	84,498	50,000	730,000	104,534	44,392	50,000	301,000	141,109	86,770	50,000	730,000	136,876	76,468	50,000	550,000
Total Square Feet	1,634	618.1	576	4,815	1,430	456.1	616	3,275	1,614	613.1	738	4,815	1,557	573.3	634	3,934
Parcel Lot Acreage	0.300	0.383	0.0352	4.998	0.648	0.840	0.0682	4.990	0.212	0.179	0.0490	4.651	0.471	0.645	0.0493	4.966
Age	49.83	25.52	1	100	49.66	24.91	1	100	59.79	21.93	1	100	44.23	23.49	1	100
Total Number of Rooms	6.419	1.376	3	12	5.863	1.101	3	9	6.435	1.367	4	11	6.633	1.607	3	12
Total Number of Bathrooms	1.660	0.691	1	4.500	1.525	0.566	1	3.500	1.541	0.661	1	4.500	1.751	0.669	1	4
Sale Year	2,008	3.692	2,002	2,015	2,008	3.638	2,002	2,015	2,008	3.674	2,002	2,015	2,008	3.677	2,002	2,015
Garage (0/1)	0.850	0.357	-	-	0.604	0.489	-	-	0.967	0.178	-	-	0.590	0.492	-	-
Fireplace (0/1)	0.426	0.494	-	-	0.407	0.491	-	-	0.463	0.499	-	-	0.433	0.496	-	-
LakeAdj (0/1)	0.00607	0.0776	-	-	0.0108	0.103	-	-	0.00115	0.0339	-	-	0.0212	0.144	-	-
NearLake (0/1)	0.0881	0.284	-	-	0.150	0.357	-	-	0.0481	0.214	-	-	0.214	0.410	-	-
DistanceToLake (100s)	44.94	30.70	0	99.99	35.23	27.34	0	99.96	53.51	29.48	0	99.99	22.54	23.92	0	99.82
Algae	4.924	10.89	1	129.4	8.896	10.13	1	52.22	1.005	0.113	1	10.65	23.84	19.30	1	88.67
WaterDischarge	8,923	2,951	3,350	16,726	8,958	2,946	3,350	16,726	8,898	2,926	3,350	16,726	8,947	3,045	3,350	16,726
Fishing License (0/1)	0.163	0.369	-	-	0.219	0.414	-	-	0.121	0.327	-	-	0.243	0.429	-	-
Boating License (0/1)	0.0527	0.224	-	-	0.0605	0.238	-	-	0.0337	0.181	-	-	0.101	0.302	-	-
Income (1000s)	84.83	31.18	10.81	583.7	67.04	11.43	36.42	127.7	86.58	36.23	10.81	583.7	86.65	24.24	42.27	225.7

	Lorain (N=24254)				Lucas (N=11902)				Ottawa (N=4196)			
Variable Name	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Purchase Price	179,784	98,202	50,000	595,000	95,423	41,073	50,000	375,000	170,004	103,606	50,000	650,000
Total Square Feet	1,921	668.1	744	3,887	1,361	411.6	677	3,945	1,598	602.7	576	3,940
Parcel Lot Acreage	0.328	0.341	0.0352	4.998	0.236	0.390	0.0650	4.912	0.359	0.445	0.0373	4.049
Age	27.97	24.00	1	98	58.55	21.75	1	100	40.18	24.88	1	100
Total Number of Rooms	6.878	1.447	4	11	5.895	1.073	4	10	5.954	1.485	3	11
Total Number of Bathrooms	2.061	0.723	1	4	1.316	0.466	1	3.500	1.752	0.687	1	4
Sale Year	2,008	3.732	2,002	2,015	2,007	3.537	2,002	2,015	2,008	3.708	2,002	2,015
Garage (0/1)	0.929	0.257	-	-	0.908	0.289	-	-	0.603	0.489	-	-
Fireplace (0/1)	0.339	0.473	-	-	0.207	0.405	-	-	0.458	0.498	-	-
LakeAdj (0/1)	0.00458	0.0675	-	-	0.00521	0.0720	-	-	0.0863	0.281	-	-
NearLake (0/1)	0.0940	0.292	-	-	0.0619	0.241	-	-	0.340	0.474	-	-
DistanceToLake (100s)	41.15	30.05	0	99.99	56.27	30.65	0	99.99	22.86	30.74	0	99.81
Algae	1.595	1.750	1	16.80	24.24	11.52	1	129.4	26.09	24.23	1	111.5
WaterDischarge	8,939	3,007	3,350	16,726	8,878	2,827	3,350	16,726	9,041	3,083	3,350	16,726
Fishing License (0/1)	0.185	0.388	-	-	0.192	0.394	-	-	0.223	0.416	-	-
Boating License (0/1)	0.0606	0.239	-	-	0.0658	0.248	-	-	0.149	0.356	-	-
Income (1000s)	95.30	28.21	31.61	165.7	64.84	11.06	25.20	143.8	116.4	40.57	56.14	221.6

Table 3: Second Stage Summary Statistics

	(N=4,553)			
Variable Name	Mean	Std Dev	Min	Max
Discharge	15,438	3,915	7,542	24,190
Water Temperature	24.52	1.053	18.84	27.26
Fishing License (0/1)	0.226	0.418	-	-
Boating License (0/1)	0.119	0.324	-	-
Composite Good (1000s)	69.88	91.50	0.950	2,416

Table 4: Water Quality Marginal Bid Function Estimates

Cobb-Douglas Specification (Log-Log)			
Variable	OLS	Market IV	Discharge
LogAlgae	-1.032***	-1.028***	-1.202***
	(0.0212)	(0.0343)	(0.0965)
Fishing License (0/1)	-0.0347*	-0.0349*	-0.0240
	(0.0185)	(0.0212)	(0.0264)
Boating License (0/1)	0.0336	0.0330	0.0649*
	(0.0231)	(0.0262)	(0.0332)
Log (Composite Good)	0.0859***	0.0853***	0.118***
	(0.0189)	(0.0189)	(0.0221)
Observations	4,553	4,553	4,553
First Stage F-Test	-	96.75	74.31
Semi-Log Specification			
Variable	OLS	Market IV	Discharge
Algae	-0.0765***	-0.121***	-0.132***
	(0.00801)	(0.0117)	(0.0143)
Fishing License (0/1)	-0.0537*	-0.0329	-0.0277
	(0.0318)	(0.0617)	(0.0631)
Boating License (0/1)	-0.00955	0.0870	0.111
	(0.0399)	(0.0814)	(0.119)
Composite Good (1000s)	5.13e-05	0.000516*	0.000631**
	(0.000258)	(0.000287)	(0.000288)
Observations	4,553	4,553	4,553
First Stage F-Test	-	20.58	43.00

Notes: ***, **, * indicates significance at the 1%, 5% and 10% level respectively. Bootstrapped standard errors have been clustered at the tract level. All other structural variables from the first stage have been included within each regression. We suppress these results, however, for brevity.

Table A1: First-stage Hedonic Results Full Set

VARIABLES	Boom Toledo	Boom Erie	Boom Cleveland	Boom Ashtabula	Bust Toledo	Bust Erie	Bust Cleveland	Bust Ashtabula
LogAlgae*(LakeAdj + NearLake)	-0.0469***	-0.0485***	-0.0396*	-0.0172	-0.0148	-0.0726***	0.0224	-0.0303**
	(0.0134)	(0.0152)	(0.0228)	(0.0174)	(0.0159)	(0.0179)	(0.0212)	(0.0142)
Totalrooms	-0.00256	0.0120***	0.0102***	0.0156***	-0.00439	0.00625	0.0117***	0.0193*
	(0.00218)	(0.00292)	(0.00132)	(0.00520)	(0.00290)	(0.00486)	(0.00130)	(0.00958)
Totalbaths	0.0453***	0.112***	0.0525***	0.0707***	0.0508***	0.0688***	0.0559***	0.0804***
	(0.00607)	(0.0131)	(0.00316)	(0.0166)	(0.00613)	(0.0122)	(0.00367)	(0.0221)
Sqft (100s)	0.0524***	0.0372***	0.0327***	0.0675***	0.0555***	0.0499***	0.0431***	0.0666***
	(0.00243)	(0.00825)	(0.00132)	(0.00523)	(0.00266)	(0.00808)	(0.00138)	(0.00955)
Acres	0.118***	0.0924**	0.158***	0.114***	0.0766***	0.0834**	0.151***	0.0683**
	(0.0180)	(0.0368)	(0.0117)	(0.0241)	(0.0199)	(0.0357)	(0.0113)	(0.0260)
Age	-0.00679***	-0.00512***	-0.0106***	-0.00463***	-0.0101***	-0.0109***	-0.0115***	-0.00645***
	(0.000593)	(0.00116)	(0.000403)	(0.00144)	(0.000595)	(0.00207)	(0.000424)	(0.00213)
Fireplace(0/1)	0.0308***	0.0754***	0.0253***	0.0762***	0.0270***	0.0668***	0.0347***	0.0541***
	(0.00457)	(0.0121)	(0.00292)	(0.0135)	(0.00597)	(0.0134)	(0.00320)	(0.0107)
Garage(0/1)	0.0963***	0.0285*	0.0498***	0.0471***	0.0549***	-0.000205	0.0509***	0.0205
	(0.00817)	(0.0147)	(0.00568)	(0.0109)	(0.0167)	(0.0114)	(0.00637)	(0.0147)
Basement(0/1)	0.0877***	0.107***	0.0658***	0.0855***	0.0774***	0.0419	0.0677***	0.0757***
	(0.00605)	(0.0169)	(0.00500)	(0.0131)	(0.00776)	(0.0300)	(0.00518)	(0.0149)
Pool(0/1)	0.0430***	0.0328	0.0500***	0.0463***	0.0587***	0.0890**	0.0495***	0.0358
	(0.00746)	(0.0384)	(0.00606)	(0.0127)	(0.00791)	(0.0396)	(0.00702)	(0.0212)
CentralAC(0/1)	0.00798	0.0632***	0.0290***	-0.0614***	0.0159**	0.0678***	0.0257***	-0.0518***
	(0.0139)	(0.0154)	(0.00240)	(0.0108)	(0.00776)	(0.0180)	(0.00275)	(0.0129)
Stories	0.0253**	-0.0198	-0.0202***	-0.00210	0.00225	-0.00978	-0.0126**	-0.0374*
	(0.0112)	(0.0206)	(0.00548)	(0.0136)	(0.00980)	(0.0207)	(0.00558)	(0.0202)
Distancetolake(100s)	0.000588*	-0.000449	-2.26e-05	0.000326	0.000220	-0.000762	0.000326	-3.94e-05
	(0.000345)	(0.000496)	(0.000294)	(0.000264)	(0.000362)	(0.000606)	(0.000281)	(0.000332)
LakeAdj	0.581***	0.597***	0.591***	0.560***	0.584***	0.794***	0.589***	0.476**
	(0.0885)	(0.0714)	(0.0317)	(0.0407)	(0.0828)	(0.0955)	(0.0505)	(0.180)
NearLake	0.215**	0.370***	0.145***	0.395***	0.178**	0.445***	0.188***	0.413***
	(0.0862)	(0.0871)	(0.0305)	(0.0485)	(0.0751)	(0.0841)	(0.0342)	(0.0481)
NearLake*Distancetolake(100s)	-0.000376*	-0.000567***	-0.000372***	-0.000983***	-0.000357**	-0.000443**	-0.000448***	-0.000978***
	(0.000208)	(0.000161)	(6.04e-05)	(0.000163)	(0.000153)	(0.000177)	(7.53e-05)	(0.000126)
Sqft Squared (10000s)	-0.000492***	-0.000264	-0.000145***	-0.00113***	-0.000484***	-0.000527***	-0.000284***	-0.000933***
	(6.46e-05)	(0.000222)	(2.56e-05)	(0.000135)	(6.10e-05)	(0.000177)	(2.68e-05)	(0.000252)
Acres Squared	-0.0192***	-0.0112	-0.0249***	-0.0172***	-0.00885**	-0.0103	-0.0247***	-0.00894*
	(0.00411)	(0.00902)	(0.00257)	(0.00509)	(0.00415)	(0.00891)	(0.00240)	(0.00509)
Age Squared	3.51e-05***	2.08e-05*	7.35e-05***	1.35e-05	6.67e-05***	7.13e-05***	7.89e-05***	2.84e-05
	(5.70e-06)	(1.17e-05)	(4.20e-06)	(1.45e-05)	(5.97e-06)	(1.97e-05)	(4.59e-06)	(1.96e-05)
Constant	10.87***	10.91***	11.35***	10.56***	11.08***	11.18***	11.24***	10.80***
	(0.0629)	(0.0865)	(0.0305)	(0.0972)	(0.0639)	(0.0981)	(0.0355)	(0.116)
Tract FE	123	17	553	24	108	17	511	24
Month FE	11	11	11	11	11	11	11	11
Year FE	5	5	5	5	7	7	7	7
Observations	14,581	1,678	58,222	1,447	11,308	1,590	50,686	1,196
R-squared	0.867	0.795	0.884	0.761	0.871	0.761	0.866	0.718

Notes: ***, **, * indicates significance at the 1%, 5% and 10% level respectively. Standard Errors have been clustered at the tracr level.

Table A2: Robustness to First-Stage Functional Form Specification

Semi-Log Specification			Level-Level Specification		
Commuting Zone	Boom (2002 - 2007)	Bust (2008 - 2015)	Commuting Zone	Boom (2002 - 2007)	Bust (2008 - 2015)
Ashtabula	-0.000936	-0.00337	Ashtabula	-266.3	-589.3
	(0.00172)	(0.00265)		(280.8)	(345.0)
Cleveland	-0.0132**	0.00497	Cleveland	-2,069**	114.7
	(0.00538)	(0.00687)		(925.4)	(1,840)
Erie	-0.00370***	-0.00424***	Erie	-766.4***	-775.1***
	(0.00111)	(0.00140)		(222.0)	(263.0)
Toledo	-0.00172**	-0.00131**	Toledo	-194.4	-341.8**
	(0.000728)	(0.000579)		(157.6)	(161.6)

Notes: ***, **, * indicates significance at the 1%, 5% and 10% level respectively. Standard Errors have been clustered at the tract level. All regressions include tract, year and month fixed effects.

Table A3: Robustness to Various Temporal and Spatial Fixed Effects

No Fixed Effects			Tract Fixed Effects			Tract and Year Fixed Effects		
Commuting Zone	Boom (2002 - 2007)	Bust (2008 - 2015)	Commuting Zone	Boom (2002 - 2007)	Bust (2008 - 2015)	Commuting Zone	Boom (2002 - 2007)	Bust (2008 - 2015)
Ashtabula	-0.0283	-0.0348***	Ashtabula	-0.0180	-0.0322**	Ashtabula	-0.0160	-0.0273**
	(0.0198)	(0.00992)		(0.0176)	(0.0130)		(0.0173)	(0.0121)
Cleveland	-0.101***	-0.0437	Cleveland	-0.0390**	0.00267	Cleveland	-0.0402*	0.0203
	(0.0306)	(0.0581)		(0.0193)	(0.0214)		(0.0230)	(0.0215)
Erie	-0.0632**	-0.0799***	Erie	-0.0464***	-0.0741***	Erie	-0.0479***	-0.0715***
	(0.0256)	(0.0213)		(0.0155)	(0.0184)		(0.0153)	(0.0185)
Toledo	-0.0236	-0.0162	Toledo	-0.0450***	-0.0161	Toledo	-0.0468***	-0.0143
	(0.0346)	(0.0245)		(0.0133)	(0.0160)		(0.0135)	(0.0156)

Notes: ***, **, * indicates significance at the 1%, 5% and 10% level respectively. Standard Errors have been clustered at the tract level in all specifications.

Table A4: Algae, Discharge Correlation Table

	Discharge	Algae	Log(Algae)
Discharge	1.00		
Algae	0.19	1.00	
Log(Algae)	0.23	0.87	1.00