NBER WORKING PAPER SERIES

EVALUATING RUBIN'S CAUSAL MODEL FOR MEASURING THE CAPITALIZATION OF ENVIRONMENTAL AMENITIES

H. Allen Klaiber
V. Kerry Smith

Working Paper 14957
http://www.nber.org/papers/w14957

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2009

Partial support for this research was provided ASU Decision Center for a Desert City. Thanks are due H. Spencer Banzhaf for providing an initial version of the GAMS code to implement the assignment model. Thanks are also due Nick Kuminoff, Chris Timmins, and Randy Walsh for helpful discussions on several aspects of this research. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Evaluating Rubin's Causal Model for Measuring the Capitalization of Environmental Amenities
H. Allen Klaiber and V. Kerry Smith
NBER Working Paper No. 14957
May 2009
JEL No. C21,D61,Q51

ABSTRACT

This paper outlines a new framework for gauging the properties of quasi-experimental estimates of the willingness to pay (WTP) for changes in environmental and other non-market amenities. As a rule, quasi-experimental methods cannot offer alternative hypotheses to judge the quality of their quasi random assignments of treatment and control outcomes to economic agents. Their results must be judged by the explanation of the event used to construct the assignment and the counter examples offered as robustness checks for the logic of each application. This paper develops a four-step procedure for situations that rely on housing price capitalization. It is a computational analog to Chetty's [2009] call for considering the measurement objectives as part of evaluating the relevance of reduced versus structural form modeling strategies. Two diverse applications are used to establish the method's relevance for environmental problems. The first examines the value of a conversion of land cover from xeric to wet landscape. The second examines the clean-up of hazardous waste sites. We find that even when quasi-experimental methods have access to statistically ideal instruments their performance in measuring general equilibrium WTP depends on other aspects of each application.

H. Allen Klaiber
Department of Economics
Arizona State University
Tempe, Arizona 85287
allen.klaiber@asu.edu

V. Kerry Smith
Department of Economics
W.P. Carey School of Business
P.O. Box 873806
Arizona State University
Tempe, AZ 85287-3806
and NBER
kerry.smith@asu.edu

I. Introduction

Quasi-experimental techniques have been proposed as offering the best strategies for improving measures of the costs and benefits of policies that enhance environmental quality. This assertion relies on both the logical appeal of the Rubin [1974] causal model (RCM) and the persuasive arguments found in an array of recent environmental applications of that logic. Nonetheless, as Greenstone and Gayer [2009] suggest, transparency in the variation serving to identify an effect does not in itself guarantee that this strategy measures an economically consistent benefit or cost concept. To address this issue the analysis must be structured within an economic model that is consistent with the design used in quasi-experimental methodology.

An assessment of arguments for and against the RCM logic (versus structural modeling) for applied welfare analysis requires a controlled setting. This paper proposes such a framework and illustrates how it can be used. The framework checks the economic robustness of quasi experimental applications. More specifically, most applications of the RCM logic argue for their superiority to structural models largely because they assume less and focus more on what is revealed by the exogenous events identifying the outcomes analysts can observe. Their asserted advantages arise from a composite of each author's logical arguments (together with the documentation of background detective work) along with the properties of the statistics used to measure the effects of interest. Each application using quasi experimental methods for price capitalization has a set of assumptions forming a composite maintained hypothesis about *economic outcomes*.

Our proposal uses the information that must be assembled for applications involving housing price capitalization to calibrate a consistent structural model. This model provides the basis for developing a set of alternative hypotheses for judging the economic robustness of the RCM logic. To demonstrate that our suggestion is feasible we illustrate it with two "real" environmental applications. The two applications were selected to describe commonly observed, but quite different, types of environmental applications with hedonic models. The first case has a transformation in an environmental amenity that would be treated as a discrete influence to housing prices. This is represented with a change in the residential landscape for a set of subdivisions from desert (xeric) land cover to vegetative or "wet" cover with cultivated vegetation, grass, and plants requiring water.[1] The second is a discrete change in conditions due to a policy where the effect would be represented as a change in a continuous measure of the negative stigma (or risk) of a specific type of land use. This example is represented with the cleanup of landfills with hazardous substances. The result is characterized using changes in the proximity of private homes to these landfills.

In each case estimates of actual hedonic price functions are used to calibrate households' preferences. A large, spatially delineated, data base on housing

[1] As Imbens and Wooldridge [2009] suggest these are often the foci of quasi experiments. Indeed, in a more recent set of comments on critiques of the RCM Imbens [2009] argues there is a clear value in these types of evaluations, noting that: "Even if simple average effects of these interventions [as binary treatments] are not directly answering the questions about plausible economic policies, they are often closely related to the effects of such policies and therefore are viewed as quantities of interest" (p.6). The bracketed phrase and a correction for a typo were inserted. "The" was originally before "answering" in the quoted statement.

transactions in Maricopa County, Arizona is used to define a fixed supply of houses that is relevant for each application. This dimension of our example is important because it illustrates how we convey to an abstract evaluation method the specific features of the houses and neighborhoods along with the associated spatial correlations that would be present with each application. Each situation exploits the logic of an assignment model to represent the locational equilibria (see Wheaton [1974]) before and after the simulated exogenous change in environmental conditions[2]. The results of the assignment solutions are used to estimate the hedonic price equilibrium with and without the exogenous changes to the amenities involved. These data provide the basis for evaluating the quasi experimental logic against a wide array of alternatives.

Our results indicate Greenstone-Gayer cautions were warranted. Three conclusions emerge highlighting the potential importance of developing an evaluation strategy to judge how well quasi experimental methods measure the incremental benefits of a policy. First, for the use of a discrete change in landscape identifying the exogenously "treated" houses that switch from dry to wet (or the reverse), conventional cross section hedonic models dominate hedonic difference models with or without ideal instruments that exactly identified treated and control

[2] A concern with the assignment approach to representing the changes in the hedonic equilibrium has been demonstrating the existence of unique equilibria in both the base and control solutions of each paired experiment. Our normalization strategy assumes the parameter for income in the indirect utility function is unity for all households. The remaining parameters are structured to be heterogeneous and vary based on what is estimated in each example, as we explain below. This effect and the single crossing property assure that the highest income household would be assigned to the "best" house. This logic defines a unique starting point for computing the equilibrium prices in the base control solutions. It reduces concerns about the uniqueness of the two equilibria. While we do not have a formal proof of uniqueness, the logic we are applying is consistent with what would be needed to develop one (see the logic outlined in the Epple et.al [1993]).

households. The instrumental variable (IV) estimator, whether using hedonic price or a difference specification, tends to overestimate the mean general equilibrium willingness to pay (WTP). OLS hedonic models, by contrast, underestimate the WTP, but the errors are generally less than ten percent.

Second, in the case of the hypothetical policies intended to represent hazardous waste cleanups as changes in a continuous measure of a stigma or risk, the quasi-experimental methods – whether hedonic price or difference models - provide superior estimates of the GE willingness to pay. Finally, the heterogeneity in households' preferences together with the role of sorting responses to exogenous changes in local amenities imply that judgments about the relative performance of conventional cross-section versus quasi-experimental methods depend on the sub-sample selected for these evaluations. Thus our findings for willingness to pay estimates are consistent with Imbens [2009] recent summary of the objectives and accomplishments of the RCM literature.[3]

Our analysis is developed in five sections after this introduction. Section two outlines our proposal and explains how the quasi-experimental framework has been applied to evaluate environmental policy based on housing prices. Section three describes the specific details of the calibration template for evaluating the RCM's logic for non-market valuation. Sections four and five present our results. The first tracks an evaluation using alternative hypotheses that assume we observe how

[3]His objective is to respond to criticisms of what seemed to others to be an exclusive reliance on the RCM as the only reliable approach to evaluating policy. He observes in describing what the literature has found that: "An important insight is that in settings with heterogeneous effects, instrumental variables do not identify the average effect of the treatment. However, as shown by IA (Imbens and Angrist) instrumental variables methods do identify the average treatment effect for a well defined subpopulation…"(p.4)

people sort in response to the exogenous changes in amenities. The second assumes the evaluation would be based on hypotheses derived from tracking the price changes of the same houses before and after a policy change that is considered as a treatment. The last section summarizes our findings and the potential for wider use of calibrated models in judging the robustness of quasi experimental applications.

II. Evaluating Quasi-Experimental Strategies for Non-Market Valuation

Recently, Imbens and Wooldridge [2009] have decomposed the logic of the RCM into three elements: (a) the distinction between potential and realized outcomes; (b) the properties of assignment mechanisms that "select" treated and control entities; and (c) the potential for interaction and / or general equilibrium effects that might link treatment and control entities through market or non-market processes. Such links might confound analysts' ability to isolate the source of variation in the condition applied to treated subjects. We exploit the Imbens-Wooldridge structure in designing the features of the examples we use to describe our methodology.

A. Using Calibrated Preferences to Frame Alternative Hypotheses for an RCM Application

Our framework has four elements. First, we follow the Cropper et al [1988] adaptation of the Koopmans-Beckman-Wheaton description for how equilibrium prices in a housing market could be computed by interactively solving an assignment problem. Second, the actual data underlying each proposed quasi experiment is used for model calibration. Most environmental applications measuring economic benefits or costs of a policy (as opposed to health related outcomes) are likely to involve housing transactions and actual site specific effects of the policies. This process assures the spatial correlations in these features, observed in any particular sample of the housing transactions that would be involved in estimating the price responses to a policy, will be embedded in the evaluation. Preferences are calibrated by linking the parameters in the preference function used in defining the bid functions for the assignment problem to the results from estimated hedonic price functions. Third, alternatives relevant to the specific details of each quasi experimental application are framed as a set of scenarios that might influence price outcomes. By using the actual data these alternatives match the spatial attributes of the actual sample of housing units selected for each proposed quasi experiment. The alternative hypotheses are then derived as different simulations using the calibrated preferences for a set of households generated to be consistent with the calibrated preference heterogeneity and a sample of housing units spatially located to mimic the effects of each policy. Finally, the computed equilibrium prices for baseline and treatment scenarios in each case are used to estimate different statistical models for each scenario. In this context it is possible to judge how the features of each alternative affect the performance of a quasi experimental strategy for estimating the benefits (or the costs) of a policy that is expected to influence housing prices.

B. Concepts Measured

For most applications associated with environmental policy questions economists are asked to provide estimates of individuals' willingness to pay for changes in amenities. If we assume V_i (m_i, p, q) is the i[th] individual's indirect utility function with m_i the income, p a vector of prices for private goods, and q a vector of non-market environmental resources available to this person, then equation (1) provides one definition for a person's general equilibrium willingness to pay, WTP_i. For this definition we assumed the change in q, Δq, is large enough to change relative prices for private goods (from p to $\tilde{p}$) but not income.

$$V_i(m_i - WTP_i, \tilde{p}, q + \Delta q) = V_i(m_i p, q) \quad (1)$$

Analysts need to have estimates of the *WTP_i's* and costs before the policy is undertaken. Because real experiments with random assignment are usually technically and / or politically impossible, we must "make do" with observed variation in q across different places or time periods. The challenge is to use observable variables, such as the housing prices, that might reveal how differences in q motivated people to select one place over another and what this implies for the tradeoffs each would make for Δq. The new focus on quasi-experimental methods has argued these selections should also seek to mimic, as nearly as possible, information that allows the analyst to distinguish the tradeoff of interest from other objectives that also motivate sorting in response to differences in spatial amenities.

There are two aspects of adapting the Imbens – Wooldridge [2009] logic to this task. The first receives virtually all of the attention in the literature and involves the internal validity of the estimates measuring the treatment effect. The second focus, a key element in our evaluation template, arises because what we observed as the "outcome measure" is not *WTP*. In many cases it is a housing (or land) price (or rent).

Greenstone and Gayer effectively summarize how the use of an instrumental variable can address the problems posed by internal validity and the estimation concerns for conventional approaches for estimating the capitalization of site specific amenities into housing prices. As a result we focus on the economic concept that is measured using estimates of the price capitalization.

The interpretation of capitalization measures has a long history in economics. Scotchmer's early papers [1985, 1986] make the key points directly. For non-marginal changes in spatially delineated amenities she noted that "…even in the simplest case of homogenous populations, neither the long-run nor short-run benefits is a simple difference in hedonic prices." (p. 63) The problem in linking price capitalization with benefit measurement is not confined to short run adjustment. Kannemoto [1985] demonstrated that price differences overestimate long-run benefits.

The literature on non-market valuation has established that the marginal willingness to pay can be recovered from hedonic price models. However, estimates are subject to the identification concerns raised by the quasi-experimental literature that are an important focus of the Greenstone-Gayer discussion. Nonetheless, if identification is resolved, as Scotchmer notes, "At best the observer makes local observations, localized at the endogenous choices of the consumer." (p. 66) Her examples highlight how heterogeneous income among consumers, as well as complementarity (or substitution) relationships in preferences for the amenity, the amount of land, or other spatially delineated attributes confound the relationship between price measures and willingness to pay. Equally important, other differences among consumers' preferences distinguish the appropriate short run measure of benefits from price capitalization measures. Indeed, even in the case when one postulates that the "correct" hedonic price function implies a constant marginal price, this does not mean the underlying marginal willingness to pay function will be constant.

Greenstone and Gallagher's [2008] use of price capitalization for land improvements to estimate the benefits of those improvements illustrates how the issues raised by Scotchmer are important to the economic meaning of what is measured by the RCM logic. Figure 1 reproduces part of their figure 1b for the case of a perfectly inelastic supply of residential land. Assume an improvement takes place in the site specific amenities relevant to this land. The demand for land would increase (in the example represented in the figure from D_0 to D_1). Greenstone and Gallagher assume the effect is small enough that the hedonic price function does not change and they argue the gain is measured as P_0 P_1 DC –"…the mean change in prices times the number of residential parcels…." (p.965) Their argument and the implicit assumptions of the diagram are subtle. Conventional practice would assume the gain from the improvement in site specific amenities would be measured by the increase in the consumer surplus fromP_0 AC to P_0 BDC or ABDC. The change in land prices from P_0 to P_1 affects how this gain is divided between residents and landowners. Greenstone and Gallagher's argument is that the gain is perfectly captured with the price capitalization. This outcome is consistent with their diagram due to the parallel shift of the demand function (and the area of the parallelogram P_0 P_1 DC)[4]. If we alter the assumed effect of the site specific amenity on the demand for land –assuming the demand pivoted at A, instead of a parallel shift, the correspondence between price capitalization effects and the benefits of the amenity change no longer holds. This difference illustrates in simple terms the Scotchmer arguments[5].

These results are recognized by those proposing structural and reduced form models based on hedonic price functions. What is at issue in these discussions is the "size" of the exogenous change providing the identifying variation, the interrelationships between the attributes of houses in preferences and the spatial correlations inherent in the supply of homes for each application. Considering the first of these, the exogenous change linked to the environmental service or amenity

[4] We are grateful to Chris Timmins for reminding us of this relationship that is central to using the diagram to illustrate the arguments Scotchmer's arguments.

[5] A closely related set of issues are raised in Starrett [1981] analysis of how to interpret price capitalization measures. He provides two conditions that guarantee the benefits of an exogenous change or project will be translated to land rents thru the intensive margin. The first he describes a requiring the benefits be intramarginal. Or in the terminology of revealed preference approaches to non-market valuation he assumes weak complementarity holds for households in the boundary region. Second he assumes residents do not sort within the town according to their relative preference. Neither of these assumptions will be satisfied by our framework. Sorting mitigatres capitalization. pp.313-314.

of interest, that is also inducing the price change, the variation needs to be sufficient (and exogenous) to assure the analysis has successfully identified a casual change in outcome of interest. However, large changes, such as those associated with past, exogenous, policy outcomes, can alter the *economic* interpretation of the measure attributed to the treatment effect. Price capitalization, as we suggested, is not guaranteed to measure either the marginal or the incremental (for large changes) willingness to pay. Nor does it relate to either a short run or long run measure of the benefits associated with a given policy change. Its interpretation depends on the assumptions we make about how the amenity is measured; how it contributes to individual utility; how diverse preferences for the amenity are across households being represented with the market effects of the amenity change; and how the markets, in the case of housing price capitalization, come to equilibrium. To address these interpretation issues we need an analytical structure with a set of maintained assumptions for each of the factors contributing to the equilibrium market adjustment. This structure is what we propose. It is a computational analog to the rubric that Chetty [2008] outlined for basing the choices between reduced form and structural models on the ability of each to measure the economic concept sought in each application[6]. His strategy uses an analytical framework for policy interactions that alter prices in accepted ways.

For environmental policies there are different types of connections between the non-market services and the private goods (i.e. the land and housing) bundled with them. As a result, the prices for these private goods are affected in a variety of

[6] He highlighted the importance of a set of strategic assumptions assuring that the reduced form estimates involved offer a sufficient economic statistic.

ways by the policies of interest. Thus, in these cases, an interpretation of what can be measured is more complicated and, in large part, specific to each application. Analysts do not know what would be a robust specification for the correlation of the important non-market amenities with other site and housing attributes. Our strategy addresses these limitations by using the actual details of each application to calibrate a model for the process that allows outcomes to be simulated. We can then evaluate whether quasi-experimental methods with housing price capitalization can be interpreted as a sufficient statistic for the general equilibrium willingness to pay for the policies involved.

C. The Assignment Problem

The formal definition for a housing market equilibrium using the assignment logic maintains the indirect utility function in equation (1) can be redefined in terms of a vector of housing attributes, A_j, a vector of household characteristics, C_i, with j an index for houses and i the index for households as in equation (2).

$$v_{ij} = V_i(m_i - B_{ji}, A_j, C_i) \tag{2}$$

$B_{ji}(V_i)$ is household i's bid for house j with its utility level at v_{ij}. Assume $X_{ji} = 1$ if household i occupies house j and $X_{ji} = 0$ otherwise. An assignment equilibrium arises when there is (for a given, fixed supply of N houses and fixed number, N, households) a set of utilities $\nu^* = (\nu_1^*, \nu_2^*, \ldots \nu_N^*)$ and housing prices $(P_1^*, P_2^*, \ldots P_N^*)$ with the NxN matrix X such that the four equations in (3) are satisfied.

$B_{ji}(v_i^*) = P_j^*$ if $X_{ji} = 1$ for all j and i

$B_{ji}(v_i^*) \leq P_j^*$ if $X_{ji} = 0$ for all j and i

$\sum_{i=1}^{N} X_{ji} = 1 \quad j=1, \ldots, N$

$$\sum_{j=1}^{N} X_{ji} = 1 \quad i=1, \ldots, N \tag{3}$$

The first component of equation (3) equates the equilibrium price for each house to the maximum willingness to pay of the household occupying it. The second equation states that no one is willing to pay more for a house than the person assigned to the house. The last two equations imply that all households have houses and all houses are occupied, respectively.

Our calibration of preferences, selection of the supply of housing units, and authentic spatial delineation of each policy assure that our computation analog offers information to judge the performance of quasi-experimental methods in each application as part of judging their robustness. The specifics of each of our examples are discussed in the next section. Each requires two solutions to the assignment model. The first corresponds to a baseline condition defined by the actual spatial distribution of the amenity (or disamenity) associated with the sample of houses selected for each problem. The second corresponds to changes in the spatially delineated amenity that are associated with discrete policies.

Our design allows for household specific preference parameters – based on applying the Brown and Rosen [1982] logic for each application. Because each set of preference parameters is assumed to represent a household we can track the re-sorting of these agents among houses from baseline to policy solutions. This tracking allows two separate interpretations of how a quasi-experimental analysis might take place. The first matches each household recording their baseline and policy scenario housing assignments and associated prices. The second matches each house keeping track of the differences in the households (and their characteristics) assigned to each house in the baseline and policy equilibrium solutions.

For the first of these matches, with the household held constant, it is clear that households' preferences are the same and the challenge is to use different housing choices made by each household following the change to recover their willingness to pay for the environmental change. In the second type of match, with the house the same, different households may be present in the baseline and policy solutions. As a result, the challenge for estimation is to ask how the capitalization relates to each household's willingness to pay for the environmental change.

The first of these matches corresponds more closely to applications of the quasi-experimental framework in labor and health related applications than it does to those for environmental policies. The second is comparable (except for the level of aggregation) to the hedonic applications to air quality (Chay and Greenstone [2005]) and the cleanup of hazardous waste landfills (Greenstone and Gallagher [2008]).

These contrasting uses of the benchmark and policy samples highlight the reason why Imbens and Wooldridge discuss the importance of distinguishing general heterogeneity in the unit effect of a treatment from heterogeneity that is causal. For benefit measurement we often assume households have different willingness to pay for changes in amenities. In the context of models that use capitalization measures to estimate an average effect, the equilibrium sorting of households influences what can be recovered and the Imbens' argument (cited earlier) for the importance of distinguishing the local average treatment effects.

In the more typical case where the analyst tracks the house – the households are different! This is the issue explicitly raised in the conclusions from Scotchmer's analysis. We observe the re-valuation of the properties due to the effect of the change on the equilibrium but to measure the incremental benefits associated with the change we need to be able to recover the willingness to pay of the household who occupies the house after the change.

Some applications argue the change is small in relation to the market so households don't move. This is the implicit assumption in figure 1. It is also what is assumed in cases where the marginal price from a hedonic price function is used to estimate marginal willingness to pay (and linear approximations used for the benefit measures for large changes). While this logic may be a reasonable characterization of the relationship between the effect estimated with a hedonic price equation and the marginal willingness to pay, it does not consider the ability of the statistical model to identify the actual effect of the amenity. This latter concern is the focus of the quasi-experimental literature. Simultaneity and omitted variables confound the ability to isolate these effects. Under that argument an exogenous change is needed to distinguish the effect of the amenity from the endogenously determined matching that lead to the cross sectional hedonic function. This conflict is one motivation for developing our computational analog to Chetty's framework. Within such a framework it is possible to evaluate the relative importance of the statistical limitations in interpreting the welfare properties of capitalization measures for large changes in spatially delineated amenities versus the econometric limitations in assuming estimates from a snapshot of the equilibrium isolate reliable estimates for the economic tradeoffs.

III. Development of the Evaluation Framework

A. Estimation of Hedonic Price Functions

Before getting into the specific details of each application, it is important to acknowledge that each is a "story." Our primary interest is illustrating our evaluation method. Our approach provides insight into what might be the expected error bounds for a benefit interpretation of the capitalization measures derived from each quasi-experiment. Thus we seek to assure the relative magnitudes of the calibrated parameters are consistent with the data and what would be observed in practice. We also want the supply structure of houses to represent the correlations in housing and location specific attributes. Our estimates of each hedonic price function are used to define the "true" value of the preference parameters.

The first of our examples involves the tradeoffs households are prepared to make to live in "wet" or irrigated landscapes as compared to desert (xeric) landscapes in metropolitan Phoenix. Given the scarcity of water, this question has attracted increasing policy interest. Land cover in the Phoenix area is predominantly xeric due to the high summer temperatures and limited rainfall. Nonetheless for historic reasons, due in part to the allocation of water rights to agriculture, a large number of residential neighborhoods consist of predominantly "wet" landscaping. Recent research (see Smith et al [2009]) has shown that households are willing to make significant tradeoffs to reduce minimum summer nighttime temperatures in the Phoenix metropolitan area. In addition, Brazel et. al. [2007] have shown that water intensive land cover is associated with cooler nighttime temperatures.

We use these results to motivate our first application. Using land cover data developed at the Central Arizona Project Long Term Ecological Research Site (CAP-LTER), together with the Stefanov et al. [2001] classification system, we matched land cover designation to individual housing parcels. Because temperature is also related to land-cover, we developed predictions for minimum summer temperatures. Appendix A outlines the details. Our prediction process assures that the temperature measure is correlated with land cover. Both are specified in the hedonic price function and treated as separate arguments of the household preference function.

Our estimates for the hedonic price function are based on housing sales from 1980 through 2004 in Maricopa County (before the "housing bubble" in the area). We estimate a semi-log specification for the price equation using housing attributes – square feet of living space, lot size in acres, number of stories, age of home, presence of a pool, presence of a garage, and two landscape-related variables – a dummy variable for the wet / xeric classification and our estimate of the July minimum temperature for each parcel. We also measured the distance to the central business district. The model included fixed effects for the year of the home sale. There are 398, 200 observations and as one might expect (with this large a sample) we are able to estimate the housing and site attributes with considerable precision. Table 1 reports summary statistics for the sample as well as our estimates of the first stage hedonic model we use for the landscape application.

Our second application was designed to consider a continuous measure for the disamenity associated with hazardous waste sites. We selected the inverse of distance to the closest site with hazardous substances as a measure of the disamenity. This selection was partially motivated by the recent Greenstone-Gallagher study calling into question most past hedonic analysis and because the Phoenix area offers an ideal location for studying this issue. The Phoenix area contains no less than seven sites on the National Priorities List (NPL). Our analysis considered three of these sites: Motorola 52nd street, 19th Avenue Landfill, and Indian Bend Wash. Appendix B provides a brief description of the three sites. As shown in figure 1, the Motorola and Indian Bend sites are located relatively close to each other, while the 19th Avenue Landfill site is located away from other NPL sites.

We use these geographic features to vary the effects of cleanup on our measure of the disamenity effects of proximity to a hazardous waste site.[7] For this application we limit our sample of housing sales in Phoenix to the years 1990 to 1999. This selection assures that the seven hazardous waste sites were listed as NPL sites and known to the public. There were 242,827 single family residential transactions during this period. In addition to the housing attributes used for the landscape application, we included city and block group fixed effects to account for other spatially varying unobservable factors that could impact housing prices, as well as temporal fixed effects. Table 2 reports the sample characteristics and the estimates for the first stage semi-log specification in this case.

A. Preference Estimates and Simulating Household Heterogeneity

Following Brown and Rosen [1982], if we assume a common preference structure together with independent identifying variation distinguishing marginal price and virtual price equations for attributes, hedonic price functions can be used to estimate preference parameters[8]. Equation (4) provides the preference specification of our example. It is the Generalized Leontief function expressed in terms of housing attributes (the A_{jk}'s), household income (m_i), and the rental price of housing type *k* (P_k).

[7] This design is intended to reflect the concerns Nancy Bockstael raised with the Greenstone-Gallagher use of census aggregates in attempting to detect the effects of cleanup of hazardous waste sites on housing values.

[8] Here we are using the algebraic logic and not requiring the exogenous variation distinguishing marginal price and virtual price functions, since the objective is to calibrate a preference function for gauging the robustness of each quasi experiment to different alternative hypotheses.

$$\nu_{ik} = \sum_{j=1}^{L} \alpha_j A_{jk}^{1/2} + \beta (m_i - P_k)^{1/2} \tag{4}$$

The virtual price function for an attribute, for example A_1, (with Π_{A1} the virtual price for A_1)is then given as in equation (5).

$$\Pi_{A_1} = \frac{\nu_{ikA_1}}{\nu_{ikm}} = \left(\frac{\alpha_1}{\beta}\right)\left(\frac{m - P_k}{A_{1k}}\right)^{1/2} \tag{5}$$

Or taking the logarithm of both sides we have equation (6).

$$\ln \Pi_{A_1} = \ln\left(\alpha_1 / \beta\right) + 1/2 \ln(m - P_k) - 1/2 \ln \Pi_{1k} \tag{6}$$

In general the parameter β cannot be separately identified from the α_j's[9]. In our case we normalized it to unity. We estimate the preference parameters for all of the attributes jointly, using the estimates from the marginal price equations for each

[9] This same logic could be adapted with alternative preference specifications. We selected the generalized Leontief because it has advantages over the translog in representing substitution patterns between the attributes of differentiated goods when these substitution elasticities are relatively small. In a related application Kuminoff, Parameter and Pope [2009] (KPP) evaluated the sensitivity of the Cropper Deck and McConnell[1988] conclusions about the properties of alternative hedonic specifications to omitted variables. The KPP study considers whether fixed effect and temporal differences alter the guidance offered by Cropper et al. KPP found their conclusions on the effects of omitted variables and spatial fixed effects were not influenced by the functional form used to characterize household preferences. While our objectives are different, their findings suggest that consideration of a wide array of functional specification may not need to be a component of the evaluation rubric we are proposing for quasi experimental applications.

hedonic function. The resulting estimates for the α_j's provide the basis for calibrating the preference functions for each of our applications. The estimated covariance matrix for the estimates of the α's provides the basis for introducing heterogeneity in preferences across the households in our simulation. Our estimate for household income is derived from the Census and corresponds to the block group mean assigned to each house within a particular block group. Housing price is annualized at a rate of 11%. Appendix C reports the specific estimates for the second stage hedonic preference estimation. We considered all of our experiments using the estimated heterogeneity as well as greater levels of heterogeneity, induced by scaling the covariance structure by a factor of ten. This increase did not alter any of our findings. As a result, our models are based on the heterogeneity estimated for the sample in each application.

B. Samples of Houses

Both of our examples are based on samples of 1000 houses from the same Phoenix area used for the hedonic models. For the first experiment, we identified 23,000 unique subdivisions from the Maricopa County assessor GIS files. For each subdivision, we calculate the percentage of homes classified as either "wet" or desert and restrict attention to subdivisions with a minimum of 25 single family residential parcels. For each of those subdivisions, we define a subdivision as wet if at least 90% of the houses within it are classified wet and we define a desert subdivision as having no wet houses. Using the reduced sample of houses with only wet and dry subdivisions, 20 houses were selected from each of 25 wet and 25 xeric subdivisions.[10] By selecting subdivisions that are either wet or xeric, rather than just selecting 1000 houses randomly, our sample reflects unobserved spatial correlation between distinct types of subdivisions that would be expected in an actual application of the quasi-experimental methodology to estimate the economic effects of a landscape amenity on housing prices.

Our strategy for the second experiment was also structured to portray the spatial array of houses in relation to NPL sites in a way that reflects proximity in Maricopa County. We focus on three hazardous waste sites that we identified earlier as important to our experiments. 500 houses were randomly selected based on being located within 3 miles of both Indian Bend Wash and the Motorola 52nd Street plant. The remaining 500 houses are randomly drawn from within 3 miles of the 19th Avenue Landfill and are not within three miles of any other hazardous waste site. Furthermore, our selection criteria assured that we compiled 250 houses whose closest site was Indian Bend and 250 houses whose closest site was Motorola. As our preference structure assumes the disamenity effect is for the nearest hazardous waste site, we do not allow for effects from the more distant site even though it is within 3 miles of each house.

C. Design of Experiments

[10] Only wet houses were selected from wet neighborhoods.

As noted above the landscape treatment assumes a discrete change in conditions. Half of the baseline wet homes switch to desert landscape and half the xeric to wet. The instruments in this case correspond to dummy variables identifying the specific homes switching from wet to dry or the reverse. The second case involving the cleanup of hazardous waste sites can describe a wider array of alternative implications for the resulting dis-amenity effect influencing housing prices. Here the non-market effect is represented through a continuous variable (i.e. the inverse distance to the closest site). As a result, the calibrated preference specification for the hazardous waste case was used in two separate ways. Both hazardous waste experiments share a common baseline with different assumed treatment effects. In the first experiment we specify that the Motorola site present in the baseline hedonic price function was cleaned up leading to a change in distance to the closest site for a subset of the houses. The houses whose initial closest site was Motorola become closest to the Indian Bend site after the cleanup. Thus the quality change is represented as an increase in the distance to the closest site with hazardous substances. Figure 2 displays the positioning of the two sites along with the overlap in the areas with homes that could be affected by the cleanup of one of the sites. This overlap is the basis for a cleanup altering the site which is closest to a subset of the homes (i.e. some of the ones closet to Motorola become closer to Indian Bend when the Motorola site is cleaned up).

The second treatment maintains a policy that simultaneously changes two sites and thus the new prices would reflect sorting in response to both changes as they affect different houses in the set of 1000 contributing to the equilibrium. One of the changes involves the assumed cleanup of the Motorola site, as before. For those houses close to the Motorola site, their closest site is once again the Indian Bend site. The second cleanup is based on the 19th Avenue Landfill. To maintain an isolated NPL site for a comparison following the cleanup, we divide the 500 houses selected near the 19th Avenue site into two groups of 250 houses, chosen randomly. In effect, this converts the single, isolated 19th Avenue Landfill site into two isolated sites with an identical structure for the underlying unobserved spatial correlation. The policy cleans up only one of these sites, reducing the dis-amenity effect of proximity to zero. Thus, while there is a differential effect for each of the houses close to this site (because in the baseline they are all different distances from the site), after the treatment some houses no longer experience an effect of any hazardous waste site influencing the preferences of those simulated households living near the cleaned up site.

The analysis challenge in the first cleanup case arises because a subset of the houses close to Indian Bend in the treated sample experienced an improvement. The analyst using the *ex post* cross sectional sample would not know which ones. For the second case distance changes mean different things depending on which site was cleaned up. While the preference specification treats them as affecting distance to the disamenity and thus they would have equivalent effects, this need not be the case and the analyst does not have this information when specifying instruments for separating each potential policy change. This is important because we can define different types of instruments to reflect the exogenous identifying information associated with the treatment that would be added as part of the logic of the quasi experimental method.

For the case of the experiments measuring the cleanup effects for hazardous waste sites, we have several different instruments. In the first of these experiments, cleaning a single site, it would be the identification of which of the homes close to Indian Bend in the treated sample were initially close to Motorola and therefore received the cleanup. For the second example, this variable along with the identification of the houses close to the isolated site that was cleaned could be used as instruments. These two instruments can be compared to the situation where we simply have a dummy variable identifying that a house is close (in the baseline sample) to a site that was "cleaned-up". The issue we consider is whether this distinction enhances the properties of the quasi experimental estimates.

D. Reduced Form Models

Our experimental design allows two different ways of interpreting the treatment components of each of our experiments. Each has different implications for the reduced form models used to estimates the effects of a treatment on equilibrium housing prices. The first matches the baseline and treatment price equilibrium by the simulated household. This formulation, as we noted, is more consistent with the objective of measuring the willingness to pay for each change but implies we observe the household in different homes. For hedonic applications, to our knowledge, this level of resolution simply does not exist in any of the existing applications.[11] The second approach matches houses and is more consistent with the practice adopted for environmental applications. It implies that potentially different households will occupy each house in the benchmark and treated solutions. For each situation a somewhat different set of models is likely to characterize what would be used in practice to implement the quasi-experimental method. We will not attempt to consider every possible alternative. Instead, a representative set of models are used: single cross sectional models, difference models, and instrumental variable models used to represent the quasi-experimental methods. In practice the specifications of this later group can be variations on one of the first two types. The results reported in the next two sections were selected to document a set of general conclusions that emerge from our experiments.

E. What is the Standard?

The final issue to be considered in evaluating our models concerns what we assume is the true value to be measured. Our experiments assume each household has different preference parameters, income, and thus will have a different willingness to pay for the environmental changes represented in each experiment. As a result, our standard could be the average of these values. Alternatively, if we assume the policy is a change in a site-specific amenity – we might use the "true" willingness to pay of those observed to be directly affected by the change. This distinction is important because the matched samples reflect the policy's effects in

[11] This matching is more characteristic of the applications in labor economics.

different ways. When we consider the same households in the baseline and treated circumstances we could consider individuals selecting houses with the improved state. By contrast, with matched houses we could consider willingness to pay of households in the affected houses before or after the exogenous change. This is an important element in the Kuminoff and Pope [2009] discussion of the results of what can be expected from empirical models of capitalization. Hedonic models estimated with the baseline sample reveal marginal willingness to pay for small changes in attributes before the “large” exogenous treatment, while those after are for the prices defining the new equilibrium. This result is consistent with the point raised by Schotchmer and identified in Kuminoff and Pope’s extension to Epple’s [1995] analytical derivation of the hedonic price function. To account for these issues, our summary includes both sample means and sub-sample results to allow for comparisons.

IV. Results – Match on Individuals

As our discussion to this point has suggested, there are a number of aspects of the property value capitalization framework that distinguish it from the traditional one used to describe quasi-experimental methods. In this case, none of the three Imbens-Wooldridge attributes exactly “fit” the conventional setting. The potential outcome that is measured, even under ideal conditions, does not correspond to what we would like to estimate. The treatment is usually assigned to houses (or groups of houses) while estimates of the tradeoffs are what households would make. These tradeoff measures should be the focus if the objective is to measure the benefits from a policy.

Everyone in the market "experiences" the treatment through the interactions that define a hedonic equilibrium. Assuming the household in a treated house is the only one experiencing the effect ignores this general equilibrium process. Equally important stratification at spatial thresholds can arise as a byproduct of sorting to a new equilibrium. For these reasons we evaluated the two sets of equilibrium prices in each experiment in two ways – (a) matching the simulated households with information on the houses they selected in the baseline and treatment equilibria and (b) matching the houses with the information on which households occupy them in these two situations. This section describes our findings for the case of matching the households before and after the simulated treatments for each of our three experiments.

Our preference specification assures a closed form expression for the willingness to pay measure associated with each policy experiment. In each case we assume the willingness to pay is for the change in the environmental service we seek to measure with the treatment effect. We adopt a general equilibrium perspective and view each change as a discrete transformation in the conditions. Thus, for the landscape case the tradeoff is for an irrigated (or wet) set of conditions and the general equilibrium dimension arises because of re-sorting, a new set of house prices, and an income effect to homeowners that arise due to the price changes in their initially assigned houses.

The two experiments involving sites with hazardous substances use a continuous index to represent the dis-amenity effect. They involve the inverse of the distance to the closest site. For the first experiment we assume policy leads to the cleanup of the Motorola site. This change will alter the closest site for those houses closest to the Motorola site. As noted earlier, after the simulated clean-up they will be closest to the Indian Bend site. Inverse distance enters household preferences so this policy implies a different change for each house. Moreover, with sorting, households can relocate. Some of the households closest to Indian Bend following the cleanup of Motorola will be the same ones that were closest to Indian Bend before the simulated policy (i.e. in the baseline case). Others will not have moved houses but directly experience the policy. Still others will re-locate into houses that are now closer to Indian Bend. Each group of households will experience different types of changes along with price changes implied by the new equilibrium.

The second hazardous waste policy involving the clean-up of multiple landfills has two types of changes. For some houses, closest to the cleaned up component of the 19th Avenue Landfill, the disamenity effect is removed by the treatment. In these cases the inverse distance measure is zero after the policy for these houses[12]. For houses closest to the Motorola site, the change is the same as with the first hazardous waste experiment. Cleanup implies these houses will now be closest to the Indian Bend site. Both site cleanups contribute in the same way to preferences – a change in inverse distance. In addition to accounting for the initial

[12] Our preference specification does not treat this case differently. Some studies have argued that the elimination of a stigma or dis-amenity has a larger effect than simply reducing the disamenity effect to zero. To consider this type of effect would require a different preference specification.

cleanups, households can relocate thereby adding to the complexity associated with the general equilibrium perspective which takes explicit account of re-sorting in response to large changes.

The simultaneous changes to the two sites are considered as one experiment because the analyst could decide to represent these types of changes differently. We discuss the results from each experiment separately using the individual matched sample and then offer a comparative summary at the end of this section.

A. Landscape Findings

Table 3 presents the mean, median, and thresholds for the 5 and 95 percentiles for the true willingness to pay and the same statistics for each model's estimates of the willingness to pay. These findings are reported for the overall sample and for a sub-sample confined to households living in homes with irrigation after the change. Both linear and semi-log specifications are used for cross sectional hedonic, difference, and the instrumental variable estimates. The first column in each of our results tables identifies the true value for the willingness to pay and then the model specification used for each of the models used. The second column describes the sample used to construct the summary statistics. The third provides the specification used for the empirical model and the fourth identifies the instruments used in the IV models. The remaining columns provide the summary statistics. For the semi-log estimates with qualitative independent variables we use the Halvorsen-Palmquist [1980] interpretation for the estimated coefficient for the dummy variable along with Goldberger's [1968] suggested correction for bias with logarithmic transformations.

The implications are direct and robust in that they hold for a wide range of models and sub-samples. The cross sectional hedonic models, whether on the baseline sample or using the equilibrium prices after the change, offer the best estimates of the mean willingness to pay. The errors with hedonic models range from 1.8% to 5.7% using the full sample means. All of the hedonic models with instruments intended to mimic the quasi experimental logic in this case are inferior to these simple cross sectional strategies and yield estimates for the mean WTP that exceed the true mean by twenty-nine percent or more for the full sample. There is some relative improvement in the difference and instrumental variable models for the sample in the irrigated houses in the treatment equilibrium. This improvement seems to arise because the true value for mean willingness to pay is larger for this group and the estimates remain about the same. The simple hedonic models remain superior but in this case understate the true value by about four percent with the treatment sample offering the best estimate.

The models considered include specifications using price differences in linear and semi-log form, as well as instrumental variable models using both specifications with several instrument definitions. We refer to one formulation as involving a "general" instrument because it assumes knowledge of whether a house changed its state between baseline and treated samples. It does not identify whether the change was from a desert to an irrigated sub-division or the reverse. An unrestricted set of instruments assumes that these two changes can be distinguished. In general, models based on the unrestricted information are superior.

The true willingness to pay measures span a wide range and none of these reduced form models is able to track individual heterogeneity in the true values. At best, they offer imprecise measures of average willingness to pay. The bottom line conclusion from this experiment would be that a hedonic price model estimated under baseline conditions without adjustment for potential endogeneity provides a superior estimate for the overall mean willingness to pay. A hedonic applied to the treatment sample is superior in estimating willingness to pay for households selecting irrigated homes after the treatment. Nonetheless, the difference in before and after hedonic estimates is small and substantially less than the error in the quasi-experimental estimates.

B. Cleanup of One Hazardous Waste Site

Table 4 reports the same summary statistics for the “true” willingness to pay as well as several models accounting for the cleanup of the Motorola site. Our design implies 250 houses experience the cleanup. About forty percent of the households in these affected homes (85 of 250) move to new locations. Thus, re-sorting is an important element in understanding the results. Of those who move the majority (246 of the 250) select homes that are close to Indian Bend, the closest landfill after the simulated cleanup. This re-allocation of simulated households is important when we consider analysis of the results matched by house. In this case identifying the house affected by the policy does not assure the household in that house will be the same before and after the change. This difference will not have a large effect on the *marginal* willingness to pay but can be expected to affect the use of these results to approximate the general equilibrium willingness to pay for site cleanup.

For this experiment the quasi-experimental methods, using the first difference specification that includes a fixed effect identifying whether the initial location is a house affected by the policy together with the changes in other housing attributes, offers the best estimates for mean willingness to pay. This method exploits an ideal instrument that assumes knowledge of whether each individual initially lived in a house affected by the policy. The linear specification has about a 4.5% overstatement. With the semi-log specification the error for the model's estimate of the mean general equilibrium WTP increases to 8.6%. The conventional cross section hedonic and first difference models (ignoring the instrument) understate the willingness to pay and are inferior estimates with errors exceeding sixteen percent.

One might argue that comparisons for the full sample are less relevant than for the households in the homes affected by the policy after the simulated cleanup. This argument would be consistent with a focus on local average treatment effects. Their choices reveal greater concern about proximity to sites with hazardous substances. The second set of results in Table 4 present the findings for this subsample. As expected, the "true" general equilibrium willingness to pay is larger but the relative ranking of methods does not change. As with the full sample the relative superiority of the quasi-experimental method relies on both the instrument and the fixed effect identifying whether the house was one impacted by the cleanup policy.

C. Cleanup of Two Hazardous Waste Sites

The second hazardous waste example introduces a number of complexities for the evaluation problem. The policy is assumed to impact two sites in different ways. As a result, the GE sorting responses are more complex. In these cases half of the houses near the 19th Avenue Landfill are no longer impacted by the negative effects of a site with hazardous waste. Those near the Motorola site experience the same type of change as in our one site example. That is, with the Motorola site cleaned they are now close to the Indian Bend site. Thus, households can sort into houses that are not close to a hazardous waste site or they can alter their proximity to a site. In the one-site scenario every house had some disamenity consequences from proximity to a waste site. The true GE benefit measure is capturing the change from the baseline distance to the 19th Avenue Landfill to a situation with no effect on household preferences

Table 5 summarizes the results. Consider first the implications of adding sites with the prospect of avoiding the disamenities due to sites with hazardous waste for the market equilibrium. Over 200 of the original households in these cleaned 19th Avenue houses remained when they faced the opportunities posed with altered distances, housing attributes, and prices. Thirty-eight moved accepting proximity to a site in exchange for other desirable attributes – housing attributes or prices that were available as a result of the new equilibrium alternatives. In this scenario we have two GE willingness to pay measures: a distance change and the elimination of the effect of a site (a distance change to zero). These are compared for the full sample and the two sub-samples corresponding to households in houses initially experiencing either the change in the Motorola cleanup or those experiencing the 19th Avenue cleanup.

An important issue in interpreting of these findings is the selection of the sample used in evaluating performance. Using the standard of households experiencing the policy, the quasi-experimental models with the difference specification, whether linear or semi-log, using one instrument (identifying houses that experience any cleanup of a site) provide superior estimates. The case of two instruments has approximately comparable performance in a few cases with semi-log models. In these situations the estimates tend to understate the true GE willingness to pay. By contrast, the hedonic model using transactions prices before the policies overestimate the GE willingness to pay for the 19th Avenue Landfill sample with errors of 38 percent and 32 percent for linear and semi-log specifications, respectively. For the sample associated with the Motorola site cleanup the estimates understate the effect and are larger in absolute magnitude. The other cross-sectional hedonic models yield larger errors. If we were to use the full sample estimates of WTP and the sub-sample means for the true values as the standard, the conclusions would be the same.

Overall, it appears that the advantages of "ideal" instruments in quasi-experimental methods are realized with continuous measures of the environmental amenity. When the change in site specific conditions is discrete, as with our example of xeric versus wet landscaping, the cross sectional hedonic (before or after the simulated policy change) provided superior estimates of the general equilibrium willingness to pay. In contrast, for both of the examples involving hazardous waste sites when cleanup led to changes in the continuous indicator for the disamenity the quasi-experimental models using the price differences and accounting for changes in other housing attributes provided superior estimates of the GE willingness to pay.

V. Results for Matched Houses

Most quasi-experimental analyses measuring the benefits of environmental policies are closer to a match of control and treated samples based on the houses in each category. As a result, this section considers the same three experiments but re-analyzes the results based on matching the houses in the baseline sample to those in the treated sample. As noted earlier, different households can be in each house for the two samples. This difference must be considered in determining which household's willingness to pay is used in evaluating the performance of each model. We propose to use the household selecting the house after the change. As a result, the mean values reported for the "true" willingness to pay are associated with the households occupying the houses after the exogenous treatment.

A. Landscape Findings

Table 6 reports the findings for the matched house landscape models. Using either the full sample or the sample of houses with irrigated landscape, the quasi-experimental estimates with the discrete indicator of a change to an irrigated status as an instrument are closer to the mean willingness to pay of households who ultimately select the houses in the treated sample. For the full sample the semi-log quasi experimental specification provides the best estimates while the linear quasi experimental specification is preferred in the subsample of wet houses. Under the linear specification for the full sample and the semi-log specification for the wet sample, the simple cross section provides superior estimates with errors of less than 6% compared to over 10% with the quasi experimental estimates. Overall these findings indicate a simple cross section hedonic based on equilibrium prices after the change provides estimates comparable to the quasi experimental IV models and doesn't require the consistent identification of the "correct" instrument.

B. Cleanup of One Hazardous Waste Site

Table 7 provides estimates for the cleanup of the Motorola site using the house matched sample so each price change is for the same house. In this case only the proximity to the hazardous waste site changes between baseline and treated samples. We follow the same convention and use the willingness to pay of the household in each house after the cleanup. The overall sample and a subset of the houses near the Motorola site are each considered. The results are consistent regardless of the sample with the conventional first difference model (for both specifications linear and semi-log) providing the best estimate of mean GE willingness to pay using the final household as the standard. While the differences between the quasi-experimental and this simple approach can be small – a 9% error with the conventional first difference versus 12% with the quasi-experimental IV estimator for the full sample, the ranking is consistent for all samples and model specifications. As a result, in this case a quasi experimental estimation strategy offers little advantage over conventional first difference models that do not attempt to isolate an instrument for the source of the "treated" houses experiencing cleanup outcomes.

C. Cleanup of Two Hazardous Waste Sites

Table 8 provides the estimates for the analysis matching houses before and after the simulated policy cleaning up two sites. Here only the sub-samples of houses affected by each change provide the basis for evaluation. For the Motorola site the quasi-experimental difference models provide mean GE willingness to pay estimates closest to the "true" mean for willingness to pay for households who select the cleaned house after the change, with errors of 9% for the linear and 12% for the semi-log. For the case of eliminating the disamenity effects of the 19th Avenue Landfill site, the quasi-experimental difference is again the best with the conventional cross-sectional difference and hedonic models using prices after the change resulting in much larger errors.

Considering all models for both of the hazardous waste cleanup scenarios the results for the matched houses suggest the quasi-experimental methods are generally superior to simple cross-sectional hedonic or difference models. This conclusion relies on selecting the ideal instrument and using the willingness to pay of households who select the houses after the treatment.

VI. Implications

Under ideal conditions (where we know how to correctly measure environmental amenities) our results imply conventional hedonic practice can yield estimates closer to the true mean general equilibrium willingness to pay. However in other cases the quasi experimental estimates for the overall mean as well as for special sub-groups would be superior. Quasi experimental estimates tend to be superior when the policy changes are complex. The performance gain can be substantial. This conclusion is relevant to the environmental applications where hedonic property value studies have been conducted using <u>the same</u> house before and after a policy change and the proxy measure for the amenity service is continuous. This is true even though different households will occupy the house before and after the change providing the source of a treatment and the associated identifying variation. The estimated price differential exploiting the exogenous treatment as an instrument provides a measure of the <u>mean</u> general equilibrium willingness to pay for the change in the amenity resulting from that treatment for those households occupying the treated houses following the change. The performance gain of this estimate over simple hedonic approaches varies with application but can be large.

For a discrete condition that shifts preferences – wet versus xeric – the advantage does not appear present. Indeed, in this case we have more than simply a direct example of how our proposed method would work. That is, the simulation of market equilibria under different circumstances confirms that our cross sectional hedonic price function used in calibration would have been the best choice to estimate the tradeoffs households would make for wet landscapes in their sub-divisions. As a result, we can conclude that the hedonic estimates of the willingness to pay for green landscape are likely to provide a reliable gauge of the value of "wet" landscape. We found that the analysis can rely solely on the cross sectional variation across subdivision in xeric versus wet conditions.

The performance of conventional versus quasi experimental methods is also not as clear for the case of cleanup of a single hazardous waste site (with matched house sample). The preferred estimator can in this case depend on the specification selected for the hedonic price equation.

When we consider samples tracking households the challenge is greater because households can alter the mix of attributes and the complementarity or substitution between these attributes. In addition, preferences can influence the ability to successfully control for the interacting effects that result from the exogenous treatment. At a conceptual level these factors were the reasons for Starrett's conditions guaranteeing capitalization measured benefits of external changes to a community (see note #5) as well as for Schotchmer's general point concerning the difficulty in measuring general equilibrium willingness to pay with a hedonic price equation. It is also what we observed with the matched individual results. Quasi-experimental methods were not always best in terms of their estimates for the mean GE willingness to pay.

Our focus was primarily on the mean estimates. The true values for WTP displayed considerable heterogeneity and all of the approaches failed to capture these differences. All the methods are reduced forms so this result should not be surprising. Quasi-experimental methods acknowledge the difficulties in estimating conditional means that attempt to capture the role of observable sources of this heterogeneity.

Nonetheless these findings need to be interpreted carefully in relation to the studies with actual data that motivated our inquiry. They assume ideal instruments are known. We did not consider other cases because there are no clear guides to distinguish good from "poor" instruments. Thus, with strict adherence to the quasi-experimental admonition that analysts must be prepared to devote considerable effort (the proverbial equivalent of the "shoe leather") in identifying instruments, our results confirm some of the current enthusiasm for this strategy where the equilibrium prices are influenced by a complex sorting response to a policy change.

There are some important caveats to these general conclusions. First, as we noted earlier, our standard was a mean effect not differential values based on observable attributes of households. Second, it assumed the analyst is working with the *actual prices* for each house not summary statistics for sets of houses, such as the means or medians used in Chay and Greenstone and in Greenstone and

Gallagher. Third, our results rely, as we noted above, on knowledge of the ideal instrument and correct specifications (i.e. no omitted variables) for the hedonic functions, including the measure for the non-market amenity. Finally, it assumed the spatial linkages connecting houses and households to environmental amenities are known. With these caveats our proposal offers a new method consistent with the logic developed in the Chetty and the Imbens and Wooldridge frameworks to evaluate when structural and quasi experimental methods are likely to be best. It is directly relevant to a large number of environmental applications when the empirical models must use price capitalization to recover the tradeoffs that motivate general equilibrium adjustment of households in response to changes in spatially delineated amenities.

An important final question is how do our specific results relate to other situations with different types of spatial interconnections? In effect, what general implications can be extracted from these detailed case studies designed to match real world conditions in a specific metropolitan area? The best answer we can offer is that we have demonstrated this is the *wrong* question to ask. One of the points of this paper is that numerical generalizations are unlikely to be effective in addressing the issues of when quasi experimental methods provide advantages over conventional hedonic price functions in measuring willingness to pay thru housing price capitalization. Instead, we need a framework that can be undertaken for each application. We have demonstrated this strategy is feasible. Our computational rubric offers a tractable approach to conducting an assessment with no more information than is routinely available with quasi experiments involving housing price capitalization. By following our proposed four step method each new quasi-experimental application can be used to calibrate preferences within an application-specific assignment study and use this template to establish the likely error bounds. Formal statistical decision rules could also be derived by repeating the process in a Monte Carlo framework[13].

[13] Our approach to this point has focused on using the estimated covariance matrix for the calibrated parameters as the basis for inducing preference heterogeneity rather than a reflection of estimation uncertainty. Any evaluation of the performance of tests or our computational rubric would require a full specification of the sources of error in the equilibrium price vectors used to gauge the effects of changes in spatially delineated environmental amenities.

References

Beatty, Timothy K.M. and Jay P.Shimshack, 2008, "School Buses, Diesel Emissions, and Respiratory Health", unpublished paper, Tulane University, September.

Brown, J.M and H. Rosen 1982, "On the Estimation of Structural Hedonic Price Models," Econonmetrica, Vol. 50 (): 765-768.

Chay, Kenneth Y. and Michael Greenstone, 2005, "Does Air Quality Matter: Evidence from the Housing Market" Journal of Political Economy, Vol. 113 (2): 376-424.

Chetty, Raj, 2008, "Sufficient Statistics for Welfare Analyses: A Bridge Between Structural and Reduced-Form Methods" NBER Working Paper 14399, October.

Cropper, M.L., Deck, N. Kishor and K.E. McConnell, 1993, "Valuing Product Attributes Using Single Market Data: A Comparison of Hedonic and Discrete Choice Approaches', Review of Economics and Statistics, Vol. 75, ( ): 225-232.

Cropper, M.L., L.B. Deck and K.E. McConnell, 1988, "On the Choice of Functional Form for Hedonic Price Equations', Review of Economics and Statistics, Vol. 70: (November): 668-75.

Duflo, E., M., Greenstone and R. Hanna, 2007, "The Health and Economic Impacts of Indoor Air Pollution: Evidence from a Randomized Study in Orissa, India" unpublished working paper, Department of Economics, MIT.

Epple, Dennis, 1987, "Hedonic Prices and Implicit markets: Estimating demand and Supply Functions for Differentiated Products" Journal of Political Economy, Vol.95,(February): 59-80.

Epple, Dennis, Radu Filimon, and Thomas Romer, 1993, "Existence of Voting and Housing Equilibrium I a System of Communities with Property Taxes", Regional Science and Urban Economics, Vol. 31,( ):585-610.

Ferraro, P, J., C. McIntosh, and M. Ospina, 2007, "The Effectiveness of the U.S. Endangered Species Act: An econometric Analysis Using Matching Methods" Journal of Environmental Economics and Management Vol54(3):245-261.

Greenstone, Michael and Justin Gallagher, 2008, "Does Hazardous Waste Matter? Evidence from Housing Market and Superfund Program" Quarterly Journal of Economics, Vol. 123 (August): 951-1004.

Greenstone, Michael and Ted Gayer, 2009, "Quasi-Experimental and Experimental Approaches to Environmental Economics" Journal of Environmental Economics and Management, Vol. 57,(January): 21-44.

Imbens, Guido M. ,2009, "Better LATE than Nothing: Some Comments on Deaton(2009) and Heckman and Urzua(2009)" NBER working paper No. 14896, April.

Imbens, Guido M. and Jeffrey M. Wooldridge, 2009, "Recent developments in the Econometrics of Program Evaluation" Journal of Economic Literature Vol. 47, (March): ,also NBER working paper # 14251, August, 2008.

Kiel, Katherine A. and Michael Williams, 2007, "The Impact of Superfund Sites on Local Property Values: Are All Sites the Same?" Journal of Urban Economics, Vol 61, (1):170-192.

Kuminoff, Nicholai V., Christopher F. Parmeter, and Jaren C. Pope, 2009, "Avertin Hidden Danger: Developing Hedonic Specifications that Mitigate Omitted Variable Bias " unpublished paper Virginia Tech March.

Kuminoff, Nicholai V. and Jaren C. Pope, 2008, "Capitalizational and Welfare Measurement in the Quasi Experimental Hedonic Model" unpublished paper Virginia Tech, December.

Pope, Jaren C., 2008, "Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model," Land Economics, Vol. 84 (4): 551-572.

Rubin, D. B. 1973, "The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies" Biometrika, Vol 29. 185-203

Rubin,D.B.,1974, "Estimating Casual Effects of Treatments in Randomized and Non-Randomized Studies" Journal of Educational Statistics , Vol 66,688-701.

Rubin, D.B. 2006 Matched Sampling for Casual Effects (Cambridge,U.K. Cambridge University Press)

Scotchmer, Suzanne, 1985, "Hedonic Prices and Cost/Benefit Analysis" Journal of Economic Theory, Vol.37,(1):55-75.

Scotchmer, Suzanne, 1986, "The Long-run and Short-run Benefits of Environmental Improvement" Journal of Public Economics, Vol.30, (June):61-83

Smith, V. Kerry, Eric M. Moore, and Sharon L. Harlan, 2009, "Will Households Mitigate Urban Heat Islands?" unpublished working paper under revision,CEESP Arizona State University, April.

Starrett, David A., 1981, "Land Value Capitalization in Local Public Finance" Journal of Political Economy, Vol.89, (April): 306-327.

Stefanov, William L., Michael S. Ramsey, and Phillip R. Christensen, 2001,"Monitoring Urban land Cover Change: An Expert System Approach to Land Cover Classification of semiarid to Arid Urban Centers" Remote Sensing of Environment, Vol. 77, 173-185.

Wheaton, William C., 1974, "Linear Programming and Locational Equilibrium: The Herbert Stevens Model Revisited" Journal of Urban Economics, Vol. 1 (3): 278-287.

Table 1: Landscape calibration first-stage hedonic

Variable	Hedonic Estimates: Estimate	Std Err	t-stat	Summary Statistics: Mean	Std Dev	Min	Max
Constant / Price[a]	8.010	0.027	297.330	16,048	10,583	1,870	184,250
Square feet (100s)	0.040	0.000	327.600	17.81	6.42	6.00	45.00
Acres	0.149	0.002	69.120	0.22	0.22	0.05	19.96
Stories	-0.182	0.001	-127.530	1.14	0.34	1.00	4.00
Bathrooms	0.140	0.001	128.420	2.43	0.81	0.50	6.00
Age	-0.003	0.000	-80.080	17.71	15.92	1.00	85.00
Pool	0.112	0.001	110.300	0.26	0.44	0.00	1.00
Garage	0.125	0.002	61.520	0.95	0.22	0.00	1.00
Wet	0.070	0.001	61.370	0.18	0.38	0.00	1.00
July min temp	-0.006	0.000	-17.970	78.75	2.01	67.16	86.30
CBD distance	0.002	0.000	21.750	14.13	5.94	0.51	42.84
Year Dummy Variables							
1981	-0.277	0.013	-20.770				
1982	-0.402	0.015	-26.360				
1983	0.102	0.009	11.790				
1984	0.149	0.008	17.720				
1985	0.204	0.008	24.930				
1986	0.238	0.008	29.730				
1987	0.228	0.008	28.490				
1988	0.226	0.008	28.250				
1989	0.236	0.008	29.310				
1990	0.222	0.008	27.810				
1991	0.214	0.008	27.000				
1992	0.229	0.008	29.180				
1993	0.293	0.008	37.440				
1994	0.353	0.008	45.680				
1995	0.432	0.008	55.980				
1996	0.501	0.008	65.470				
1997	0.570	0.008	74.700				
1998	0.624	0.008	82.440				
1999	0.695	0.008	92.070				
2000	0.767	0.008	101.670				
2001	0.820	0.008	108.790				
2002	0.870	0.008	115.550				
2003	0.963	0.008	127.060				
2004	1.049	0.008	137.790				
Statistics							
# Observations	398,200						
Adjusted R^2	0.7751						

[b]Price is based on 11% of sale value and corresponds to summary statitics

Table 2: Hazardous waste calibration first-stage hedonic[a]

	Hedonic Estimates			Summary Statistics			
Variable	Estimate	Std Err	t-stat	Mean	Std Dev	Min	Max
Constant / Price[b]	8.803	0.061	145.350	14,558	9,063	1,870	151,085
Square feet (100s)	0.032	0.000	228.740	18.46	6.49	6.00	45.00
Acres	0.098	0.002	47.650	0.24	0.30	0.05	20.00
Stories	-0.084	0.002	-52.420	1.14	0.34	1.00	4.00
Bathrooms	0.054	0.001	43.570	2.53	0.81	0.50	6.00
Age	-0.006	0.000	-70.340	14.46	14.60	1.00	80.00
Pool	0.066	0.001	59.710	0.29	0.45	0.00	1.00
Garage	0.053	0.002	21.890	0.96	0.20	0.00	1.00
NPL inverse distance	-0.021	0.006	-3.220	0.15	0.18	0.02	14.04
CBD distance	0.001	0.001	0.550	14.72	5.98	0.51	72.17
Year Dummy Variables							
1991	-0.005	0.003	-1.880				
1992	0.026	0.003	9.690				
1993	0.098	0.003	37.470				
1994	0.167	0.003	65.920				
1995	0.240	0.003	95.580				
1996	0.299	0.002	121.490				
1997	0.355	0.002	145.610				
1998	0.415	0.002	173.430				
1999	0.487	0.002	204.360				
Statistics							
# Observations	242,827						
Adjusted R^2	0.8396						

[a]City fixed effects not reported

[b]Price is based on 11% of sale value and corresponds to summary statitics

Table 3: Landscpae WTP using matched individuals

Model	Sample	Specification	Instruments[a]	Estimated WTP Mean	Median	5%	95%	% Error[b]
True WTP	Full			467.4	474.9	191.8	720.1	0.0%
CS Before	Full	Linear		476.0	476.0	476.0	476.0	1.8%
CS After	Full	Linear		494.2	494.2	494.2	494.2	5.7%
Difference	Full	Linear		358.6	358.6	358.6	358.6	-23.3%
IV - CS After	Full	Linear	Wet	641.9	641.9	641.9	641.9	37.3%
IV - Difference	Full	Linear	General	611.1	611.1	611.1	611.1	30.7%
IV - Difference	Full	Linear	Wet	603.8	603.8	603.8	603.8	29.2%
IV - Difference	Full	Linear	Wet, Dry	602.5	602.5	602.5	602.5	28.9%
CS Before	Full	Log-Linear		479.9	476.5	465.3	504.0	2.7%
CS After	Full	Log-Linear		492.8	490.1	478.3	516.1	5.4%
Difference	Full	Log-Linear		362.0	360.0	351.3	379.0	-22.6%
IV - CS After	Full	Log-Linear	Wet	639.5	636.0	620.7	669.7	36.8%
IV - Difference	Full	Log-Linear	General	622.6	619.1	604.3	651.9	33.2%
IV - Difference	Full	Log-Linear	Wet	615.1	611.7	597.0	644.1	31.6%
IV - Difference	Full	Log-Linear	Wet, Dry	613.5	610.1	595.5	642.4	31.3%
True WTP	Wet Only			518.3	520.4	225.0	750.2	0.0%
CS Before	Wet Only	Linear		476.0	476.0	476.0	476.0	-8.2%
CS After	Wet Only	Linear		494.2	494.2	494.2	494.2	-4.6%
Difference	Wet Only	Linear		358.6	358.6	358.6	358.6	-30.8%
IV - CS After	Wet Only	Linear	Wet	641.9	641.9	641.9	641.9	23.9%
IV - Difference	Wet Only	Linear	General	611.1	611.1	611.1	611.1	17.9%
IV - Difference	Wet Only	Linear	Wet	603.8	603.8	603.8	603.8	16.5%
IV - Difference	Wet Only	Linear	Wet, Dry	602.5	602.5	602.5	602.5	16.2%
CS Before	Wet Only	Log-Linear		483.8	481.2	466.1	507.7	-6.6%
CS After	Wet Only	Log-Linear		496.6	493.7	479.2	519.2	-4.2%
Difference	Wet Only	Log-Linear		364.7	362.6	352.0	381.4	-29.6%
IV - CS After	Wet Only	Log-Linear	Wet	644.4	640.7	621.9	673.8	24.3%
IV - Difference	Wet Only	Log-Linear	General	627.3	623.7	605.4	655.9	21.0%
IV - Difference	Wet Only	Log-Linear	Wet	619.8	616.2	598.1	648.0	19.6%
IV - Difference	Wet Only	Log-Linear	Wet, Dry	618.2	614.6	596.6	646.3	19.3%

[a]Wet = Dry to wet switch; Dry = wet to dry switch; General = Any switch, regardless of direction

[b]Error computed based on mean willingness to pay

Table 4: Single hazardous waste model performance using matched individuals

Model	Sample	Specification	Instruments[a]	Estimated WTP Mean	Median	5%	95%	% Error[b]
True WTP	Full	Linear		1.9	0.0	-7.9	15.6	0.0%
CS Before	Full	Linear		0.7	0.0	-5.7	7.5	-62.4%
CS After	Full	Linear		1.3	0.0	-10.9	14.3	-28.5%
Difference	Full	Linear		1.4	0.0	-11.1	14.7	-26.8%
IV - CS After	Full	Linear		20.3	0.0	-163.7	215.4	975.1%
IV - Difference	Full	Linear	Switch	2.0	0.0	-15.9	20.9	4.5%
IV - Difference	Full	Linear	Mot	1.3	0.0	-10.6	13.9	-30.5%
CS Before	Full	Log-Linear		0.7	0.0	-5.5	7.5	-62.5%
CS After	Full	Log-Linear		1.4	0.0	-10.6	14.4	-27.6%
Difference	Full	Log-Linear		1.4	0.0	-10.8	14.6	-26.5%
IV - CS After	Full	Log-Linear		20.4	0.0	-158.7	216.1	984.6%
IV - Difference	Full	Log-Linear	Switch	2.0	0.0	-15.9	21.6	8.6%
IV - Difference	Full	Log-Linear	Mot	1.3	0.0	-10.4	14.1	-29.1%
True WTP	Motorola	Linear		6.1	4.3	-7.2	21.7	0.0%
CS Before	Motorola	Linear		2.1	1.8	-3.0	7.8	-64.6%
CS After	Motorola	Linear		4.1	3.4	-5.7	14.9	-32.8%
Difference	Motorola	Linear		4.2	3.5	-5.9	15.2	-31.2%
IV - CS After	Motorola	Linear		61.4	51.7	-86.2	224.0	911.1%
IV - Difference	Motorola	Linear	Switch	6.0	5.0	-8.4	21.8	-1.7%
IV - Difference	Motorola	Linear	Mot	4.0	3.3	-5.6	14.5	-34.7%
CS Before	Motorola	Log-Linear		2.2	1.8	-3.0	8.0	-64.5%
CS After	Motorola	Log-Linear		4.2	3.4	-5.8	15.4	-31.5%
Difference	Motorola	Log-Linear		4.2	3.4	-5.9	15.7	-30.4%
IV - CS After	Motorola	Log-Linear		62.3	50.7	-86.6	230.9	926.4%
IV - Difference	Motorola	Log-Linear	Switch	6.2	5.1	-8.7	23.1	2.8%
IV - Difference	Motorola	Log-Linear	Mot	4.1	3.3	-5.7	15.1	-32.9%

[a]Switch = Indicates individual's moving into a house initially Motorola; Mot = Indicates households initially living nearest Motorola

Table 5: Multiple hazardous waste model performance using matched individuals

				Estimated WTP - Distance Reduction					Estimated WTP - Site Removed				
Model	Sample	Specification	Instruments[a]	Mean	Median	5%	95%	% Error[b]	Mean	Median	5%	95%	% Error[b]
True WTP	Full	Linear		9.8	3.7	-8.9	43.0	0.0%	41.3	41.8	16.3	66.6	0.0%
CS Before	Full	Linear		3.4	1.8	-6.4	14.7	-65.5%	38.8	37.8	19.2	58.2	-6.1%
CS After	Full	Linear		6.8	3.6	-12.9	29.4	-30.8%	23.5	21.4	13.9	42.3	-43.2%
Difference	Full	Linear		1.7	0.9	-3.2	7.2	-83.0%	13.7	12.5	8.1	24.8	-66.7%
IV - CS After	Full	Linear	Land, Mot	102.6	54.9	-195.2	445.3	947.4%	355.3	323.3	210.5	640.6	760.1%
IV - Difference	Full	Linear	Switch	10.3	5.5	-19.6	44.7	5.2%	35.7	32.5	21.1	64.3	-13.6%
IV - Difference	Full	Linear	Land, Mot	8.6	4.6	-16.3	37.1	-12.7%	29.6	27.0	17.5	53.4	-28.3%
CS Before	Full	Log-Linear		3.3	1.8	-6.1	14.0	-66.7%	37.7	36.8	18.4	56.0	-8.8%
CS After	Full	Log-Linear		6.9	3.8	-12.9	29.7	-29.1%	24.2	22.3	14.2	43.0	-41.5%
Difference	Full	Log-Linear		1.5	0.8	-2.8	6.4	-84.7%	13.9	12.8	8.1	24.7	-66.5%
IV - CS After	Full	Log-Linear	Land, Mot	102.0	55.9	-189.1	436.5	941.3%	354.8	327.1	208.2	631.8	758.9%
IV - Difference	Full	Log-Linear	Switch	10.7	5.9	-19.8	45.7	9.1%	37.2	34.3	21.8	66.2	-10.0%
IV - Difference	Full	Log-Linear	Land, Mot	8.8	4.8	-16.3	37.7	-10.1%	30.7	28.3	18.0	54.6	-25.8%
True WTP	Motorola	Linear		7.4	5.0	-7.7	25.4	0.0%					
CS Before	Motorola	Linear		2.4	1.8	-3.6	8.8	-67.2%					
CS After	Motorola	Linear		4.9	3.6	-7.3	17.7	-34.3%					
Difference	Motorola	Linear		1.2	0.9	-1.8	4.3	-83.9%					
IV - CS After	Motorola	Linear	Land, Mot	73.5	54.9	-109.8	268.4	893.7%					
IV - Difference	Motorola	Linear	Switched	7.4	5.5	-11.0	27.0	-0.2%					
IV - Difference	Motorola	Linear	Land, Mot	6.1	4.6	-9.2	22.4	-17.1%					
CS Before	Motorola	Log-Linear		2.4	1.8	-3.6	9.0	-67.3%					
CS After	Motorola	Log-Linear		5.1	3.8	-7.6	19.1	-30.4%					
Difference	Motorola	Log-Linear		1.1	0.8	-1.6	4.1	-85.0%					
IV - CS After	Motorola	Log-Linear	Land, Mot	75.5	56.3	-111.4	279.9	921.7%					
IV - Difference	Motorola	Log-Linear	Switched	7.9	5.9	-11.7	29.3	7.1%					
IV - Difference	Motorola	Log-Linear	Land, Mot	6.5	4.9	-9.6	24.2	-11.7%					
True WTP	Landfill	Linear							31.1	29.7	12.1	50.8	0.0%
CS Before	Landfill	Linear							42.8	46.5	19.1	59.3	37.7%
CS After	Landfill	Linear							21.7	17.3	13.7	42.5	-30.2%
Difference	Landfill	Linear							12.7	10.2	8.0	24.9	-59.1%
IV - CS After	Landfill	Linear	Land, Mot						328.4	262.3	207.4	643.6	956.9%
IV - Difference	Landfill	Linear	Switched						33.0	26.3	20.8	64.6	6.1%
IV - Difference	Landfill	Linear	Land, Mot						27.4	21.9	17.3	53.7	-11.9%
CS Before	Landfill	Log-Linear							41.0	44.9	18.1	56.7	32.1%
CS After	Landfill	Log-Linear							22.0	17.9	13.9	43.0	-29.1%
Difference	Landfill	Log-Linear							12.6	10.3	7.9	24.7	-59.4%
IV - CS After	Landfill	Log-Linear	Land, Mot						323.4	263.6	203.6	631.8	940.7%
IV - Difference	Landfill	Log-Linear	Switched						33.9	27.6	21.3	66.2	9.0%
IV - Difference	Landfill	Log-Linear	Land, Mot						27.9	22.8	17.6	54.6	-10.1%

Table 6: Landscpae WTP using matched houses

Model	Sample	Specification	Instruments[a]	Estimated WTP Mean	Median	5%	95%	% Error[b]
True WTP	Full			467.4	474.9	191.8	720.1	0.0%
CS - After	Full	Linear		494.2	494.2	494.2	494.2	5.7%
Wet to Dry - Difference	Full	Linear		618.5	618.5	618.5	618.5	32.3%
Any Wet - Difference	Full	Linear		36.0	36.0	36.0	36.0	-92.3%
IV - CS After	Full	Linear	Wet	523.0	523.0	523.0	523.0	11.9%
CS - After	Full	Log-Linear		488.5	485.8	474.2	511.6	4.5%
Wet to Dry - Difference	Full	Log-Linear		624.4	620.9	606.1	653.9	33.6%
Any Wet - Difference	Full	Log-Linear		30.9	30.8	30.0	32.4	-93.4%
IV - CS After	Full	Log-Linear	Wet	460.9	458.3	447.3	482.6	-1.4%
True WTP	Wet Only			518.3	520.4	225.0	750.2	0.0%
CS - After	Wet Only	Linear		494.2	494.2	494.2	494.2	-4.6%
Wet to Dry - Difference	Wet Only	Linear		618.5	618.5	618.5	618.5	19.3%
Any Wet - Difference	Wet Only	Linear		36.0	36.0	36.0	36.0	-93.1%
IV - CS After	Wet Only	Linear	Wet	523.0	523.0	523.0	523.0	0.9%
CS - After	Wet Only	Log-Linear		492.2	489.4	475.1	514.7	-5.0%
Wet to Dry - Difference	Wet Only	Log-Linear		629.2	625.6	607.2	657.9	21.4%
Any Wet - Difference	Wet Only	Log-Linear		31.2	31.0	30.1	32.6	-94.0%
IV - CS After	Wet Only	Log-Linear	Wet	464.4	461.7	448.1	485.5	-10.4%

b Error based on median true WTP after

Table 7: Single hazardous waste model performance using matched houses

Model	Sample	Specification	Instrument	Estimated WTP Mean	Median	5%	95%	% Error[b]
True WTP	Full			1.9	0.0	-7.9	15.6	0.0%
CS - After Distance Change	Full	Linear		1.3	0.0	0.0	9.9	-28.5%
CS - After Distance Total	Full	Linear		1.2	0.0	0.0	8.6	-37.9%
Difference	Full	Linear		1.7	0.0	0.0	12.6	-9.3%
IV - Difference	Full	Linear	Mot	1.6	0.0	0.0	12.1	-12.9%
IV - CS After	Full	Linear	Mot	2.9	0.0	0.0	21.4	54.2%
CS - After Distance Change	Full	Log-Linear		1.4	0.0	0.0	10.1	-27.6%
CS - After Distance Total	Full	Log-Linear		1.1	0.0	0.0	8.2	-41.1%
Difference	Full	Log-Linear		1.7	0.0	0.0	12.7	-8.7%
IV - Difference	Full	Log-Linear	Mot	1.7	0.0	0.0	12.3	-11.5%
IV - CS After	Full	Log-Linear	Mot	2.7	0.0	0.0	19.8	41.8%
True WTP	Motorola			7.7	5.8	0.0	22.2	0.0%
CS - After Distance Change	Motorola	Linear		5.4	3.4	0.4	16.0	-29.9%
CS - After Distance Total	Motorola	Linear		4.7	3.0	0.3	13.9	-39.1%
Difference	Motorola	Linear		6.8	4.4	0.5	20.4	-11.1%
IV - Difference	Motorola	Linear	Mot	6.6	4.2	0.5	19.5	-14.6%
IV - CS After	Motorola	Linear	Mot	11.6	7.4	0.8	34.6	51.2%
CS - After Distance Change	Motorola	Log-Linear		5.5	3.5	0.4	17.1	-29.0%
CS - After Distance Total	Motorola	Log-Linear		4.4	2.8	0.3	13.9	-42.2%
Difference	Motorola	Log-Linear		6.9	4.4	0.5	21.6	-10.5%
IV - Difference	Motorola	Log-Linear	Mot	6.7	4.3	0.5	20.9	-13.2%
IV - CS After	Motorola	Log-Linear	Mot	10.7	6.8	0.8	33.5	39.1%

Table 8: Multiple hazardous waste model performance using matched houses

				Estimated WTP - Motorola					Estimated WTP - Landfill				
Model	Sample	Specification	Instruments[a]	Mean	Median	5%	95%	% Error[b]	Mean	Median	5%	95%	% Error[b]
True WTP	Full			9.8	3.7	-8.9	43.0	0.0%	41.3	41.8	16.3	66.6	0.0%
CS - After Not Site Specific	Full	Linear		6.8	0.0	0.0	27.2	-30.8%	6.8	0.0	0.0	27.2	-83.6%
CS - After Site Specific	Full	Linear		5.6	0.0	0.0	22.6	-42.5%	-4.2	0.0	-16.8	0.0	-110.1%
Difference	Full	Linear		7.1	0.0	0.0	28.5	-27.5%	7.6	0.0	0.0	30.4	-81.7%
IV - Difference	Full	Linear	Mot, Clean	10.7	0.0	0.0	42.9	9.1%	10.7	0.0	0.0	42.9	-74.1%
IV - CS After	Full	Linear	Mot, Clean	11.9	0.0	0.0	47.6	21.1%	11.9	0.0	0.0	47.6	-71.3%
CS - After Distance Change	Full	Log-Linear		6.9	0.0	0.0	27.4	-29.1%	6.9	0.0	0.0	27.4	-83.2%
CS - After Distance Total	Full	Log-Linear		-1.9	0.0	-7.7	0.0	-119.8%	-10.7	0.0	-42.2	0.0	-125.8%
Difference	Full	Log-Linear		6.9	0.0	0.0	27.1	-30.0%	7.6	0.0	0.0	30.2	-81.5%
IV - Difference	Full	Log-Linear	Mot, Clean	10.7	0.0	0.0	42.4	9.6%	10.7	0.0	0.0	42.4	-74.0%
IV - CS After	Full	Log-Linear	Mot, Clean	12.1	0.0	0.0	47.8	23.4%	12.1	0.0	0.0	47.8	-70.7%
True WTP	Motorola			8.3	6.5	0.0	23.0	0.0%					
CS - After Distance Change	Motorola	Linear		5.7	3.6	0.4	16.9	-31.2%					
CS - After Distance Total	Motorola	Linear		4.7	3.0	0.3	14.1	-42.9%					
Difference	Motorola	Linear		6.0	3.8	0.4	17.7	-27.9%					
IV - Difference	Motorola	Linear	Mot, Clean	9.0	5.7	0.6	26.7	8.5%					
IV - CS After	Motorola	Linear	Mot, Clean	9.9	6.3	0.7	29.6	20.4%					
CS - After Distance Change	Motorola	Log-Linear		6.0	3.8	0.4	18.9	-27.2%					
CS - After Distance Total	Motorola	Log-Linear		-1.7	-1.1	-5.3	-0.1	-120.3%					
Difference	Motorola	Log-Linear		5.9	3.8	0.4	18.6	-28.1%					
IV - Difference	Motorola	Log-Linear	Mot, Clean	9.3	5.9	0.7	29.2	12.4%					
IV - CS After	Motorola	Log-Linear	Mot, Clean	10.5	6.7	0.7	32.8	26.6%					
True WTP	Landfill								31.6	30.3	11.8	51.3	0.0%
CS - After Distance Change	Landfill	Linear							21.4	17.3	13.7	41.9	-32.2%
CS - After Distance Total	Landfill	Linear							-13.2	-10.7	-25.9	-8.5	-141.9%
Difference	Landfill	Linear							23.9	19.3	15.3	46.8	-24.3%
IV - Difference	Landfill	Linear	Mot, Clean						33.8	27.3	21.6	66.1	7.0%
IV - CS After	Landfill	Linear	Mot, Clean						37.5	30.3	24.0	73.3	18.7%
CS - After Distance Change	Landfill	Log-Linear							21.8	17.7	13.8	41.9	-31.1%
CS - After Distance Total	Landfill	Log-Linear							-33.5	-27.2	-64.4	-21.3	-205.9%
Difference	Landfill	Log-Linear							24.0	19.5	15.2	46.1	-24.2%
IV - Difference	Landfill	Log-Linear	Mot, Clean						33.6	27.4	21.4	64.8	6.5%
IV - CS After	Landfill	Log-Linear	Mot, Clean						37.9	30.8	24.1	72.9	20.0%

Figure 1: Greenstone and Gallagher Illustration of Price Capitalization

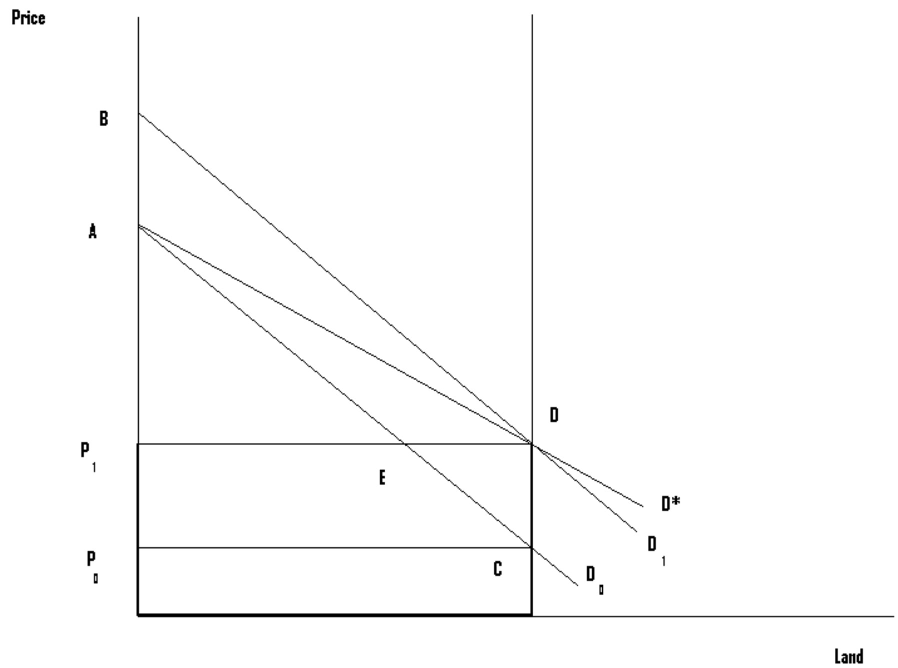
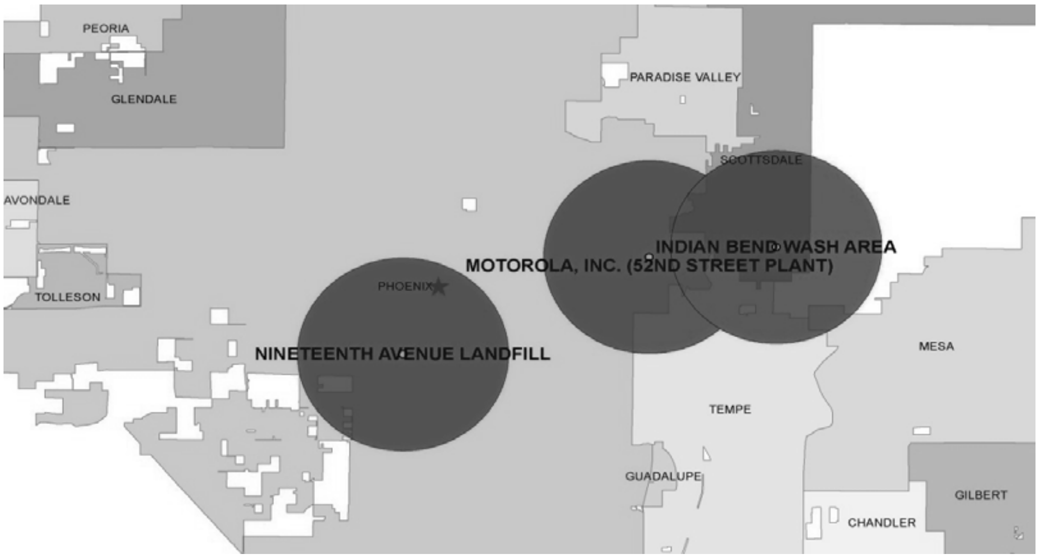


Figure 2: Hazardous Waste Sites



Appendix A. Estimating minimum temperature

Estimating a model similar to Brazel et al. [2007], who found that water intensive land cover is associated with cooler nighttime summer temperatures; we use actual temperature monitoring data along with land cover data to generate predictions of July minimum low temperatures for each block group in the study area. The source of land cover data is satellite interpreted images using the Stefanov et al [2001] classification system. This system analyzes differences in reflectivity to assign one of 12 land cover types to a 30x30 meter square. The land cover types include cultivated vegetation, cultivated grass, vegetation, fluvial and lacustrine sediments (canals), water, undisturbed, disturbed soil with agricultural water rights, compacted soil, disturbed (commercial/industrial), disturbed (asphalt and concrete), disturbed (mesic residential), and disturbed (xeric residential).

To predict temperatures, we estimate a model relating low temperatures to neighborhood land cover. We use daily observed low temperatures at 20 NOAA monitoring stations located across Phoenix for the years 2000 through 2008 and attach those temperature monitors to any Census 2000 block group centroid located within 1 mile of each monitor. In total, this results in 308 block groups assigned actual temperature data. For each block group we calculate the average low temperature for each month forming our dependent variable. Our independent variables consist of household income reported by the Census 2000 at the block group, elevation, the diversity of land cover types, the percentage asphalt, and the percent vegetative cover in each block group. In addition, we include a series of year and monthly dummy variables.

Results from this regression are reported in table A.1 and show that vegetative land cover is predictive of lower summertime temperatures as is the percent of water, higher elevation, and diversity of types of land cover. Asphalt is associated with higher temperatures, as expected.

Using the regression results relating land cover to temperature, we predict the 2005 July low temperatures in each block group across the Phoenix region. By construction, this temperature measure is correlated with land cover, both of which we specify to enter the "true" utility function for households. Summary statistics for the predicted July low temperatures are reported in table A.2. In total, we predict temperatures for 2,113 block groups with July low temperatures ranging from a low of just over 72 degrees to a high of nearly 86.5 degrees. The mean block group predicted low temperature is slightly over 80 degrees.

Table A.1: Minimum temperature regression

Variable	Estimate	Std Err	t-stat
Constant	44.743	0.483	92.650
% Vegetative	-8.711	0.566	-15.400
% Asphalt	5.546	0.490	11.320
% Water	-10.675	2.310	-4.620
Elevation	-0.005	0.000	-13.170
Diversity	-0.038	0.041	-0.920
Monthly Dummy Variables			
February	0.432	0.234	1.850
March	6.485	0.234	27.690
April	10.992	0.231	47.520
May	21.244	0.233	91.120
June	28.077	0.231	121.680
July	37.097	0.233	159.090
August	35.817	0.233	153.650
September	28.002	0.239	117.160
October	18.925	0.240	78.780
November	5.944	0.239	24.890
December	-1.683	0.241	-6.990
Year Dummy Variables			
2001	0.202	0.212	0.950
2002	-0.324	0.207	-1.560
2003	0.716	0.205	3.490
2004	0.121	0.207	0.590
2005	0.777	0.206	3.780
2006	0.332	0.208	1.600
2007	1.185	0.212	5.600
2008	-0.080	0.239	-0.340
Statistics			
# Observations	38,424		
Adjusted R^2	0.6766		

Table A.2: Predicted 2006 July low temperatures

Temperature Prediction	
Block Groups	2113
Mean	80.56
Std Dev	1.22
Min	72.05
Max	86.45

Appendix B. Description of hazardous waste sites

The Indian Bend Wash site was proposed for listing on the National Priorities List (NPL) in December, 1982 and received final listing on September, 1983. The site was listed due to contaminated ground water and includes 12 square miles of land stretching from Scottsdale to Tempe. As a result of the contamination of ground water, six city wells were closed. Over 350,000 people live in the contaminated area. The site received "construction completed" status in 2006 but has yet to be fully deleted from the active NPL list.

The second hazardous waste site we focus on is the Motorola 52nd Street plant. This site was proposed for listing in June, 1988 and received final listing in December, 1989. To date, the site has not received construction complete status and remains an active cleanup site. This site is located on a former semiconductor manufacturing plant and encompasses 90 acres in the midst of a residential and commercial area. As a result of a leaking underground storage tank, groundwater and soil were contaminated. The contaminated water has spread several miles underground, but is not being used for drinking water, but resulted in the closure of several wells.

The third and final site we examine is the Nineteenth avenue landfill which was proposed for listing in December, 1982 and received final listing in September, 1983. The site was deleted in September, 2006. The 213 acre landfill is located in an industrial area adjacent to the Salt River. Within six miles of the site live over 16,000 people with the closest people located only 1/3rd of a mile away. As with the previous two sites, this site is responsible for contaminated ground water and has been made worse intermittently due to flooding of the nearby Salt River which had breached areas of the closed landfill. Unlike the other two sites, there are no residential wells located in the immediate vicinity of the landfill. Cleanup of the site ultimately cost of 22 million dollars.

Appendix C. Second-stage hedonic preference parameter estimates

Table C.1: Preference calibration second-stage hedonic

	Landscape			Hazardous Waste		
Variable	Estimate	Std Err	t-stat	Estimate	Std Err	t-stat
Square feet (100s)	1.828	0.007	276.580	1.834	0.006	291.610
Acres	0.803	0.003	302.660	0.591	0.002	285.920
Stories	-3.294	0.007	-455.330	-1.780	0.004	-471.370
Bathrooms	2.613	0.009	298.320	1.288	0.004	320.140
Age	-0.120	0.000	-257.260	-0.223	0.001	-303.920
Pool	2.045	0.005	434.570	1.313	0.003	405.220
Garage	3.178	0.006	495.070	1.559	0.003	501.870
Wet	1.333	0.003	474.880	n/a	n/a	n/a
July min temp	-0.967	0.002	-520.850	n/a	n/a	n/a
NPL inverse distance	n/a	n/a	n/a	-0.157	0.000	-489.130
CBD distance	0.115	0.000	316.660	0.038	0.000	344.120