COMBINING REVEALED AND STATED PREFERENCE DATA TO ESTIMATE PREFERENCES FOR RESIDENTIAL AMENITIES: A GMM APPROACH[^]

Daniel J. Phaneuf

North Carolina State University

Laura O. Taylor

North Carolina State University

John B. Braden

University of Illinois

Version Date: 23 June 2010

Preliminary/Incomplete Draft: Please do not cite or distribute further

[^] This study was supported in part by grant no. GL-96553601 from the U.S. Environmental Protection Agency/Great Lakes National Program Office, Cooperative States Research Education and Extension Service, U.S. Department of Agriculture/ Illinois Agricultural Experiment Station projects MRF 470311 and ILLU-470-316, and Illinois-Indiana Sea Grant project AOM NA06 OAR4170079 R/CC-06-06. Any opinions, interpretations, conclusions, and recommendations are entirely the responsibility of the authors and do not necessarily reflect the views of the aforementioned sponsors and individuals.

1. INTRODUCTION

In this paper we propose a new means of combining revealed and stated preference data in the context of property value models for non-market valuation. We show how stated preference information obtained from a conjoint experiment, and revealed preference information based on market transactions, can be combined using a generalized method of moments (GMM) estimator. In particular, we propose using a moment condition matching the predicted marginal willingness to pay from a first stage hedonic model to the marginal willingness to pay formula implied by the conjoint specification. This moment is coupled with other moments implied by the assumptions used in the conjoint model, to produce estimates of preference parameters that reflect the strengths of each data source. We demonstrate our method using an application valuing remediation of a contaminated site in Buffalo, NY, and find evidence in support of estimates arising from our approach.

The starting point for our proposal is the first stage hedonic property value model. This model is well established in environmental economics due to its ability to provide estimates of households' marginal willingness to pay for an amenity using readily available data on residential housing transactions. The attractiveness of valuation estimates arising from the hedonic model is rooted in their connection to a large and consequential market decision on the part of households. The first stage hedonic model, however, is notably lacking in its ability to provide valuation measures for discrete changes in an amenity. For this the second stage hedonic model is needed, but agreement on how best to implement a second stage of estimation for a single market has proven elusive. This has caused researchers to examine stated preference methods – particularly choice experiments – as a means of obtaining the variability needed to estimate households' preferences for residential amenities. While this approach has considerable appeal, as with all stated preference methods concerns about hypothetical bias cannot be readily dismissed.

Property value applications provide a good example of a larger issue in non-market valuation regarding revealed preference (RP) and stated preference (SP) data. It is often the case that observed behavior and natural variation in the environment can be used to obtain good estimates of behavioral functions under baseline conditions. However, in many cases there is insufficient variation to consistently estimate shifts in behavioral functions arising from non-marginal discrete changes in environmental conditions. Stated preference methods are well suited for generating variation in behavior and environmental conditions through designed experiments, but they typically lack a connection to consequential behavioral outcomes. This observation is the basis for a large literature examining different ways that RP and SP data can be productively combined (see Whitehead *et al*., 2008, for a review). One strand of this literature focuses explicitly on the notion that RP data can be used to calibrate preferences at baseline conditions, while SP data is used to estimate the size of movements away from the revealed baseline. Von Haefen and Phaneuf (2008) provide an example of this logic for combining RP and SP data to estimate random utility maximization models of recreation behavior. In this paper we describe a strategy for combining RP and SP data in a model of residential location that is similar in concept but distinct in its approach.

Our combined RP/SP approach envisions a data environment in which a conjoint experiment is conducted in a market for which home sales transactions information is simultaneously gathered. We constructed a database of this type to measure values held by households in Buffalo, NY, for remediation of a contaminated aquatic site near the eastern end of Lake Erie (Braden *et al*., 2006). In particular, the data set contains transactions data and GIS information needed to relate home sales prices to measures of distance to the contaminated site. These data are used to estimate the hedonic price equation, from which we produce estimates of households' marginal willingness to pay for proximity to the site at baseline conditions. A sub-sample of households who purchased a home was also asked to complete a survey. The survey included a conjoint experiment in which respondents were asked to weigh hypothetical houses that differ in price, size, and distance to the contaminated site against their actual purchase. Our approach uses the two data sources to quantify the tradeoffs people are willing to make between housing price and distance to the contaminated site, while also assuring to the extent possible that our preference estimates are grounded in the reality of baseline market conditions.

We find that our RP/SP model provides estimates of preference parameters that are qualitatively similar in sign and significance to the conjoint only (SP) model, but that their magnitudes imply substantially different welfare measures. In particular, marginal willingness to pay estimates from the SP model are larger than comparable measures from our proposed model. Likewise, non-marginal change welfare measures based on increasing the distance to the contaminated site result in economically significant differences between the two models. Comparisons to estimates from the hedonic model provide evidence of our proposed model's ability to calibrate the hypothetical conjoint responses to a consequential market baseline. This, combined with the ability to conduct discrete change welfare analysis, suggests our GMM approach to combining RP and SP data produces a characterization of preferences that is preferred to either the RP or SP approach in isolation.

The remainder of this paper is organized as follows. In Section 2 we describe the conceptual basis for the hedonic model, and then discuss how this has motivated both RP and SP approaches to estimating preferences for residential amenities. In Section 3 we describe how a GMM estimation approach can be used to combine the two data sources. Section 4 describes the Buffalo, NY, data that provides the basis for demonstrating our estimation proposal. Section 5 describes our empirical specifications and estimation results, and Section 6 concludes the paper.

2. HEDONIC MODEL

To motivate our discussion, consider the standard property value hedonic model as reviewed by Palmquist (2005). In this framework a residential property is completely characterized by the variables x and q, where x is a J dimensional vector of property characteristics that is broadly defined to include all structural, parcel, and neighborhood attributes, and q is an environmental (dis)amenity. For purposes of exposition, we assume q is a scalar good. The market price of a house is determined by an equilibrium price schedule $P(x,q)$ – the hedonic price function – that arises via the interaction of all buyers and sellers in the market. A household participates in the market by choosing the levels of attributes x and q to maximize utility, subject to its budget constraint and the price schedule. Formally, households solve the problem

$$\max_{x,q,z} U(x,q,z;h) \quad s.t. \quad y = z + P(x,q), \tag{1}$$

where z is the strictly positive numeraire good, y is income, and h summarizes the household's characteristics, such as family size and composition. Focusing specifically on q, the first order conditions for this problem lead to the familiar result that the household selects the level of q to equate its marginal rate of substitution between q and the numeraire to the implicit price of q:

$$\frac{\partial U(\cdot)/\partial q}{\partial U(\cdot)/\partial z} = \frac{\partial P(\cdot)}{\partial q}. \tag{2}$$

Since $\partial U(\cdot)/\partial z$ is the marginal utility of income, the first order conditions imply that, in equilibrium, households select q to equate their marginal willingness to pay for q to the marginal implicit price of q. This is the primary result upon which much of revealed preference analysis of property value markets is based. We return to this point below.

We first, however, gain further conceptual insight by examining Rosen's (1974) bid function $b(x,q,y,h,\bar{u})$ for a representative level of utility $\bar{u}$, which is implicitly defined by

$$U\left[x,q,y-b(\cdot);h\right] = \bar{u}. \tag{3}$$

The bid $b(\cdot)$ is the maximum amount that the household would (and could) pay for a house with attributes x and q, given its income and characteristics, while holding utility fixed at $\bar{u}$. Differentiating (3) with respect to q allows us to relate the bid function to the marginal rate of substitution at any point q:

$$\frac{\partial b(x,q,y,h,\bar{u})}{\partial q} = \frac{\partial U(\cdot)/\partial q}{\partial U(\cdot)/\partial z}. \tag{4}$$

Bockstael and McConnell (2007) note that the properties of $b(\cdot)$ imply that $\partial b(\cdot)/\partial q$ is not a function of income, and so we can rewrite (4) as

$$\pi^q(x,q,h,\bar{u}) \equiv \frac{\partial b(x,q,y,h,\bar{u})}{\partial q} = \frac{\partial U(\cdot)/\partial q}{\partial U(\cdot)/\partial z}, \tag{5}$$

where $\pi^q(x,q,h,\bar{u})$ is the marginal willingness to pay (compensated inverse demand) function for q. Connecting this back to the first order conditions, we see that in equilibrium

$$\pi^q(x^0, q^0, h, u^0) = \frac{\partial P(x^0, q^0)}{\partial q}, \tag{6}$$

where x^0 and q^0 are the household's observed choices for the attributes, and u^0 is the level of utility obtained. The primary objective of empirical models of household residential choice is to obtain an estimate of all or parts of $\pi^q(x,q,h,\bar{u})$ for individual households, since this function contains the information needed to conduct welfare analysis for changes in q. In particular, the willingness to pay (compensating variation) for a discrete improvement in q from q_0 to q_1 is given by

$$\begin{aligned} WTP &= \int_{q^0}^{q^1} \pi^q(x^0, q, h, u^0) dq \\ &= b(x^0, q^1, y, h, u^0) - b(x^0, q^0, y, h, u^0) \\ &= b(x^0, q^1, y, h, u^0) - P(x^0, q^0). \end{aligned} \tag{7}$$

This *WTP* measure is what Bockstael and McConnell (2007) refer to as 'pure willingness to pay', since it holds the person at his current location (as opposed to allowing adjustment through mobility), and reflects only the preference effect net of any change in the actual price paid. In what follows we describe the typical revealed and stated preference techniques that have been used to estimate *WTP* as defined by this expression.

Revealed Preference Estimation

The model described thus far is usually presented as the conceptual basis for using observed housing market transactions, coupled with information about the households who purchased the homes, to estimate $\pi^q(x,q,h,\bar{u})$ in two steps. The first step involves combining transaction prices and the attributes of properties to estimate $P(x,q)$ based on the econometric model

$$p_i = f(x_i, q_i, \theta, \varepsilon_i), \quad i = 1, \ldots, I, \tag{8}$$

where (p_i, x_i, q_i) are the observed sale price and attributes for property i, $f(\cdot)$ is a functional specification for the price schedule, θ is a vector of parameters to be estimated, and ε_i is a disturbance term. This is the first stage of hedonic estimation, and it is ubiquitous in applied analysis for several reasons. Perhaps most importantly, equation (6) suggests that an estimate of $P(x,q)$ alone can provide a measure of each household's marginal willingness to pay for q at their observed choice. Thus, with few assumptions and good econometric work, we obtain a point on household i's marginal willingness to pay curve based on

$$\pi_i^q(q_i, x_i, h_i, u_i^0) = \frac{\partial \hat{f}(x_i, q_i, \hat{\theta})}{\partial q_i} = \hat{p}_{q_i}, \quad i = 1, \ldots, I, \tag{9}$$

where we use $\hat{p}_{q_i}$ to denote an estimate of the marginal implicit price of q for person i.

To estimate the full marginal willingness to pay curve we need to conduct a second stage of estimation. This involves estimating an econometric model of the form

$$\hat{p}_{q_i} = g(q_i, y_i, h_i, \gamma, \eta_i), \quad i = 1, \ldots, I, \tag{10}$$

where the left hand side variable is predicted from the first stage of estimation, $g(\cdot)$ is a specification for the ordinary inverse demand for q function, γ is a vector of parameters to be estimated, and η_i is a disturbance term. This is the second stage of hedonic estimation and, as discussed by Palmquist (2005), it is fraught with conceptual and econometric challenges. The largest of these is known as the identification problem. Note that for non-linear specifications of $f(\cdot)$ there is cross-sectional variation in $\hat{p}_{q_i}$ and the right hand side variables (q_i, h_i, y_i), so equation (10) can in principle be estimated. However, the sample of size I does not in general contain enough information to trace out a function specific to household i. This is best seen by looking at Figure 1. The top panel shows a cross section of the hedonic price function relating q to $P(x,q)$ for given values of x. Equilibrium outcomes for two distinct households 1 and 2 are shown at the points of tangency between their respective bid functions and the hedonic price function. Specifically, household 1 locates at point a and consumes q_1, and household 2 locates at point b and consumes q_2. The lower panel shows how points a and b correspond to single points on each of two inverse demand for q functions. Note that for each household, $\pi^q(\cdot)$ is traced out as the slope of its bid function as q changes. For household 1 we observe point a' on $\pi^q(\cdot,h_1)$ but nothing more; likewise for household 2 we observe only b' on $\pi^q(\cdot,h_2)$. For this example with I=2, points a' and b' represent the sample of data referred to in equation (10). The identification problem arises because the regression fits a function such as the dashed line in the lower panel, which is *not* the inverse demand curve for either of the sampled households without additional, and usually strong, assumptions. Thus the main challenge in using a second stage regression to estimate $\pi^q(x,q,h,\bar{u})$ is that a single housing market does not provide adequate variability in price/quantity space, since each household reveals only one price/quantity outcome.

A large literature is devoted to examining solutions to this and other difficulties with the second stage hedonic model, yet to date there is no agreed upon solution for robust estimation of $\pi^q(\cdot)$ using data gathered for a single market area. This presents a dilemma for non-market valuation using revealed preference models of property markets. Specifically, while the first stage of estimation usually provides a solid estimate of households' baseline marginal willingness to pay, researchers typically rely on approximations of unknown quality, rather than second stage estimates, to obtain the value of discrete changes in q.

Stated Preference Estimation

An alternative to the revealed preference approach is to use stated preference methods to elicit households' preferences for q as related to their choice of residential location. This has typically been done using choice experiment, or conjoint, analysis in which surveyed households are presented with hypothetical choices between homes of different configurations, and asked to indicate their preferred option. Examples of this method are provided by Earnhart (2000, 2002), Braden et al. (2004), and Chattopadlhyay et al. (2005). A typical procedure is to ask respondents to compare their current home to a hypothetical home in which the attributes of interest – e.g. the home price and level of q – are experimentally designed to vary away from their baseline levels, while all other attributes remained fixed. This gives rise to a discrete choice model in which the utility from a particular choice c is

$$U_{ic} = V\left(x_{ic}, q_{ic}, y_i - P(x_{ic}, q_{ic}), h_i; \beta\right) + \varepsilon_{ic}, \quad c = 0,1 \quad i = 1,...,I, \tag{11}$$

where $V(\cdot)$ is the observable component of utility that person i gets from choosing option c, β is a vector of parameters characterizing utility, and we have substituted out for z_{ic} using the budget constraint $y_i = z_{ic} + P(x_{ic}, q_{ic})$. The random variable ε_{ic} accounts for the unobserved component of preferences, and it is assumed to have a known distribution. Finally, in what follows we use c=0 to denote the actual home the person purchased, and c=1 to denote a hypothetical home against which the actual home is compared. Under the assumption that people select the option with the greatest utility, we can use maximum likelihood to recover estimates of β and thereby a characterization of $V(\cdot)$. We discuss estimation of this model in detail in the following section.

Once we obtain a characterization of $V(\cdot)$, welfare analysis is relatively straightforward. The marginal willingness to pay for q at baseline conditions is found by differentiating (11) with respect to q to obtain

$$\frac{\partial V\left(x_{i0},q_{i0},y_i-P(x_{i0},q_{i0}),h_i;\beta\right)}{\partial q}=\frac{\partial V\left(x_{i0},q_{i0},y_i-P(x_{i0},q_{i0}),h_i;\beta\right)}{\partial z}\times\frac{\partial P(x_{i0},q_{i0})}{\partial q}, \tag{12}$$

which we can rewrite as

$$\pi_i^q(x_{i0},q_{i0},h_i,u_i)=\frac{\partial V\left(x_{i0},q_{i0},y_i-P(x_{i0},q_{i0}\cdot),h_i;\beta\right)/\partial q}{\partial V\left(x_{i0},q_{i0},y_i-P(x_{i0},q_{i0}),h_i;\beta\right)/\partial z}=\frac{\partial P(x_{i0},q_{i0})}{\partial q}. \tag{13}$$

Note that the model once again suggests that the baseline marginal willingness to pay is equal to the marginal implicit price of q in the market. In the stated preference approach, however, we calculate its magnitude from the utility parameter estimates rather than an estimate of the price schedule. Indeed, the right hand side of (13) is not available in a purely SP study. This is an important distinction that further illustrates how RP and SP approaches rely on different information sources to predict similar quantities.

The value of a discrete change in q is found by integrating the marginal willingness to pay function over the relevant range of q:

$$WTP_i=\int_{q_{i0}}^{q_{i1}}\pi_i^q(x_{i0},q,h_i,u_i)dq=\int_{q_{i0}}^{q_{i1}}\frac{\partial V\left(x_{i0},q,y_i-P(x_{i0},q),h_i;\beta\right)/\partial q}{\partial V\left(x_{i0},q,y_i-P(x_{i0},q),h_i;\beta\right)/\partial z}dq. \tag{14}$$

For the special case of a model that is linear in $y_i-P(x_{ic},q_{ic})$ and q_{ic}, the welfare measure reduces to the familiar expression

$$
\begin{aligned}
CV_i &= \frac{1}{\beta_y}\left\{V\left[x_{i0}, q_{i1}, y - P(x_{i0}, q_{i1}), h_i; \beta\right] - V\left[x_{i0}, q_{i0}, y - P(x_{i0}, q_{i0}), h_i; \beta\right]\right\} \\
&= \frac{\beta_q}{\beta_y}\left[q_{i1} - q_{i0}\right] - \left[P(x_{i0}, q_{i1}) - P(x_{i0}, q_{i0})\right].
\end{aligned} \tag{15}
$$

where β_y is the coefficient on the budget constraint (i.e., the marginal utility of income) and β_q is the coefficient on *q*. This definition of *CV* corresponds to the gross willingness to pay, since it includes an adjustment for the change in purchase price. The pure willingness to pay corresponding to equation (7) is the first term in the expression.

The stated preference approach is attractive in that, via the designed experiment, respondents reveal tradeoffs between different levels of *q* and home prices. In this sense it solves the fundamental dilemma of second stage hedonic estimation in the RP context, because it delivers the variability in price/quantity space that is needed to characterize household-specific marginal *and* non-marginal change values for *q*. However, like all stated preference exercises, respondents do not bear the real consequences of their choices. This potential for hypothetical bias may therefore give one pause when interpreting estimates arising from purely SP methods, and it motivates our combined RP/SP approach.

3. A Combined RP/SP Approach using GMM

Our proposal is based on equation (13), which illustrates how the RP and SP approaches overlap conceptually but differ empirically. Note in particular that both sides of the equation show the baseline marginal willingness to pay for *q*; this equality links the conjoint and first stage hedonic approaches to a common underlying model of preferences. The two approaches, however, obtain this estimate in different ways: the RP estimate is based on analyzing market transactions, and the SP estimate is based on hypothetical tradeoffs. The former is almost certainty a better baseline estimate, while the coefficient estimates from the latter expand the range of measurement possibilities to include analysis of non-marginal changes.

To see how the RP and SP data can be combined, consider first the estimating equations for the conjoint model. For ease of notation and exposition let $W_{ict}=(x_{ict}, q_{ict})$ and assume that $U_{ict}=W_{ict}\cdot\beta+\varepsilon_{ict}$, where W_{ict} can include interactions between q and household characteristics, K is the dimension of β, ε_{ic} is distributed type I extreme value, and t indexes the different choice situations the person faces in the survey. In this case the probability $\Pr_{itc}$ of observing a particular choice has a simple closed form, and the sample log-likelihood function is given by

$$LL(\beta)=\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{c=0}^{1} y_{itc}\times\log\Pr_{itc}, \tag{16}$$

where I is the number people, T is the number of choice occasions each person faces, and $y_{itc}=1$ if person i chooses house c on question t, and zero otherwise. The value of β that maximizes (16) is the maximum likelihood estimator.

Equivalently, we can interpret the estimator arising in (16) as a method of moments (MM) estimator. Define the $K\times1$ score vector $s_{it}(\beta)$ for an observation indexed (i,t) as

$$s_{it}(\beta)=\sum_{c=0}^{1} y_{itc}\frac{\partial\ln\Pr_{itc}}{\partial\beta}, \tag{17}$$

and recall that the first order conditions for the maximum likelihood estimator are

$$g_{SP}=\sum_{i=1}^{I}\sum_{t=1}^{T} s_{it}(\beta)=0. \tag{18}$$

Equation (18) defines a just-identified MM estimator, in which the sums of the scores for the sample serve as the K moments. For future reference we refer to these K moments as g_{SP}.

Consider now adding an additional moment condition based on equation (13). In particular, define

$$g_{RP}=\sum_{i=1}^{I}\left(\frac{\partial V\left(x_{i0},q_{i0},y_i-p_{i0},h_i;\beta\right)/\partial q}{\partial V\left(x_{i0},q_{i0},y_i-p_{i0},h_i;\beta\right)/\partial z}-\frac{\partial\hat{f}(x_{i0},q_{i0},\hat{\theta})}{\partial q}\right)=0 \tag{19}$$

as a moment condition, where p_{i0} is the actual price individual i paid for a home. Note that this moment relates the prediction of baseline marginal willingness to pay for person i in the SP model to the prediction for the person's marginal willingness to pay as given by the RP model. Assuming that we have effectively estimated the latter for the full sample using a first stage hedonic model, g_{RP} can be computed as a function of data and the K unknown parameters. With equation (19) we now have a collection of K+1 moments with which to estimate K unknowns, and our model is over-identified. In this case a GMM estimator is appropriate, which is defined as the value of β that minimizes

$$Q(\beta) = \frac{1}{I} \cdot \begin{bmatrix} g_{SP} \\ g_{RP} \end{bmatrix}' \cdot W_I \cdot \begin{bmatrix} g_{SP} \\ g_{RP} \end{bmatrix}, \tag{20}$$

where W_I is a $(K+1)\times(K+1)$ dimension weighting matrix that is generally unknown. The feasible, two step GMM estimator for β as described by Cameron and Trividi (2005) is used to compute our combined RP/SP estimates for β.

Relationship to Other Literature

Our proposal is related to two strands of literature, the most obvious being the large body of work on combined RP and SP models. In particular, it is akin to early work by Cameron (1992) and Kling (1997) that focused on merging RP and SP information from different decision margins – recreation trips and contingent valuation in their cases – to estimate a single preference function. It is also related to more recent research by Whitehead *et al*. (2010) and von Haefen and Phaneuf (2008), which views the joint use of RP and SP data as a means of exploiting different sources of variability for a single inference task. Because we are asserting (rather than testing) that the property value and conjoint data are generated by the same underlying behavioral process, our approach is distinct from the strand of literature that uses joint models of common decision margins to test the validity of one or the other elicitation method (e.g. Azevedo *et al.*, 2003).

Our approach is also related to the literature on second stage estimation of hedonic models. Purely RP strategies for solving the identification problem are based on using multiple markets that are spatially distinct to generate the exogenous variability needed for estimation. In our case the additional variability comes from the choice experiment rather than additional RP datasets. However, many of the same notions apply. As in Bajari and Kahn (2005), our strategy relies on first estimating a flexible hedonic price function and then matching predictions for baseline marginal willingness to pay to a parametric specification for the inverse demand function.

4. Application

We investigate the performance of our combined RP/SP strategy via an application examining household's willingness to pay to avoid proximity of their primary residence to an aquatic hazardous waste site. The 1987 Amendments to the Great Lakes Water Quality Agreement between the US and Canada designated 43 sites in the Laurentian Great Lakes, and their tributaries, as Areas of Concern (AOC). A common feature of these areas is the presence of toxic chemicals – notably polychlorinated biphenyls (PCBs) – known to cause cancer and neurological defects in humans and to bio-accumulate in aquatic food webs. Since the 1987 Amendments, only one US and two Canadian AOCs have been delisted. The remaining remedial activities on the US side alone are expected to cost between $1.5 billion and $4.5 billion (Great Lakes Regional Collaboration, 2005). There is considerable interest in discerning whether further expenditures on cleanup will produce benefits consonant with the costs.

Our analysis focuses on the Buffalo River, NY, AOC, which is shown in Figure 2. The area consists of a commercial harbor and a 6.2 mile segment of the river running eastward from its terminus into Lake Erie. The AOC is flanked by a large industrial complex, which is in decline and contains many brown fields. Nevertheless, there are private homes nearby – the 2000 Census counted 52,628 single-family homes within five miles of the AOC. Our objective is measure how proximity to the AOC affects the market value of these private homes, and what the value to homeowners would be of remediation. For this purpose both real estate transactions and survey data were collected, in which the latter provides both characteristics of households who purchased a home and the results of a conjoint experiment. We explain these two sources of data in turn.

Real Estate Data

Our analysis uses sales of single family, owner occupied homes that occurred between January 2002 and December 2004. The data were collected by Braden et al. (2006) and initial, separate analyses of the RP and SP data are reported by Braden et al. (2008). The present study is the first effort to combine the data for joint estimation as well as to explore the potential to use a GMM estimator in this context. The sample is limited to properties that lie within five linear miles of any point along the Buffalo River AOC. The study area encompasses most of the City of Buffalo, all of Lackawanna, and portions of Cheektowaga, Hamburg, and West Seneca as well as two smaller municipalities.

Two primary databases were combined to characterize homes sales in our study area. The first comes from local tax assessors, and contains sales prices (normalized to 2004 dollars), transaction dates, and property characteristics that include: lot size; square feet of living area; age of primary structure; and miscellaneous housing characteristics. The top section of Table 1 displays the names, definitions, and summary statistics for these variables. The second database describes spatial features of the properties that sold in our study area. Variables that are of particular interest include: proximity of the house to the AOC; proximity of the house to other location-specific (dis)amenities, such as the shoreline of Lake Erie, local parks, transportation networks, and employment districts; and spatial units such as census tract and block, and school district. The proximity measures were created for each parcel using a GIS map of the Buffalo area. The lower sections of Table 1 show the names, descriptions, and summary statistics for these variables. In particular, the summaries show that 47% of the sales in our sample occurred north of the Buffalo River. This distinction becomes important when we discuss our estimation results. Also, the mean distance to the AOC is approximately three miles. Other summaries we examined indicated that 12 percent and 16 percent of homes north and south of the river, respectively, lie within 1.5 miles of the AOC.

Table 1 also indicates that there are 118 additional dummy variables for use in the analysis, each representing a census tract in which a property is located. Census tracts are designed to be relatively homogeneous with respect to population characteristics, economic status, and living conditions. By including census tract identifiers, our analysis non-parametrically controls for infrastructure and demographic factors that influence home choices and prices across space. These spatial fixed effects help eliminate confounding between our distance measures of interest (e.g. miles to the AOC) and other factors that may be correlated with these distances but not included in our explanatory variables.

Survey Data

Based on the home sales data, Braden et al. (2006) randomly selected 850 households that purchased a transacted property, each of whom was sent a survey.[1] Among these 315 were returned; excluding 63 undeliverable surveys, the response rate was 40.7%. Of the returned surveys, 281 were sufficiently complete to be of use for this analysis. The survey was designed to complement the real estate market data. Four categories of information were collected: verification of current home characteristics; measurement of respondent attitudes regarding the AOC; responses to conjoint questions; and household demographic information. The conjoint questions asked respondents to imagine that additional homes had been on the market during their recent home-buying experience. Hypothetical homes were then sequentially offered. Respondents were asked whether, at the time of purchase, they would have preferred the hypothetical home to the home they actually bought. A representative choice question is shown in Figure 3.

In order to focus respondents' attention on variables of interest and to make the choices as concrete as possible, the hypothetical homes were described as being identical to the current home, aside from four designed attributes. Table 2 summarizes the designed attributes. These were chosen to focus on trade-offs between private aspects of homes (sale price; square feet of living area), and spatial aspects of the neighborhood (distance to the AOC; condition of the AOC). Values for the attributes were expressed in relation to the home/location as it existed at the time of purchase. For sale price and square feet of living area, the designed levels are proportions of the price and home size of the property actually purchased. For proximity, Braden et al. (2006) used nominal deviations from current distance, and asked

[1] The survey instruments were developed with assistance from the University of Illinois Survey Research Laboratory (SRL) and in cooperation with the Great Lakes Program, University at Buffalo. [1] Early versions were assessed by focus groups held at a public library branch in West Seneca, NY, in early 2005. Advanced versions were pretested in Spring 2005. For the final survey, respondents could either mail back a completed questionnaire or complete an equivalent instrument using the Zoomerang.com commercial survey website. Approximately nine percent of the responses were received online.

respondents to imagine the river being closer (further) to (from) their home without changing other features of the neighborhood. The environmental condition of the river was varied qualitatively, with toxic pollution increasing, decreasing, staying the same, or being completely eliminated. The four attributes with four levels each suggest there are 4^4=256 possible combinations of hypothetical homes. The resulting factorial design varies attribute levels to achieve orthogonally in the explanatory variables, thereby maximizing the efficiency of parameter estimates (Montgomery, 2000). Sixty-four unique choice alternatives resulted from the design. Eight survey versions were created so that each version contained eight choice tasks, each comparing one of the hypothetical homes to the respondent's actual purchase.

5. RESULTS

Hedonic Model

Our modeling approach involves first estimating the hedonic price function. Using the transactions and spatial data described above, Braden *et al*. (2006, 2008) estimated the price/AOC distance gradient using several parametric specifications and variable interactions. For our purposes a flexible specification in *AOC* is desirable, in order to obtain the highest level of variability in house-specific estimates of marginal willingness to pay supportable by the data. Flexible specifications are also consistent with Ekeland *et al*. (2004), who advocate a non-parametric approach to hedonic estimation for theoretical reasons. To obtain flexibility while maintaining transparency and ease of estimation we use the specification

$$price_i = \theta_0 + h(\ln(AOC_i), N_i, \theta_1, \theta_2, \theta_3, \theta_4) + \theta' x_i + \varepsilon_i, \qquad (21)$$

where AOC_i is the distance from property i to the area of concern (measured in tenths of miles), N_i is a dummy variable that takes the value one when the property is north of the Buffalo River, and x_i is a vector of other control variables thought to influence the sale price of the property. The term $h(\cdot)$ is a non-linear function in N_i and the natural logarithm of AOC_i, which we construct using a restricted cubic spline with three knots (Harrell, 2001). In particular,

$$h\left(\ln(AOC_i), N_i, \theta_1, \theta_2, \theta_3, \theta_4\right) = \theta_1 \ln(AOC_i) + \theta_2 \ln(AOC_i) \times N_i + \theta_3 \times \tilde{h}\left(\ln(AOC_i)\right) + \theta_4 \times \tilde{h}\left(\ln(AOC_i)\right) \times N_i, \tag{22}$$

where

$$\tilde{h}(\cdot) = (\ln(AOC_i) - t_1)_+^3 - \frac{(\ln(AOC_i) - t_2)_+^3 (t_3 - t_1)}{(t_3 - t_2)} + \frac{(\ln(AOC_i) - t_3)_+^3 (t_2 - t_1)}{(t_3 - t_2)}, \tag{23}$$

(t_1, t_2, t_3) are the locations of the three knots, and $(u)^3_+$ means the expression is u^3 if $u>0$ and zero otherwise. Harrell (2001, p.21) suggests that empirical percentiles of the variable to be transformed are the optimal knot locations, and that with three knots the percentiles should be (0.10, 0.50, 0.90). For the observed distribution of ln(*AOC*), these correspond to 2.46, 3.49, and 3.82, respectively.

Our use of a restricted cubic spline provides for a flexible, non-linear relationship between housing prices and *AOC* that is smooth (the splines connect at the knots) and well-behaved in its tails (the effect is linear for values of ln(*AOC*) below 2.46 and above 3.82). Also, it nests a specification that is linear in the logarithm of AOC_i. Selected coefficient estimates obtained using ordinary least squares on equation (21) with I=3,474 are shown in Table 3. Estimates for the full set of parameters are given in appendix Table A1. The values for θ_1, θ_2, θ_3,and θ_4 are difficult to interpret numerically, but some intuition can be gained from statistical tests. Note first that $\theta_2=\theta_4=0$ implies there is no difference in the effect of the disamenity between properties located north and south of the river; this restriction is rejected in our model (F=3.21, p-value=0.04). Since θ_2 is individually insignificant the difference is due to the higher order terms in $h(\cdot)$. Furthermore, a test of $\theta_3=\theta_4=0$ reveals that the relationship between ln(*AOC*) and housing prices is non-linear (F=6.11, p-value=0.002). Taken together, our estimates suggest the gradient of ln(*AOC*) is complex in its dependence on multiple spatial factors. In some ways this is not surprising, particularly for properties north of the Buffalo River. Figure 2 shows that a major highway corridor runs between the AOC and much of the northern subsample. Crisscrossing rail lines and an industrial area also lie between the AOC and many of the northern properties. These physical aspects appear to act as a buffer between the residential real estate market to the north and the AOC, particularly beyond certain distances. This notion is reinforced by Figure 4, which depicts the level relationship between ln(*AOC*) and property values for properties north and south of the AOC, holding all else constant. The slopes of these lines suggest that the gradient of ln(*AOC*) is positive and similar for all properties out to a distance of approximately 1.5 miles (i.e. ln(*AOC*)=2.75). Beyond this distance the gradient for properties south of the AOC stays positive, while it levels off and becomes negative for properties north of the AOC. Thus for small distances from the disamenity marginal willingness to pay is positive for both types of properties, but it is larger for those south of the river.

In our linear-log specification the marginal willingness to pay for a small change in distance to the AOC is

$$\hat{p}_{q_i} = \frac{\partial price_i}{\partial \ln(AOC_i)} \times \frac{1}{AOC_i}, \quad i = 1,\ldots,I. \tag{24}$$

Table 3 includes a summary of predictions for baseline marginal willingness to pay (*MWTP*), based on (24). The sample average is −\$78 (median=\$490.52), but this confounds properties north and south of the AOC. The figure for properties south of the river is \$682 (median=\$663), which on average is a relatively small percentage of a home's value. For small distances to the AOC, however, the effect on home values for all properties becomes more substantial. For example, the *MWTP* for homes south of the river, located less than 0.3 miles from the AOC is on average over 5% of the purchase price. In general our estimates imply economically significant effects on property values for houses near the AOC, but that the effect decreases fairly rapidly as the distance increases.

SP and RP/SP Models

We turn now to the conjoint model. For our demonstration we consider a specification for utility given by

$$\begin{aligned} U_{itc} &= \beta_0 R_{itc} + \beta_1 \ln(AOC_{itc}) + \beta_2 \ln(AOC_{itc}) \times N_i + \beta_3 \ln(AOC_{itc}) \times y_i + \beta_4 \ln(AOC_{itc}) \times K_i + \\ &\quad \beta_5 \ln(size_{itc}) + \beta_6 \ln(size_{itc}) \times y_i + \sum\nolimits_{e=1}^{3} \beta_{7e} E_{itc}^{e} + \sum\nolimits_{e=1}^{3} \beta_{8e} E_{itc}^{e} \times \ln(AOC_{ict}) + \beta_9 price_{itc} + \varepsilon_{itc}, \end{aligned} \tag{25}$$

where R_{itc}=1 if choice c is the house actually purchased and zero otherwise, K_i is the number of people living in the household, E_{itc}^{e} is a dummy variable indicating the environmental condition of the AOC, and

ε_{itc} is distributed type I extreme value. The marginal willingness to pay for a change in distance at baseline conditions given this specification for utility is

$$\pi_i^q(x_{i0}, AOC_{i0}, y_i, h_i, u_i) = -\frac{1}{\beta_9 \times AOC_{i0}}\left(\beta_1 + \beta_2 \times N_i + \beta_3 \times y_i + \beta_4 \times K_i\right), \quad i = 1, \ldots, I. \qquad (26)$$

Note that, by including interactions between ln(AOC_{itc}) and household characteristics we allow the inverse demand for *AOC* to be conditional on the type of household that chose to occupy the house.

Maximum likelihood estimation results using only the SP data are shown in Table 4. The parameter estimates are generally sensible. For example, $\beta_1>0$ indicates that respondents prefer their baseline house relative to the hypothetical house, all else equal. Consistent with the first stage hedonic estimates, we also find that house size and distance from the AOC are positive housing attributes, though we do not find a statistically significant estimate for the latter. The interactions between distance to the AOC and household characteristics are also not statistically significant. It is particularly noteworthy that $\beta_2\approx 0$, suggesting that respondents from north and south of the Buffalo River do not systematically differ in how they trade off distance from the AOC and other housing attributes. This finding somewhat contradicts the market-based evidence from the hedonic model, in that it does not account for any non-linear effects. We also find that differences in income and household structure do not significantly shift preferences for the distance attribute. Finally, our findings regarding preferences for the environmental condition of the river are intuitive. The left out category is *baseline pollution*, so we expect the coefficients on *partial cleanup* and *full cleanup* to be positive and increasing in magnitude, and the coefficient on *additional pollution* to be negative. We find that *full cleanup* has a positive effect relative to the baseline and *additional pollution* a negative effect. The estimate for *partial cleanup* is statistically not different from zero. This suggests people are indifferent between the site as it is and efforts to partially remediate, but they would be willing to pay to obtain a full remediation.

The estimation results using the combined RP/SP GMM estimator are also shown in Table 4. Qualitatively the estimates do not differ much from their SP-only counterparts, but there are important differences in coefficient magnitudes and statistical significance. Formally we cannot compare the values for the marginal utilities between the two models, since the scale of utility may not be equal in the two specifications. However, given that the estimate for the price coefficient is similar for the two models, we can gain some informal intuition through direct parameter comparisons. Most notably, the level estimate for the coefficient on ln(*AOC*) is larger and statistically significant in the GMM model, and the interaction with family size is (nearly) significantly negative. Not surprisingly, the estimates not related to the distance to the AOC are largely unchanged in the GMM model, due most likely to the orthogonal nature of the experimental design.

Welfare Effects

The relevant comparisons between the models is not coefficient estimates, but rather what the estimates imply about households willingness to pay for changes in features of the AOC. To shed light on this, the bottom panel of Table 4 compares point estimates of marginal willingness to pay at baseline conditions for the two models for a representative household, at two different family sizes and distances from the AOC. It is notable how the SP and RP/SP estimates differ from each other, with the SP estimates being larger than their RP/SP counterparts. These point estimates suggest the SP model alone may not be matching the market-based baseline result, and that the combined model succeeds to some degree in calibrating the conjoint estimates to the baseline conditions.

The main advantage of the SP and RP/SP models relative to the first stage hedonic model is that they are both capable of delivering estimates of the willingness to pay for discrete changes in the conditions or existence of the AOC. As noted by Bockstael and McConnell (2007), however, there are different welfare measures we can use depending on whether or not households can move in response to a shock, and how we treat changes in prices. For our SP and RP/SP conjoint models, it is necessary to assume that households stay in their current home (i.e. they are not mobile), since the experimentally designed choice set does not characterize the true collection of homes a household might consider in the case of a move. This assumption is typical in most property value welfare measurements, given the difficulty of predicting counterfactual moves. Thus, the proper formula for welfare measurement using the utility function parameter estimates is given in equation (15), rather than by the log-sum expected utility formula used in other logit contexts. Note that this formula includes the price change effect, as well as the preference-based effect of the change in q. Conceptually, when households are renters the price change is a transfer from households to landlords, since the rental rate of the property rises when q increases. For the home owning households that constitute our sample, the change in price is a capital gain. In some instances we may want to include the price change in our welfare measure; however, doing so requires we have available a mechanism for predicting how market prices changes with changes in q. This is generally not available for the SP only approach. Thus, in reporting our welfare measures we focus on the first term in equation (15) – the pure willingness to pay – and separately predict the price changes using the hedonic model.

Table 5 contains point estimates for our counterfactual analysis of discrete changes in the AOC. We examine a discrete change in distance to the AOC that implies no house is closer than 4 miles to the site[2]. This is a common way to proxy the elimination of a disamenity, and it provides a direct means of comparing welfare measures from the conjoint data to predicted price changes from the hedonic estimates. The table shows household-level willingness to pay (*WTP*) estimates, broken out by different baseline distances to the AOC. In particular, we consider the value of our remediation proxy for representative households living at five different distances from the contaminated site under current conditions. For both models the *WTP* decreases as the baseline distance from the contaminated site increases, as one would expect. The magnitude of the welfare effect, however, is substantially smaller when we use the RP/SP estimates. Point estimates from the SP model are roughly four times are large as comparable estimates from the RP/SP model. Table 5 also displays predicted price changes from the hedonic model, computed using estimates from the full hedonic data set. For homes located close to the AOC the price change estimate is an order of magnitude larger than what we find with the two conjoint models. This difference, however, does not say anything about the quality of the welfare measures arising from either the SP or the RP/SP model, since price change predictions are measures of *WTP* only

[2] Computationally we do this by setting the new value for *AOC* to four for all survey respondents with $AOC_i \leq 4$ at the observed baseline.

in special cases. Indeed, for quality increases standard theory holds that the price change is an upper bound on the pure willingness to pay for the improvement, but the degree of overestimation is not generally known. One interpretation of our findings is that, for homes located close to the AOC at baseline conditions, predicted price changes are not a very good approximation to the true welfare effects.

6. Discussion

What are the advantages and disadvantages of our proposal relative to other options for measuring discrete change welfare effects using property value models? Consider first the disadvantage, which is that it is relatively data intense, since an SP survey is needed along with a property value database. Recall, however, that any effort to measure household-level preferences requires household level data, usually obtained via a survey. This includes the second stage hedonic model. Thus the extra data collection cost lies in the inclusion of a conjoint exercise; the fixed costs of a survey need to be borne in any case. Our sense is that the type of conjoint experiment a researcher would need to include in the survey is comparatively simple, and outside of the particular amenity being examined, could be standardized – thereby substantially reducing development costs for individual studies. Designing the SP questions for the particular amenity of interest would, or course, be study specific. The need to coordinate this with the amenity data collected for the hedonic model might also be viewed as a disadvantage of our approach. However, scoping exercises to determine how and which environmental conditions affect behavior – regardless of whether one uses an SP, RP, or combined technique – is a critical component of any non-market valuation study and not unique to our approach, except perhaps in degree.

There are three advantages to our approach, beyond what we have already described. First, econometric innovations (including quasi-experimental methods) and the rich home sales databases suggest marginal willingness to pay estimates from the first stage hedonic model are of high and increasing quality. Our approach allows analysts to leverage this progress by coupling it to conjoint models, which provide greater flexibility in the valuation measures that can be provided. Second, the environment in which the SP data is collected favors its use as we have proposed. Evidence suggests (citation) that people who are experienced with the good under consideration generally are able to accurately express their preferences via an SP survey. Our approach envisions targeting people who recently purchased a home – and hence have experience with the commodity – for the survey. Thus we combine the accurate baseline characterization from the hedonic model with SP data that is arguably gathered under near ideal circumstances. Finally, like maximum likelihood, our approach requires nothing more than numerical optimization to implement. Since starting values can be obtained from a conjoint maximum likelihood estimation routine, our sense is that the computational burden of our GMM model is comparable to maximum likelihood and small relative to the potential benefits.

7. References

Azevedo, C., J. Herriges, and C. Kling. 2003. Combining revealed and stated preferences: consistency tests and their interpretations. *American Journal of Agricultural Economics* 85: 525-537.

Bajari, P. and M. Kahn. 2005. Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business and Economic Statistics* 23: 20-33.

Bockstael, N. and Braden, J. B., A. A. Patunru, S. Chattopadhyay and N. Mays. 2004. Contaminant Cleanup in the Waukegan Harbor Area of Concern: Homeowner Attitudes and Economic Benefits. *Journal of Great Lakes Research*, 30(4): 474-491.

Bockstael, N. and K. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preference*: *A Theoretical Guide to Applied Practice*, Springer, Dordrecht, Netherlands.

Braden, J.B., L.O. Taylor, D. Won, N. Mays, A. Cangelosi, and A.A. Patunru, 2006. Economic Benefits of Sediment Remediation. Report to the Great Lakes National Program Office, U.S. Environmental Protection Agency under award GL-96553601.

Braden, J.B., L.O. Taylor, D. Won, N. Mays, A. Cangelosi, and A.A. Patunru. 2008. Economic Benefits of Remediating the Buffalo River, NY Area of Concern. *Journal of Great Lakes Research*, 34 (4): 631-648.

Cameron, A.C. and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, NY.

Cameron, T. 1997. Combining contingent valuation and travel cost data for the valuation of nonmarket goods. *Land Economics* 68: 302-317.

Chattopadhyay, S., J. B. Braden and A. Patunru. 2005. Benefits of Hazardous Waste Cleanup: New Evidence From Survey- and Market- Based Property Value Approaches. *Contemporary Economic Policy*, 23(3):357-375.

Earnhart, D., 2001. Combining Revealed and Stated Preference Methods to Value Environmental Amenities at Residential Location. *Land Economics*, 77(1): 12-29.

Earnhart, D., 2002. Combining Revealed and Stated Data to Examine Housing Decisions Using Discrete Choice Analysis. *Journal of Urban Economics,* 51: 143-169.

Ekeland, I., J. Heckman, and L. Nesheim. 2004. Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(1): S60-S109.

Harrell, F. 2001. *Regression modeling with applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.

Kiel, K.A. and M. Williams. 2007. The Impact of Superfund Sites on Local Property Values: Are All Sites the Same? *Journal of Urban Economics*, 61: 170-192.

Kling, C. 1997. The gains from combining travel cost and contingent valuation data to value nonmarket goods. *Land Economics* 73: 428-439.

McCluskey, J.J. and G.C. Rausser. 2003. Stigmatized Asset Value: Is It Temporary or Long-term? *Review of Economics and Statistics*, 85: 276-285.

Rosen, R. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82: 34-55.

Palmquest, R. 2005. Property Value Models. Chapter 16 in *Handbook of Environmental Economics*, K.G. Maler and J. Vincent, editors, Elsevier.

Von Haefen, R. and D. Phaneuf, 2008. Identifying demand parameters in the presence of unobservables: a combined RP and SP approach. *Journal of Environmental Economics and Management* 56: 282-98.

Whitehead, J., D. Phaneuf, C. Dumas, J. Herstine, J. Hill, and B. Buerger. 2010. Convergent validity of revealed and stated preference behavior with quality change: a comparison of multiple and single site demand. *Environmental and Resource Economics*, in press.

Whitehead, J., S. Pattanayak, G. Van Houtven, and B. Gelso. 2008. Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science. *Journal of Economic Surveys* 22(5): 872-908.

Table 1: Variable Descriptions and Summaries for Property Value Data

Category/Variable	***Description***	***Summary***	
House		Mean	Std Error
saleprice	sales price of parcel in 1000's of 2004 dollars	100.01	74.35
acres (acres2)	acreage of parcel (number of acres squared)	0.184	0.241
age (age2)	age of home at time of sale (age of home squared)	60.82	29.11
sfla	square feet of living area	1,543.43	635.08
bedrooms	number of bedrooms	3.22	0.81
fullbaths	number of full-bathrooms	1.25	0.56
halfbaths	number of half-bathrooms	0.34	0.49
grade_ab	dummy variable =1 if tax assessor assigns a quality grade of “a” or “b” (on a scale of a, b, c, d, e, with a being the highest quality)	0.07	-
grade_de	dummy variable =1 if tax assessor assigns a quality grade of “d” or “e”	0.03	-
grade_c	dummy variable =1 if tax assessor assigns a quality grade of “c”	0.9	-
cape	dummy variable =1 if home is described as a cape-cod style	0.21	-
colonial	dummy variable =1 if home is described as a colonial style	0.09	-
oldstyle	dummy variable =1 if home is described as “old-style”	0.41	-
otherstyle	dummy variable =1 if home is not among three above	0.29	-
fullbasement	dummy variable =1 if the home has a full basement	0.85	-
fireplace	dummy variable =1 if the home has at least one fireplace	0.23	-

Table 1 continued

Location		Total (I=3474)	% Total
Buffalo_N	dummy variable =1 if the parcel is in the City of Buffalo, north of AOC	1,041	29.97
Buffalo_S	dummy variable =1 if the parcel is in the City of Buffalo, south of AOC	427	12.29
Cheektowaga/Sloan	dummy variable =1 if the parcel is located in Cheektowaga or Sloan	794	22.86
West Seneca	dummy variable =1 if the parcel is located in West Seneca	881	25.36
Lackawanna	dummy variable =1 if the parcel is located in Lackawanna	261	7.51
Hamburg/Blasdell	dummy variable =1 if the parcel is located in Hamburg or Blasdell	70	2.01
North	dummy variable =1 if the parcel is located north of the Buffalo River	1,633	47.01
Census Tract ID	A series of dummy variables indicating which of 118 census tract in which each property is located.		
Proximity		**Mean**	**Std Error**
cbd	Distance to the central business district	5.01	1.71
delpark	Distance to Delaware Park	4.93	2.71
park	Distance to the closest park	0.55	0.34
rail	Distance to the closest segment of a rail line	0.57	0.4
stream	Distance to the nearest stream, other than the AOC	1.57	1.71
airport	Distance to the Buffalo Airport	5.79	1.99
hws	Distance to the nearest hazardous waste site	0.66	0.31
hwy	Distance to the nearest point on a major highway	0.77	0.46
hwyx	Distance to the nearest highway interchange	0.95	0.46
shore	Distance to the shoreline of Lake Erie	3.7	1.74
AOC	Distance to the AOC	3.05	1.27

Table 2: Conjoint Experiment Design

Attribute	***Description***	***Levels***			
House size	Size of home expressed as percent change from purchased home.	25%	15%	No Change	-15%
AOC Condition	Environmental condition of the AOC expressed qualitatively.	Full cleanup	Partial cleanup	No Change	More Pollution
Proximity to AOC	Distance to the Buffalo River expressed as change in distance from purchased home.	2 mi closer	1 mi closer	No Change	2 mi further
Home Price	Purchase price of home expressed as percent change from the price of home actually purchased.	30%	15%	No Change	-10%

Table 3: Selected Results from First Stage Hedonic Model

Parameter Estimates	Estimate	Standard Error	t-statistic
θ_1 - $\ln(AOC)$	6057.41	2582.888	2.35
θ_2 - $\ln(AOC)\cdot N_i$	-416.0771	8066.107	-0.05
θ_3 - $h(\ln(AOC))$	11173.25	6143.156	1.82
θ_4 - $h(\ln(AOC))\cdot N_i$	-41683.52	16653.64	-2.5
	I=3,474; R^2=0.830		
Marginal WTP Estimates	Sample Average MWTP	Average Percent of Value	Observations
All properties (I=3473)	-$78.76	0.21	3474
Properties with $N_i = 0$	$682.49	0.9	1841
Properties with $N_i = 1$	-$936.97	1.47	1633
Properties with:			
AOC_i<0.3	$2,397.91	5.39	40
0.3< AOC_i<0.5	$1,252.57	3.10	41
0.5 < $dAOC_i$<1.0	$690.37	1.37	185
1.0 < AOC_i<1.5	$411.97	0.72	225
1.5 < AOC_i<2.0	$266.66	0.36	374
AOC_i > 2.0	-$284.03	0.63	2609

Table 4: Estimates from SP and RP/SP Models

Parameter Estimates	**SP Only Model**		**RP/SP GMM Model**	
	Estimate	*t-stat*	*Estimate*	*t-stat*
β_0 - R	0.940	8.420	1.015	9.049
β_1 - ln(AOC)	0.128	1.074	0.248	4.288
β_2 - ln(AOC)·N	0.015	0.178	-0.012	-0.153
β_3 - ln(AOC)·y^a	0.010	0.933	-0.014	-1.539
β_4 - ln(AOC)·K	-0.029	-0.922	-0.059	-3.149
β_5 - ln($size$)	2.734	4.700	2.765	4.917
β_6 - ln(size)·y^a	0.164	2.068	0.143	1.939
β_{71} - E(more pollution)	-1.541	-5.365	-1.532	-5.611
β_{72} - E(partial cleanup)	-0.139	-0.674	-0.193	-0.942
β_{73}- E(full cleanup)	0.627	3.350	0.568	3.080
β_{81}- E(more pollution)·ln(AOC)	0.204	2.144	0.243	2.661
β_{82}- E(partial cleanup)·ln(AOC)	0.007	0.100	0.061	0.840
β_{83}- E(full cleanup)·ln(AOC)	-0.074	-1.168	-0.007	-0.107
β_9 - $price^a$	-0.356	-7.538	-0.315	-7.412

[a]All dollar-denominated variables are measured in \$10,000 increments at 2003 levels.
Estimation includes 2160 observed choice outcomes from 281 survey respondets.

MWTP Estimates	**AOC = 0.5, y = \$40K, N=0**		**AOC = 1.5, y = \$40K, N=0**	
	K=2	***K=3***	***K=2***	***K=3***
SP Only Model	\$614.22	\$451.81	\$204.74	\$150.60
RP/SP GMM Model	\$477.94	\$101.56	\$159.31	\$33.85

Table 5: Non-Marginal Change Welfare Estimates for Different Baseline Distances to AOC

	Baseline Distance to AOC				
	$AOC = 0.3$	$AOC = 0.6$	$AOC = 1.0$	$AOC = 1.5$	$AOC = 2.0$
SP Model	$5,851.57	$4,285.71	$3,131.72	$2,215.75	$1,565.86
RP/SP Model	$1,315.27	$963.31	$703.93	$498.04	$351.96
Hedonic price change	$15,954.00	$10,087.00	-$4,886.00	-$1,038.00	$963.00

[a] Computed for income = 40K and family size = 3

Table A1: Parameter Estimates from Hedonic Model (spatial fixed effects not reported)

Structure Characteristics			*Proximity Variables*			*Locations Dummy Variables*		
Name	Estimate	Std Error	Name	Estimate	Std Error	Name	Estimate	Std Error
acres	28,780.50	9,581.68	lncbd_n	-7,776.46	16,224.79	Buffalo_S	-216,600.70	153,286.60
acres squared	-8,324.73	3,337.79	lndelpark_n	-17,781.55	6,243.66	Cheektowaga	-143,145.30	143,669.00
age	941.00	293.84	lnpark_s	355.58	896.34	West Seneca	-218,308.90	153,035.50
age squared	-12.12	3.95	lnrail	48.14	1,040.81	Lackawanna	-212,695.20	153,193.20
age cubed	0.04	0.02	lnrail*north	2,115.71	1.41	north	-177,492.20	343,868.90
sfla	28.60	3.23	lnstream	-1,573.92	1,028.11			
sfla_north	6.10	4.40	lnstream*north	3,805.32	0.61	*AOC Variables*		
bedrooms	-2,401.56	1,302.97	lnairport	-40,432.70	17,554.79	Name	Estimate	Std Error
fullbaths	16,647.38	3,058.02	lnairport*north	29,006.67	-0.55	$\ln(AOC)$	6057.41	2582.888
halfbaths	6,073.35	1,728.43	lnhws	2,730.68	1,411.05	$\ln(AOC)\cdot N_i$	-416.0771	8066.107
grade_ab	65,393.44	6,826.95	lnhws*north	3,933.30	1.41	$h(\ln(AOC))$	11173.25	6143.156
grade_de	-3,753.88	2,795.42	lnhwy	-3,740.37	2,780.02	$h(\ln(AOC))\cdot N_i$	-41683.52	16653.64
cape	-8,725.07	1,647.78	lnhwy*north	5,568.40	-2.35			
colonial	14,177.28	3,756.28	lnhwyx	3,061.14	1,718.65			
oldstyle	-20,902.92	4,045.25	lnhwyx*north	2,881.50	0.59			
fullbasement	1,847.12	1.58	lnshore	-13,023.43	8,297.20			
fireplace	4,576.65	1,889.37	lnshore*north	14,907.35	2.29			

Figure 1: Identification problem in second stage hedonic estimation

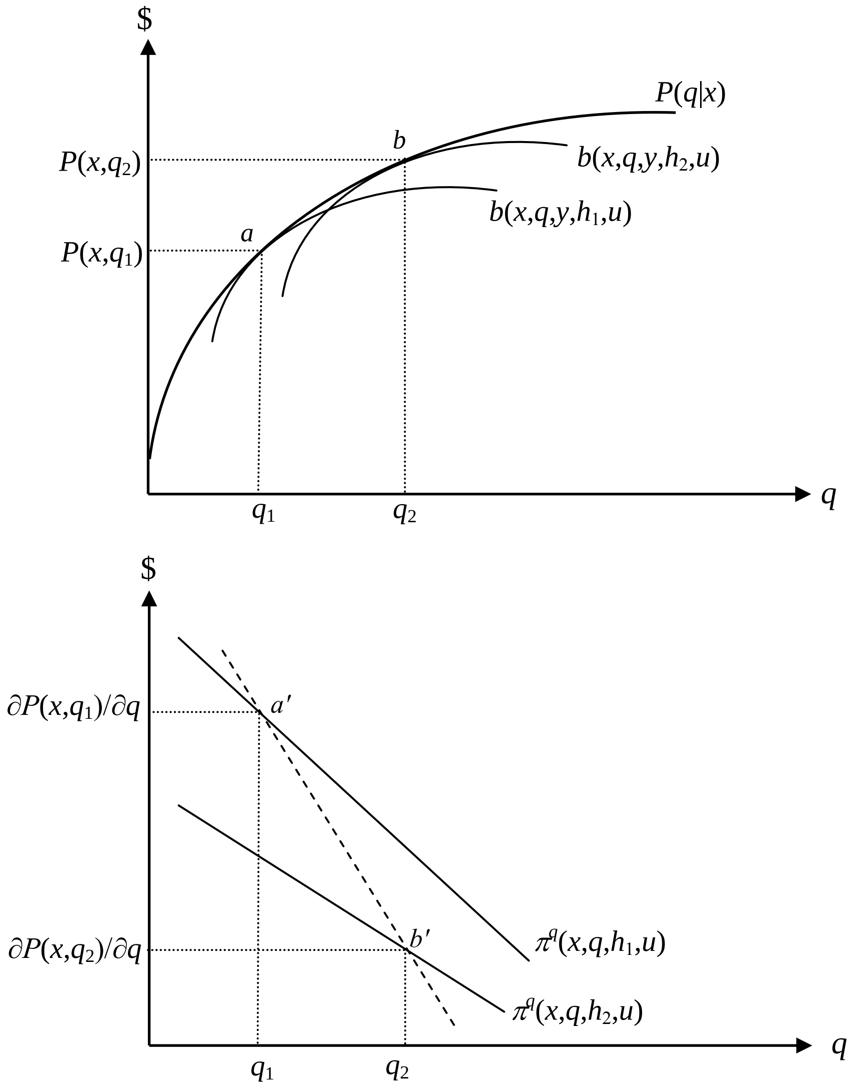


Figure 2: Map of Buffalo River, NY, Area of Concern

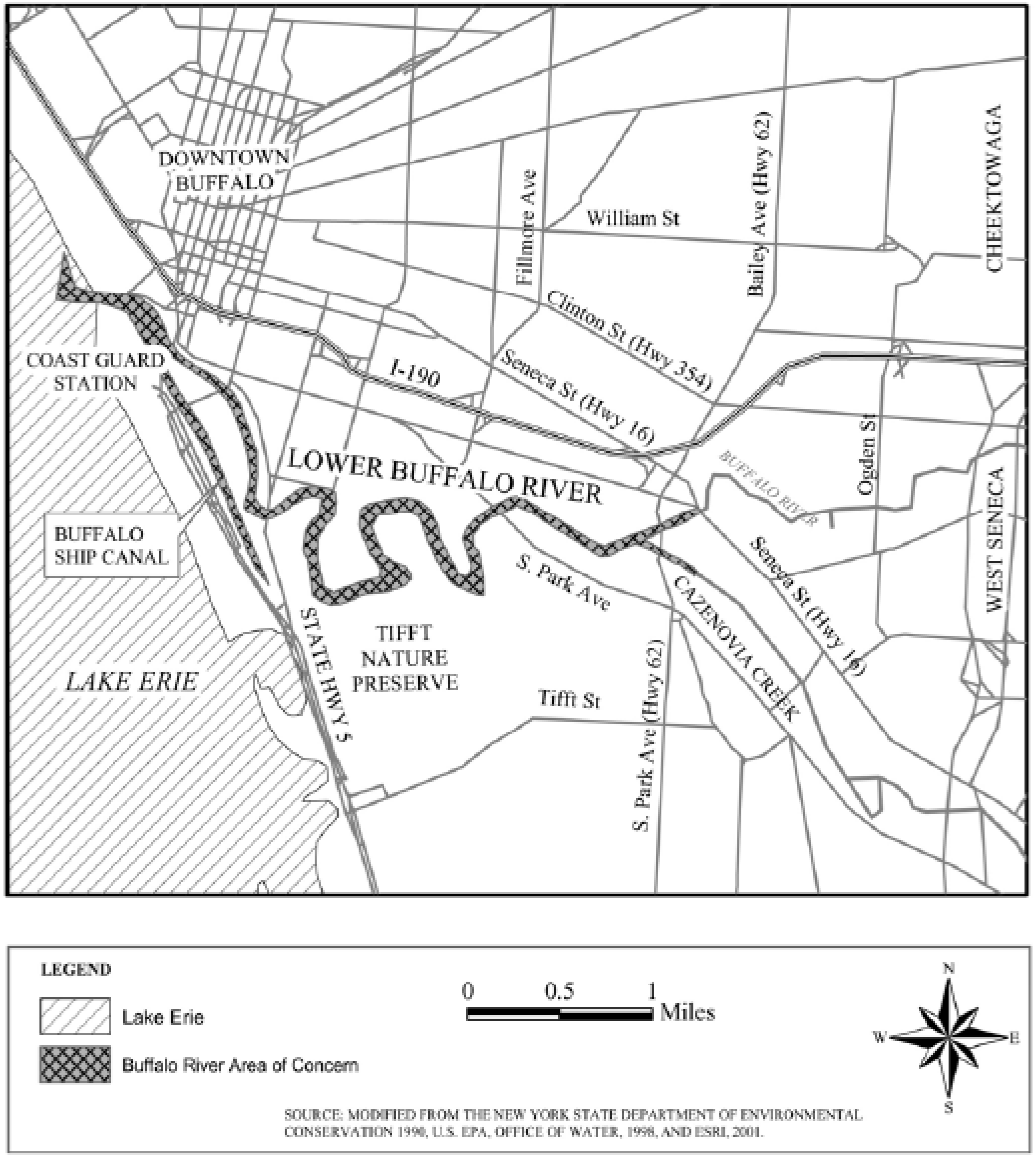


Figure 3: Example Choice Question

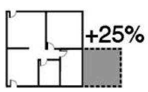
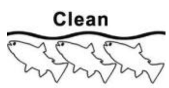
Imagine your current home modified as follows:				***Your Choice:*** (check one)
House size	**Environmental condition of Lower Buffalo River**	**Proximity to Lower Buffalo River**	**Home price**	1☐ **Modified home** 2☐ **Current home**
				

Figure 4: Partial effect of ln(*AOC*) on price