Effects of environmental zoning on household sorting: empirical evidence and ecological implications

Bill Provencher (rwproven@wisc.edu)
Dave Lewis (dlewis2@wisc.edu)
Joe Schoen (schoen1@wisc.edu)
Department of Agricltural and Applied Economics
University of Wisconsin –Madison

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August 1, 2007

Draft May 30, 2007

I. Introduction

The purpose of this paper is to examine empirically the extent to which households with heterogeneous preferences over the environmental attributes of residential locations sort themselves across a landscape. The application is to households on inland lakes in northern Wisconsin. We examine whether such variables as lake water clarity, the level of shoreline development, lake size, distance to a town with major services, the presence of public access to the lake, shoreline development restrictions, and other variables influence the sorting process.

There are two reasons for examining this question. The first concerns the management of a landscape: if a landscape is characterized by household sorting, then government agencies concerned with the health of the local ecosystem can tailor education/enforcement to take advantage of this sorting. For instance, finding that households with no knowledge about an invasive species such as Eurasian Milfoil tend to locate on a particular type of lake –large lakes with public access, for instance – suggests that educational efforts should be directed towards property owners on such lakes.

Second, to the extent that heterogeneous households sort themselves across a landscape, and the location decision is correlated with other household behaviors, then land use policies designed to protect ecosystems may have a significant indirect, heterogeneous effect on ecosystems, by grouping households with similar behaviors impacting the ecosystem. For example, development restrictions may have the effect of sorting “eco-friendly” households onto different lakes than “eco-indifferent” households.

Our hypothesis is that lakes where development restrictions are especially strict are settled by households that do such things as maintain and promote native vegetation, maintain the forest canopy on their property, leave course woody material (important to the biotic structure of a lake) in the water, and take extra precautions to avoid accidentally bringing invasive species in a lake. Moreover, to the extent that land use policies promote the sorting process, either directly (because for some households land use restrictions are an important attribute of a property) or indirectly (households respond to the landscape effects of land use restrictions), then these land use restrictions may promote local collective action.

The next section provides an overview of the econometric model used in the analysis. Section III illustrates the model and its potential for informing land management and land use policy with a simple application to the location decisions of households with lakeshore property in Vilas County, Wisconsin. Section IV discusses a number of conceptual and econometric issues to be addressed in future analyses using the dataset.

II. General Statement of the Econometric Model

We formulate a model in which the utility gained from a particular residential location is a function of both neighborhood attributes and attributes that are specific to the particular residential parcel. Neighborhood attributes include environmental attributes. The question at hand is whether –and to what extent –households sort across neighborhoods in response to both the environmental attributes of the neighborhoods, and the land use policies designed to protect the attributes. Our application involves choosing a lakefront parcel, and so in the remainder of this discussion, to make things concrete, we cast the household decision problem as one of choosing from among a set of lakeshore parcels.

The rental price of a parcel on lake *j* is given by the hedonic price function,

$$p_k = S_k + \alpha_X \mathbf{X_j} + \alpha_Z \mathbf{Z_k} + \nu_k \ , \tag{1}$$

where S_k is the value of parcel improvements (primarily, the residence itself); X_j is a set of observed characteristics of the lake; $\mathbf{Z}_k$ is a set of observed land characteristics of the parcel; and ν_k is a stochastic term capturing price effects unobserved by the analyst. We specify a linear form for the hedonic price function to simplify the analysis. Later in the paper we discuss briefly the implication of a nonlinear hedonic model. It deserves, mention, though, that in previous hedonic analyses of lakeshore property, Papenfus and Provencher (2006) found that a linear model generates results similar to a variation of a log-linear model.

There are several ways to represent household heterogeneity in a model of neighborhood choice. The approach we use here is latent class (finite mixture) analysis. Provencher and Moore (2006) argue that the choice of latent class modeling vs. random parameters logit (mixed logit) depends on the analyst's maintained assumptions about the nature of preference heterogeneity. Neither approach is theoretically superior to the other, and neither encompasses the other. We use latent class analysis to simplify the discussion about household sorting and its implications. By labeling each class according to its prominent features, it becomes convenient to talk about household "types": the preferences of each type, the demographic features of each type, the implications of each type for management of the lake system, and so forth.

The money-metric utility obtained by household *i* in preference class *m*, on parcel *k* on lake *j*, is denoted by,

$$U_{ik}^{m} = \left(Y_i - p_k\right) + \beta_S^m S_k + \beta_X^m \mathbf{X_j} + \beta_Z^m \mathbf{Z_k} + \varphi_{ik}, \tag{2}$$

where Y_i is household income and φ_{ik} denotes effects known to the household but unobserved by the analyst.

Substitution of (1) into (2) yields,

$$U_{ik}^{m} = Y_i + \left(\beta_S^m - 1\right) S_k + \gamma_X^m \mathbf{X_j} + \gamma_Z^m \mathbf{Z_k} + \varphi_{ik} - \upsilon_k, \tag{3}$$

where $\gamma^m = \left\{\gamma_X^m, \gamma_Z^m\right\}$ is the difference $\gamma^m = \beta^m - \alpha^m$.

In our application, we do not observe S_k, but we do observe the assessed value of structural improvements on the property. We assume that property tax assessments of improvements are correct up to a factor of proportionality and an iid error, in which case we can write,

$$S_k = \phi S_k^T + \xi_k \tag{4}$$

If tax assessments are correct on average, then $\phi = 1$. Substituting (4) into (3) generates,

$$\begin{aligned} U_{ik}^{m} &= Y_i + \left(\beta_S^m \phi - \phi\right) S_k^T + \gamma_X^m \mathbf{X_j} + \gamma_Z^m \mathbf{Z_k} + \varepsilon_{ik} \\ &= Y_i + \gamma_S^m S_k^T + \gamma_X^m \mathbf{X_j} + \gamma_Z^m \mathbf{Z_k} + \varepsilon_{ik} \end{aligned} \tag{5}$$

where $\varepsilon_{ik} = \varphi_{ik} - \upsilon_k + \left(\beta_S^m - 1\right)\xi_k$. We assume ε_{ik} is distributed iid extreme value.

Given membership in group *m*, the problem of the household faced with choice set *K* is,

$$\begin{aligned} & Max\left\{U_{ik}^{m}\right\}_{K} \\ & = Y_i + Max\left\{V_{ik}^{m}\right\}_{K} \end{aligned}, \tag{6}$$

where

$$V_{ik}^{m} = \gamma_S^m S_k^T + \gamma_X^m \mathbf{X_j} + \gamma_Z^m \mathbf{Z_k} + \varepsilon_{ik} \, . \tag{7}$$

The probability that household *i* chooses property *k* , conditional on preferences *m*, is given by,

$$p\left(k \mid \mathbf{X}, \mathbf{Z}, S; \gamma^m, \sigma\right) = \frac{e^{V_{ik}^m}}{\sum_{l=1}^{K} e^{V_{il}^m}} \tag{8}$$

The unconditional probability of an observed location choice is the sum, over all preferences classes, of the product of the probability of the choice conditional on preferences *m*, and the probability of holding preferences *m*. The probability of membership in preference class *m* is denoted by,

$$\pi\left(m \mid \mathbf{W}_i; \phi^1, \phi^2, ..\phi^M\right) = \frac{e^{\phi^m \mathbf{W}_i}}{\sum_{j=1}^{M} e^{\phi^j \mathbf{W}_i}}, \tag{9}$$

with $\phi^M = 0$. The variables $\mathbf{W}_i$ include variables that predict preference class membership. The likelihood of the location decision *k* by household *i* is then,

$$\begin{aligned} L_i\left(k_i\right) &= \sum_{m=1}^{M} \pi\left(m \mid \mathbf{W}_i; \phi^1, \phi^2, ..\phi^M\right) \cdot p\left(k \mid \mathbf{X}, \mathbf{Z}, S; \gamma^m, \sigma\right) \\ &= \sum_{m=1}^{M} \left[\frac{e^{\phi^m \mathbf{W}_i}}{\sum_{j=1}^{M} e^{\phi^j \mathbf{W}_i}} \cdot \frac{e^{V_{ik}^m}}{\sum_{l=1}^{K} e^{V_{il}^m}} \right] \end{aligned} . \tag{10}$$

The log-likelihood function for the sample is the sum, over all sample households, of $\ln L_i(k_i)$.

At this juncture, one conceptual issue and two estimation issues deserve mention. The conceptual issue concerns the framing of the location decision as a static decision problem. In reality, we would expect that the location decision has a strong dynamic component due to the high cost of relocating, and in fact, including land use restriction variables, as we do in our application, essentially concedes this point, as the effects of such variables lie in their impact on *future* states of the world and future decisions. With this in mind, our model is best understood to be a reduced-form approximation of a dynamic decision problem, and for this reason we avoid welfare analysis with the model because preference parameters are no doubt conflated with parameters defining expectations over future values of relevenat state variables, such as the market price of a property, lakeshore development density, and lake water quality.

Estimation Issue: Parameter identification

The most obvious estimation issue concerns parameter identification. The parameter vector β_W cannot be estimated with a linear model. The parameters comprising γ^m -- β^m and α^m –can be identified only if the prices p_k are observed –they are not –or if the hedonic function is separately estimated. In fact, if either the utility function or the hedonic function were nonlinear, estimation of the hedonic price function, or observation of actual market prices, would be necessary, because in such a case substitution of the hedonic price function into the utility function would not generate the convenient form in (3).

We have the data to conduct a hedonic analysis to separately identify the underlying parameters, and we have additional survey data that may also serve the purpose of parameter identification. But in this preliminary analysis, separate identification of the parameters β^m and α^m is not a priority, because the main issue at hand is the sorting of heterogeneous households, for which identification of γ^m is sufficient.

Estimation issue: The household choice set

Suppose the analyst is able to specify the neighborhoods *J* from which the choice set faced by the household is drawn. There remains the problem of actually specifying the *K* properties available to the household at the time the location choice is made. It is possible, albeit expensive, to construct the choice set from public archives of property sales. An alternative approach is to approximate the decision problem by including in the choice set the actual location chosen and a random draw of locations in the neighborhoods *J*. Using the properties of the Gumbel distribution, it can be shown that if the set of location choices used in the actual decision is a fairly large random draw from the parcels in J neighborhoods, and the set of randomly chosen locations used in the estimation is relatively small, then the probability that the actual choice remains the best choice in the estimation is close to 1.

This is not quite the same problem examined by McFadden in his seminal 1978 paper on housing choice. In the problem examined in that paper, the analyst knows the full set of choices faced by the household, but draws a subset of choices from the full set in estimation, to reduce the cost of estimation. In the problem presented here, the analyst does not actually know the full choice set faced by the household, but is able to specify the neighborhoods (however defined) from which the actual choice set is a random draw. The basic logic is that randomly drawing an alternative choice set in estimation gives correct parameter estimates if the alternative choice set used in estimation is relatively small compared to the actual choice set. This is the approach taken in the empirical analysis of the next section. The choice set for each household in the sample is comprised of the property actually chosen by the sample household, and a random draw of 30 other properties in the dataset. A more thorough presentation of the implications of this approach to specifying the choice set is left to later drafts of the manuscript

An important consequence of the approach is that by construction –and by economic theory –a latent class model with a single class will generate no statistical significance for any of the parameters γ. Quite simply, because every property in the sample is occupied by a household in the sample, and all households have the same preferences by assumption, then by construction no variable in the econometric specification of the utility function can "explain" the observed choice. This result is consistent with economic theory; if all individuals are the same, then the price gradient must equal the utility gradient. In the current context with linear utility and hedonic price functions, $\gamma = \beta - \alpha = 0$.

When there exist two or more preference classes, it is no longer true that the coefficients of V^m must "zero-out". This is shown in Figure 1, where we consider the case of two preference classes with different preferences for water quality. The hedonic price gradient for water quality must lie between the utility gradient for the types, lying closer to the dominant type in the population (i.e., the type with the largest share of the population). If initially (out of equilibrium) the price gradient coincided with the WTP_2 curve, households in preference class 1 would attempt to move onto lakes with higher water quality, bidding up the implicit price of water quality. If instead the initial position of the price gradient coincided with WTP_1, households of preference class 2 would attempt to move onto lakes with lower water quality, bidding down the price premium for water quality. As it stands in Figure 1, the only reason households in preference class 2 are on the lake system is that other attributes of lakeshore living sufficiently compensate for the net welfare loss they suffer from paying more for water quality than they would prefer.

An important point to keep in mind from this example is that a negative sign on an element of γ^m does not indicate that an attribute generates negative utility. Rather, it indicates that households in the preference class place a lower value on the attribute than the implicit price. Moreover, with two preference classes it is not theoretically possible for *both* classes to have a positive sign for any element of γ^m; the implicit price must lie between their respective MWTP's for the attribute in question.

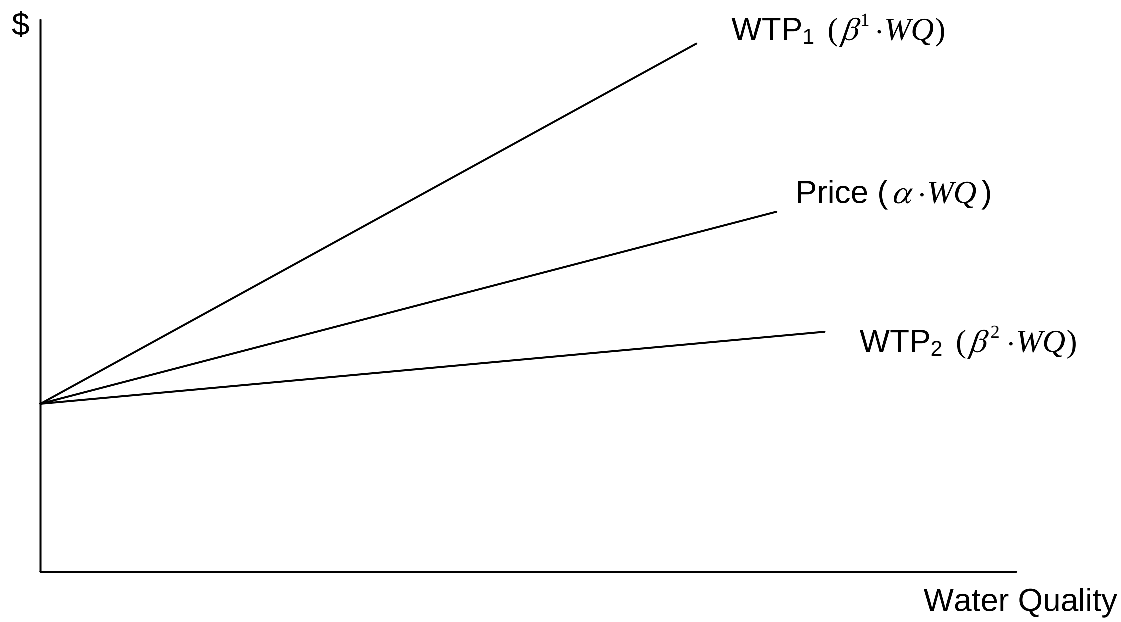

Figure 1.

III. An Application to the Choice of Lakeshore Residence

We apply this model to the location choices of a sample of lakeshore property owners in Vilas County, Wisconsin, a popular vacation destination with the largest concentration of freshwater lakes in the world. The unit of observation is the household. The data set used in this analysis consists of 502 households that purchased a lakeshore property in Vilas County between 1990 and 2004. Table 1 enumerates the variables used in the analysis. These data were derived from a number of sources:

- County tax rolls provided data on the assessed value of structural improvements to a property, mainly residences (STRUCTURE).
- An Internet survey of lakeshore residents conducted during the summer of 2005 provided information on household activities (the variables CANOE, SKI),

household knowledge (KNOWZONE, KNOWWQ, KNOWMIL), and demographics (INCOME, SEASONAL). A mail survey was used to obtain responses for households without Internet access, with a total response rate of approximately 50%.

- GIS maps were the source of calculations of the distance of a property's lake to the nearest town with major services (DISTANCE), as well as calculations of the size and shoreline frontage of a property (ACRES, FRONTAGE), the measure of publicly-owned shoreline (PUBLIC), and the measure of shoreline development density (DENSITY).
- The Wisconsin DNR provided data on lake size and whether a lake had public access (LAKESIZE, ACCESS).
- The University of Wisconsin Environmental Remote Sensing Center (ERSC) provided data on lake water quality (WQ). This data was obtained from satellite imagery correlated with field samples.
- The variable ZONING was constructed from town ordinances and the Vilas County Lake Classification, an ordinance adopted in 1999.

Table 1. Variables Used in Estimation

Variable	Description
S_k^T : STRUCTURE	Assessed value of structural improvements to property *k*.
X_j :	
WQ	Water quality; Secchi depth in feet
LAKESIZE	Surface are of the lake in acres
ACCESS	A dummy variable taking a value of 1 if the lake has public boat access, and 0 otherwise
DISTANCE	Distance to the nearest town with major services (Minoquoa or Eagle River) in miles
DENSITY	The average number of parcels on the lake per 1000 feet of private frontage
PUBLIC	The proportion of the lake's shoreline owned by County, State, or Federal government

MFR·DENSITY	An interaction between DENSITY and the minimum frontage requirement (MFR) for the lake at the time the property was purchased by the respondent. Interacting the MFR with DENSITY provides a measure of the effective impact of the MFR on future development.
Z_k	
FRONTAGE	The shoreline frontage of property *k*, in feet
ACREs	Acreage of property *k*
W_i	
KNOW_ZONE	A dummy variable taking a value of 1 if the respondent knew the minimum frontage requirement on the lake at the time of purchase of the property, as self-reported in the survey
KNOW_WQ	A dummy variable taking a value of 1 if the respondent knew the water quality of the lake at the time of purchase of the property, as self-reported in the survey
KNOW_MIL	A dummy variable taking a value of 1 if the respondent reports in the survey that he knows whether Eurasion Milfoil –an aquatic invasive plant species –is currently present in his lake. The variable takes a value of 1 if the respondent answers that Milfoil is present or not present, and 0 if the respondent says he does not know if the species is present.
CANOE	A dummy variable taking a value of 1 if the respondent canoed during the 2004 summer season
SKI	A dummy variable taking a value of 1 if the respondent water skis or jetskis (uses a personal motorized watercraft) at least once during the 2004 summer season
INCOME	Respondent annual household income, in dollars
PRIMARY	A dummy variable taking a value of 1 if the property is the primary residence of the respondent, as reported in the survey.

The variables used as arguments in the utility function are those used in a recent hedonic analysis of lakeshore property prices (Papenfus and Provencher 2005). The variables used to condition preference class membership, W_i, are preliminary, and represent three types of variables hypothesized to condition class membership: demographic variables (INCOME, SEASONAL), behavioral variables (CANOE, SKI), and knowledge/attitudinal variables (KNOW_ZONE, KNOW_WQ, KNOW_MIL).

Estimation Results

In latent class analysis there is no classical test statistic to determine the correct number of preference classes. In this analysis we restrict estimation to two groups for expositional reasons, though in current work we are expanding the number of preference classes. Results were obtained using the E-M Algorithm, with a BFGS gradient algorithm used in the M (maximization) step.

Results are presented in Table 2. They indicate that in a two-class model, the two classes are differentiated by the modesty of their properties and the remoteness of their lakes. In particular, individuals in the first preference class tend to prefer smaller properties with smaller structures, on lakes far from town and with no public access. We dub such individuals “Thoreaus”, and label the second class, “Socials” It is important to emphasize again that a negative sign on a variable does not indicate that the variable generates negative utility; the analysis does not emphasize, for instance, that individuals in the first preference class place negative value on structural improvements. Rather, the negative sign on STRUCTURE indicates they value such improvements at a rate lower than the market rate. In particular, assuming that structural assessments are accurate, the data indicate that households in the first preference class are willing to pay $.72 for every dollar of improvements, while households in the second preference class are willing to pay $1.32 per dollar of improvements. To the extent that residential structures are underassessed –a distinct possibility –the values of β_S^m are closer to 1 for both preference groups.

Interestingly, with this model we find no heterogeneity in the value placed on development density, zoning restrictions, lake size, water quality, and public shoreline. A working hypothesis underlying the paper is that environmental zoning, such as a lakes’s minimum frontage requirement (MFR), facilitates the sorting process and therefore engenders differential ecosystem dynamics, insofar as households with different preferences over lake attributes behave differently with respect to the lake ecosystem. Although this preliminary analysis finds no preference heterogeneity over lakeshore density and MFRs, it does invite one to speculate that Thoreaus behave differently than Socials in ways relevant to lake ecology (consider, for instance their apparent preference for smaller dwellings), and this in turn suggests that lakes favored by Thoreaus –remote lakes –will evolve differently than those dominated by Socials.

The conditioning variables indicate that compared to the Socials, the Thoreaus have lower incomes, are more likely to canoe and less likely to jet ski, are more likely to be seasonal residents, and are more likely to know about their lake's zoning and state of Milfoil infestation. This last point illustrates the potential for such models to inform lake management policies; overall, education/outreach about milfoil infestations is best directed to lakes close to town with public access, not only because such lakes probably receive relatively high boat traffic, but because residents of these lakes appear to be relatively uninformed about Milfoil.[1]

Table 2. Estimation Results

	Variable	**Coefficient Estimate**	**Standard Error**	**t-stat**
Preference Class 1	FRONTAGE	-0.00925	0.04383	-0.21099
	DISTANCE	0.243423	0.079101	3.077366
	WQ	0.041547	0.046628	0.891033
	ACRES	-0.02567	0.014783	-1.73608
	ACCESS	-0.09495	0.157612	-0.6024
	PUBLIC	-0.49334	0.345973	-1.42595
	LAKESIZE	-0.01002	0.03467	-0.28901
	STRUCTURE	-0.28287	0.07963	-3.55234
	DENSITY	0.340108	0.635101	0.535517
	MFR·DENSITY	-0.24795	0.349416	-0.70961

[1] This statement is perhaps too strong, as it implies the inference that property owners who do not know whether Eurasian Milfoil is on their lake are generally uninformed about this invasive species. Obviously, a survey with more detailed questions about lake ecosystem knowledge would generate more reliable conclusions.

Preference Class 2	FRONTAGE	-0.00103	0.036802	-0.02789
	DISTANCE	-0.87292	0.172067	-5.07312
	WQ	0.013304	0.079968	0.166372
	ACRES	0.032784	0.017111	1.915934
	ACCESS	1.22147	0.479675	2.54645
	PUBLIC	0.680325	0.393026	1.730991
	LAKESIZE	-0.02508	0.052236	-0.48014
	STRUCTURE	0.318992	0.070075	4.552131
	DENSITY	-1.0849	1.057698	-1.02571
	MFR·DENSITY	0.268906	0.629618	0.427094
Conditioning Variables	CONSTANT	8.171296	1.718149	4.755873
	KNOW_WQ	-0.9974	0.715002	-1.39496
	KNOW_ZONE	10.65646	2.492451	4.275497
	KNOW_MILFOIL	3.128488	0.812737	3.849324
	SKI	-17.0388	3.475941	-4.90192
	CANOE	16.07207	3.216935	4.996081
	INCOME	-24.393	4.491409	-5.43104
	PRIMARY	-15.7555	3.332257	-4.72817

IV. Conclusion

In this paper we present a preliminary analysis of whether and how spatial variation in environmental attributes affects the residential sorting of households with heterogeneous preferences. An important implication of such sorting arises if variation in preferences over environmental attributes is correlated with household activities affecting the local ecosystem, such as the replacement of native vegetation with lawns, and the removal of course woody habitat from a lake. In this case the sorting process may engender differential evolution of local ecosystems (lakes) with the same initial ecological state. The model presented here has the potential to statistically examine this issue, and therefore holds promise for understanding the behavioral implications of land use policies designed to protect local ecosystems. By facilitating the grouping of different types of households onto different lakes, for instance, lakeshore zoning policies may engender differentiation in the ecological evolution of lakes beyond what would be expected from the zoning policies themselves.

References

Papenfus, Michael, and Bill Provencher. “A hedonic analysis of environmental zoning”. Working Paper (2006).

Provencher, Bill, and Rebecca Moore. “A Discussion of ‘Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model’”, *Environmental and Resource Economics*, 34:117-124 (2006).