Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits

Michael R. Moore*, Jonathan P. Doubek†, Hui Xu**, and Bradley J. Cardinale††

Abstract

We estimate the effect of lake water quality on residential housing prices using cross-sectional data on water clarity from 113 lakes across the United States. An instrumental-variables approach is developed to address potential endogeneity bias in the water-clarity variable. Three lake-based physical variables serve as instrumental variables: total nitrogen concentration in lake water, total phosphorus concentration in lake water, and water temperature near the lake surface. The econometric results include three methodological findings: the instruments are valid and strong; the ordinary-least-squares estimate of the coefficient on the water-clarity variable is unbiased; and water clarity is the attribute of water quality valued in the housing market. The estimated water-clarity effect shows that a one-tenth of a meter change in water clarity leads to a one-percent change in housing price, or an elasticity of 0.20 at mean clarity. Coupling this effect with estimates of an ecological production function, a lake-specific benefit index is developed that shows the effect on housing values of bringing lakes into compliance with the US Environmental Protection Agency's regional recommendations for phosphorous concentration in lake waters.

JEL Codes: Q51, Q53, Q57

Keywords: residential housing market, instrumental variables regression, ecological-economic modeling, national phosphorus recommendations

Corresponding author: Michael R. Moore, School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109. E-Mail: micmoore@umich.edu, tel: (734) 647-4337

* micmoore@umich.edu. School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109
† jdoubek@lssu.edu. School of Natural Resources & Environment and Center for Freshwater Research and Education, Lake Superior State University, Sault Ste. Marie, MI 49783
** Hui.Xu@anl.gov. Energy Systems Division, Argonne National Laboratory, Lemont, IL 60439
†† bradcard@umich.edu. Cooperative Institute for Great Lakes Research, School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109

1. Introduction

Water quality is a pressing environmental issue in the United States despite the prolonged period of improvement that followed passage of the 1972 Clean Water Act (Keiser and Shapiro 2019a; National Research Council 2004 2009; USEPA 2009). While a few cases have garnered much of the public attention (e.g., Chesapeake Bay and northern Gulf of Mexico), over 40 percent of U.S. inland lakes were experiencing fair to poor water conditions in 2007 (USEPA 2010), and a 2012 assessment found little change in conditions (USEPA 2016), with more U.S. lakes becoming eutrophic (Stoddard et al. 2016). In large areas of the country, nutrient flows of phosphorus and nitrogen intersect with major zones of agricultural production (Steffen et al. 2015). Given the ongoing intensification of agricultural production to meet food demand while complying with the federal biofuel mandate, fertilizer use and related nutrient runoff into U.S. freshwater systems are only likely to increase in the absence of new regulations (National Research Council 2011). In this paper, we analyze the effect of lake water clarity on residential housing prices in a study of U.S. lakes and then develop an integrative ecological-economic tool on nutrients, water quality, and economic benefits.

Environmental economists have long applied hedonic price theory (Rosen 1974) to study the effect of environmental quality on the residential housing market. While air quality has been studied more than water quality (Olmstead 2010), a recent spate of hedonic price studies have examined various aspects of water quality, including: grant-funded water treatment infrastructure under the Clean Water Act (Keiser and Shapiro 2019b); groundwater contamination risk from shale gas development (Muehlenbachs, Spiller, and Timmins 2015); invasive species in lakes (Horsch and Lewis 2009; Zhang and Boyle 2010); water quality in Chesapeake Bay (Klemick et al. 2018; Walsh et al. 2017); coastal water quality in Florida (Bin and Czajkowski 2013; Bin et al. 2017), and inland lake water quality in different states and regions of the United States (Tuttle and Heintzelman 2015; Walsh et al. 2011; Wolf and Klaiber 2017). The first three studies in this list apply panel data methods, while the remaining eight apply cross-sectional methods.

Beginning with Small (1975), omitted variables bias has been an ongoing concern when using cross-sectional methods in hedonic price studies (e.g., Chay and Greenstone 2005). Kuminoff, Parmeter, and Pope (2010) examine spatial fixed effects as a way to control for unobserved neighborhood characteristics in hedonic regressions with cross-sectional data. Based on findings of reducing or nearly eliminating bias in their analyses, they recommend spatial fixed effects as a strategy for use with cross-sectional methods. Several of the recent hedonic studies apply spatial fixed effects, some using census tracts as the spatial unit (e.g., Tuttle and Heintzelman 2015; Wolf and Klaiber 2017) and others using lakes (Horsch and Lewis 2009; Zhang and Boyle 2010), cities (Bin et al. 2017), and regional districts (Bin and Czajkowski 2013). At the same time, Kuminoff, Parmeter, and Pope (2010) also warn that spatial fixed effects are not perfect controls. For example, census-tract fixed effects do not control for variation across census blocks within a census tract.[1] Lake fixed effects, similarly, may be poorly aligned, spatially, with neighborhoods and school districts. In addition, spatial fixed effects do not control for omitted variables related to structural features of the house.[2] In short, the ability of spatial fixed effects to mitigate omitted variables bias is generally an untestable empirical question.

Measurement error in the environmental quality variable creates a similar endogeneity concern in hedonic price studies. Frequently overlooked, Keiser (2019) raises this issue in a related context – the effect of water quality on recreation demand at U.S. lakes – and addresses the issue with an instrumental variables estimator.[3] The relevant question, for our purposes, is how to define the water-quality variable that enters the consumer's demand function for residential property near a lake. A homebuyer, for example, might be influenced by one or more temporal variations of a single water-quality variable such as water clarity: clarity at the time of the sale; clarity during the summer recreation season; average clarity over the last year; or average clarity over a longer time horizon, say three years or even ten years. In addressing this issue, the main approach applied in the hedonic price literature is to estimate OLS regressions with several different definitions of water clarity and, then, to apply the different estimates to derive a set of distinct implicit prices of water clarity (e.g., Michael, Boyle, and Bouchard 2000; Walsh et al. 2017; Klemick et al. 2018). Following

[1] Bayer, Ferreira, and McMillan (2007) develop a model of individual neighborhood choice and use the census block, not the (larger) census tract, as the spatial unit to represent a neighborhood.
[2] In studies with panel data on individual housing transactions (e.g., Muehlenbachs, Spiller, and Timmins 2015), the repeat sales approach is being used such that parcel fixed effects control for time invariant characteristics at the smallest possible spatial scale, i.e., they control for both structural features of the house and spatial amenities and disamenities of the neighborhood.
[3] Keiser (2019) finds that the negative effect of phosphorus pollution on recreation demand is an order of magnitude larger in absolute value when estimated with an instrumental variables estimator rather than OLS.

Keiser (2019), an alternate approach is to posit that, for both conceptual and empirical reasons, any water-clarity variable comes with measurement error, and one can address the resulting endogeneity bias with an instrumental variables estimator. We take this alternate approach.

We apply hedonic price theory to the residential housing market in a study of U.S. inland lakes. The study applies a unique database that combines water-quality data from the 2007 National Lakes Assessment (NLA) and data on housing market transactions at properties near 113 lakes in 32 U.S. states. Following Keiser and Shapiro's (2019b) study of water quality, we analyze the 32 states as a single market in the main results. Water clarity is the variable of interest, and it is represented using a standard measure, Secchi disk depth.[4]

We develop an instrumental-variables approach to address concerns related to endogeneity bias in estimating the effect of water clarity. Three physical variables serve as instruments for water clarity: water temperature near the lake's surface, total phosphorus concentration in lake water (hereafter P concentration), and total nitrogen concentration in lake water (hereafter N concentration). Keiser and Shapiro (2019b) note that, while omitted variables are relevant to the study of air quality, their relevance to the study of water pollution and the housing market is unknown.

A second focus of the analysis is to couple the hedonic price results with an ecological production function to develop an integrated ecological-economic framework. Here we apply the NLA data to estimate an ecological production function for water clarity as a function of lake water temperature and the two nutrient concentrations. We integrate the ecological results and the hedonic price results to derive a benefit index that shows, by each study lake, the percentage increase in housing prices from achieving federally prescribed, ecoregion-based recommendations for P concentrations in lake waters. The coupled ecological and economic models provide a new water-quality application of the ecosystem services paradigm (Keeler et al. 2012).

[4] A Secchi disk is an 8-inch disk with alternating black and white quadrants. As a measurement device, the disk is lowered into the water of a lake until the observer can no longer see it. This depth of disappearance, called the Secchi disk depth, is the common measure of water clarity.

We develop three main findings. One, the OLS estimator does not suffer from endogeneity bias with respect to the variable of interest, water clarity. The three instrumental variables are physical variables that correlate with water clarity through phytoplankton production (Lorenzen 2003; Beaulieu, Pick, and Gregory-Eaves 2013), thus making them good candidates for instruments. Each instrumental variables regression generates three tests related to water clarity and the instruments: overidentifying restrictions, endogeneity, and weak instruments. The tests of overidentifying restrictions develop strong evidence indicating that the instrumental variables are valid. Valid instruments are necessary for endogeneity tests. Next, the endogeneity tests develop strong evidence that water clarity is not an endogenous variable in the OLS regressions. Lastly, the instrumental variables are not weak instruments according to results from the first-stage regressions of the instrumental variables estimator. Our study is the first application of an instrumental variables estimator in a hedonic study of water quality with cross-sectional data.[5]

Two, water clarity has a positive and statistically significant impact on housing price. The preferred coefficient estimate on water clarity is 0.099, or a one-meter change in Secchi disk depth results in a 9.9-percent change in housing price given the log-linear form. This scales linearly in meters versus percentage change, such that a 0.1-meter change in *SECCHI* generates a 0.99-percent change in price. The impact is relatively inelastic, with an elasticity of 0.20 evaluated at mean water clarity and 0.14 evaluated at median water clarity. We find evidence that clarity's effect is robust to a variety of different model assumptions and variable definitions.

Three, with the coupled hedonic price results and ecological results, we derive for each study lake the incremental benefit of achieving the lake's recommended P concentration. Seventy-five of the 113 lakes have excess P concentrations. For these noncompliant lakes, the percentage increase in housing prices ranges from 0.4% to 9.2%, with a mean of 3.3%. We also demonstrate how this method can inform a local water quality issue by evaluating the change in the housing stock's value at a single lake in Indiana.

[5] Zhang, Boyle, and Kuminoff (2015) apply an instrumental-variables approach to address a related topic, the endogeneity problem in identifying amenity demand parameters from the hedonic price model's second stage. We focus solely on the first stage of the hedonic model.

2. Econometric Model and Data

2.1 Hedonic Price Regression. In the context of the housing market, the hedonic price function (HPF) represents an equilibrium between housing prices and characteristics.[6] Tangencies between seller offer functions and buyer bid functions determine the equilibrium. The HPF is expressed as housing price (HP) as a function of characteristics, or

(1) $$HP = f(z_1, z_2, \ldots, z_n)$$

where the $z_1, z_2, \ldots, z_n$ include structural characteristics of a house, local public good characteristics of the neighborhood, and environmental quality characteristics. In our application to lake water quality, three variables are of primary interest in explaining variation in transaction prices. Water quality of a nearby lake is measured using the water-clarity variable *SECCHI* (Secchi disk depth). *LKFRNT* is a binary variable that represents whether the property is lakefront property. *LK_DSTNC* measures the property's distance in feet to the lake. We form a log-linear regression equation:

(2) $$\ln HP_{gm} = \alpha + \beta_1 SECCHI_g + \beta_2 LKFRNT_{gm} + \beta_3 LK\_DSTNC_{gm} + \boldsymbol{x}_{gm}\boldsymbol{\beta} + \boldsymbol{z}_g\boldsymbol{\gamma} + \mu_{gm}$$

$$g = 1,\ldots,G;\ m = 1,\ldots,M_g.$$

where g is a lake (i.e., a cluster), m is a residential property, and M_g is the number of properties sampled near lake g. Explanatory variables, other than the three primary variables, include $\boldsymbol{x}_{gm}$, which is a $1 \times k$ vector of variables that vary both within and across lakes, and $\boldsymbol{z}_g$, which is a $1 \times l$ vector of variables that vary only at the lake level. The variables $\boldsymbol{x}_{gm}$ are structural features of the house (Table 1). The variables $\boldsymbol{z}_g$ are individual lake characteristics (other than water quality $SECCHI_g$) and neighborhood characteristics that serve as local public goods nearby individual lakes (Table 1). In the regressions, year fixed effects and quarter-of-the-year fixed effects are included in all specifications (the year 2010 and the first quarter are the omitted fixed effects). The year fixed effects control for annual trends in the housing market during the period from which the housing transactions data are taken, 2010-2013. Quarter-of-the-year fixed effects capture intrannual cyclical movements in the housing market. In one specification, state fixed effects are included for the 21 states in which more than one lake is sampled.

[6] More insight into this approach is available in Davis (2011), Greenstone and Gallagher (2008), Freeman, Herriges and Kling (2014), and Muehlenbachs, Spiller, and Timmins (2015).

Equation (2) shows that the estimated effect of *SECCHI*, β_1, is identified from cross-sectional variation in water clarity at the 113 study lakes.

The composite error term μ_{gm} in equation (2) consists of three components,

(3) $$\mu_{gm} = c_g + \theta q_{gm} + \epsilon_{gm}$$

where c_g is an unobserved cluster effect, θq_{gm} is an unobserved factor, and ϵ_{gm} is the idiosyncratic error. The cluster effect represents arbitrary within-lake correlation, which could occur due to an omitted lake or neighborhood characteristic that serves as a local public good (e.g., local school quality). Because such a characteristic is nonrival and spatially invariant within the relevant area, we model this as a cluster effect rather than as spatial autocorrelation. We also posit that, with cross-sectional data, a variable q_{gm} is unobserved (i.e., an omitted variable, or variables), and hence its effect θq_{gm} enters the error term. The omitted variable could be a house characteristic such as whether the lake is the household's source of drinking water.

The error term in (3) shows two complications for which the estimation accounts. First, with the omitted variable, the variable of interest, $SECCHI_g$, is potentially an endogenous variable. We thus develop an instrumental-variables approach. Second, with lakes serving as clusters, we estimate cluster-robust variance-covariance matrices and apply cluster-robust inference. The data are consistent with the (desirable) case of the number of clusters g substantially exceeding the typical number of individual observations within the clusters M_g (Angrist and Pischke 2009; Wooldridge 2010). The number of clusters equals 113, i.e., the number of lakes at which we sample residential housing transactions. The mean number of observations within the clusters is 12.9 housing properties and the standard deviation is 16.8. Note that these two corrections – an IV approach and cluster-robust standard errors – are known to pose challenges for inference by generating larger standard errors on coefficient estimates. We will see, later in the paper, that they are not detrimental to the results.

Lastly, the *SECCHI* variable could suffer from measurement error relative to the theoretically correct variable. As described in the Introduction, several arguments can be made for different temporal definitions of *SECCHI* in the hedonic price function. The NLA data are for 2007, such that *SECCHI* potentially could serve as a proxy for average water clarity prior to the housing market transactions in 2010-2013. Our approach thus is to recognize that our *SECCHI* variable comes with measurement error relative to the true value in the function and to address the corresponding potential endogeneity bias with an instrumental-variables approach (as in Keiser 2019). [7]

2.2 Data. The U.S. EPA's 2007 NLA deployed a standardized method of data collection to produce the first large-scale systematic study of the physical, chemical, and biological characteristics of over 1,000 lakes across the continental U.S. (USEPA 2010). The 113 lakes in our study were selected because of their nearly complete data on phytoplankton biovolume, from which three variables were constructed (concentration of total phytoplankton biovolume, concentration of cyanobacterial biovolume, and concentration of green algal biovolume). Other candidate lakes were excluded if the cell biovolume data were missing for more than one-third of the relevant phytoplankton taxa (Doubek et al. 2015). Analysis with phytoplankton-based variables is important for understanding their role relative to water clarity and N and P concentrations.

Table 1 describes all variables and provides descriptive statistics. Additional details follow here.

Housing data. To develop data on residential housing, we used the Zillow® database on actual residential property transactions for 2010 through 2013. Individual houses are spatially explicit at www.zillow.com, including houses for which transaction data are available. We first prepared a spatial representation of each lake with a 0.1 mile GIS buffer around the lake. This distance was chosen to select houses in proximity to the lakeshore.[8] Transactions within the buffer were then recorded. The mean number of transactions was 12.9 per lake, and the median was seven. Only eleven lakes (of 113) had 30 transactions or more. The final sample consists of 1,462 housing transactions.

[7] Keiser (2019) develops an interesting approach of explicitly examining the validity of instrumental variables for phosphorus concentrations in lake waters by regressing measurement error estimates on the instrumental variables, upstream concentrations of phosphorus.

[8] A similar buffer distance is used in other NLA studies to assess relationships of land use and lake water quality (e.g., Doubek et al. 2015; Read et al. 2015).

In addition to sales price and transaction date, the data on house characteristics included square footage, number of bedrooms, number of bathrooms, and year built. We identified geographic coordinates for each house using Geocoding API, and then recorded whether the house was a lakefront property and calculated the distance to the lake using the distance tool on Google maps. The sales prices were converted to real terms in 2013 dollars using the House Price Index of the Federal Housing Finance Agency.

The effect of *SECCHI* is identified from cross-sectional variation at the 113 study lakes (equation (2)). Thus, the number of transactions is not a limiting factor in the estimation even though the sample of 1,462 transactions is small relative to studies with data purchased from vendors.[9] Furthermore, like Zhang, Boyle, and Kuminoff (2015), our analysis is limited to parcels on or near inland lakes, and thus we are not attempting to estimate the price gradient for houses beyond the 0.1 mile buffer.

Water quality data. The NLA data on lake water quality included Secchi disk depth, total N concentration, total P concentration, water temperature at a subsurface depth of 0.1 meter in the middle of the lake, and the three measures of various phytoplankton biovolumes. The NLA applied a stratified random sampling design to lakes in the contiguous U.S. that were at least 1 meter depth and had a surface area of at least 0.04 km^2. Researchers followed a consistent protocol across lakes to sample physical (e.g., water temperature), chemical (e.g., total N) and biological (e.g., phytoplankton biovolume) conditions of each lake. The data were collected between May and September in 2007, on a date that varies by lake. About 10% of the lakes were sampled more than once, but we include data only from the first visit to keep this consistent across the 113 study lakes. Data from the NLA are available at: http://water.epa.gov/type/lakes/lakessurvey_index.cfm.[10]

An important point is that the NLA measurements of Secchi disk depth for the lakes in our study were quite similar over time. This is relevant as the lake data were collected earlier than the housing data. Data from the second NLA in 2012 only recently became publicly available, i.e., they were not available at the start of this project. Thirty-eight of the 113 study lakes that were sampled

[9] Other recent studies (e.g., Zhang, Boyle, and Kuminoff 2015) use relatively small samples of house transactions.
[10] Although the NLA is now scheduled for every five years, Keiser and Shapiro (2019a) note that data on water pollution are relatively limited, especially in contrast to air pollution in which some data are reported hourly.

in 2007 had repeat sampling in the 2012 assessment. For these lakes, the Pearson correlation coefficient for the *SECCHI* variables is 0.85. For all lakes sampled in both years – a total of 376 – the correlation coefficient is 0.84. In short, the correlations are relatively high and provide evidence of relative stability over time, which lends credibility to use of the 2007 data as a proxy for average water clarity in this study. We include a robustness check in Section 3.2.3 that makes a rigorous regression-based comparison of the 2007 and 2012 *SECCHI* data.

Lake characteristics. The NLA data also include lake surface area, lake perimeter, the area of developed land, and the area of agricultural land (both within a 200-meter zone around each lake). The land area data are originally from the USGS National Land Cover Dataset.

Neighborhood characteristics. Data on neighborhood characteristics near each lake were collected at the zip code level from the U.S. Census (using factfinder2.census.gov). These included socioeconomic data (e.g., median income, education, and race), population density, proportion of population below the poverty level, proportion of residential property that is rental housing, and mean commuting time to work.

Data for two climate variables, annual heating degree days and annual precipitation, were obtained from the National Climatic Data Center (http://www.ncdc.noaa.gov/cdo-web/datasets). NCDC computed annual climate normals for the 30-year period, 1981 to 2010. Data were collected at the zip code level for each lake if available at this resolution. If data were not available at the zip code level, they were collected at the county level (if there were multiple stations present in a county, data were averaged across all stations). The climate variables are included because research shows that climate affects geographical sorting and thus housing demand (Albouy et al. 2016).

3. Empirical Results

3.1 Main Results. We estimate equation (2) using ordinary-least-squares (OLS) regressions and instrumental variables generalized-method-of-moments (IV-GMM) regressions. In the main results, we report on eight regressions that explore different variable specifications across the OLS and IV-GMM regressions (Table 2). The specifications include four different combinations of house characteristics, lake and neighborhood characteristics, and state fixed effects. These include four variables for structural characteristics of the house; fourteen variables for public-good characteristics of the neighborhood and lake; a flexible functional form for the house covariates; and state fixed effects for the 21 states with more than one lake in the study.

The results show that lake water clarity, the variable *SECCHI*, has a statistically significant impact on lake housing prices near 113 lakes across the United States. The estimated coefficients on *SECCHI* range across a fairly narrow band from 0.099 to 0.154, with all coefficients being highly statistically significant (Table 2). This occurs despite use of cluster-robust standard errors to account for heteroskedasticity and cluster effects by lake; cluster-robust standard errors typically are much higher than their robust counterparts (Angrist and Pischke 2009).

To develop a preferred estimate of the coefficient on *SECCHI*, we first need to address potential endogeneity of *SECCHI* in light of the cross-sectional approach. Three variables serve as instrumental variables for *SECCHI*: P concentration in the water (*PHSPHRS*), N concentration in the water (*NITROGEN*), and water temperature just below the lake surface (*TEMP)*. These three physical variables correlate with water clarity through phytoplankton production (Beaulieu, Pick, and Gregory-Eaves 2013; Lorenzen 2003), thereby making them good candidates for instruments.[11] Each IV-GMM regression generates three types of tests related to *SECCHI* and its potential instruments: a test of overidentifying restrictions, an endogeneity test, and a test for weak instruments (Table 3).

We develop both conceptual arguments and empirical evidence that the three instrumental variables satisfy the exclusion restriction. We first argue that N concentration, P concentration, and lake temperature do not directly enter as arguments of the hedonic price function. Concentrations of N and P are imperceptible to humans, given that they are tasteless, colorless, and odorless when

[11] Lakes can be either N or P limited, or co-limited by N and P, depending on many factors such as land use, nutrient stoichiometry, and type of prevalent primary producers present. Historically, P-limitation by phytoplankton was considered the more common limiting reagent (e.g., Schindler 1977; Smith 1983). Recently, however, studies and evidence suggest co-limitation (e.g., Xu et al. 2010; Filstrup et al. 2018; Lewis et al. 2020). Without information on nutrient limitation for each lake, we are not able to determine whether lakes are N limited, P limited, or co-limited by N and P. We thus included both N and P as instrumental variables.

dissolved in water. In addition, the regressions include two land-use variables, percentages of the land base within 200 meters of the lake that are developed land and agricultural land, respectively. These two variables belong in the hedonic price regression as local characteristics that affect housing demand. They also represent typical vectors of N and P inputs into water bodies through fertilizer runoff, and thus they reduce the potential for lake concentrations of N and P to correlate with the regression's error term.

Water temperature of the lake (the third instrumental variable), similarly, is likely unknown to a prospective homebuyer, rather than an observed data point. A homebuyer would simply observe whether and how a lake was being used for recreation. The hedonic price regression does include a variable for air temperature (average annual heating degree days) to account for climate. Albouy et al. (2016) find that average annual heating degree days, and other climate variables, factor into residential location decisions. Once again, with air temperature in the regression, the potential for water temperature to correlate with the regression's error term is reduced.

Next, the empirical evidence on the instruments comes from a test of overidentifying restrictions. As context here, the three instruments could correlate with the error term of the regression. For example, P concentration, N concentration, and lake temperature affect phytoplankton biovolume (including cyanobacterial biovolume); if phytoplankton is an omitted variable from the hedonic price regression, then this is another reason why the instruments could correlate with the error term.[12] The test of overidentifying restrictions examines this. With the IV-GMM estimator, the test of overidentifying restrictions applies Hansen's *J* statistic, which is distributed as χ^2 with degrees of freedom equal to the degree of overidentification. As reported in Table 3, the relatively high *p* values indicate that the null hypothesis cannot be rejected in specifications (1)-(3). This provides suggestive evidence that the variables *PHSPHRS*, *NITROGEN*, *TEMP* are valid instruments. The statistical evidence also suggests that *NITROGEN* and *TEMP* continue as valid instruments in specification (4).[13] Valid instruments are necessary for endogeneity tests.

[12] In section 3.2.1, we report hedonic price regressions with both *SECCHI* and three phytoplankton variables to provide evidence that water clarity, and not the phytoplankton attributes, is the water-quality attribute valued by consumers in the housing market.

[13] Wooldridge (2010, p. 135) issues a caution with the overidentification test. He provides an example in which one of the instruments is endogenous (not exogenous), and the full and reduced set of instruments is

The endogeneity tests on *SECCHI* apply the *C* test, which is distributed as χ^2 with one degree of freedom. The high *p* values indicate that the null hypothesis – that *SECCHI* may be treated as exogenous – cannot be rejected. To provide supporting evidence for the test results, note that the coefficients on *SECCHI* in the paired OLS and IV-GMM regressions are very close in magnitude (Table 2), e.g., 0.099 in column 4a and 0.101 in column 4b of the table. The results show convincingly that *SECCHI* is not an endogenous variable in the OLS regressions.

Lastly, the instrumental variables are not weak instruments according to the first-stage regressions of the IV-GMM estimator. The estimated coefficients on the instruments are highly significant, and have the expected signs, in explaining variation in *SECCHI* (Appendix Table 1); this is one test for valid instruments. In joint tests, the instruments add substantial explanatory power to the regressions, and the values of the *F* statistics in general exceed the rule-of-thumb that weak instruments have values lower than 10 (Staiger and Stock 1997) (Table 3). In addition, the value of Shea's partial R^2 statistic exceeds 0.47 in specifications (1)-(3) and 0.28 in specification (4). The IV-GMM results thus are not affected by issues associated with weak instruments.

In light of these results, the preferred estimated coefficient on *SECCHI* is 0.099; this is the OLS estimate that utilizes the entire set of covariates (Table 2, column 4a). The coefficient from the related IV-GMM regression is very similar, 0.101 (column 4b), and is also unbiased, yet its standard error is almost twice as large.

Given the log-linear form of the hedonic price function, housing price (and marginal willingness-to-pay) increases at an increasing rate with lake water clarity. The coefficient of 0.099 implies that a one-unit change in water clarity (i.e., a one-meter change in *SECCHI*) causes a 9.9-percent change in housing price. This scales linearly in meters versus percentage change, such that a 0.1-meter change in *SECCHI* generates a 0.99-percent change in price. For context, the values for

asymptotically biased in similar ways, such that the null hypothesis of the overidentification test is (falsely) not rejected.

mean/minimum/maximum *SECCHI* in the sample are 2.1/0.2/9.5 meters, thus making a one-meter change fairly substantial.[14]

The coefficient of 0.099 combines with the mean housing price to generate an estimated change of $3,971 in mean price for a 0.1-meter change in *SECCHI*. This also translates into an estimated change of $12,104 for a one-foot change in *SECCHI*. For comparison, a study of central Florida estimates a $5,595 change in mean lakefront property value for a one-foot change in *SECCHI* (Walsh, Milon, and Scrogin 2011). A study of southern Florida estimates a $36,070 change in mean waterfront property value for a one-percent change in water clarity (Bin and Czajkowski 2013). Water clarity in that study is measured as the ratio of Secchi disk depth to bottom depth at points in the St. Lucie River Estuary of southern Florida. A related study (Bin et al. 2017) estimates a $2,614 change in mean waterfront property value for a one-percent change in a water quality index. Their index is a percentage value based on data from four water-quality measures: water clarity, dissolved oxygen, pH, and salinity. Overall, the results suggest that water clarity is an economically significant factor in explaining housing prices.

The preferred estimate, 0.099, also generates relatively inelastic elasticities of 0.20 evaluated at mean *SECCHI* and 0.15 evaluated at median *SECCHI*. These elasticities are similar in magnitude to other hedonic price estimates. For lakefront property, an elasticity of 0.13 is estimated for *SECCHI* in the central Florida application (Walsh, Milon, and Scrogin 2011). Similarly, for waterfront property on Chesapeake Bay, elasticity estimates are highly inelastic, ranging from -0.033 to -0.156 (Walsh et al. 2017; Klemick et al. 2018). Their negative numbers are expected, as their water-quality variable is a measure of water-column light attenuation, which essentially is the inverse of water clarity. With air pollution, the elasticity of housing price with respect to total suspended particulates ranges from -0.20 to -0.35 in a national study (Chay and Greenstone 2005). In the Bay Area of California, three housing price elasticities are evaluated at the medians of three air pollutants (Bajari, et al. 2012); the particulate matter elasticity ranges from -0.076 to -0.084; the sulfur dioxide elasticity ranges from -0.18 to -0.22; and the ground-level ozone elasticity ranges

[14] The range of *SECCHI* covers a spectrum of ultra-oligotrophic to hypereutrophic lakes.

from -0.60 to -0.64. The evidence thus shows a pattern of strongly inelastic responses of housing prices to water and air quality.

Two other variables of interest – lakefront property (*LKFRNT*) and distance to the lake (*LK_DSTNC*) – have estimated coefficients that are reasonable in magnitude, highly statistically significant, and robust to alternate specifications (Table 2). We again rely on the OLS estimates that utilize the entire set of covariates as the preferred estimates (column 4a). Lakefront property receives a 42.5 percent premium over non-lakefront property. Housing price declines by 6.4 percent for each 100 feet of distance between the house and lake, holding *LKFRNT* (and other variables) constant. These estimates are comparable to estimates from a hedonic price study of lakes in Adirondack Park in New York State (Tuttle and Heintzelman 2015).

The estimated coefficients on the four variables for the structural features of a house have the expected signs and are statistically significant (Appendix Table 2). Most of the estimated coefficients on the fourteen variables for the public-good characteristics of the lake and neighborhood have the expected signs, and half of them are statistically significant.

3.2 Robustness Checks on *SECCHI*. In the robustness checks, we continue the pattern of estimating regressions based on OLS and IV-GMM estimators, with specifications (1) through (4) continuing to represent four combinations of house characteristics, lake and neighborhood characteristics, and state fixed effects. We report three robustness checks related to the *SECCHI* variable here. In Appendix B, an additional check on *SECCHI is* reported that investigates whether *SECCHI* exerts a diminishing marginal effect on housing price. Appendix B also reports results from a two-stage least squares (2SLS) estimator as a comparison to the IV-GMM estimator.

3.2.1 Do water-quality variables related to phytoplankton affect housing prices? Environmental concerns with water quality focus on the relationship between nutrient runoff and biological growth, including outcomes such as harmful algal blooms and hypoxic conditions. Using NLA data, we formed three variables related to the biomass of phytoplankton in the lakes: concentration of total phytoplankton biovolume in the lake water (*PHPLNKTN*), concentration of cyanobacterial biovolume (*CYANBCTR*), and concentration of green algal biovolume (*GRNALGAE*). While homeowners clearly value water clarity, do homeowners perceive and value the biological attributes independently of water clarity, i.e., do one or more of these three variables affect housing price after controlling for *SECCHI*?

We included the three biovolume variables iteratively in hedonic price regressions using OLS. We report results only for specifications (1) and (4) of the covariates (Table 4), although the results are consistent across all four specifications. The estimated coefficients on the biovolume variables were not statistically significant, while the coefficients on *SECCHI* continued to be highly significant and very similar in magnitude to the main results, i.e., results without the biovolume variables.

As a second check on this, we repeated the regressions with the three biovolume variables while using an IV-GMM specification. The regressions use the same instrumental variables as above: N concentration, P concentration, and water temperature. The estimated coefficients on the biovolume variables were not statistically significant, while the coefficients on *SECCHI* continued to be highly significant and very similar in magnitude to results without the biovolume variables. Overidentification tests showed that, once again, the instruments are valid.[15]

The robustness checks with the biovolume variables raise a question about how to include different aspects of water quality in a hedonic regression. Concerns about harmful algal blooms have increased both in the United States and other countries (Ho, Michalak, and Pahlevan 2019). Wolf and Klaiber (2017) develop a variable that applies a cyanobacteria toxicity threshold above which use of the water is dangerous. In a study of inland lakes in Ohio, they find that the variable exerts a negative effect on housing prices for parcels on or near a lake. An open question is whether their regression specification should include a water-clarity variable such as Secchi disk depth to counteract potential omitted variables bias. In our analysis, an OLS regression with cyanobacterial biovolume (*CYANBCTR*) and *without SECCHI* finds a negative and significant coefficient on *CYANBCTR*. However, the *CYANBCTR* coefficient is not statistically different from zero when *SECCHI* is included in the regression (Table 4). This suggests that potential homebuyers perceive water quality in terms of clarity rather than the biovolume concentrations of phytoplankton in a lake.

[15] These results are available from the authors upon request.

3.2.2 Does the marginal effect of SECCHI vary regionally? Several previous studies assume a national housing market when estimating hedonic price functions (e.g., Chay and Greenstone, 2005; Bayer, Keohane, and Timmins 2009; Davis 2011; Keiser and Shapiro 2019b). We make the same assumption, but examine it by estimating regressions in which the marginal effect of *SECCHI* varies by state (Table 5). We first identify the three states with the largest number of study lakes and property transactions: Florida has 10 lakes and 151 observations; Indiana has 25 lakes and 238 observations; and Washington has 9 lakes and 178 observations. In the regressions, estimated coefficients on the variables *SECCHI*FLORIDA*, *SECCHI*INDIANA*, and *SECCHI*WASHINGTON* are not statistically different from zero, and the estimated coefficients on *SECCHI* are quite similar to estimates in the benchmark regressions reported as the main results. We conclude that, within our sample, the marginal effect of *SECCHI* does not vary regionally.[16]

3.2.3 Do results change when using SECCHI data from the 2012 National Lakes Assessment? The timing of the NLA (2007) relative to the housing transactions (2010-2013) raises a question about whether the *SECCHI* data from 2007 represent lake water quality for homebuyers in the 2010-2013 period. Here we compare regression results from the 2007 and 2012 NLA.[17] The main results rely on water-quality data from 113 lakes that were sampled in the 2007 NLA. Of those 113 lakes, 38 lakes were both sampled again in the 2012 NLA and had the requisite water-quality data. Our comparison focuses on two sets of OLS regressions using data from the 38 lakes that are common to the two periods. This dataset includes 600 observations on housing transactions at the 38 lakes. This comparison holds constant all variables, observations, and lakes in the compared regressions except for the *SECCHI* variable, whereby one regression uses the 2007 *SECCHI* variable and its paired regression uses the 2012 *SECCHI* variable. This comparison isolates performance of the two *SECCHI* variables.

Table 6 reports results for the *SECCHI* variables in the two sets of regressions. Specifically, the two sets of regressions are identical in all regards (house, lake, and neighborhood covariates; the

[16] It would be informative to repeat this type of analysis with larger samples of housing transactions and lakes.
[17] We thank reviewers for suggesting this approach.

various fixed effects; and clustering on the lakes for cluster-robust standard errors) except for using a variable based on 2007 *SECCHI* data in the first set of regressions and a variable based on 2012 *SECCHI* data in the second set.

The results are very similar across the two years. For each specification (i.e., for each column in the table), the coefficient estimates on the *SECCHI* variables are very similar in magnitude and statistical significance. This is particularly true in specification (4) with the full set of controls – the coefficient estimates are 0.085 and 0.089; the standard errors are 0.029 and 0.031; and the coefficients are both significant at the 0.01 percent level. We conclude that the *SECCHI* variables are sufficiently stable to generate very similar regression results across the two years. This lends credibility to our main results using 2007 NLA data.

4. Economic Benefits with Coupled Ecological-Economic Results

In this section, we estimate an ecological production function to explain variation in *SECCHI* across lakes and then couple the ecological results and hedonic price results to develop a benefit index for achieving EPA-prescribed P concentrations in lake waters. This approach of integrating models to understand the value of managing P concentrations addresses a research gap identified by Garnache et al. (2016).[18]

We begin by first explaining the context in which hedonic price results can measure the economic benefits of environmental quality changes. The gradient of a bid function at the tangency with the hedonic price function represents a household's marginal willingness-to-pay (MWTP) for an individual characteristic (Rosen 1974). For a localized amenity, the hedonic price function serves as an approximate measure of MWTP for a nonmarginal change in environmental quality (Freeman, Herriges, and Kling 2014).[19] Lake water quality is such a localized amenity. The housing stock near lakes is a very small percentage of the national housing stock. Water quality, moreover, might

[18] Garnache et al. (2016, 1348) write, "The critical need here is for empirical applications that produce results directly relevant to policy by estimating site-specific values and connecting them to P [phosphorus]." Our approach is consistent with their recommendation.
[19] The localized amenity argument was applied recently in national studies with circumstances of spatially distinct changes in environmental quality, including power plant openings (Davis 2011), toxic plant openings and closings (Currie et al. 2015), and Superfund site remediation (Greenstone and Gallagher 2006).

change at only some lakes in the country. Our preferred estimate of the effect of *SECCHI*, 0.099, thus is appropriate for measuring the economic benefits of water-clarity changes at U.S. lakes.

4.1 Ecological Regression Model. We estimate an ecological production function in which *SECCHI* is explained by P concentration in the water (*PHSPHRS*), N concentration in the water (*NITROGEN*), and water temperature near the lake surface (*TEMP*).[20] Data from the 113 study lakes and a Cobb-Douglas form for the production function are applied in an OLS regression:

$$\text{(4)} \qquad ln\ SECCHI_g = \alpha + \beta_1 ln\ PHSPHRS_g + \beta_2 ln\ NITROGEN_g + \beta_3 ln\ TEMP_g + \epsilon_g$$

$$g = 1,\ldots,G,$$

where *g* is a lake. The parameter of interest is the estimated coefficient on *ln PHSPHRS*, which is $\hat{\beta}_1$= 0.49. This response is relatively inelastic in that a 1% change in P would result in a 0.49% change in *SECCHI*. The estimation procedure and results for equation (4) are described in Appendix Table 3 and the accompanying material in Appendix A.

4.2 Coupled Results: Economic Valuation of Phosphorus Standards. To link the valuation results to a management action, we couple the ecological results with the hedonic price results to evaluate the economic benefit of a potential reduction in P concentration in lake waters. In the potential action, P concentration in individual lakes is reduced to meet ambient lake water quality recommendations made by the U.S. EPA (2000). As a basis for this, we match our individual study lakes with recommended P concentrations for lakes in Level III Subregions of the U.S. EPA Ecoregions. That is, each lake is matched with a P concentration for its respective Level III Subregion.[21] Comparison of the recommended P concentration and the actual P concentration (from the NLA data) shows that: 36 lakes from our study already meet or exceed the recommendations; 75 lakes need to reduce concentrations; and recommended phosphorus concentrations are not available for two lakes. For the noncompliant lakes, the percentage decrease in P concentration to meet the recommendations ranges from 6% to 97% (mean = 63%).

[20] We include the two nutrient variables in the regression model based on the same rationale for using both N and P concentrations as instrumental variables: without information on nutrient limitation by lake, we are not able to determine whether lakes are N limited, P limited, or co-limited by N and P. See footnote 11 for more detail.
[21] A spreadsheet that reports, by lake, actual P concentrations from the NLA and the matching P concentration in the respective Level III Subregion is available from the authors upon request.

We derived a benefit index that shows, by lake, the percentage increase in housing values (i.e., the incremental benefit) that would result from achieving the recommended P concentrations. The index is computed *for each lake* by applying three relationships. (1) Using the ecological result on $\hat{\beta}_1$, the percentage decrease in P concentration to meet the recommended level translates into a percentage increase in *SECCHI*. (2) This converts to an absolute increase in *SECCHI* based on the baseline *SECCHI* measurement from the NLA. (3) Using the hedonic price result on the effect of *SECCHI* on price, the *SECCHI* increase translates into a percentage increase in housing prices.

The benefit index shows that, for the 75 noncompliant lakes, the percentage increase in housing values ranges from 0.4% to 9.2%, with a mean of 3.3%. Panel A of Figure 1 plots the frequency of percentage increase for the 111 lakes, i.e., including the 36 zeroes for the compliant lakes. Panel B of Figure 1 maps the benefit index for the study lakes.

The methodology described above can also be used to inform a local water quality issue. To illustrate, we identified (using Google Maps) a population of 216 residential properties within the 0.1 mile buffer zone of Big Barbee Lake in northeastern Indiana, where water quality concerns have been identified (Bosch et al. 2015). Using Zillow® estimates for each property, the total value of these properties equals $39.837 million. A 52% reduction in P concentration in the lake is necessary to meet the EPA-recommended concentration of 10.0 µg/L. This reduction would increase *SECCHI* from the baseline of 1.85 meters to the improved 2.32 meters. This would translate into a 4.7% increase in housing values, i.e., 4.7% is the benefit index's level for Big Barbee Lake. When aggregated across the housing stock, the 4.7% increase equals $1.872 million. The projected $1.872 million in benefits could be compared to the costs of reducing P loadings to inform management policy at Big Barbee Lake. More generally, the methodology could be applied to the 75 noncompliant lakes and, with adequate data, to any U.S. lake.

5. Discussion and Conclusion

This research advances the literature in three ways. *First*, we developed the first instrumental-variables approach for assessing endogeneity bias when applying the hedonic price model to lake water quality and the housing market.[22] In principle, unobserved variables may covary with water clarity and housing price, e.g., variables such as the type of recreational infrastructure or level of recreational activity on a lake. While several water-quality studies apply spatial fixed effects to control for omitted variables, whether such fixed effects succeed in remedying bias is unknown. The additional concern about measurement error as a second source of bias in the context of lake water quality (Keiser 2019) only strengthens support for an approach that directly addresses potential endogeneity.

The instrumental-variables approach generated one main finding: despite the use of cross-sectional data, the OLS estimator yielded unbiased estimates of the effect of water clarity on housing price. The effect of water clarity was identified by the cross-sectional variation in Secchi disk depth at the 113 study lakes, such that the sample size of housing transactions was not a shortcoming in the analysis. The approach also developed suggestive evidence that three variables – P concentration in lake water, N concentration in lake water, and water temperature – were valid and strong instrumental variables. Because natural experiments and quasi-experiments are difficult to find with water pollution (Keiser and Shapiro, 2019a), future research can follow our approach to identifying the hedonic price function by using cross-sectional variation in an environmental quality variable in tandem with instrumental-variables methods. While our broad geographic scale breaks new ground on hedonic studies of lake water quality, a study with a larger number of lakes and a larger spatial area around each lake (i.e., observations beyond the 0.1 mile buffer of this study) could produce a truly national study of the economic value of lake water quality.

Second, we isolated water clarity as the attribute of lake water quality that is valued by lake-based property owners. The ecosystem services literature emphasizes that economic valuation should focus on a valued attribute of an ecosystem service (Boyd and Banzhaf 2007; Polasky and Segerson 2009; Keeler et al. 2012), i.e., an attribute that enters directly as an argument of the consumer demand function. For example, Keeler et al. (2012) argue that water clarity is the attribute valued by lakeshore homeowners. In previous research, however, valuation studies (e.g., Bockstael,

[22] Zhang, Boyle, and Kuminoff (2015) apply an instrumental-variables approach to address a related topic, the endogeneity problem in identifying amenity demand parameters from the hedonic price model's second stage. We focus solely on the first stage of the hedonic model.

McConnell, and Strand 1989; Egan, et al. 2009) have used water-quality variables that measure attributes that likely would be unknown to homeowners or recreationists, attributes such as N or P concentrations in the water, which are invisible to the typical consumer. Their use could have resulted in specification errors in the regression estimation.

Here, we showed in Section 3.2.1 that the estimated coefficients on the water-clarity variable are consistently robust to regression specifications that also included three biophysical variables associated with lake water quality, including green algal biovolume, cyanobacterial biovolume, and phytoplankton biovolume. Environmental concerns with lake water quality focus on the scientific relationship between nutrient runoff and biological growth, including outcomes such as harmful algal blooms from high concentrations of cyanobacteria. In the robustness checks, the estimated coefficients on the three biophysical variables are statistically *insignificant* while the estimated coefficients on the water-clarity variable are highly statistically significant and similar in magnitude to the coefficients in the main results. That is, homeowners appear not to value these attributes independently of water clarity. This finding provides empirical evidence that the biophysical variables do not affect housing price directly. In addition, these regressions suggest that water clarity could be an important omitted variable in a hedonic study of the effect of a cyanobacteria toxicity threshold on residential housing prices near inland lakes (Wolf and Klaiber 2017). Our results suggest that future work needs to focus on variables that potential homeowners both directly perceive and care about when purchasing a house.

Third and last, we developed a framework that merged an ecological production function and an economic valuation function, thus providing a new application of the ecosystem services paradigm (e.g., Keeler et al. 2012). Scientists and managers frequently use water quality metrics such as N and P concentration, while the public actually values different attributes such as water clarity. Our framework thus creates a tool to integrate economic valuation with policy prescriptions on nutrient concentrations, thereby providing a basis for improved decision-making on water quality and other ecosystem services (Bateman et al. 2011; Egan et al. 2009; Garnache et al. 2016). With the framework, we derived a benefit index that showed, by lake, the percentage increase in housing prices from achieving U.S. EPA-recommended P concentrations in the 113 study lakes. The benefit-index results provide new insight into the spatial heterogeneity of economic benefits as a basis to consider a spatially differentiated policy on water quality (Keiser and Shapiro 2019b).

While we developed a coupled framework, a fully comprehensive approach to studying phosphorus pollution would require linking four modeling frameworks: farmer decision-making on phosphorus use and management, watershed modeling, an ecological production function for lake water quality, and an economic valuation model. Implementing such an integrated framework would best be accomplished at the watershed scale. This is a topic for future research.

More research is also needed to understand the role of monitoring and measurement of the various dimensions of water quality. The sampling protocol followed in the NLA – described as “snapshot sampling” – has been deployed in Europe in addition to the U.S. to capture large-scale associations at the continental-level, despite the limitation of only having one or a few samples per lake (e.g., Stoddard et al. 2016; Mantzouki and Ibelings 2018; Mantzouki et al. 2018). At the same time, Keiser (2019) demonstrated that the NLA has measurement error relative to a long-term monitoring dataset of lakes in the state of Iowa. Understanding tradeoffs in monitoring approaches and their implications for water-quality valuation is a topic on which economics can inform natural science and resource management.

A final caveat to our study is that the valuation estimates are only partial measures of the benefits of improved lake water quality. They count benefits accruing to lake residents from ecosystem services that depend solely on water clarity, such as lake aesthetics and lake recreation opportunities. They do not, however, count nonuse benefits and benefits derived by other users of the lakes – most prominently, people who travel to lakes for various recreation activities. The economic benefits of improved water quality are sure to rise even higher as future research develops more comprehensive estimates of water-quality benefits at lakes across the United States.

References

Albouy D, Graf W, Kellogg R, Wolff H (2016) Climate amenities, climate change, and American quality of life. *J Asso Env Res Economists* 3(1):205-246.

Angrist JD, Pischke J-S (2009) *Mostly Harmless Econometrics* (Princeton University Press, Princeton, NJ).

Bajari P, Fruehwirth JC, Kim KI, Timmins C (2012) A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: the price of pollution. *Amer Econ Rev* 102(5):1898-1926.

Bateman IJ, Mace GM, Fezzi C, Atkinson G, Turner K (2011) Economic analysis for ecosystem service assessments. *Environ Resource Econ* 48:177-218.

Bayer P, Ferreira F, and McMillan R (2007) A unified framework for measuring preferences for schooling and neighborhoods. *J Pol Economy* 115(4):588-638.

Bayer P, Keohane N, Timmins C (2009) Migration and hedonic valuation: The case of air quality. *J Environ Econ Mgmt* 58:1-14.

Beaulieu M, Pick F, Gregory-Eaves I (2013) Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. *Limnol Oceanogr* 58:1736-1746.

Bin O, Czajkowski J (2013) The impact of technical and non-technical measures of water quality on coastal waterfront property values in South Florida. *Marine Res Econ* 28(1):43-63.

Bin O, Czajkowski J, Li J, Villarini G (2017) Housing market flucutations and the implicit price of water quality: empirical evidence from a South Florida housing market. *Environ Resource Econ* 68:319-341.

Bockstael NE, McConnell KE, Strand IE (1989) Measuring the benefits of improvements in water quality: The Chesapeake Bay. *Marine Res Econ* 6(1):1-18.

Bosch N, et al. (2015) Blue-green algae in northern Indiana lakes: an analysis of the algal toxin, mycrosystin, over 2010-2013 in lakes of Kosciusko County, Indiana. Working paper, Center for Lakes and Streams, Grace College.

Boyd J, Banzhaf S (2007) What are ecosystem services? The need for standardized environmental accounting units. *Ecol Econ* 63:616-626.

Chay KY, Greenstone M (2005) Does air quality matter? Evidence from the housing market. *J Pol Econ* 113(2):376-424.

Currie J, Davis L, Greenstone M, Walker R (2015) Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *Amer Econ Rev* 105(2):678-709.

Davis LW (2011) The effect of power plants on local housing values and rents. *Rev Econ and Stat* 93(4):1391-1402.

Doubek JP, Carey CC, Cardinale BJ (2015) Anthropogenic land use is associated with N-fixing cyanobacterial dominance in lakes across the continental United States. *Aquat Sci* 77:681-694.

Egan KJ, Herriges JA, Kling CL, Downing JA (2009) Valuing water quality as a function of water quality measures. *Am J Agric Econ* 91:106-123.

Filstrup CT, Wagner T, Oliver SK, Stow CA, Webster KE, Stanley EH, and Downing JA (2018) Evidence for regional nitrogen stress on chlorophyll *a* in lakes across large landscape and climate gradients. *Limnology and Oceanography* 63:S324–S339.

Freeman, AM III, Herriges JA, Kling CL (2014) *The Measurement of Environmental and Resource Values*, 3rd Edition (RFF Press, Washington, DC).

Garnache, C, Swinton, SM, Herriges, JA, Lupi, F, and Stevenson RJ (2016) Solving the phosphorus pollution puzzle: Synthesis and directions for future research. *Am J Agric Econ* 98:1334-1359.

Greenstone M, Gallagher J (2008) Does hazardous waste matter? Evidence from the housing market and the Superfund program. *Quart J Econ* 123(3):951-1003.

Ho, JC, Michalak AM, Pahlevan N (2019) Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* 574:667-670.

Horsch EJ, Lewis DJ (2009) The effects of aquatic invasive species on property values: Evidence from a quasi-experiment. *Land Econ* 85(3):391-409.

Keeler BL et al. (2012) Linking water quality and well-being for improved assessment and valuation of ecosystem services. *Proc Nat Acad Sci* 109(45):18619-18624.

Keiser DA (2019) The missing benefits of clean water and the role of mismeasured pollution. *J Assoc Env Res Economists* 6(4):669-707.

Keiser DA, Shapiro JS (2019a) Burning water to crystal springs? U.S. water pollution regulation over the last half century. *J Econ Perspectives* 33(4):51-75.

Keiser DA, Shapiro JS (2019b) Consequences of the U.S. Clean Water Act and the demand for water quality. *Quart J Econ* 134(1):349-396.

Klemick H, Griffiths C, Guignet D, Walsh P (2018) Improving water quality in an iconic estuary: An internal meta-analysis of property value impacts around the Chesapeake Bay. *Environ Res Econ* 69:265-292.

Kuminoff NV, Parmeter CF, Pope JC (2010) Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *J Env Econ Mgmt* 60:145-160.

Lewis, ASL, Kim BS, Edwards HL, Wander HL, Garfield CM, Murphy HE, Poulin ND, Princiotta SD, Rose KC, Taylor AE, Weathers KC, Wigdahl-Perry CR, Yokota K, Richardson DC, and Bruesewitz DA (2020) Prevalence of phytoplankton limitation by both nitrogen and phosphorus related to nutrient stoichiometry, land use, and primary producer biomass across the northeastern United States. *Inland Waters*. DOI: doi: 10.1080/20442041.2019.1664233

Lorenzen MW (2003) Use of chlorophyll-Secchi disk relationships. *Limnol Oceanogr* 25:371-372.

Mantzouki E, Beklioğlu M, Brookes JD, de Senerpont Domis LN, Dugan HA, Doubek JP, Grossart H-P, Nejstgaard JC, Pollard AI, Ptacnik R, Rose KC, Sadro S, Seelen L, Skaff NK, Teubner K, Weyhenmeyer GA, Ibelings BW (2018) Snapshot surveys for lake monitoring, more than a shot in the dark. Front Ecol Evol 6:10.3389/fevo.2018.00201.

Mantzouki E, Ibelings BW (2018) The principle and value of the European multi lake survey. *Limnol Oceanogr Bull* 27:82-86.

Michael HJ, Boyle KJ, Bouchard R (2000) Does the measurement of environmental quality affect implicit prices estimated from hedonic models? *Land Econ* 76(2):283-298.

Muehlenbachs L, Spiller E, Timmins C (2015) The housing market impacts of shale gas development. *Amer Econ Rev* 105(12):3633-3659.

National Research Council, Committee on Assessment of Water Resources Research (2004) *Confronting the Nation's Water Problems* (The National Academies Press, Washington DC).

National Research Council, Committee on Water Resources Activities at the U.S. Geological Survey (2009) *Toward a Sustainable and Secure Water Future* (The National Academies Press, Washington DC).

National Research Council, Committee on Economic and Environmental Impacts of Increasing Biofuels Production (2011) *Renewable Fuel Standard: Potential Economic and Environmental Effects of U.S. Biofuel Policy* (The National Academies Press, Washington DC).

Olmstead, S (2010) The economics of water quality. *Rev Env Econ Policy* 4(1):44-62.

Polasky S, Segerson K (2009) Integrating ecology and economics in the study of ecosystem services: some lessons learned. *Annu Rev Resour Econ* 1:409-434.

Read EK, Patil VP, Oliver SK, Hetherington AL, Brentrup JA, Zwart JA, Winters KM, Corman JR, Nodine ER, Woolway RI, Dugan HA, Jaimes A, Santoso AB, Hong GS, Winslow LA, Hanson PC, Weathers KC (2015) The importance of lake-specific characteristics for water quality across the continental United States. *Ecol Appl* 25:943-955.

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in perfect competition. *J Polit Econ* 82:34-55.

Schindler, DW (1977) Evolution of phosphorus limitation in lakes. *Science* 195: 260–262.

Small KA (1975) Air pollution and property values: a further comment. *Rev Econ Stat* 58:105-107.

Smith, VH (1983) Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* 221: 669–671.

Staiger D, Stock J (1997) Instrumental variables regression with weak instruments. *Econometrica* 65:557-586.

Steffen W, et al. (2015) Planetary boundaries: guiding human development on a changing planet. *Science* 347(6223):736-746.

Stoddard JL, Van Sickle J, Herlihy AT, Brahney J, Paulsen S, Peck DV, Mitchell R, Pollard AI (2016) Continental-scale increase in lake and stream phosphorus: are oligotrophic systems disappearing in the United States? *Environ Sci Technol* 50:3409-3415.

Tuttle CM, Heintzelman MD (2015) A loon on every lake: a hedonic analysis of lake water quality in the Adirondacks. *Resour Energy Econ* 39:1-15.

U.S. Environmental Protection Agency (2009) *An Urgent Call to Action: Report of the State-EPA Nutrient Innovations Task Group*.

U.S. Environmental Protection Agency (2010) *National Lakes Assessment: A Collaborative Survey of the Nation's Lakes*, EPA 841-R-09-001, Washington, DC.

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

U.S. Environmental Protection Agency, Office of Water (2000) Ecoregional Nutrient Criteria for Lakes & Reservoirs; various ecoregions. https://www.epa.gov/nutrient-policy-data/ecoregional-nutrient-criteria-lakes-reservoirs.

Walsh PJ, Milon JW, Scrogin DO (2011) The spatial extent of water quality benefits in urban housing markets. *Land Econ* 87(4):628-644.

Walsh PJ, Griffiths C, Guignet D, Klemick H (2017) Modeling the property price impact of water quality in 14 Chesapeake Bay counties. *Ecol Econ* 135:103-113.

Wolf D, Klaiber HA (2017) Bloom and bust: Toxic algae's impact on nearby property values. *Ecol Econ* 135:209-221.

Wooldridge, JM (2009) *Introductory Econometrics* (South-Western Cengage Learning; Mason, OH).

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

Xu, H, Paerl HW, Qin B, Zhu G, and Gao G (2010) Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnology and Oceanography* 55: 420–432.

Zhang C, Boyle KJ (2010) The effect of an aquatic invasive species (Eurasian watermilfoil) on lakefront property values. *Ecol Econ* 70:394-404.

Zhang C ,Boyle KJ, Kuminoff NV (2015) Partial identification of amenity demand functions. *J Env Econ Mgmt* 71:180-197.

Figure 1, Panel A. Frequency distribution of the benefit index for 111 study lakes (mean = 2.2%; minimum = 0%; maximum = 9.2%). The index estimates, by lake, the percentage increase in housing prices after a hypothetical total phosphorus reduction to meet the recommended phosphorus concentration for the lake. Thirty-six lakes meet the recommendation and thus are at 0%. Recommended phosphorus concentrations are not available for two lakes.

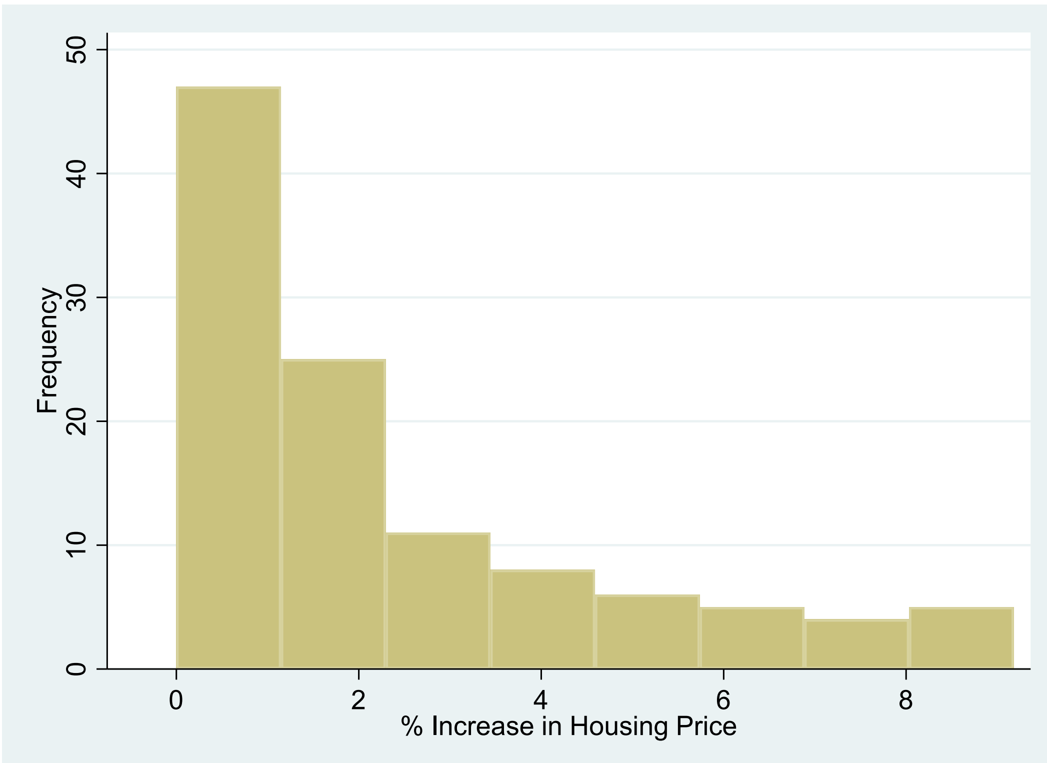

Figure 1, Panel B. Map of United States shows variation, by lake, in benefit index

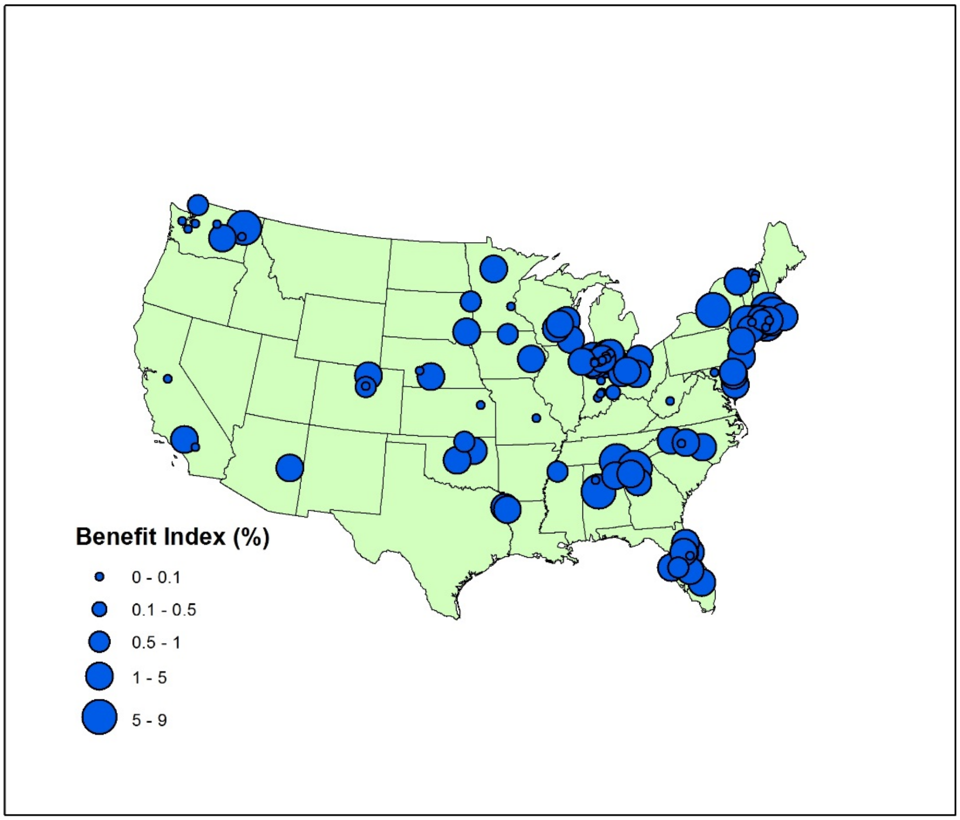

TABLE 1
Variable Definitions and Descriptive Statistics

Variable	Description	Units	Mean	St. dev.	Maximum	Minimum
Housing Price	Transaction price of house in 2013$	$	401,146	534,757	5,160,907	4,148
ln Housing Price	Log-transformed transaction price	$	12.40	0.97	15.5	8.3
SECCHI	Water clarity	meters	2.1	1.9	9.5	0.2
LKFRNT	Whether house was lakefront	0 or 1	0.6	0.5	1.0	0.0
LK_DSTNC	House distance to lake	feet	234	145	530	10
Instrumental variables:						
NITROGEN	Total nitrogen	µg/L	972	1066	6672	37
PHSPHRS	Total phosphorus	µg/L	69	124	819	1
TEMP	Lake water temperature	˚C	25.9	3.0	32.2	19.4
House covariates:						
BDRMS	Number of bedrooms	---	3.2	1.0	9.0	1.0
BTHRMS	Number of bathrooms	---	2.4	1.1	8.0	0.5
SQFT	Square footage of house	ft^2	2221	1265	10271	192
YR_BUILT	Year house was built	---	1972	25	2012	1850
Lake covariates:						
LK_AREA	Area of the lake	km^2	40.2	192.0	1674.9	0.1
LK_PRMTR	Perimeter of the lake	km	49.9	120.6	962.2	1.2
Neighborhood covariates:						
CMMT_TM	Commuting time to work	minutes	25.2	5.3	40.2	13.3
PVRTY	Proportion of population below poverty level	%	11.1	6.5	36.9	1.5
EDUCTN	Proportion of individuals with bachelor degree or above	%	24.4	13.8	4.2	76.1
MED_INCM	Medium income	$	56813	19584	159713	26313
%AFRAMER	Proportion African-American	%	5.6	10.4	0.0	59.3
RENTAL	Proportion rental properties	%	24.4	10.8	4.7	66.4

POP_DNSTY	Population density	person/acre	0.9	1.6	0.01	8.8
UNEMPLD	Proportion unemployed	%	8.7	4.0	29.1	2.1
DVLPD_LND	Proportion developed land	%	16.2	15.3	77.2	0.0
AG_LND	Proportion agricultural land	%	21.5	16.5	77.2	0.0
CLMT_HDD	Long-term average of annual heating degree days	˚F	5334	2028	9787	459
CLMT_PRCP	Long-term average of annual precipitation	hundredths of inches	4227	1236	7684	805
Other water-quality variables:						
PHPLNKTN	Total phytoplankton biovolume	$\mu m^3/mL$	4.3×10^6	8.5×10^6	7.0×10^7	1.6×10^4
CYANBCTR	Total cyanobacterial biovolume	$\mu m^3/mL$	1.5×10^6	3.3×10^6	2.1×10^7	739
GRNALGAE	Total green algal biovolume	$\mu m^3/mL$	2.3×10^6	7.4×10^6	6.9×10^7	114

Notes: N = 1,462 for house attribute data, including Housing Price, ln Housing Price, LKFRNT, LK_DSTNC, BDRMS, BTHRMS, SQFT, and YR_BUILT. N = 113 (one observation per lake) for water quality, lake, and neighborhood variables (i.e., all other variables).

TABLE 2

Effects of Secchi Disk Depth, Lakefront, and Distance-to-Lake on *Ln Housing Price*

	OLS (1a)	IV-GMM (1b)	OLS (2a)	IV-GMM (2b)	OLS (3a)	IV-GMM (3b)	OLS (4a)	IV-GMM (4b)
SECCHI	0.151***	0.154***	0.114***	0.117***	0.121***	0.126***	0.099***	0.101***
	(0.019)	(0.027)	(0.016)	(0.026)	(0.016)	(0.025)	(0.019)	(0.037)
LKFRNT	0.322***	0.362***	0.383***	0.397***	0.390***	0.401***	0.425***	0.425***
	(0.061)	(0.058)	(0.055)	(0.053)	(0.057)	(0.055)	(0.057)	(0.056)
LK_DSTNC	-0.00035	-0.00049**	-0.00062***	-0.00063***	-0.00058**	-0.00058***	-0.00064***	-0.00064***
	(0.00025)	(0.00023)	(0.00023)	(0.00022)	(0.00022)	(0.00021)	(0.00017)	(0.00017)
House covariates	-------- Yes ---------		-------- Yes --------		--------- Yes ---------		--------- Yes ---------	
Lake and neighborhood Covariates	-------- No ---------		-------- Yes --------		--------- Yes ---------		--------- Yes ---------	
Flexible form of house Covariates	-------- No --------		-------- No --------		--------- Yes ---------		--------- Yes ----------	
State fixed effects	-------- No --------		-------- No --------		--------- No --------		--------- Yes ----------	
Year fixed effects	-------- Yes --------		-------- Yes --------		--------- Yes ---------		-------- Yes ---------	
Quarter fixed effects	-------- Yes --------		-------- Yes --------		--------- Yes ---------		-------- Yes ---------	
R^2	0.616	0.614	0.674	0.673	0.688	0.688	0.731	0.731

Notes: *N* = 1,462 observations. **OLS** represents ordinary-least-squares regression. **IV-GMM** represents instrumental-variables generalized-method-of-moments regression with the potential endogenous variable *SECCHI*. Three variables serve as instruments: *NITROGEN* (nitrogen), *PHSPHRS* (phosphorus), and *TEMP* (lake temperature); except in regression (4b), in which only *NITROGEN* and *TEMP* are valid instruments. Regressions develop cluster-robust estimates, with 113 lakes as clusters. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. The flexible functional form of house covariates includes quadratics, cubics, and interactions of the variables as controls.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 3
Tests of *SECCHI* and Its Instrumental Variables

	Specification of IV-GMM regression			
	House covariates (1)	Add: lake and neighborhood covariates (2)	Add: flexible form of house covariates (3)	Add: state fixed effects (4)
(a) Tests of overidentifying restrictions (Hansen's *J* statistic)				
	$\chi^2(2) = 3.441$	$\chi^2(2) = 1.883$	$\chi^2(2) = 2.183$	$\chi^2(1) = 0.015$
	($p = 0.179$)	($p = 0.390$)	($p = 0.336$)	($p = 0.903$)
(b) Endogeneity tests (*C* tests)				
	$\chi^2(1) = 0.041$	$\chi^2(1) = 0.012$	$\chi^2(1) = 0.038$	$\chi^2(1) = 0.010$
	($p = 0.839$)	($p = 0.914$)	($p = 0.845$)	($p = 0.921$)
(c) Tests of weak instruments				
	$F(3,113) = 8.51$	$F(3,113) = 18.02$	$F(3,113) = 17.56$	$F(2,113) = 18.58$
	(prob>F = 0.00)	(prob>F = 0.00)	(prob>F = 0.00)	(prob>F = 0.00)

Notes: Regressions estimated using the instrumental-variables generalized-method-of-moments estimator and cluster-robust variance-covariance matrices. Three variables serve as instruments: *NITROGEN* (nitrogen), *PHSPHRS* (phosphorus), and *TEMP* (lake temperature); except in regression (4), in which only *NITROGEN* and *TEMP* are valid instruments. Cluster-robust estimates developed using 113 lakes as clusters. Year fixed effects and quarter fixed effects included in regressions.

TABLE 4
Effects of *SECCHI* and Other Water-Quality Variables on *Ln Housing Price*

	(1)			(4)		
	(a)	(b)	(c)	(a)	(b)	(c)
SECCHI	0.151***	0.151***	0.151***	0.096***	0.094***	0.098***
	(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)
PHPLNKTN	-6.13e-11	---	---	-4.06e-09	---	---
	(3.59e-09)			(3.13e-09)		
CYANBCTR	---	1.32e-09	---	---	-9.59e-09	---
		(1.23e-08)			(8.57e-09)	
GRNALGAE	---	---	8.12e-10	---	---	-3.21e-09
			(4.72e-09)			(3.12e-09)
House covariates	- - - - - - - - - - - Yes - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		
Lake and neighbor-hood covariates	- - - - - - - - - - - No - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		
Flexible form of house covariates	- - - - - - - - - - - No - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		
State fixed effects	- - - - - - - - - - - No - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		
Year fixed effects	- - - - - - - - - - - Yes - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		
Quarter fixed effects	- - - - - - - - - - - Yes - - - - - - - - - - -			- - - - - - - - - - - Yes - - - - - - - - - - -		

Notes: *N* = 1,462 observations. OLS estimator. Specification (2) and Specification (3) of the covariates (as defined in the text and several other tables) are not reported since they do not add insight to the topic. Cluster-robust standard errors are reported in parentheses below the coefficient estimates.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 5
State-based Effects of *SECCHI* on *Ln Housing Price*

	(1)	(2)	(3)	(4)
SECCHI	0.153***	0.128***	0.138***	0.111***
	(0.017)	(0.023)	(0.021)	(0.023)
*SECCHI*FLORIDA*	-0.118	0.407	0.358	0.251
	(0.310)	(0.278)	(0.268)	(0.478)
*SECCHI*INDIANA*	-0.068	-0.029	-0.041	-0.072
	(0.062)	(0.066)	(0.063)	(0.065)
*SECCHI*WASHINGTON*	-0.017	-0.020	-0.030	-0.007
	(0.037)	(0.030)	(0.028)	(0.040)
House covariates	Yes	Yes	Yes	Yes
Lake and neighborhood covariates	No	Yes	Yes	Yes
Flexible form of house covariates	No	No	Yes	Yes
State fixed effects	No	No	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes

Notes: *N* = 1,462 observations. OLS estimator. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. Florida has 10 lakes and 151 observations. Indiana has 25 lakes and 238 observations. Washington has 9 lakes and 178 observations.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 6
Effects of Secchi Disk Depth (*SECCHI*) on *Ln Housing Price*,
OLS Regressions using 2007 NLA Data versus 2012 NLA Data

	(1)	**(2)**	**(3)**	**(4)**
SECCHI	0.210***	0.090***	0.103***	0.085***
2007 NLA Data	(0.045)	(0.031)	(0.025)	(0.029)
SECCHI	0.186***	0.070*	0.084**	0.089***
2012 NLA Data	(0.047)	(0.035)	(0.034)	(0.031)
House covariates	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -
Lake and neighborhood covariates	- - - - No - - - -	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -
Flexible form of house covariates	- - - - No - - - -	- - - - No - - - -	- - - - Yes - - - -	- - - - Yes - - - -
State fixed effects	- - - - No - - - -	- - - - No - - - -	- - - - No - - - -	- - - - Yes - - - -
Year fixed effects	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -
Quarter fixed effects	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -	- - - - Yes - - - -

Notes: *N* = 600 observations. **OLS** represents ordinary-least-squares regression. Regressions develop cluster-robust estimates, with 38 lakes as clusters. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. The flexible functional form of house covariates includes quadratics, cubics, and interactions of the variables as controls.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

ONLINE APPENDIX MATERIALS

Appendix A: Main Regressions: Additional Details

(a) First-stage regression results. One requirement for an instrumental variable is that, conditional on the other exogenous variables, it is correlated with the endogenous variable. From an ecological perspective, the three candidates for instruments – *NITROGEN*, *PHSPHRS*, and *TEMP* – increase net primary production in a lake ecosystem by providing nutrients and ambient growing conditions. In the case of a lake, net primary production involves phytoplankton production, and water clarity declines with an increase in phytoplankton production. The expectation, then, is that the estimated coefficients on *NITROGEN*, *PHSPHRS*, and *TEMP* will be negative in a regression that explains variation in *SECCHI*.

We report first-stage results based on the procedure (described in the main text and reported in Table 2) of sequentially adding covariates to the IV-GMM regressions to develop four standard specifications for the analysis (Appendix Table 1). In the first-stage regressions, the estimated coefficients on *NITROGEN*, *PHSPHRS*, and *TEMP* are negative and statistically significant at the 1 percent level in the first three specifications. In the fourth specification, the coefficient on *PHSPHRS* is not statistically significant. Column (4) thus reports the results with the two remaining instruments, in which the estimated coefficients on *NITROGEN* and *TEMP* are negative and statistically significant at the 1 percent level. These results demonstrate that the instruments meet the requirement of correlation with *SECCHI*.

In the main text, additional test results demonstrate that the variables are valid instruments and are not weak instruments. Thus, the IV-GMM estimator is overidentified with *SECCHI* as a potential endogenous variable and with the high-quality instruments in *NITROGEN*, *TEMP*, and, in several cases, *PHSPHRS*.

(b) OLS: complete results. The main text and Table 2 report hedonic price results for only three variables, *SECCHI*, *LKFRNT*, and *LK_DSTNC*. In the appendix, we describe results on other variables and report estimated coefficients for the entire set of exogenous variables in Appendix Table 2.

The expectation is that the estimated coefficients on the house covariates will be positive in the hedonic price regression. For example, *YR_BUILT* increases as the house was built more recently, and a newer house should come with a price premium. As reported in Appendix Table 2, estimated coefficients on the house covariates are positive and statistically significant in the OLS regressions for specifications (1) and (2). (We primarily report OLS results in Appendix A due to the finding that exogeneity could not be rejected.) Specifications (3) and (4) apply a flexible functional form for the house covariates that includes quadratics, cubics, and interactions of the variables. We do not report those results as their purpose is to provide many control variables, not to provide informative parameter estimates.

Estimated coefficients on 7 of the 14 lake and neighborhood variables are statistically significant in the OLS regressions (Appendix Table 2). Housing price declines in *LK_AREA*, suggesting that homeowners prefer smaller lakes to larger lakes; yet it increases in *LK_PRMTR*, suggesting that homeowners pay a premium for an irregular shoreline. *AG_LAND* near the lake results in a discount. We model socioeconomic characteristics of the neighborhood as local public goods. Housing price increases in the educational (*EDUCTN*) and median income (*MED_INCM*) composition of the neighborhood. Housing price also increases in *CLMT_HDD*, which implies that housing price decreases as temperature increases given the definition of *CLMT_HDD*. Curiously, it decreases in the rate of rental property in the neighborhood (*%RENTAL*).

Of note, we did not estimate regressions with a flexible functional form for the lake and neighborhood covariates since these variables did not perform as strongly as the house covariates.

(c) Ecological production function: estimation results: Appendix Table 3 reports results of the OLS estimates of the ecological model using a Cobb-Douglas functional form. The results show robustness in the parameter of interest, the estimated coefficient on the ln(phosphorus) variable. The coefficient is estimated at 0.49 in specifications without and with the ln(temperature) variable. With the addition of three state fixed effects, the coefficient is estimated at 0.51. In Section 4 of the main text, we apply 0.49 in the quantitative analysis related to lake-specific phosphorus reductions because it is the more conservative estimate.

Appendix B: Additional Robustness Checks

Appendix B reports two additional robustness checks on the *SECCHI* variable along with a robustness check on the IV-GMM estimator.

(a) Does SECCHI exert a diminishing marginal effect on housing price? A conjecture in the literature is that water quality may have strongly diminishing marginal benefits (Keeler et al., 2012). To investigate this, we estimate the hedonic price function using a model with *Housing Price* (not *Ln Housing Price*) as the dependent variable and with *SECCHI* entering with linear and squared terms (Appendix Table 4). Use of a quadratic function (i.e., a squared term for a variable in addition to a linear term) in regression modeling is a standard approach for evaluating the possibility of curvature of a functional relationship. According to a well-known econometric textbook (Wooldridge 2009, p. 192), "Quadratic functions are also used quite often in applied economics to capture decreasing or increasing marginal effects."

The estimated coefficients on the *SECCHI* and *SECCHI***SECCHI* variables are not statistically significant in any specification. Our case, with water clarity as the measure of water quality, provides no support for diminishing marginal benefits in the hedonic price model.

For additional context, we also estimate (but do not report) specifications with *Housing Price* (not *Ln Housing Price*) as the dependent variable and *SECCHI* entering with only a linear term. The estimated coefficients on *SECCHI* in these specifications are positive and statistically significant at the 1 percent level.

(b) Robustness check on the IV-GMM estimator. Results from IV-GMM regressions, to this point, have been reported to represent the instrumental-variables approach. The two-stage-least-squares estimator (2SLS) is an alternate IV estimator, and the expectation is for a strong concordance between the 2SLS and IV-GMM results. Comparative results are reported for the three main variables of interest, *SECCHI*, *LKFRNT*, and *LK_DSTNC* (Appendix Table 5). The estimated coefficients and standard errors in the paired IV-GMM and 2SLS regressions are very similar, and they are virtually identical in specification (4) with all the controls in place. The estimated coefficients are clearly robust to the 2SLS estimator.

APPENDIX TABLE 1
First-Stage Results for Instrumental Variables in IV-GMM Regressions

	(1)	(2)	(3)	(4)
NITROGEN	-0.00050***	-0.00054***	-0.00054***	-0.00055***
	(0.00017)	(0.00011)	(0.00011)	(0.00013)
PHSPHRS	-0.00344***	-0.00361***	-0.00364***	-----
	(0.00122)	(0.00088)	(0.00087)	
TEMP	-0.3163***	-0.3606***	-0.3574***	-0.3268***
	(0.0795)	(0.0722)	(0.0730)	(0.0689)
House covariates	Yes	Yes	Yes	Yes
Lake and neighbor-hood covariates	No	Yes	Yes	Yes
Flexible form of house covariates	No	No	Yes	Yes
State fixed effects	No	No	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
R^2	0.546	0.711	0.717	0.843

Notes: *N* = 1,462 observations. Dependent variable is *SECCHI*. IV-GMM is the instrumental-variables generalized-method-of-moment estimator. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. *PHSPHRS* is not a valid instrument in specification (4). The flexible functional form of house covariates includes quadratics, cubics, and interactions of the variables as controls.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

APPENDIX TABLE 2
OLS Results:
House, Lake, and Neighborhood Covariates

	(1)	(2)
SECCHI	0.151***	0.114***
	(0.019)	(0.016)
LKFRNT	0.322***	0.383***
	(0.061)	(0.055)
LK_DSTNC	-0.00035	-0.00062***
	(0.00025)	(0.00023)
House covariates:		
BDRMS	0.102***	0.056**
	(0.035)	(0.028)
BTHRMS	0.168***	0.148***
	(0.035)	(0.030)
SQFT	0.00023***	0.00023***
	(0.00005)	(0.00004)
YR_BUILT	0.00228*	0.00319***
	(0.00132)	(0.00113)
Lake and neighbor-hood covariates:		
LK_AREA	---	-0.00040***
		(0.00009)
LK_PRMTR	---	0.00069***
		(0.00024)
DVLPD_LND	---	0.0045
		(0.0033)
AG_LND	---	-0.0147**
		(0.0069)
CMMT_TM	---	0.0119
		(0.0072)
PVRTY	---	0.0058
		(0.0087)
EDUCTN	---	0.0066*
		(0.0036)
MED_INCM	---	8.93e-06***
		(2.88e-06)
%AFRAMER	---	-0.5185
		(0.5571)
RENTAL	---	0.9591*
		(0.4858)
POP_DNSTY	---	-44.450
		(163.279)
UNEMPLD	---	-0.0047
		(0.0127)
CLMT_HDD	---	0.00004

		(0.00003)
CLMT_PRCP	---	1.73e-06
		(0.00002)
Flexible form of house covariates	No	No
State fixed effects	No	No
Year fixed effects	Yes	Yes
Quarter fixed effects	Yes	Yes
R^2	0.616	0.674

Notes: *N* = 1,462 observations. Dependent variable is *Ln Housing Price*. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. Specifications (3) and (4) are not reported because "flexible form of house covariates", while useful as controls, generates results that are not informative.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

APPENDIX TABLE 3

OLS Results: Cobb-Douglas Ecological Function

Intercept	2.78***	7.56***	5.62***
	(0.44)	(1.38)	(1.72)
LN_PHOSPHORUS	-0.49***	-0.49***	-0.51***
	(0.06)	(0.06)	(0.06)
LN_NITROGEN	-0.13	-0.08	-0.03
	(0.09)	(0.09)	(0.09)
LN_TEMP	---	-1.57***	-1.03**
		(0.42)	(0.51)
Florida FE	---	---	-0.33**
			(0.14)
Indiana FE	---	---	-0.13
			(0.11)
Washington FE	---	---	0.20
			(0.13)
R^2	0.72	0.75	0.76
N	113	113	113

Notes: Dependent variable is *LN_SECCHI*. The observations are from the 113 study lakes. *FE* variables add state fixed effects for states with several lakes (Florida = 10; Indiana = 25; Washington = 8). Robust standard errors reported in parentheses.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

APPENDIX TABLE 4
Effect of Quadratic Form for *SECCHI* on *Housing Price*

	(1)	(2)	(3)	(4)
SECCHI	42,786	18,593	15,245	26,670
	(39,495)	(33,334)	(30,497)	(40,272)
*SECCHI*SECCHI*	5,864	6,800	6,587	4,258
	(5,588)	(4,287)	(3,974)	(4,433)
House covariates	Yes	Yes	Yes	Yes
Lake and neighbor-hood covariates	No	Yes	Yes	Yes
Flexible form of house covariates	No	No	Yes	Yes
State fixed effects	No	No	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes

Notes: *N* = 1,462 observations. Dependent variable is *Housing Price*, not *Ln Housing Price*. OLS estimator. Cluster-robust standard errors are reported in parentheses below the coefficient estimates. House covariates in this case include *LKFRNT* and *LK_DSTNC*.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

APPENDIX TABLE 5
Comparing 2SLS and IV-GMM Regression Results

	2SLS (1A)	**IV-GMM** (1B)	**2SLS** (2A)	**IV-GMM** (2B)	**2SLS** (3A)	**IV-GMM** (3B)	**2SLS** (4A)	**IV-GMM** (4B)
SECCHI	0.175***	0.154***	0.120***	0.117***	0.128***	0.126***	0.102***	0.101***
	(0.029)	(0.027)	(0.027)	(0.026)	(0.025)	(0.025)	(0.037)	(0.037)
LKFRNT	0.320***	0.362***	0.383***	0.397***	0.390***	0.401***	0.424***	0.425***
	(0.063)	(0.058)	(0.055)	(0.053)	(0.056)	(0.055)	(0.056)	(0.056)
LK_DSTNC	-0.00038	-0.00049**	-0.00062***	-0.00063***	-0.00057***	-0.00058***	-0.00064***	-0.00064***
	(0.00024)	(0.00023)	(0.00023)	(0.00022)	(0.00022)	(0.00021)	(0.00017)	(0.00017)
House covariates	-------- Yes ---------		-------- Yes --------		--------- Yes ---------		--------- Yes ---------	
Lake and neighborhood covariates	-------- No ---------		-------- Yes --------		--------- Yes ---------		--------- Yes ---------	
Flexible form of house covariates	-------- No --------		-------- No --------		--------- Yes ---------		--------- Yes ----------	
State fixed effects	-------- No --------		-------- No --------		--------- No --------		--------- Yes ----------	
Year fixed effects	-------- Yes --------		-------- Yes --------		--------- Yes ---------		-------- Yes ---------	
Quarter fixed effects	-------- Yes --------		-------- Yes --------		--------- Yes ---------		-------- Yes ---------	
R^2	0.614	0.614	0.674	0.673	0.688	0.688	0.731	0.731

Notes: *N* = 1,462 observations. Dependent variable is *Ln Housing Price*. **2SLS** is the two-stage-least-squares estimator. **IV-GMM** is the two-stage instrumental-variables generalized-method-of-moments estimator. The potential endogenous variable is *SECCHI*. Three variables serve as instruments: *NITROGEN* (nitrogen), *PHSPHRS* (phosphorus), and *TEMP* (lake temperature); except in regression (4B), in which only *NITROGEN* and *TEMP* are valid instruments. Cluster-robust standard errors are reported in parentheses below the coefficient estimates.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

--- This page is intentionally left blank. ---