THE VALUE OF WATER AS AN URBAN CLUB GOOD: A MATCHING APPROACH TO HOA-PROVIDED LAKES

Joshua K. Abbott[1*] and H. Allen Klaiber[2]

[1]Arizona State University, School of Sustainability, Global Institute of Sustainability, PO Box 875502, Tempe, AZ 85287, joshua.k.abbott@asu.edu, 480-965-5528

[2] The Pennsylvania State University, Department of Agricultural Economics and Rural Sociology, 208A Armsby Building, University Park, PA 16802, aklaiber@psu.edu, 814-865-0469

Authorship is equally shared.

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting Pittsburgh, Pennsylvania, July 24-26, 2011

THE VALUE OF WATER AS AN URBAN CLUB GOOD: A MATCHING APPROACH TO HOA-PROVIDED LAKES

Joshua K. Abbott[1*] and H. Allen Klaiber[2]

[1]Arizona State University, School of Sustainability, Global Institute of Sustainability, PO Box 875502, Tempe, AZ 85287, joshua.k.abbott@asu.edu, 480-965-5528

[2] The Pennsylvania State University, Department of Agricultural Economics and Rural Sociology, 208A Armsby Building, University Park, PA 16802, aklaiber@psu.edu, 814-865-0469

Authorship is equally shared.

*Corresponding Author

WORKING PAPER, DO NOT CITE WITHOUT PERMISSION

Abstract

Urban lakes located in arid environments require large quantities of water to maintain their water levels, with much of this water associated with high opportunity costs. Many of these lakes are manmade and provide various amenities to surrounding residents. In this paper we use matching techniques to recover the average capitalized value of lakes to surrounding communities and differentiate between community members and adjacent households to recover heterogeneous treatment effects. Importantly, we consider the role of both unobservable and observable features of matching to recover heterogeneous capitalization across lake communities. Our results suggest that the capitalized value of lakes to community residents is highly heterogeneous and ranges from an annual value of -$29 to +$20 per homeowner per acre foot of water. These results suggest that small changes in water pricing could remove the surplus benefits of lakes to community residents.

Keywords: Matching; Treatment effects; Urban lakes; Capitalization

The Value of Water As An Urban Club Good: A Matching Approach To HOA-Provided Lakes

1. Introduction

The ecosystem services accorded by water in urban areas are of great policy relevance due to the increasing concentration of human societies in urban areas, the acceleration of growth in arid and semi-arid cities dependent upon often strained groundwater and surface water supplies, and the uncertain effects of climate change on the level and variability of these supplies. In addition to its value in indoor private consumption [36], water serves as an input into features of urban and suburban landscapes such as greenspace, water features, private landscaping, and swimming pools. These amenities vary substantially in the range of services they provide, the excludability of these services across the landscape, and in the degree of public/private involvement in their provision and maintenance.

A prominent trend in many areas, particularly high-growth arid regions in the U.S, has been the increasing provision of such amenities by home owners' associations (HOAs) [33]. From 1980 to 2000, half the new housing in the United States occurred in developments governed by a neighborhood association with more than 50 million Americans now living in such communities [35]. While HOAs serve a range of governance functions, they are frequently tasked with providing, maintaining and regulating access to a range of "common" landscape amenities while assessing members (typically through membership dues) for the cost of their maintenance.[1] As such, HOAs serve as an institutional means to the provision and maintenance of club goods [15, 16] since many of their associated amenities provide services that are excludable to residents (either by explicit enforcement or by distance or geographic barriers) and are largely non-rival in their utility to residents.[2] In many cases the amenities that are provided may closely resemble, and to some degree substitute for, amenities provided by local and regional governments such as local parks and sporting facilities.

Despite the potential for these clubs to enhance the welfare of their members, the "private" provision of amenities by HOAs may generate meaningful public spillovers. In the domain of water consumption, if, as is frequently the case, water is supplied to developments at a subsidized rate below its social cost then community landscaping may be excessively skewed toward water-intensive features relative to what is socially optimal.[3] Therefore, there may be compelling social reasons to address the internal provision of amenities by HOAs through constraints at higher levels of governance via such tools as water pricing and zoning policies. As a first step toward understanding the rationale and potential for such constraints it is first important to gain a sense of the benefits to residents of water usage in residential landscaping. These values can then be compared to the marginal social cost of supply to assess the efficiency of current water use patterns within communities. Furthermore, comparing these values to the pricing of water to communities can provide evidence of the distribution of the surplus or excess burden across community members and the general public.

As a first step in this pursuit, we examine households' valuation of water as conveyed through capitalization to houses located within communities with artificial, community lakes within the Phoenix metropolitan area. Lake communities are a prominent and high-volume consumer of water. In many cases this water derives from potable (after treatment) surface flows with potentially high opportunity costs. The Phoenix area also obtains much of its water supply from the Colorado River and other surface water flows, and faces long-run vulnerabilities as the effects of climate change are likely to reduce water availability and increase its variability. These long run constraints are currently not reflected in the water pricing structures employed by local municipalities. Therefore, the study of urban lakes in this arid environment serves as a useful case study for considering the valuation of water in landscape amenities with an eye toward assessing its current efficiency and the degree to which the prevalence of such development may be responsive to changes in land-use and water policies.

The substantive objective of our work is to recover the average surplus value of the water embodied in HOA-provided lakes to community residents – the benefits accruing to residents over and above the costs of maintenance. To accomplish these objectives we assemble a large database of housing transactions data, detailed GIS layers on lakes and lake community boundaries, housing and neighborhood characteristics and other amenities. The exclusive nature of the services provided by the lakes as well as their patchy distribution over the landscape allows us to conceptualize their services as falling into two additively-separable categories: 1) those accruing to HOA residents and 2) those accruing uniquely to HOA residents in lake-adjacent properties. This ability to define binary "treatment" and "control" groups lends itself to the use of matching methods [23]. Conceptually, we match non-adjacent properties within HOAs to observationally similar properties outside the HOA to determine the average capitalization effect of lake access. We also match lake-adjacent properties within HOAs to non-adjacent properties in the same community to estimate the average additional value of adjacency. We then convert these estimated capitalization effects into equivalent rental rates and use data on lake surface area and evaporation rates to attribute this annualized value to the annual flow of water required to maintain the lakes on an annual basis.

In addition, our paper makes a number of methodological advances. First, we eschew a hedonic regression approach for the use of matching methods. This approach allows us to flexibly control for selection on observable housing and neighborhood characteristics without requiring the explicit formulation of a potentially highly-nonlinear hedonic price function. It also allows for generalized interactions between the amenities of interest and observable characteristics as well as for heterogeneous treatment effects both within and across lake communities. Second, we broaden the selection-on-observables assumptions implicit in matching to consider the importance of unobservable spatial and temporal factors. We consider the effects of "matching on unobservables" by incorporating both "hard" and "soft" elements of matching within temporal and spatial windows – drawing an analogy between the matching and panel data literature [5, 6].

2. BACKGROUND

The Phoenix metropolitan area has experienced rapid population growth and housing development in recent decades, with a population of over 4.1 million inhabitants in 2010 from a population in the 1950s of only 300,000. To meet the demand of this development for water in the midst of a desert climate yielding approximately 7 inches of precipitation annually, the area relies upon a far-flung infrastructure network to store and deliver surface water from the Salt and Verde River watersheds, water from the Colorado River, and, particularly in fringe areas, groundwater [27]. While conservation has dramatically reduced water use in Phoenix, per-capita usage remains high with residential use accounting for two-thirds of water demand [11]. Of this demand, approximately 74% occurs through outdoor use [32], primarily for swimming pools and landscape watering [31].

A relatively small but conspicuous aspect of outdoor water demand in Phoenix is reflected in planned communities built around substantial private manmade lakes, with approximately 40 large scale developments fitting this description. While scattered throughout the metropolitan area, they are especially common in the southeast near canal infrastructure (see Fig. 1).[4] These lakes are built by developers and maintained by HOAs who raise funds for their maintenance by charging residents higher HOA dues than in non-HOA communities. The amount of this premium varies widely based upon the number of units in the community, the quality of water used and the nature of amenities provided but is often several times the fee of a non-lake HOA and may be hundreds of dollars a month [10].

The nature of services provided by the lakes varies widely across communities. Boating, fishing[5] and other non-contact forms of recreation are commonly allowed (albeit with a range of restrictions) while water-skiing and swimming are permissible in a small number of cases where the source of the water is of adequate quality. In addition to these benefits, lakes may grant aesthetic benefits from views, privacy or temperature moderation.[6] A common feature of these benefits is that they are largely exclusive to the membership of the HOA. Most community bylaws explicitly forbid non-residents from using the lakes (except as a guest to a member) and lakes are rarely located on the exterior of communities. Furthermore, even within the community, there are typically variations in the degree of exclusivity across members due to the fact that only a fraction of properties are adjacent to the water, creating the potential for heterogeneous capitalization across homes.

While they vary widely in size and allowable uses, an average lake community has approximately 38 acres of surface area, and these lakes are fairly shallow, typically 8 to 12 feet in depth [8]. Importantly, warm temperatures lead to extremely high evaporation rates – approximately 6 feet per year [8]. Therefore, a lake will typically require that 50-75% of its volume be replaced each year for its maintenance. This dramatically increases the water demand of a community; the annual evaporation from an average lake community is nearly 230 acre-feet – enough to supply the water needs of over 1650 single-family residential users. Of course the opportunity cost of this water use depends to a large degree upon the source and quality of the water. Older lake communities located near to delivery infrastructure for surface flows often use surface water supplied by the Salt River Project that can be used (after treatment) for drinking water. However, newer lake communities (which are often located along the southern suburban fringe) often utilize reclaimed water or, if the community lacks access to surface supplies, groundwater. While reclaimed water may have a lower opportunity cost than drinkable supplies, its use precludes a lake being used for swimming or skiing purposes.

3. Data

In order to establish the average value of water in urban lake communities we first created a GIS layer of lakes and other water features utilizing a combination of land use data from the Maricopa Association of Governments and remote sensing data, verified by visual spot-checking using satellite mapping products. From this information we are able to calculate the surface area of water in each water feature. We then built a database of candidate lake communities and associated these water features with these communities. In doing this we had to account for the fact that lakes are found in a variety of settings throughout the metropolitan area, and are often correlated with a variety of related land uses or community types that may confound our ability to isolate the effects of the lake itself. We immediately eliminated lakes contained in public parks from our sample as well as lakes that are integrated within golf courses or lake communities that primarily serve retirement residents. We also limit our attention to lake communities that primarily serve single-family residential customers. Finally, we eliminated a small number of communities located in the extreme exurban fringe since it was felt that they were in a substantially different housing market. This process left us with a sample of 29 candidate lake communities (Figure 1).

Given this initial sample, we then utilized information on subdivision names from assessor data combined with community maps to establish precise spatial boundaries for lake community membership. These boundaries provide a discrete measure of "treatment" given that houses inside the boundaries have the right to access the lakes and responsibilities of paying for their maintenance while those outside the bounds are likely excluded from most of the lakes' services.[7] In other words, spatial boundaries and the boundaries of club membership coincide precisely. Altogether, these communities contained approximately 1145 acres of lakes with a (conservative) combined volume of 9200 acre-feet and estimated annual evaporation of 6870 acre-feet – a rate of loss that can meet the demand of approximately 50,000 Phoenix residential consumers.

While this reflects our ideal population of lake communities, we further restrict our sample in order to ensure that our identification strategy captures the value of the lakes themselves rather than other bundled aspects of particular lake communities that may differ markedly from typical "control" communities in ways that are not easily controlled for. In particular, some communities offer unique amenities such as extensive clubhouse facilities, sports courts or are heavily integrated with park space (both public and private). Other lake communities are highly clustered together such that separating the effects of community membership from broader neighborhood spillovers is problematic. Still other lake communities are associated with custom, ultra-luxury housing or large-lot development for which suitable control communities are scarce (particularly within the same geographic area).[8] After omitting these communities, we have 15 communities left in our sample, representing 437 acres of lakes and more than 2,600 acre-feet of annual evaporation. While this dramatically restricts our sample, we feel this truncation is justifiable on the basis that our objective is not to recover the value of community membership itself but to credibly recover the value of the water associated with lake communities. The communities in our final sample and their summary characteristics are presented in Table 1.

In order to build a dataset of housing transactions and characteristics within and outside lake communities, we combine data purchased from a private data vendor, Dataquick, with data from the Maricopa County assessor and restrict our attention to 2002 to 2005 sales containing a complete list of housing characteristics. In addition to these characteristics, we also observe the sales price and sale date of each house as well as a unique identifier which allows us to link these data to GIS assessor parcel maps and assign a set of spatial coordinates to each house. We supplement these parcel-specific variables with key neighborhood variables including variables relating to the amount of subdivision-provided open space (including measures of adjacency, distance to closest open space and overall provision[9]) [5, 6], golf courses (including adjacency, a dummy reflecting whether the subdivision is in proximity to a golf course, and a distance variable for each house to the nearest course), and a dummy indicating the elementary school attendance zone associated with a parcel [14].

4. METHODOLOGY

To estimate the average value of water in maintaining the benefits associated with urban lakes we take the approach of first finding the average change in capitalization associated with lake community membership and subsequently finding the additional capitalization associated with adjacency. Before detailing our estimation methods, we make two important points. First, we hypothesize that the benefits from lakes within a community are likely heterogeneous due to the fact that some lots are adjacent to the water and may derive unique value from this proximity. We therefore posit that the value to adjacent properties of lakes is additively separable into two components: a component that is shared (in the sense of having the same conditional expectation) between adjacent and non-adjacent properties and a component that is unique to adjacent properties. Second, our measure of the "price" of the house does not include the cost of commitments attached to ownership of the house. Most importantly, the price does not reflect the present discounted value of the home ownership fees paid by residents. To the extent that these fees are substantially higher in lake communities than in otherwise similar non-lake communities (which anecdotally they are) any hedonic estimator utilizing the sales prices will tend to undervalue the amenities in the lake community. However, our indications are that the differences in dues are primarily tied to the extra maintenance costs associated with the lakes themselves. Therefore, a comparison of prices between lake and non-lake communities remains informative in that it yields estimates of the surplus value to homeowners above and beyond their contribution to the maintenance of the amenity itself. A negative or zero price differential across otherwise identical lake and non-lake properties is therefore not necessarily indicative of zero value for the lake amenities but may reflect that the associated costs of the amenities exhausts or overwhelms any surplus value to homeowners.

Both membership in a lake community and adjacency to a lake within a particular neighborhood are fundamentally discrete amenities. In asserting this we acknowledge that particular characteristics or the expected future services of amenities which consumers utilize in forming their bid functions may have continuous aspects; however, without substantial information we are unable to unlock these latent factors and therefore are tasked with valuing the bundled attributes in a consistent manner. In other words, we can perceive membership and adjacency in each community as a discrete "treatment". Our econometric task is then to recover the average treatment effects on the treated (ATT) homes.

A vast array of approaches is available for estimating average treatment effects [9, 23]. A central concern of all these methods is that the assignment to the treatment group may be correlated with observable or unobservable variables which themselves have a role in determining the outcome of interest (i.e. the price of a house). For instance, individuals that desire to live in lake communities may (either from income effects or due to heterogeneous preferences) also prefer homes with larger lots. Failure to acknowledge this difference in comparisons between the treatment and control groups will yield biased ATT estimates. Coping with omitted variable bias and other forms of endogeneity is a dominant theme in the hedonic pricing literature [5, 25, 37].[10]

A common approach to valuing discrete amenities in the hedonic literature is to specify a (typically, but not necessarily, parametric) regression function controlling for observable aspects of a house and its neighborhood while entering a dummy variable into the hedonic price function for the discrete amenity in an additively separable fashion. While the coefficient on this dummy variable is the ATT in such a model, this approach relies on strong identifying assumptions. In particular deriving the ATT from a single parameter in a hedonic regression requires that the virtual price of the amenity in the underlying "true" hedonic price function not be dependent on the values of other observable characteristics, an assumption that is false in all but the most restrictive of scenarios for underlying utility functions and supply processes [26].[11]

As a flexible alternative to restrictive regression approaches, we employ a tractable matching estimator commonly used in the program evaluation literature [24]. The fundamental challenge for the estimation of ATT, adopting the language of Rubin's potential outcome framework [41], is that we consistently fail to observe the counterfactual outcome for treated observations (i.e. we fail to observe the price of lake community houses if they weren't located in lake communities). Matching estimators impute this missing potential outcome from the reservoir of potential "control" observations and then average over the differences between the matched pairs to find the ATT. This matching occurs in one of two ways. The first employs the estimated propensity score [40], the probability of a unit with particular observable characteristics receiving the treatment, to provide this match. Propensity score methods have seen fairly extensive use in papers on land use and conversion issues [13, 28-30, 34]. The second approach, which has seen broad use in the program evaluation and labor economic literatures [e.g. 17], is to match treated observations with the average outcome of a chosen number of their nearest neighbors from the control group, where "nearness" is determined by the distance of the observable covariate vectors along some norm. Both methods provide valid ATT estimates under two assumptions. The first, known variously as "conditional independence", "ignorability" and "selection on observables", is that the assignment to either the treatment or control group is independent of factors that affect the magnitude of the outcome variable after matching along observable factors.[12] The second is that the observable aspects of the treated and control observations exhibit sufficient overlap.

We adopt the nearest neighbor matching approach here. While not necessarily superior to propensity score approaches, this technique has much to recommend it in our context. In particular, given that we wish to estimate heterogeneous treatment effects over 15 different communities and adjacency within each community, this approach provides credible inference without the considerable number of difficult specification decisions involved in achieving consistent estimates of the propensity score for each treatment [4, 23]. Notwithstanding this advantage it has been demonstrated that matching estimators in which the number of matches is fixed as the sample size grows exhibit a bias due to imperfect matching of all covariates that does not disappear as the sample size increases – rendering the estimates inconsistent [2]. Furthermore, the large-sample variance of the estimator has not been well established and traditional bootstrap approaches to estimating the standard errors are invalid [3].

Fortunately, Abadie and Imbens [2] have developed a version of the simple matching estimator which adjusts the imputed counterfactual outcome from matching using a regression estimator of the conditional mean outcome for control observations. This estimator is not only consistent but also robust to misspecification of the regression equation [2]. Furthermore, a consistent estimate of the large-sample variance is available that accommodates general forms of heteroskedasticity [1, 2]. A recent Monte Carlo study demonstrates that this estimator has attractive properties in terms of bias, mean-squared-error and the coverage rate of confidence intervals [4]. Finally, these estimators are easily implemented within Stata [1].

Despite their attractiveness, matching estimators (like regression) rely heavily upon the validity of the "selection on observables" assumption for their consistency. However, the importance of controlling for potentially correlated unobservable heterogeneity is now well established in the hedonic pricing literature. Two forms of heterogeneity are particularly important: 1) time-varying heterogeneity, including the potential for market adjustment within the sample frame [26], and 2) unobserved spatial heterogeneity, particularly difficult-to-measure "neighborhood" variables [37] such as school quality and local public good provision that are shared by a number of proximate properties [7]. Each is problematic when using matching methods since transactions that are ideal matches based on observable housing characteristics may differ markedly in terms of these neighborhood variables or occur on either side of a shock to a hedonic equilibrium.

We utilize a number of methods to immunize our matching estimates to unobservable heterogeneity. To account for time-varying effects over our multi-year sample of transactions we first deflate housing prices by a monthly price index for the local housing market to purge prices of common market trends. We then include the date of sale as a covariate in the matching algorithm to control for seasonality or other temporal effects apart from the overall appreciation in our sample period. We take special care to purge our estimates of correlated spatial effects, utilizing both “soft” and “hard” techniques to increase the probability of matching on important spatial heterogeneity as well as observable variables. In the “soft” category we include both the latitude and longitude of transactions and indicator variables for school attendance zone (an important aspect of neighborhood choice with demonstrated capitalization effects [14]) among the matching variables. We supplement this approach by truncating the matching sample to neighborhoods within 1, 2 and 3-mile buffers of each lake community. To the extent that properties within such buffers share many of the same unobservable neighborhood characteristics, differing across their housing prices identifies the effects of lake community membership on the basis of “within” variation – mimicking the approach of spatial fixed effects estimators [5, 12].[13] We examine the sensitivity of our estimates to using the more circumscribed buffers, paying particular attention to the potential tradeoff between controlling for unobservable heterogeneity and potential bias or lack of precision from poorer observable matches. This is, to the authors’ knowledge, the first explicit consideration of spatial and temporal omitted variables bias within the applied land use propensity score/matching literature.

5. RESULTS

The full suite of variables employed for matching are reported in Table 2. These include a number of structural characteristics that are commonly included in property valuation studies as well as some controls for highly localized amenities (i.e. golf and subdivision open space) whose importance in this real estate market has been previously demonstrated [5, 7]. We also include the aforementioned “soft” controls for spatial and temporal heterogeneity.

While we include date of sale as a matching variable to control for changes in the underlying hedonic equilibrium through time, this does not account for the overall appreciation of housing prices over the 2002 to 2005 sample period. Simply averaging over matched price differentials may lead to ambiguity of units and may skew the comparison of the estimated treatment effects across communities if the temporal volume of transactions differs across communities. Therefore, we normalize the sales prices of each house to January 2000 levels using the Case-Shiller monthly housing price index[14] for the Phoenix market prior to estimation.

Once the control and outcome variables are selected, nearest neighbor matching requires two additional decisions from the analyst. The first, assuming exact matching is impossible (which is the case here given the presence of continuous covariates), is to define a norm $\|x\|_A = (x'Ax)^{\frac{1}{2}}$ for measuring the closeness of a match in the *K* dimensional space of observable covariates. Two norms are commonly employed in the literature (although see [42]). The first, known as the Mahalanobis metric, defines *A* as the inverse sample covariance of the matched covariates. The second, employed by Abadie and Imbens [4], utilizes the diagonal entries of this matrix with the off-diagonal entries set to zero – measuring Euclidean distance on the standardized covariates so that deviations in covariates are weighted in inverse proportion to their standard errors. We employ the second metric here exclusively, but preliminary analysis suggests that our results are not significantly influenced by the matching metric. Preliminary results did suggest, however, that our estimates were sensitive to weak matching on two key characteristics that previous hedonic work has shown to be extremely important in determining the price of a house – the number of bathrooms and number of rooms. By trying to match along competing dimensions, the matching algorithm sometimes compromised this critical dimension of the match. We therefore increased the weighting on these two discrete covariates to force "exact" matching wherever possible before matching on secondary traits – a change that greatly increased the robustness of our estimates to alternative assumptions.[15]

The second major choice is the minimum number of matches to use from the control group for each treated unit. As with the choice of bandwidth in a kernel regression, expanding the number of matches involves an increase in precision at the cost of greater potential bias. This occurs because the quality of the match will typically deteriorate as the number of matches increases. While there is no firm guidance from the literature, we follow empirical practice by selecting a small number of matches, ultimately selecting four matches for each observation, which yields sufficient precision given our fairly large sample of treated properties and the large reservoir of reasonable controls for most transactions.[16] As with the matching metric, our results are not highly sensitive to this decision.

5.1 Community Membership

Table 3 shows preliminary initial estimates of the average effect of community membership on the transaction price of housing for all 15 communities under three spatial definitions of the reservoir of controls. The first control group is limited to a 1 mile buffer extending out from the boundaries of each lake community (where the buffers retain the shape of the community boundary). The second and third groups include the initial set of controls while extending an additional 1 and 2 miles respectively. As one metric of the quality of the observable matches behind each assessment we have included the percentage of exact matches obtained by the matching algorithm for the number of rooms and bathrooms. While admittedly selective, it provides a metric of the extent of successful matching that is comparable across communities and across different control groups within a community.

Comparing these estimates at the 1 mile scale it is apparent that there is substantial heterogeneity in the estimates of ATT across communities as well as marked differences in the quality of matches. This highlights the importance of addressing this problem in a heterogeneous treatment framework rather than averaging over communities in a manner which might introduce biases into our inferences. A number of communities have very high rates of exact matches (>95%); however, some are far less satisfactory. For example, Playa del Rey is a waterski community with substantially larger homes than are typically observed in its environs.

As we extend the buffer for control properties outward, we see (with the exception of Playa del Rey) that the quality of observable matches, as reflected by exact matching on number of rooms and bathrooms[17] increases substantially. In fact, if we strictly adopt the conditional independence assumptions behind our estimator, we should strongly prefer matching at 3 mile buffers to matching within 1 mile. However, in doing so we are assuming that any unobservable spatial heterogeneity that plays a role in determining housing prices is uncorrelated with the proximity of the control properties relative to the treatment community. This is a very strong assumption. By examining the sensitivity of our estimates to the scale of the control group we can assess the potential for biases due to spatial heterogeneity for each ATT estimate. In doing this we find that several communities exhibit comparatively robust estimates of the ATT across different control groups in terms of the quantitative magnitude of estimates or the persistence of insignificant findings, particularly between the 1 mile and 2 mile range. On the other hand, we find there are 4 communities – Crystal Gardens, Oakwood Lakes, Pinelake Estates and Playa del Rey – that vary substantially. These estimates also tend to be implausibly large (or excessively negative), calling into question their overall reliability. Out of these four alarming cases, three exhibit exceptionally poor observable fit at the 1 mile range; therefore, it is possible that the swings in estimates may result from reductions in bias from poor observable matches as well. Cases in which estimates are not robust over spatial scales can be resolved by matching to controls that are likely to share identical/similar unobserved traits (typically within a close spatial proximity of the treated home). However, doing this may compromise our ability to match closely along observable dimensions of homes.[18] Optimally resolving these countervailing biases in our choice of control group is difficult on an a priori basis. We therefore eliminate these communities from consideration in determining the value of water. After eliminating these properties, we find that 6 of our estimates are positive and significant at the 5% significance level, 4 are insignificant, and one is negative and significant with effects on sale values that range between -$10,000 per house to over $22,000. Importantly, these reflect the capitalized value of lake community membership without incorporating the levies from HOAs in prices. Since HOA fees are clearly presented to potential homebuyers before they purchase a house and are part of the responsibility of owning a home, the obligation to pay them should be reflected in the final transaction price. Therefore, these estimated values reflect the average value of community membership over and above the expected cost of maintenance for each homeowner. As a result, small negative values may be consistent with the lakes providing gross amenities to non-adjacent homeowners in a community.

5.2 Adjacency

Table 4 contains estimates for the average treatment effects of lake adjacency within communities. These are obtained by matching adjacent homes to observationally similar homes inside their community. Despite relatively small numbers of transactions for adjacent properties (see Table 1), adjacency is statistically and economically significant in all but a handful of cases – far more consistently so than for community membership itself. Not surprisingly, there is a substantial premium attached to the private benefits associated with adjacency – often running into the tens of thousands of dollars.

The quality of matches, as reflected by percentage of exact matches on number of rooms and bathrooms, is generally fair. However, there are some cases where lower match percentages suggest some of our estimates may reflect the influence of substantial differences in the observable characteristics of homes relative to the non-adjacent homes they are evaluated against. In the case of Playa del Rey, the estimated $100,000 premium for a house may reflect the benefits of immediate access to a ski lake, but it may also reflect the value of correlated housing characteristics such as high-end backyard entertaining fixtures or boat-lifts. Similarly, in the case of The Islands, the quality of exact matches is sufficiently low as to suggest that the $70,000 premium attributed to lake adjacency may reflect, at least in part, the provision of larger, higher-end properties along the waterfront.

5.3 The Average Value of Water

To calculate the value of water we must first determine what the appropriate volume of water is for the purposes of valuation. While we know the rough depth of lakes and can therefore calculate their volume, this volume is not the appropriate divisor in our calculations. Rather we are concerned with the amount of water required to maintain the services provided by the lakes. This is primarily reflected in the water required to counteract evaporation – a quantity of roughly 6 feet per year [8]. To resolve these annual evaporation estimates with our capitalization estimates, we must convert our capitalization estimates into equivalent annualized "rents". To do this we follow the literature by multiplying our estimates by 11% [39]. After making these conversions we divide the annualized rents by the number of acre-feet to get an estimate of the average annualized premium attributable to each home of an acre-foot of water. We then sum these values over the number of houses in the neighborhood given the non-rivalrous, club good nature of lakes in the community. Table 5 reports our estimates, showing the average value attributable to an acre-foot of water due to the adjacency premium for all adjacent properties, its shared value to all community properties, and the total of both values.

Out of 11 properties, 4 show negative average values for water at the community scale while positive values range from less than $2,000 to more than $9,000 an acre-foot. Recall that these are *surplus* values beyond homeowners' contributions from their HOA dues. Since HOA dues provide the funding for the placement of water orders to suppliers, this demonstrates that homeowners are, in most communities, willing to spend considerably more than their current rates for the services they derive from the water employed in their club goods. The mean premium associated with an acre foot of water per homeowner varies between $3 and $20 with most communities falling in the $3 to $10 range.

The value of water in adjacency is dramatic and accounts for 50% or more of the overall value of water, even though adjacent housing is typically a much smaller fraction of overall housing in the neighborhood. Table 6 presents estimates of the annualized value of water on a per-home basis. The implicit average value of water for adjacent homes as reflected in housing rents is many times that of non-adjacent properties. This has several interesting implications from a political economy perspective. First, the supply of water to lake communities by public utilities at inefficiently low prices effectively serves as a substantial subsidy to a relatively small set of water-adjacent homeowners. Second, the internal political economy of HOAs further skews this distributional consequence in that, to our knowledge, every community charges the same HOA fees to all members regardless of whether a home is adjacent to a lake or not. This may have important consequences for the willingness of communities to expend additional resources for the better maintenance of their lakes since non-adjacent homeowners are the majority of the community yet often receive fairly small surplus benefits from the lakes. Third, the excludable benefits of adjacency are sufficiently high to suggest that developers have strong incentives to manipulate lake topology and placement so as to maximize lake frontage for housing. Indeed, we find this to be the case, with lakes taking extremely deformed shapes in the majority of cases (see Figure 2). The average lake community in our initial sample has lakes with perimeters nearly five times that of a circle of equivalent surface area (a circle being the minimum-perimeter shape for a given area), and provision of significant community park space alongside lakes is extremely rare. This finding is understandable since developers rarely retain management of the community after the initial sale of homes and therefore have incentives to maximize the initial sales value of the development. However, there is evidence that this may lead to suboptimal lake design with respect to the long-run provision of services to residents since the large number of inlets can lead to poor water circulation, low dissolved oxygen, algae accumulation, fish kills and other challenges [8]. Countering these undesirable consequences often entails substantially higher maintenance costs to communities.

6. DISCUSSION

We employ a bias-corrected form of nearest neighbor matching to assess the value of water as an input into the provision and maintenance of an urban club good. This approach is unusual in that papers evaluating the value of water *quantity* rather than water quality outside of private, excludable consumption are very rare; while there are a handful of papers evaluating these questions in a recreational context [18, 19] we are not aware of any applications to housing transactions. In deriving our estimates we make a number of methodological contributions to the matching literature. Most notably, we demonstrate the importance of spatial scale in deriving the appropriate control group for our matches and suggest that omitted spatial variables may bias estimates if the scale of matching is picked incorrectly. In many cases there will be a tradeoff between expanding the reservoir of controls to increase the closeness of matches on observable structural characteristics and increasing the likelihood of potentially poor matches on coarser-grained “neighborhood” variables. We also demonstrate how the evaluation of spatially distinct treatments on an individual basis can help to uncover treatments whose effects cannot be reliably identified, helping to increase the integrity of estimates.

Ultimately we find that water in subdivision lakes has substantial value to homeowners over and above the everyday costs of maintenance. Nevertheless, a large share of the benefits of these lakes apparently accrues exclusively to a minority of homeowners with lake frontage. Our estimates are valuable in that they suggest the extent to which community water use may be elastic with respect to water pricing policies. To the extent that levies for water exceed the mean values reported in Table 6, we would no longer expect lakes to function as net amenities to community members, deterring such development in the future. Indeed, given the unequal distribution of benefits and maintenance costs between adjacent and non-adjacent homeowners, far smaller levies are necessary to turn many lakes into dis-amenities for the median homeowner.

References

[1] A. Abadie, D. Drukker, J.L. Herr, G.W. Imbens, Implementing matching estimators for average treatment effects in Stata, *The Stata Journal*, **4** (2004) 290-311.

[2] A. Abadie, G.W. Imbens, Large sample properties of matching estimators for average treatment effects, *Econometrica*, **74** (2006) 235-267.

[3] A. Abadie, G.W. Imbens, On the Failure of the Bootstrap for Matching Estimators, *Econometrica*, **76** (2008) 1537-1557.

[4] A. Abadie, G.W. Imbens, Bias-Corrected Matching Estimators for Average Treatment Effects, *J Bus Econ Stat*, **29** (2011) 1-11.

[5] J.K. Abbott, H.A. Klaiber, An Embarassment of Riches: Confronting Omitted Variable Bias and Multi-Scale Capitalization in Hedonic Price Models, *In press, Review of Economics and Statistics*, (2010).

[6] J.K. Abbott, H.A. Klaiber, Is all space created equal? Uncovering the relationship between competing land uses within subdivisions, in, Arizona State University, 2010.

[7] J.K. Abbott, H.A. Klaiber, Is all space created equal? Uncovering the relationship between competing land uses in subdivisions, *Ecological Economics*, **70** (2010) 296-307.

[8] R. Amalfi, Vice-president & limnologist, Aquatic Consulting Inc., in, 2010.

[9] J.D. Angrist, J. Pischke, Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, New Haven, 2009.

[10] G. Baker, Realtor, in, 2010.

[11] R.C. Balling, P. Gober, Climate variability and residential water use in the city of Phoenix, Arizona, *J Appl Meteorol Clim*, **46** (2007) 1130-1137.

[12] S. Banzhaf, R.P. Walsh, Do people vote with their feet? An empirical test of Tiebout's mechanism, *American Economic Review*, **98** (2008) 843-863.

[13] A. Bento, C. Towe, J. Geoghegan, The effects of moratoria on residential development: Evidence from a matching approach, *American Journal of Agricultural Economics*, **89** (2007) 1211-1218.

[14] S.E. Black, Do better schools matter? Parental valuation of elementary education, *Quarterly Journal of Economics*, **114** (1999) 577-599.

[15] J.M. Buchanan, An Economic Theory of Clubs, *Economica*, **32** (1965) 1-14.

[16] R. Cornes, T. Sandler, The theory of externalities, public goods and club goods, 2 ed., Cambridge University Press, New York, 1996.

[17] R.H. Dehejia, S. Wahba, Causal effects in, nonexperimental studies: Reevaluating the evaluation of training programs, *J Am Stat Assoc*, **94** (1999) 1053-1062.

[18] M.E. Eiswerth, J. Englin, E. Fadali, W.D. Shaw, The value of water levels in water-based recreation: A pooled revealed preference/contingent behavior model, *Water Resour Res*, **36** (2000) 1079-1086.

[19] E. Fadali, W.D. Shaw, Can recreation values for a lake constitute a market for banked agricultural water?, *Contemporary Economic Policy*, **16** (1998) 433-441.

[20] J.Y. Hahn, On the role of the propensity score in efficient semiparametric estimation of average treatment effects, *Econometrica*, **66** (1998) 315-331.

[21] J.J. Heckman, H. Ichimura, P. Todd, Matching as an econometric evaluation estimator, *Rev Econ Stud*, **65** (1998) 261-294.

[22] J.J. Heckman, H. Ichimura, P.E. Todd, Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Rev Econ Stud*, **64** (1997) 605-654.

[23] G.W. Imbens, Nonparametric estimation of average treatment effects under exogeneity: A review, *Review of Economics and Statistics*, **86** (2004) 4-29.

[24] G.W. Imbens, J.M. Wooldridge, Recent Developments in the Econometrics of Program Evaluation, *J Econ Lit*, **47** (2009) 5-86.

[25] E.G. Irwin, N.E. Bockstael, The problem of identifying land use spillovers: Measuring the effects of open space on residential property values, *American Journal of Agricultural Economics*, **83** (2001) 698-704.

[26] N.V. Kuminoff, C.F. Parmeter, J.C. Pope, Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?, *Journal of Environmental Economics and Management*, **60** (2010) 145-160.

[27] E.K. Larson, N.B. Grimm, P. Gober, C.L. Redman, The paradoxical ecology and management of water in the Phoenix, USA metropolitan area, *Ecohydrology and Hydrobiology*, **5** (2005) 287-296.

[28] X.P. Liu, L. Lynch, Do zoning regulations rob rural landowners' equity?, *American Journal of Agricultural Economics*, **93** (2011) 1-25.

[29] L. Lynch, W. Gray, J. Geoghegan, Are farmland preservation program easement restrictions capitalized into farmland prices ? what can a propensity score matching analysis tell us ?, *Rev Agr Econ*, **29** (2007) 502-509.

[30] L. Lynch, X.P. Liu, Impact of designated preservation areas on rate of preservation and rate of conversion: Preliminary evidence, *American Journal of Agricultural Economics*, **89** (2007) 1205-1210.

[31] C. Martin, Landscape water use in Phoenix, Arizona., *Desert Plants*, **17** (2001) 26-31.

[32] P.M. Mayer, W.B. DeOreo, Residential end uses of water. American Water Works Association Report 1999.

[33] E. McKenzie, Privatopia: homeowner associations and the rise of residential private government. , Yale University Press., New Haven, , 1994.

[34] D.P. McMillen, J.F. McDonald, Land values in a newly zoned city, *Review of Economics and Statistics*, **84** (2002) 62-72.

[35] R.H. Nelson, Private neighborhoods and the transformation of local government, Urban Institute Press, Washington, DC, 2005.

[36] S.M. Olmstead, The economics of managing scarce water resources, *Rev Env Econ Policy*, **4** (2010) 179-198.

[37] R.B. Palmquist, Valuing Localized Externalities, *Journal of Urban Economics*, **31** (1992) 59-68.

[38] R.B. Palmquist, Property Value Models, in: K.G. Maler, J.R. Vincent (Eds.) Handbook of Environmental Economics, Elsevier, North Holland, 2005.

[39] J.M. Poterba, House Price Dynamics - the Role of Tax Policy and Demography, *Brookings Pap Eco Ac*, (1991) 143-203.

[40] P.R. Rosenbaum, D.B. Rubin, The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, **70** (1983) 41-55.

[41] D.B. Rubin, Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *J Educ Psychol*, **66** (1974) 688-701.

[42] Z. Zhao, Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence, *Review of Economics and Statistics*, **86** (2004) 91-107.

	Total Houses		Total Transactions		Average Sales Price	
Community	Adjacent	Non-Adjacent	Adjacent	Non-Adjacent	Adjacent	Non-Adjacent
Crystal Gardens	596	309	454	238	143,641	131,153
Desert Harbor	278	1,006	114	429	197,411	140,002
Garden Lakes	234	1,982	83	846	185,895	136,559
Lago Estancia	114	299	35	121	166,308	144,612
Oakwood Lakes	62	138	19	64	266,349	200,900
Oasis at Anozira	40	347	15	110	230,264	221,912
Pecos Ranch	122	281	45	121	168,927	157,538
Pinelake Estates	98	119	120	130	323,696	275,192
Playa del Rey	33	93	17	71	413,823	294,640
Raintree Ranch	64	99	38	53	180,855	163,727
Stonebridge Lakes	119	249	38	110	189,856	157,258
The Islands	380	1,749	114	735	274,882	156,961
The Lakes	220	572	53	200	272,223	167,238
The Springs	330	329	141	160	158,005	124,611
Wind Drift	53	592	22	261	183,801	137,905

Table 1: Lake Community Sales Characteristics

House-specific variables	
	Lot size
	# bathrooms
	# rooms
	# stories
	Square footage
	Year built
	Swimming pool (0/1)
	Garage (0/1)
Neighborhood amenities	
	Golf adjacent (0/1)
	Golf community (0/1)
	Distance to closest golf
	Subdivision open space (SOS) adjacent (0/1)
	SOS distance
	Neighborhood SOS provision
Temporal controls	
	Date of sale
Spatial controls	
	Latitude (decimal degrees)
	Longitude (decimal degrees)
	School attendance zone (0/1)*
	School district zone (0/1)*

* These are =1 when a property is in the same elementary school attendance zone(s)/district of the associated lake community

Table 2: Variables Used in Matching Estimates

	1 Mile Buffer				2 Mile Buffer				3 Mile Buffer			
Community	ATT	Std	z-stat	% exact	ATT	Std	z-stat	% exact	ATT	Std	z-stat	% exact
Crystal Gardens	**52,894**	3,013	17.56	96.85	**-31,307**	4,986	-6.28	100.00	**-34,588**	6,930	-4.99	100.00
Desert Harbor	**6,831**	2,030	3.37	90.21	**6,472**	4,105	1.58	99.13	**5,635**	3,365	1.67	99.94
Garden Lakes	**3,317**	874	3.79	90.28	**10,468**	720	14.55	98.17	**4,618**	835	5.53	98.82
Lago Estancia	**3,645**	1,744	2.09	97.52	**4,733**	1,846	2.56	99.59	**-2,341**	2,225	-1.05	100.00
Oakwood Lakes	**-45,712**	5,659	-8.08	78.13	**-3,149**	5,077	-0.62	98.83	**2,793**	4,458	0.63	100.00
Oasis at Anozira	**8,426**	6,704	1.26	95.00	**9,647**	5,746	1.68	100.00	**3,447**	6,970	0.49	100.00
Pecos Ranch	**-3,379**	2,213	-1.53	100.00	**-3,149**	2,289	-1.38	100.00	**-4,914**	2,281	-2.15	100.00
Pinelake Estates	**9,245**	8,794	1.05	86.92	**-19,410**	10,606	-1.83	95.19	**-26,396**	5,594	-4.72	100.00
Playa del Rey	**35,496**	12,280	2.89	50.70	**8,707**	14,014	0.62	52.46	**-73,831**	15,233	-4.85	59.15
Raintree Ranch	**-10,046**	3,487	-2.88	100.00	**2,825**	3,298	0.86	100.00	**1,214**	3,725	0.33	100.00
Stonebridge Lakes	**-5,734**	3,507	-1.63	100.00	**5,881**	6,041	0.97	100.00	**5,528**	6,316	0.88	100.00
The Islands	**7,076**	1,323	5.35	97.38	**12,958**	1,545	8.39	99.08	**12,103**	1,675	7.22	99.66
The Lakes	**22,131**	3,710	5.96	97.75	**19,159**	3,086	6.21	99.62	**17,587**	3,081	5.71	99.75
The Springs	**-195**	1,782	-0.11	99.84	**-714**	1,786	-0.40	100.00	**-6,166**	1,761	-3.50	100.00
Wind Drift	**4,655**	1,485	3.14	99.62	**4,265**	1,358	3.14	100.00	**2,950**	1,364	2.16	100.00

Table 3: Average treatment effect on the treated for community membership

	Adjacency			
Community	ATT	Std	z-stat	% exact
Crystal Gardens	9,290	1,224	7.59	93.84
Desert Harbor	52,688	2,551	20.65	96.05
Garden Lakes	39,466	3,144	12.55	90.66
Lago Estancia	13,695	3,732	3.67	88.57
Oakwood Lakes	50,171	7,818	6.42	94.74
Oasis at Anozira	18,923	11,923	1.59	100.00
Pecos Ranch	19,008	3,648	5.21	85.56
Pinelake Estates	41,586	4,228	9.84	90.83
Playa del Rey	103,744	19,525	5.31	85.29
Raintree Ranch	3,996	3,768	1.06	94.74
Stonebridge Lake	33,797	3,537	9.56	94.74
The Islands	71,832	5,709	12.58	83.55
The Lakes	69,911	10,008	6.99	71.70
The Springs	24,571	3,836	6.41	75.00
Wind Drift	20,682	15,208	1.36	89.77

Table 4: Average treatment effect on the treated for lake adjacency

Community	Lake area (acres)	Evaporation (af/yr)	$/af·yr - Adjacency	$/af·yr - Community	$/af·yr - Overall
Desert Harbor	47.7	286	$ 5,633	$ 3,373	$ 9,007
Garden Lakes	40.5	243	$ 4,180	$ 3,327	$ 7,508
Lago Estancia	15.2	91	$ 1,885	$ 1,817	$ 3,703
Oasis at Anozira	13.0	78	$ 1,065	$ 4,588	$ 5,653
Pecos Ranch	15.6	94	$ 2,724	$ (1,599)	$ 1,125
Raintree Ranch	5.0	30	$ 936	$ (5,992)	$ (5,056)
Stonebridge Lakes	11.7	70	$ 6,290	$ (3,300)	$ 2,990
The Islands	76.9	461	$ 6,511	$ 3,593	$ 10,105
The Lakes	45.5	273	$ 6,191	$ 7,056	$ 13,247
The Springs	33.8	203	$ 4,403	$ (70)	$ 4,333
Wind Drift	5.9	35	$ 3,407	$ 9,332	$ 12,738

* Calculated assuming 11% annual rental rate

Table 5: The average value of maintenance water in lake communities

Community	$/af·homeowner·yr (Mean)	$/af·homeowner·yr (Adjacent)	$/af·homeowner·yr (Non-adjacent)
Desert Harbor	$ 7	$ 23	$ 3
Garden Lakes	$ 3	$ 19	$ 2
Lago Estancia	$ 9	$ 21	$ 4
Oasis at Anozira	$ 15	$ 38	$ 12
Pecos Ranch	$ 3	$ 18	$ (4)
Raintree Ranch	$ (31)	$ (22)	$ (37)
Stonebridge Lakes	$ 8	$ 44	$ (9)
The Islands	$ 5	$ 19	$ 2
The Lakes	$ 17	$ 37	$ 9
The Springs	$ 7	$ 13	$ (0)
Wind Drift	$ 20	$ 79	$ 14

* Calculated assuming 11% annual rental rate

Table 6: The average value of maintenance water per home in lake communities

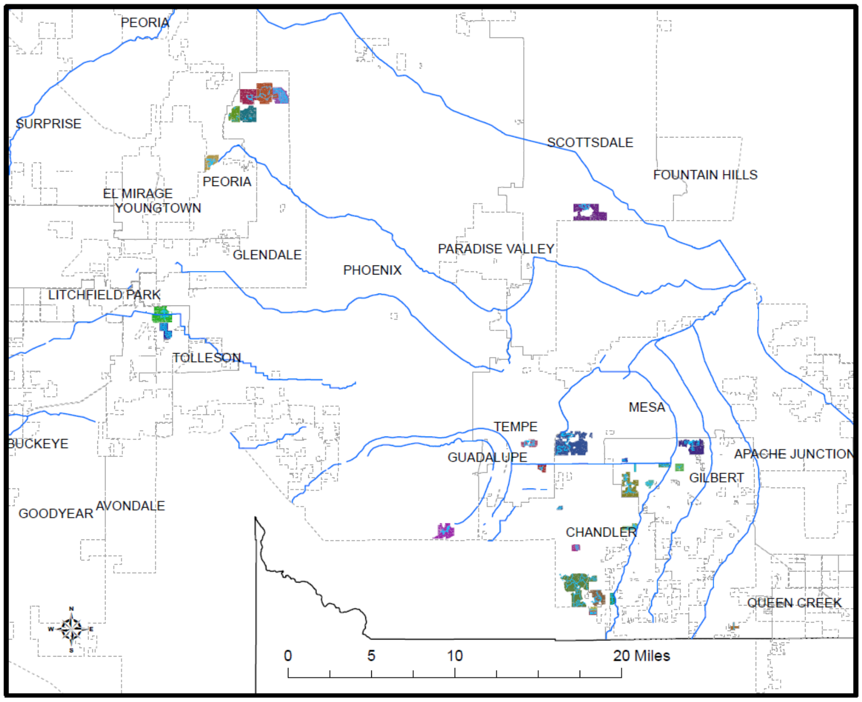

Figure 1: Locations of Lake Communities in Metropolitan Phoenix

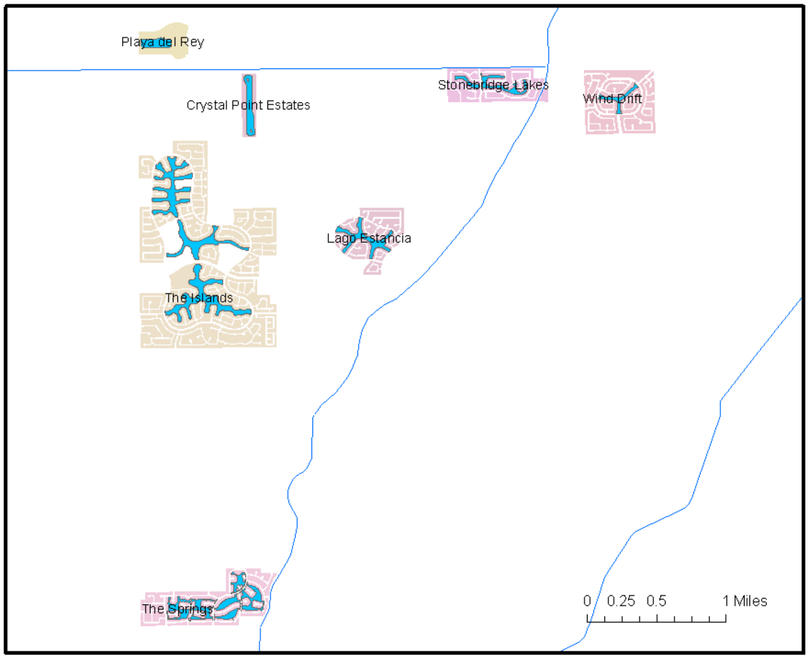

Figure 2: Representative illustration of lake topologies (rectangular lakes to the NW are ski lakes)

ENDNOTES

[1] HOAs also frequently regulate "private" landscaping and structures in order to limit the externalities associated with private decisions – a form of localized zoning. We do not consider these functions here.

[2] HOAs may also provide additional club goods by facilitating the sorting of individuals with similar characteristics (e.g. golfers or senior citizens) into communities.

[3] However, water intensive landscaping also creates a cooling effect that may alleviate the "urban heat island" beyond the bounds of the community, perhaps generating a countervailing positive externality to such development as well.

[4] Our definition of a lake community excludes lakes associated with golf course development or in which development abuts publically provided lakes or other water features.

[5] Certain fish species (i.e. white amur, common carp, goldfish) are frequently stocked as a means to control the populations of algae, grass and insect larvae in the lake. In addition, sport fish such as bass and even trout, are often stocked for recreational purposes.

[6] Lakes may contain undesirable features that make them a dis-amenity. For instance, a combination of high dissolved nutrients, high temperatures and poor water circulation may lead to algae growth and odor. A variety of biological and chemical controls are employed by contracted lake managers to limit these problems.

[7] It is possible that there could be some spillover to neighboring homes but this is likely minimal since the design of lake communities typically makes the lakes highly interior to the development.

[8] Unfortunately, most water-ski lakes (with the possible exception of Playa del Rey) are associated with these sorts of high-end development. This makes it impossible for us to thoroughly examine the difference in capitalization attributable to different levels of water quality.

[9] Overall provision is calculated as the total acreage of subdivision open space located within at least 500 feet of a home in the subdivision. Using this calculation strategy, subdivision open space located between two adjacent subdivisions would be included in the total attributable to each subdivision.

[10] See Palmquist [38] for a useful survey of the hedonic pricing literature.

[11] Regression-based approaches are capable of estimating ATT under assumptions no more restrictive than those we employ in this paper [20-22]. However, doing so requires the use of nonparametric methods.

[12] The conditional independence assumption is a generalization of the usual exogeneity assumptions for regression analysis [9].

[13] While analogous to spatial fixed effects estimation, the matching approach is far more general in that it is robust to non-additive spatial heterogeneity (i.e. interactions of spatial effects with observable characteristics).

[14] The Case-Shiller index is a quality-adjusted price index based upon repeat-sales estimation methods and is available from http://www.standardandpoors.com.

[15] We do this by up-weighting the weights for these variables by a factor of 1000 [1].

[16] In a simulation context, Abadie and Imbens [4] found that four matches performed well in terms of root mean squared error.

[17] We acknowledge that this metric does not capture all the relevant dimensions of covariate matching. We are developing parsimonious ways of measuring the quality of matching along all covariate dimensions.

[18] A similar problem occurs in spatial panel estimation where fine spatial ("within") variation in some characteristics may be sparse.