Disentangling Sources of Water Quality Variation in Property Sale Models

ACES 2024
December 11, 2024

Kristen Swedberg (ORISE @ EPA), Nathaniel H. Merrill (USEPA), Joel Corona (USEPA), Matthew T. Heberling (USEPA)

Background

- Homeowners value the water quality nearby their homes and are key beneficiaries of policies targeting improvements in water quality
- Property sale models allow economists to understand how much water quality factors into sale prices and monetize the benefits of water quality improvements to homebuyers
- These models analyze the relationship between variation in water quality and sale prices to predict how housing prices may change if water quality improved
  - $Sale\ Price = f(Water\ Quality, Control\ Variables)$
- Different sources of water quality variation can be modelled
  - Differences in water clarity between two nearby lakes (Mamun et al., 2022)
  - Changes in dissolved oxygen over time (Kuwayama et al., 2022)
  - Replacement of lead service lines (Theising, 2019)

Motivation

- To evaluate the potential benefits of policy applications, models that utilize water quality changes over time preferred.
  - The ideal scenario evaluates the impacts of policy applications ex-post (e.g. Theising, 2019)
  - A second best scenario is to isolate temporal variation in a repeat sales analysis (e.g. Kuwayama et al., 2022)
- Temporal variation can be difficult to model
  - Water quality changes resulting from policy may be small and unobservable (e.g. USEPA, 2023)
  - For some water quality measures, there may be limited variation in the short term
  - Over longer periods of time, there may be large shocks to housings markets that confound estimation
  - Repeat sales represent only a small share of all sales
- Spatial variation is simpler
  - Makes intuitive sense from the perspective of a homebuyers
  - Spatial variation may exceed year to year variation
  - More data available for added model efficiency

Motivation

- Because studies choose one path or the other it is unclear how different sources of water quality variation impact modelled results
  - If the results are consistent between both sources of variation, either can be used for policy analysis
  - If the results differ between sources of variation, then researchers may want to isolate temporal variation
- Research Aims
  - Identify spatial and temporal sources of water quality variation for different types of water bodies and measures in Southern New England (MA and RI)
  - Evaluate the effect of each source of water quality variation on property sale model estimates
  - Demonstrate the impacts on benefits estimates for potential water quality improvements

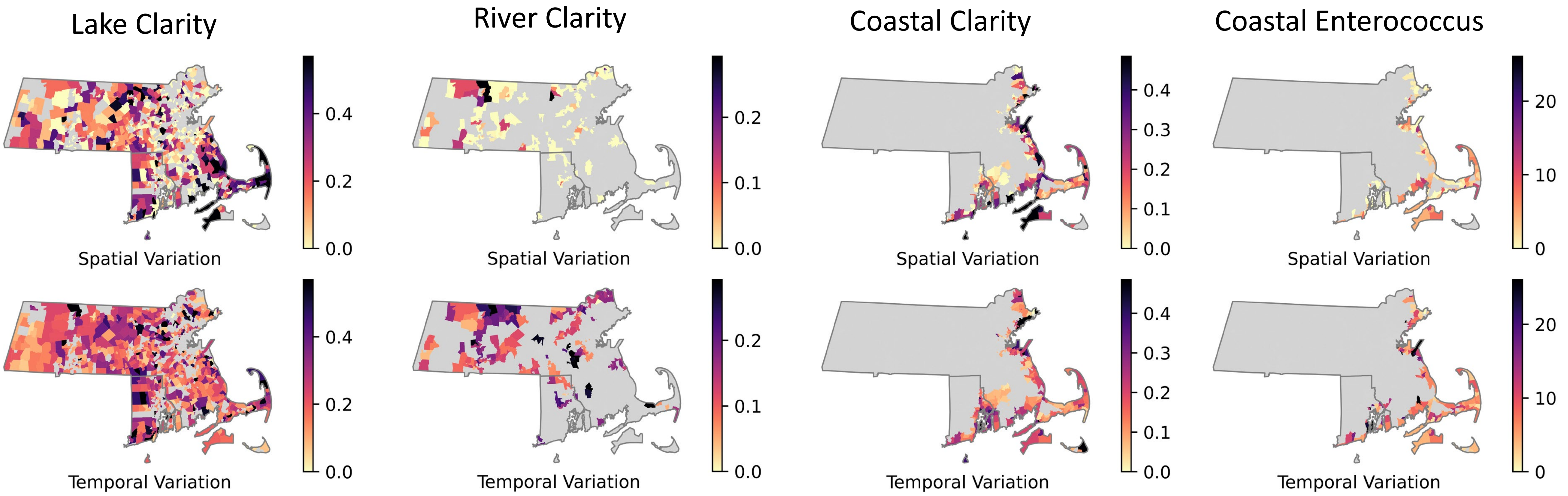
Overview

- Water Quality
  - Data
  - Sources of Variation
- Property Sale Model
  - Data
  - Model
  - Results
- Benefits Scenarios
  - Scenarios
  - Results

Water Quality Data

- Water Clarity
  - Remotely sensed (Landsat 8) observations for June-September of 2014-2020
  - Rasters summarized across individual NHD lakes (> 1ha), NHD rivers (segmented by HUC12 & > 1ha), and NOAA shorelines (buffered by ¼ mile) for each year
- Enterococcus
  - EPA BEACHES program for June-September of 2014-2020
  - Averaged across nearest NOAA shoreline each year
- Isolating Sources of Variation
  - Spatial Variation: Average water quality for each water feature from 2014-2020
  - Temporal Variation: Average water quality within Census Tracts for each year

Water Quality Variation



Standard deviations in meters for clarity measurements and colony forming units per liter (cfu/L) for enterococcus

Property Data

- Property Sales
  - Historical parcel data and sales from MASSGIS and LightBox for 2014-2020 (N = 264679)
  - 2023 HPI adjusted sale prices for single-family residential properties
  - House age, lot size, building area, number of rooms
- Neighborhood Characteristics
  - Distance to NOAA shorelines and NHD waterbodies
  - Distance to airports and townhalls from MASSGIS and RIGIS
  - Distance to highways and rails from Tiger/LINE
  - 2020 Census/ACS Attributes at block group level: population, median income, % black, % bachelors degree, median age, seasonal properties, vacant properties
- Matching to Water Quality
  - Water features within 0.25 km of parcel centroid
  - Intersecting area weighted

Property Sale Model

$$\ln(P_{it}) = \beta_0 + \beta_1 \ln(Lake_{it}) * D_{Lake_{it}} + \beta_2 \ln(River_{it}) * D_{River_{it}} + \beta_3 \ln(Coast_{it}) * D_{Coast_{it}} + \beta_4 \ln(Ent_{it}) * D_{Ent_{it}} + \boldsymbol{\beta_P P} + \gamma_m + \lambda_t + \omega_j + \varepsilon_{it}$$

- log-log specification for ambient water quality for property i in year t
  - β_{1-4} represent “elasticities” – % change in price for 1% increase in water quality measure
- $\boldsymbol{P}$ is a vector of property attributes
  - All property and neighborhood characteristics
  - Also includes dummy variables for the presence of each types of water quality data
- Fixed Effects
  - month (γ_m) and year (λ_t)
  - Spatial (ω_j): Census blockgroups

Results

Selected Regression Results by Water Quality Variation

	Spatial	Temporal
Log Lake Clarity	0.0227**	0.0194**
	(0.0106)	(0.0095)
Log River Clarity	0.0378	0.0493**
	(0.029)	(0.0226)
Log Coastal Clarity	0.1351***	0.0499***
	(0.0204)	(0.0177)
Log Coastal Enterococcus	-0.1177***	-0.0701***
	(0.0202)	(0.0137)
AIC	59215.91	59387.45

Notes: N = 264679. $^{*}p<.1$, $^{**}p<.05$, $^{***}p<.01$ Standard errors in parentheses clustered at Census tract level. All models include year, month, and Census block group fixed effects.

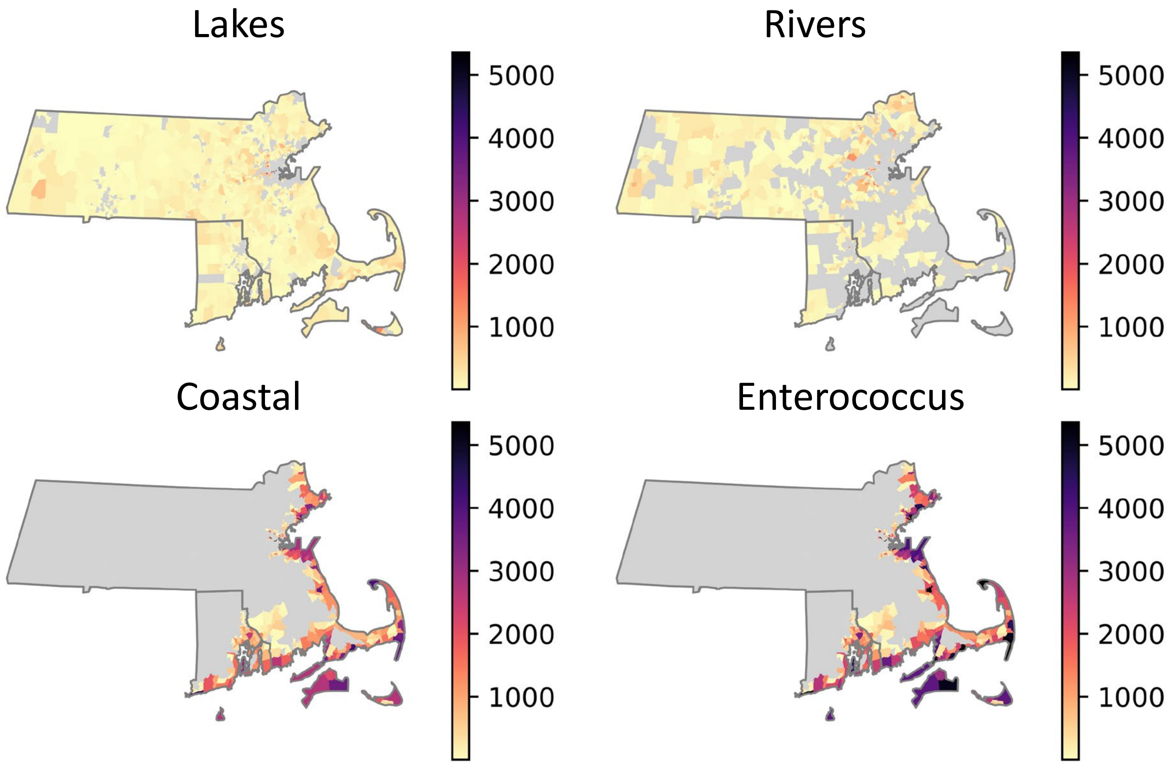
Benefits Scenarios

- Housing market capitalized values (CV) for elasticities

$$CV_{i,wq} = {}^{\%Change_{wq} * \hat{\beta}_{wq} * p_i}\Big/_{100}$$

  - $\%Change_{wq}$: predicted improvement in average water quality
  - $\hat{\beta}_{wq}$: **temporal** estimates for each water quality measure
  - p_i: most recent assessed property values
  - i: single-family property within 0.25 km of lakes/rivers/coast
  - Could be implemented in BenSPLASH by storing $\hat{\beta}_{wq}$ and p_i values, to combine with $\%Change_{wq}$ inputs
- Consider 10% improvements in all water quality measurements
  - Small enough to be considered “marginal” and falls within observed temporal variation
  - Reasonable approximation of potential benefits of uniform improvements
  - Adding up CVs across all properties for different water quality measurements accounts for heterogenous spatial distributions of properties and property values

Results



Average Capitalized Value by Census Tract for 10% Improvement

Total Benefits for 10% Improvement

	Properties	Total Benefits (Millions)
Lakes	164,623	$176
Rivers	55,298	$123
Coastal	110,100	$509
Enterococcus	110,100	$715

Conclusion

- Property sale models are a useful tool that provide insight into the potential benefits of environmental improvements to property owners
- Model results are sensitive to the source of water quality variation
  - Previous studies that rely on spatial and spatiotemporal variation to estimate benefits may not be relevant to potential policy applications
  - Researchers interested in predicting policy benefits can isolate temporal variation by averaging across spatial units when alternative methods are not feasible
- Coastal water quality valued higher than freshwater
  - Coastal enterococcus has largest capitalized values, totaling over $715 million for 10% reduction
  - Water clarity elasticities are smallest on lakes, with estimates below the national average (Mamun et al., 2023)
  - Potential sample differences between monitored lakes and lakes observed via remote sensing

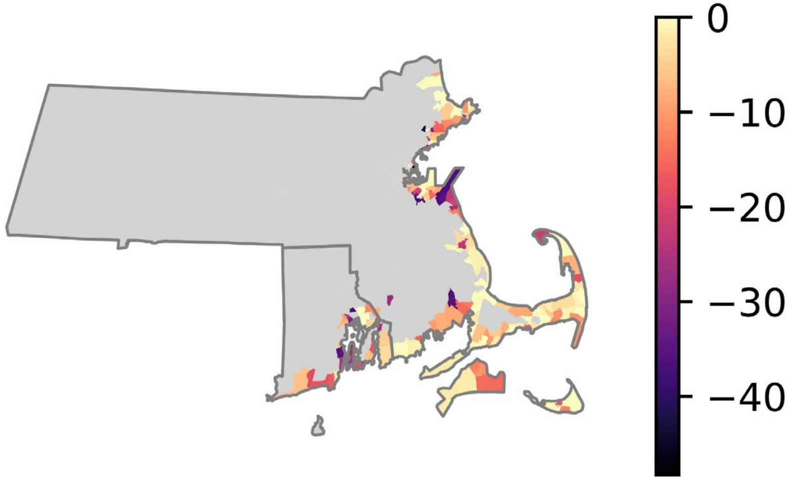
Thank you!

Questions?

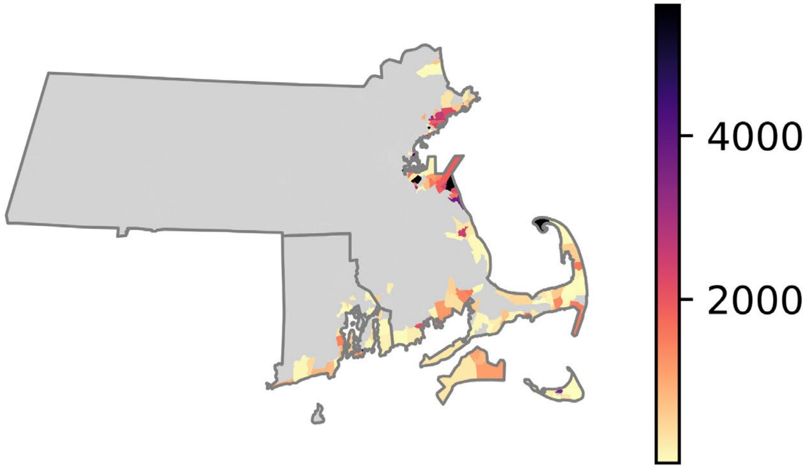
Swedberg.Kristen@epa.gov

Targeted Improvements Scenario

- Adjust seasonal enterococcus averages that exceed human health criteria (35 cfu/L) to meet limit
- Calculate percent changes in average enterococcus levels
- Total benefits = $212 Million across 20,761 properties



% Change by Census Tract



Average Capitalized Value by Census Tract

Additional Results

Spatiotemporal variation: Water quality observations correspond to specific property and year of sale

Selected Regression Results by Water Quality Variation

	Spatiotemporal		Spatial		Temporal	
	(1)	(2)	(1)	(2)	(1)	(2)
Log Lake Clarity	0.0174*	0.0144*	0.0248**	0.0227**	0.0233**	0.0194**
	(0.009)	(0.0081)	(0.0118)	(0.0106)	(0.0114)	(0.0095)
Log River Clarity	0.0472**	0.0432**	0.0487*	0.0378	0.0514**	0.0493**
	(0.0192)	(0.0188)	(0.0271)	(0.029)	(0.0245)	(0.0226)
Log Coastal Clarity	0.0931***	0.0841***	0.1379***	0.1351***	0.0618***	0.0499***
	(0.0178)	(0.0134)	(0.0268)	(0.0204)	(0.022)	(0.0177)
Log Coastal Enterococcus	-0.0752***	-0.065***	-0.1393***	-0.1177***	-0.0797***	-0.0701***
	(0.0117)	(0.0112)	(0.0202)	(0.0202)	(0.0148)	(0.0137)
AIC	59484.7	59298.76	59366.79	59215.91	59601.25	59387.45
Spatial Fixed Effects	CT	CBG	CT	CBG	CT	CBG
Property Controls	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X
Month Fixed Effects	X	X	X	X	X	X

Notes: N = 264679. *p<.1, **p<.05, ***p<.01 Standard errors in parentheses clustered at Census tract level.

Repeat Sales Results

Water Quality Observations Counts

	Full Sample	Repeat Sample
Lake Clarity	21525	3016
River Clarity	4513	551
Coastal Clarity	16598	2089
Coastal Enterococcus	3761	414
Total Sales	264679	33585

Selected Regression Results for Repeat Sale Sample

	Spatiotemporal			Spatial		Temporal		
	(1)	(2)	(3)	(1)	(2)	(1)	(2)	(3)
Log Lake Clarity	0.0306*	0.0339*	0.005	0.0395*	0.050**	0.042**	0.053***	0.045
	(0.0168)	(0.0181)	(0.0344)	(0.0205)	(0.0225)	(0.0186)	(0.0197)	(0.0456)
Log River Clarity	0.042	-0.0329	0.0412	0.0514	-0.1145	0.0822	-0.0042	-0.0168
	(0.0553)	(0.0865)	(0.1456)	(0.0874)	(0.1532)	(0.0684)	(0.0753)	(0.1735)
Log Coastal	0.0834***	0.0733***	-0.0061	0.1573***	0.1498***	0.0747**	0.0426	0.0152
Clarity	(0.0296)	(0.027)	(0.0376)	(0.0471)	(0.0472)	(0.0372)	(0.0392)	(0.0464)
Log Coastal	-0.0854***	-0.0685**	0.0096	-0.1549***	-0.1515***	-0.128***	-0.1045**	-0.0062
Enterococcus	(0.0245)	(0.0292)	(0.0267)	(0.0406)	(0.0548)	(0.0394)	(0.0444)	(0.0337)
AIC	10089.05	12107.15	25704.6	10069.28	12087.21	10087.98	12107.23	25702.74
Spatial FE	CT	CBG	P	CT	CBG	CT	CBG	P

Notes: N = 33585. *p<.1, **p<.05, ***p<.01 Standard errors in parentheses clustered at Census tract level. All models include year, month, and Census block group fixed effects.