Hedonic Equilibria, Land Value Capitalization, and the Willingness to Pay for Public Goods[†]

Nicolai V. Kuminoff
Department of Economics
Arizona State University

Jaren C. Pope
Department of Economics
Brigham Young University

This version: May, 2010

First version: December, 2008

Abstract

Models of household location choice provide a theoretical foundation for measuring the willingness to pay for public goods. The difficulty is identification. Empirical work was traditionally believed to suffer from widespread identification problems. Recent studies have revived this literature by demonstrating that quasi-experiments can provide credible estimates for the rates at which shocks to public goods are capitalized into land values. In this paper, we develop a unified framework that relates land value capitalization to the underlying concept of market equilibrium on which welfare measurement is based. The foundation for our analysis is Rosen's description of a differentiated product market with heterogeneous buyers and sellers. First we define the restrictions on preferences and technology that support a welfare interpretation for the rate at which an exogenous shock is capitalized into equilibrium prices. Then we translate those restrictions into testable conditions on micro data sets and on the design of quasi-experiments. Finally, we use the new framework to analyze the differences between: (i) hedonic estimates of the willingness to pay for improvements to public school quality from boundary discontinuity regressions in ten markets and (ii) capitalization rates for changes in test scores that occurred over the first four years of the federal No Child Left Behind program. We find that hedonic measures of the average resident's willingness to pay for improved school quality are four times as large as capitalization based measures.

[†] We thank Kelly Bishop, Richard Carson, Tom Downes, Roger von Haefen, Matt Kahn, Mike Keane, Glenn MacDonald, Gilbert Metcalf, Alvin Murphy, Ray Palmquist, Jonah Rockoff, Ed Schlee, V. Kerry Smith, Jeff Zabel, seminar participants at Arizona State University, Brigham Young University, Tufts University, University of Tennessee, Utah State University, Virginia Tech, Washington University in St. Louis, and conference participants at AEA and W2133 for valuable comments and suggestions on earlier drafts of this paper. We gratefully acknowledge support from the Lincoln Institute of Land Policy and the National Science Foundation.

I. Introduction

How can we measure the public's willingness to pay for a public good? This problem has intrigued economists for decades.[1] Recently, Chay and Greenstone (2005) proposed a novel solution—use quasi-experiments to identify the rates at which shocks to public goods are capitalized into land values. The appeal of combining a credible identification strategy with a welfare interpretation of the capitalization effect has led to a resurgence of interest in using markets for private property to assess the benefits of public programs.[2] Despite the growing importance of this methodology, the assumptions that enable us to translate capitalization effects into welfare measures have not been closely examined.

This paper uses the concept of hedonic equilibrium to investigate what land value capitalization reveals about the willingness to pay for public goods. We identify problems with using capitalization to measure willingness to pay, and we propose solutions to those problems. Our conceptual model builds on Rosen's (1974) description of the market for a differentiated product. We consider a market for housing where: (i) a house conveys a bundle of public and private goods; (ii) heterogeneous buyers and sellers make trades to maximize profits and utility; and (iii) equilibrium is described by a hedonic price function. Our point of departure from Rosen is to describe how the price function adjusts following an unexpected shock to a public good influencing the market equilibrium. Depending on the severity of the shock, adjustment may involve a movement along the hedonic price function or a change in its shape. We express the rate of change in equilibrium prices (i.e. the capitalization rate) in terms of the reduced form parameters of the price function which, in turn, depend on market primitives (preferences, income, and technology). This functional relationship reveals that, in general, capitalization rates do not identify the willingness to pay for public goods.

The scope for divergence between capitalization and welfare depends on the size of the

[1] Past proposals have included the median voter model (Bergstrom and Goodman 1973, Rubinfeld, Shapiro, and Roberts 1987) the conventional land value capitalization model (Lind 1973, Starrett 1981), the hedonic model of housing market equilibrium (Scotchmer 1985, 1986, Bartik 1987), and equilibrium sorting models of neighborhood choice (Epple and Sieg 1999, Bayer, Ferreira, and McMillan 2007).

[2] For examples see the recent quasi-experimental capitalization studies by Davis (2004), Chay and Greenstone (2005), Greenstone and Gallagher (2008), Linden and Rockoff (2008), Pope (2008), Bin, Landry, and Meyer (2009), Horsch and Lewis (2009), and Cellini, Ferreira, and Rothstein (2010).

shock and the duration of the study period. As both approach zero, the capitalization rate approaches the marginal willingness to pay (MWTP). In the limit, our model provides a conceptual foundation for Chay and Greenstone's (2005) estimator. As the size of the shock grows, so does the wedge between capitalization and welfare. The identification problem is intuitive. In a hedonic demand system, such as Epple (1987), a non-marginal shock to *any* attribute of a differentiated product will change the MWTP for *every* attribute. All of these changes are condensed into the same capitalization rate. To isolate the willingness to pay for a single attribute, more information is needed.

One way to provide the extra information is to place restrictions on the primitives of Rosen's model. Consider a non-marginal shock to a public good that is capitalized over an interval when the supply of housing is less than perfectly elastic. We prove three restrictions are both necessary and sufficient to interpret the capitalization rate as an exact measure of MWTP. First, preferences, income, and technology must be fixed over the duration of the study period. Second, utility must be separable in the public good and its demand curve must be perfectly elastic over the range of the shock. Third, the second derivative of the hedonic price function with respect to the public good must be zero. If any one of these restrictions is violated, capitalization rates may understate or overstate MWTP.

Restrictions on the primitives of Rosen's model have testable implications for the evolution of the hedonic price function. Using a linear-in-parameters specification for the price function, we derive conditions on the data under which capitalization rates will identify the average consumer's MWTP. One can identify MWTP in the pre-shock equilibrium if the hedonic gradient is constant over the duration of the study period. If this condition does not hold, one can still identify MWTP in the post-shock equilibrium if the shock (or an instrument for the shock) is orthogonal to all other variables. A key point is that randomization of an instrument can provide the extra information needed to identify post-shock MWTP in lieu of restrictions on market primitives.

In the second half of the paper, we apply our framework to the problem of measuring the willingness to pay for improving the quality of public schooling. We have assembled a unique set of micro data for the analysis. The data describe a quarter of a million individual

homes that sold in the cities and suburbs of Fairfax VA, Portland OR, Detroit MI, Los Angeles CA, and Philadelphia PA during the 2003 and 2007 school years.[3] Each observation includes the sale price of a home, its structural features, the demographic composition of its neighborhood, the local public goods available to its residents, and most importantly, measures of academic performance at the public schools to which children living in that home would have been assigned.

Most schools reported significant increases in their students' math and reading proficiency between 2003 and 2007, with the largest improvements reported by the lowest quality schools. These trends are consistent with the new incentives that school administrators faced after the No Child Left Behind Act (NCLB) took effect in 2003 (Dee and Jacob 2009, Neal and Schanzenbach forthcoming). NCLB required each state to implement a test-based accountability system for its public schools. Schools that repeatedly failed to meet targets for math and reading proficiency would face a schedule of sanctions. NCLB also required every school to publicly report the share of its students who achieve proficiency in each subject. We use these data to compare the willingness to pay for improved school quality with the capitalization of publicly reported changes in math and reading proficiency.

Our measures of willingness to pay are derived by estimating hedonic price functions in each of the ten (school year, metro area) pairings. The price functions are identified by boundary discontinuity designs that exploit the discreteness in each area's laws for assigning children to schools. Table 1 compares our main findings to the results from previous boundary discontinuity studies of Boston (Black 1999) and San Francisco (Bayer, Ferreira, and McMillan 2007). Hedonic estimates for the elasticity of property values with respect to test scores are remarkably similar across metro areas (column 1). In 2003 our estimates range from 0.12% in Fairfax to 0.27% in Philadelphia. Converting these estimates into constant year 2000 dollars reveals the average resident would be willing to pay between $422 (Detroit) and $743 (Philadelphia) for a 1% increase in test scores (column 3).

[3] After an exhaustive search over potential study regions, we concluded that Fairfax, Portland, Detroit, Los Angeles, and Philadelphia were the only metro areas with public school assignment laws, consistent reporting of test scores, and micro data on recent property sales that would allow us to develop hedonic boundary discontinuity designs at the standards set by Black (1999) and Bayer, Ferreira, and McMillan (2007).

When we repeat the boundary discontinuity analysis for 2007 we find significant changes in hedonic gradients. These changes drive a wedge between our hedonic measures of willingness to pay in 2003 and estimates based on the capitalization of changes in test scores between 2003 and 2007. Columns 3 and 4 illustrate that the two sets of estimates differ by more than 100% for the average resident in Fairfax, Portland, Detroit, and Philadelphia. Furthermore, correlation between changes in test scores and other variables drives a wedge between capitalization and willingness to pay in 2007. Aggregating our hedonic results over all five study areas reveals that the average resident's willingness to pay for a 1% increase in scores increased from $536 in 2003 to $688 in 2007. These figures are four times as large as our capitalization-based measures. We conclude that researchers must be cautious in using capitalization as the sole basis for evaluating the benefits of public programs.

Overall, our findings add to three distinct literatures. First, we define the connection between land value capitalization (Lind 1973, Starrett 1981) and hedonic equilibria (Scotchmer 1985, 1986, Bartik 1987) in the revealed preference literature on using private market outcomes to predict the willingness to pay for public goods. Second, we establish a conceptual framework for interpreting evidence on capitalization from the new quasi-experimental literature on policy evaluation (Davis 2004, Chay and Greenstone 2005, Greenstone and Gallagher 2008, Linden and Rockoff 2008, Cellini, Ferreira, and Rothstein, 2010). Finally, our boundary discontinuity results extend the literature on valuing school quality by providing the first consistent evidence on variation in willingness to pay across time and space (Oates 1969, Kain and Quigley 1975, Black 1999, Figlio and Lucas 2004, Bayer, Ferreira, and McMillan 2007).

The rest of the paper proceeds as follows. Section II briefly reviews the ideas behind hedonic and capitalization based approaches to valuing public goods. Section III presents our conceptual model and defines conditions under which capitalization rates identify willingness to pay. We translate those conditions into testable econometric restrictions in section IV. Section V describes our data, section VI reports regression results, and section VII analyzes implications for measuring willingness to pay. Finally, section VIII concludes by summarizing the problems with capitalization-based benefit measurement and the potential solutions.

II. Hedonic and Capitalization Models for Valuing Public Goods

In his seminal 1956 paper, Tiebout hypothesized that freely mobile households will reveal their preferences for public goods through the location choices they make. His reasoning influenced the development of two revealed preference techniques: the capitalization model and the hedonic property value model. Hundreds of applications of these methods over the past 40 years have contributed much of what we currently know about the willingness to pay for public goods.

Capitalization studies use data before and after a market shock to measure its effect on property values.[4] The power of this technique is the ability to simultaneously measure a change in asset values and demonstrate that the change was caused by some event. Capitalization models are routinely used by expert witnesses in litigation over private property externalities (Simons 2006). They are also used to measure the market value of risk and uncertainty (Brookshire et al. 1985). A limitation of the technique is that it lacks a precise welfare interpretation. Lind (1973) and Starrett (1981) demonstrated that, under the type of sorting behavior Tiebout envisioned, market capitalization of a large shock may understate or overstate the change in household welfare.[5]

In contrast, the hedonic property value model based on Rosen (1974) offers a theoretically consistent approach to welfare measurement. The difficulty is identification. Scotchmer (1985, 1986) proved that data from a single market are only sufficient to identify marginal values. To identify a demand curve, one must collect multi-market data on the characteristics of households and their houses, plus instrumental variables for endogenous characteristics (Bartik 1987, Epple 1987). Unfortunately, barriers to obtaining these data have stymied demand estimation.[6] The vast majority of empirical studies only aspire to recover marginal

[4] The idea for using panel data to measure how changes in quality characteristics influence housing prices dates back at least to Bailey, Muth, and Nourse (1963). Economic applications begin with Palmquist (1982).

[5] While Lind (1973) does not develop a formal utility theoretic framework, he proves that any welfare interpretation of capitalization requires there be zero consumer surplus. This effectively rules out preference-based sorting by heterogeneous agents, as Starrett (1981) later demonstrated.

[6] An alternative strategy to identify demand is to provide some information about the structure of consumer preferences. This information may consist of a parametric representation for the utility function (Epple and Sieg 1999, Bajari and Benkard 2005), separability restrictions on preferences (Ekeland, Heckman and Nesheim 2004), an assumption that the populations residing in different cities share a common distribution of

values.[7]

Even the seemingly modest task of estimating marginal values is now believed to be plagued by omitted variable bias. Chay and Greenstone (1998, 2005) characterized the problem and proposed a solution. They replaced the conventional hedonic estimator with an instrumental variables strategy that isolates how property values are affected by unexpected shocks to public goods. Their microeconometric model bridged the capitalization and hedonic literatures. It integrated a quasi-experimental version of the identification strategy from the capitalization literature with the welfare interpretation of Rosen's hedonic model.

To illustrate the basic idea, let the price of housing be expressed as $p = p(g, h, \xi)$, where g is the public good of interest, h measures all other public goods and housing characteristics observed by the analyst, and ξ represents unobserved variables. It is standard practice to specify a linear-in-parameters price function such as

$$p_1 = g_1\theta_1 + h_1\eta_1 + \varepsilon(\xi_1), \tag{1}$$

where the subscripts indicate the time period. Assuming the specification is correct, the first order conditions from Rosen (1974) allow us to interpret θ_1 as the marginal willingness to pay (MWTP) for the public good in period 1. However, θ_1 is not identified if ξ_1 is correlated with g_1 or h_1.

Now suppose p, g, and h are also measured after an unexpected shock. First-differencing the data produces a new estimator,

$$\Delta p = \Delta g\phi + \Delta h\gamma + \Delta\varepsilon, \tag{2}$$

where $\Delta d = d_2 - d_1$ for $d = [p, g, h, \varepsilon]$. If the bias from omitted variables is purged by differencing the data, (2) provides an unbiased estimator for ϕ. Alternatively, if one suspects that $E[\Delta\varepsilon \mid \Delta g, \Delta h] \neq 0$, instrumental variables may be used to develop a consistent estimator.

---

unobserved tastes (Bartik 1987), or an assumption that migration decisions do not arise from changes in individual tastes (Bishop and Timmins 2008).

[7] That said, the *number* of studies that aspire to recover marginal values is also vast. To give a rough sense of scale, there are more than 1600 citations of Rosen (1974) in the Social Science Citation Index and approximately 4000 reported by Google Scholar. Property value applications are one of (if not *the*) most frequent application. See Palmquist (2005) for a review of the literature.

Interpreted literally, ϕ is the average rate of change in property values associated with the shock to *g*. Chay and Greenstone observe that this capitalization rate will equal MWTP if the gradient of the price function in (1) is constant over time (i.e. $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$ implies $\phi = \theta_1 = \theta_2$ and $\gamma = \eta_1 = \eta_2$). Recent studies have used this result to estimate the willingness to pay for changes in cancer risk (Davis 2004), air quality (Chay and Greenstone, 2005), hazardous waste (Greenstone and Gallagher 2008), crime (Linden and Rockoff 2008, Pope 2008), open space (Bin, Landry, and Meyer 2009), invasive species (Horsch and Lewis 2009), low income housing credits (Baum-Snow and Marion 2009), and investment in education (Cellini, Ferreira, and Rothstein 2010). In all of these studies, the validity of welfare measures rests on the maintained assumption that the gradient of the hedonic price function is constant over the duration of the study. The assumption has been made for periods between 10 and 20 years, for areas ranging from a single county to the contiguous United States.

Because the price function is an equilibrium outcome generated by interactions between all of the buyers and sellers in a market, assumptions about its evolution implicitly restrict preferences and technology.

III. Hedonic Equilibria and the Capitalization of Market Shocks

This section considers the evolution of the hedonic gradient. After introducing the primitives of the hedonic model and characterizing market equilibrium, we define restrictions on preferences and technology that are sufficient to assure the gradient will be time-constant. A proof is followed by brief discussion.

A. Demand, Supply, and Market Equilibrium

Price-taking households are assumed to be free to choose a home with any combination of housing characteristics (e.g. bedrooms, bathrooms, sqft) in the neighborhood that provides their desired levels of amenities (e.g. school quality, air quality, racial composition). The utility maximization problem is

$$\max_{g,X,b} \; U(g, X, b; \alpha) \;\; \textit{subject to} \;\; y = b + P(g, X; \Theta), \tag{3}$$

where $X = [h, \xi]$. A household chooses housing characteristics, amenities, and the numeraire composite commodity (b) to maximize its utility, given its preferences (α), income (y), and the after-tax price of housing, $P(g, X; \Theta)$, which is expressed as a general parametric function of g, X, and a parameter vector, Θ. The first order conditions are

$$\frac{\partial P(g, X; \Theta)}{\partial g} = \frac{\partial U / \partial g}{\partial U / \partial b} \equiv D(g; X, \alpha, y), \tag{4a}$$

$$\frac{\partial P(g, X; \Theta)}{\partial X} = \frac{\partial U / \partial X}{\partial U / \partial b} \equiv R(X; g, \alpha, y). \tag{4b}$$

The first equality in (4a) implies that each household will choose a neighborhood that provides a quantity of g at which their marginal willingness-to-pay for an additional unit exactly equals its marginal implicit price. Assuming the marginal utility of income is constant, the second equality observes that as g varies the marginal rate of substitution defines the inverse demand curve, conditional on X. Equation (4b) states analogous first order conditions for X.

Producers in this market may include developers, contractors, and individuals selling their homes. Let $C(g, M, X; \beta)$ denote a producer's cost function, where M is the number of type-(g,X) homes they sell and β is a vector of parameters describing the producer.[8] Variation in β captures differences in costs faced by different producers. Following Rosen (1974), we treat each producer as a price taker who is free to vary the number of units they sell as well as a subset of the characteristics of each unit. For notational convenience, g is assumed to be exogenously determined.[9] In this case, the profit maximization problem is

[8] For a developer or contractor, the cost function will reflect the physical, labor, and regulatory costs of building a home. For a homeowner, the cost function will reflect their psychological attachment to the home as well as the cost of renovation.

[9] The results of this section are not altered by allowing firms to choose g or by restricting their ability to choose X. The key restriction needed to relate our model to the new empirical capitalization literature is that g may be influenced by forces that are exogenous to the model.

$$\max_{X,M} \ \pi = M \cdot P(g, X; \Theta) - C(g, M, X; \beta), \tag{5}$$

with the corresponding first order conditions

$$P(g, X; \Theta) = \frac{\partial C(g, M, X; \beta)}{\partial M}, \qquad \frac{\partial P(g, X; \Theta)}{\partial X} = \left(\frac{1}{M}\right)\frac{\partial C(g, M, X; \beta)}{\partial X}. \tag{6}$$

Producers choose M to set the offer price of the marginal home equal to its production costs, and they choose X to set the marginal per unit cost of each attribute equal to its implicit price.

Equilibrium occurs when the first order conditions in (4) and (6) are simultaneously satisfied for all households and producers. This system of differential equations implicitly defines the equilibrium hedonic price function that clears the market (Rosen 1974). It will be useful to rewrite the price function to acknowledge its dependence on model primitives,

$$P(g, X; \Theta) \equiv P[g, X(g, \mathrm{A}, \mathrm{B}); \Theta(g, \mathrm{A}, \mathrm{B})]. \tag{7}$$

Equilibrium levels of X are determined by all of the exogenous variables: g the public good of interest, $\mathrm{A}: F(y, \alpha) \sim \mathrm{A}$, a vector of parameters that describes the joint distribution of household income and preferences, and $\mathrm{B}: V(\beta) \sim \mathrm{B}$, a parameter vector describing the distribution of producer characteristics.[10] Naturally, the reduced form parameters describing the shape of the price function are also functions of the exogenous variables.

B. Necessary Conditions to Interpret the Capitalization Rate as a Measure of MWTP

Now consider two different hedonic equilibria, observed before and after an unexpected shock to g. The change in the value of a house j depends on the difference in the pre and post-shock price functions,

$$P\left[g_{2j}, X_{2j}(g_2, \mathrm{A}_2, \mathrm{B}_2); \Theta(g_2, \mathrm{A}_2, \mathrm{B}_2)\right] - P\left[g_{1j}, X_{1j}(g_1, \mathrm{A}_1, \mathrm{B}_1); \Theta(g_1, \mathrm{A}_1, \mathrm{B}_1)\right], \tag{8}$$

where the 1 and 2 subscripts denote pre and post-shock equilibria. To isolate the capitalization rate, we condition on X and divide the change in property value by the change in g,

[10] M drops out of the expression for X in (7) because it is a function of model primitives.

$$\phi_j = \frac{P\left[g_{2j};\Theta(g_2, \mathrm{A}_2, \mathrm{B}_2)\middle|X_{2j} = \overline{X}\right] - P\left[g_{1j};\Theta(g_1, \mathrm{A}_1, \mathrm{B}_1)\middle|X_{1j} = \overline{X}\right]}{g_{2j} - g_{1j}}. \tag{9}$$

This difference quotient provides a general expression for the capitalization parameter estimated in the literature.[11]

Because ϕ_j depends on two (potentially different) price functions, it is not the measure of MWTP from Rosen (1974). To interpret ϕ_j as the MWTP, we must restrict preferences and technology to assure that the capitalization rate will equal the partial derivative of the pre-shock and/or post-shock price functions. Severity of the restriction depends on the size of the shock. If the change in *g* is small, we need only restrict $\mathrm{A}_1 = \mathrm{A}_2$ and $\mathrm{B}_1 = \mathrm{B}_2$. Under this condition, the difference quotient in (9) approaches the partial derivative in (4a) as $g_{2j} - g_{1j}$ approaches zero.[12] In the limit, pre-shock MWTP equals post-shock MWTP which equals the capitalization rate. This is intuitive. An infinitesimal change in one hedonic characteristic will not alter the shape of the price function; equilibrium prices simply increase by MWTP.

In the case of a nonmarginal shock, three restrictions are needed to establish a welfare interpretation for the capitalization rate. We state this formally as

ASSUMPTION 1.

a. $\mathrm{A}_1 = \mathrm{A}_2$ and $\mathrm{B}_1 = \mathrm{B}_2$.

b. $\partial\Theta/\partial g = 0$.

c. $\partial P(g, X;\Theta)/\partial g = f(X, \Theta)$.

Condition *a* restricts preferences, income, and technology to be constant over the duration of the study.[13] It follows that supply and demand curves for each characteristic are also fixed. The last two conditions restrict the shapes of those curves. Condition *b* states that changes in *g* have no effect on the shape of the price function. If supply curves are less than perfectly

[11] *P* and *g* are typically measured in levels or logs.
[12] Proof of this statement follows immediately from the definition of a derivative.
[13] This condition can be relaxed as long as other restrictions are added to guarantee that changes in income, preferences, and technology have no effect on the shape of the equilibrium price function. More precisely, Θ must be invariant to any changes in the elements of A and B.

elastic, for example, condition *b* is satisfied if demand is perfectly elastic. Condition *c* further restricts supply and demand so that the marginal price function for *g* does not depend on *g*. If all three conditions are satisfied, it is straightforward to show that the capitalization rate in (9) must equal the MWTP in (4a).

THEOREM 1. If assumption 1 holds for a shock to *g*, then the capitalization rate, ϕ, equals the pre-shock MWTP, which equals the post-shock MWTP.

Proof. Consider any home, *j*, with characteristics $X_j = \overline{X}$ for which g_j changes from g_{j1} to g_{j2}. Since $\mathrm{A}_1 = \mathrm{A}_2$, $\mathrm{B}_1 = \mathrm{B}_2$, and $\partial\Theta/\partial g = 0$, we know that $\Theta_1 = \Theta_2$. Combining this result with the assumption that $\partial P(g, X; \Theta)/\partial g = f(X, \Theta)$ implies $f(\overline{X}, \Theta_1) = f(\overline{X}, \Theta_2)$. It follows from the Mean Value Theorem that $\phi_j = f(\overline{X}, \Theta_1) = f(\overline{X}, \Theta_2)$. The second term measures pre-shock MWTP and the third term measures post-shock MWTP, as defined by the first-order conditions from Rosen (1974). QED.

The model proposed by Chay and Greenstone (2005) provides an example. Their linear price function (1) is consistent with condition *c*, and their assumption that $\Theta_1 = \Theta_2$ implies conditions *a* and *b* are satisfied. Using these restrictions, it is a simple algebraic exercise to demonstrate that (9) returns the MWTP for *g*. Alternatively, for models that violate assumption 1 the Mean Value Theorem implies

$$\phi_j \neq \partial P(g_{1j}, X_{1j}; \Theta_1)/\partial g \neq \partial P(g_{2j}, X_{2j}; \Theta_2)/\partial g \,.$$

In this case, the direction and magnitude of the bias from misinterpreting the capitalization rate as a welfare measure will depend on the correlations in the data and the shapes of supply and demand curves.[14]

C. Discussion

We have established that the capitalization rate approaches the partial derivative of the price

[14] This dependence is easily demonstrated using a closed form expression for the equilibrium price function such as Tinbergen's (1959) linear-quadratic-normal model. Section IV demonstrates the role of correlations in the data using the standard empirical specification for the hedonic price function.

function as $A_2 \to A_1$, $B_2 \to B_1$, and $\Delta g \to 0$. Based on this limiting result, we would expect the capitalization rate to provide a good approximation to average MWTP for small shocks that can be tracked over short periods. However, recent studies have focused on large shocks and/or periods of a decade or more. For example, Chay and Greenstone (2005) measure the capitalization of large air quality improvements during the 1970s. Davis (2004) tracks the capitalization of a six-fold increase in pediatric leukemia risk. Greenstone and Gallagher (2008) estimate capitalization rates for the cleanup of hazardous waste sites over the first 20 years of the federal "Superfund" program (1980 to 2000).[15] We use these studies as examples for two reasons. First because they develop ingenious identification strategies to provide what are perhaps the most credible estimates for public good capitalization rates. Second because their exploitation of large shocks and/or long intervals means the ability to interpret their estimates as measures of average MWTP rests on the validity of assumption 1.

Consider what assumption 1 implies. At a single point in time, condition *c* requires the distribution of marginal prices for g to be degenerate once we condition on *X*. This is a special case of a linear marginal price function, which Ekeland, Heckman, and Nesheim (2004) prove is a nongeneric property of hedonic equilibrium. Even if we invoke the degeneracy restriction with the idea that it represents a linear approximation to the true price function, conditions *a* and *b* impose deeper restrictions on preferences and technology.

Recall that the hedonic gradient provides a mapping to the distribution of marginal values in the consumer population (4) and marginal costs in the producer population (6). The only theoretically-grounded restriction on this mapping that supports $\partial\Theta/\partial g = 0$ is that either the demand for *g* or its supply is perfectly elastic.[16] Utility must also be separable in *g* and *X*.

[15] Unlike the first two examples, the scope of the Superfund shock was small in the sense that only 1% of census tracts contained sites that were cleaned. The key assumptions that enable Greenstone and Gallagher to use the capitalization rate for benefit-cost analysis are that: (i) the MWTP for cleanup does not depend on the degree of contamination, and (ii) the hedonic gradient was invariant to all changes in public goods, housing characteristics, income, preferences, and construction costs that occurred in the United States over their 20-year study period.

[16] All else constant, a positive shock to *g* will decrease MWTP (changing Θ) if demand is downward sloping. It is possible to offset the change in Θ by a concomitant shock to preferences. While this type of mathematical restriction presents a technical possibility, it has no economic content and, in our opinion, does not merit serious consideration.

Otherwise, a shock to g could change the implicit prices of the elements of X. If g is the crime rate, for example, we must be willing to assume that changes in crime do not affect the willingness to pay for home security systems, fences, or proximity to city parks. These restrictions on own and cross-price elasticities are not limited to the public good of interest. They also apply to all of the elements of X that are subject to exogenous shocks. A change in the relative price of *any* hedonic characteristic violates $\partial\Theta/\partial g = 0$ and drives a wedge between MWTP and the capitalization rate for *any other* characteristic. Finally, even if the demand for every characteristic is perfectly elastic, we must still restrict their relative prices to be unaffected by changes in wealth, preferences, and construction costs that may occur during the study period.

If assumption 1 is violated, the gradient of the price function may change between the pre and post-shock observation periods. The bias associated with interpreting the capitalization rate as a measure of MWTP will depend on: (i) the shape of the price function; (ii) magnitudes of changes in the reduced-form parameters; and (iii) correlations in the data. Given a parametric representation for the price function, the capitalization bias can be expressed in terms of $\Theta_1, \Theta_2, g_1, g_2, X_1$, and X_2. We derive this relationship in the next section and use it to define testable restrictions on the data that neutralize the capitalization bias.

IV. Sufficient Conditions for Capitalization Based Welfare Measurement

Empirical studies almost always specify the price function to be linear in parameters.[17] We follow this convention and abstract from econometric complications such as measurement error and approximation error in the choice of functional form. These abstractions allow us to focus attention on the relationship between capitalization and welfare in the workhorse model of the empirical literature.[18]

We begin by repartitioning X into observed (h) and unobserved (ξ) components. Us-

[17] The prevalence of the linearity assumption is partly due to Cropper, Deck, and McConnell (1988). Working with simulated data, they found that linear approximations tended to provide better predictions for MWTP than a more flexible Box-Cox quadratic specification in the presence of unobserved variables and errors in variables.

[18] That said, one could repeat our analysis in this section under any set of assumptions about the shape of the price function and the sources of error. If the price function lacks a closed form solution, numerical methods could be used to solve for the equilibrium, as in Klaiber and Smith (2009) or Kuminoff and Jarrah (2010).

ing this partition, the linear price functions that describe market equilibria before and after an unexpected shock to g are $p_1 = g_1\theta_1 + h_1\eta_1 + \varepsilon_1(\xi_1)$ and $p_2 = g_2\theta_2 + h_2\eta_2 + \varepsilon_2(\xi_2)$.[19] Parameter subscripts recognize that the shape of the function may have been altered by the shock to g and by concomitant changes in h, ξ, $F(y,\alpha)$, and $V(\beta)$.

Subtracting the old price function from the new one yields a general time-differenced model,

$$\Delta P = (g_2\theta_2 - g_1\theta_1) + (H_2\eta_2 - H_1\eta_1) + \Delta\varepsilon . \tag{10}$$

In the special case where $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$, equation (10) reduces to the capitalization estimator from (2), $\Delta p = \Delta g\phi + \Delta h\gamma + \Delta\varepsilon$.

Applying the Frisch-Waugh Theorem, the relationship between the estimated capitalization rate ($\hat{\phi}$) and MWTP (θ_1, θ_2) can be expressed as:

$$\hat{\phi} = \theta_2 + \frac{r'g_1}{r'r}(\theta_2 - \theta_1) + \frac{r'h_1}{r'r}(\eta_2 - \eta_1) + \frac{r'\Delta\varepsilon}{r'r}, \tag{11}$$

where $r = \Delta g - \Delta h(\Delta h'\Delta h)^{-1}\Delta h'\Delta g$. The estimate for ϕ is a function of all the parameters of the true price functions that precede and follow the shock. Put differently, (11) reports what we can expect to learn about MWTP from estimating (2) when (10) is the true model.

The estimate for the capitalization rate is a function of ex-ante MWTP, ex-post MWTP, and correlations between housing characteristics. The second term on the right of the equality in (11) is a "price effect" that arises from a change in the implicit price of g between the initial equilibrium and the new equilibrium. The third term is a "substitution effect" that arises from changes in the implicit prices of other housing characteristics that affect utility and, in some sense, serve as substitutes for g. The last term reflects the bias that arises from correlation between changes in observed and unobserved variables.

Without any restrictions on the data, $\hat{\phi}$ may fall outside the range of values for MWTP defined by θ_1 and θ_2. Consider a quality improvement that decreases MWTP but has

[19] The h_1 matrix of control variables may also include a vector of ones so that η includes an intercept.

no effect on the control variables or their marginal implicit prices: $\Delta h = \eta_2 - \eta_1 = \Delta\varepsilon = 0$. In this case, (11) implies that $\hat{\phi} < \theta_2 < \theta_1$ if $\Delta g' g_1 > 0$. Alternatively, $\theta_2 < \theta_1 < \hat{\phi}$ if $\Delta g' \Delta g < -\Delta g' g_1$. It is clear that additional restrictions are needed to give the estimated capitalization rate a welfare interpretation.

Two sets of restrictions are sufficient for the capitalization model to provide an unbiased estimate of MWTP. The first set follows directly from assumption 1. If assumption 1 is satisfied, the hedonic gradient must be time-constant. Adding the usual orthogonality restriction on the error term gives us

SUFFICIENT CONDITION 1. $\theta_1 = \theta_2$, $\eta_1 = \eta_2$, and $\Delta g, \Delta h \perp \Delta\varepsilon$. (12)

Under these restrictions, equation (11) reduces to $\hat{\phi} = \theta_1 = \theta_2$. In this case the capitalization model (2) provides an unbiased estimate of ex ante MWTP which equals ex post MWTP. If estimation of single-period price functions is possible, time-constancy of the hedonic gradient can be tested.

The second set of restrictions relaxes the need for the gradient to be constant over time by adding orthogonality restrictions on the data that exploit the linearity of the model. More precisely, it can be seen from (11) that $\hat{\phi} = \theta_2$ if the following restrictions hold

SUFFICIENT CONDITION 2. $g_1, h_1, \Delta h \perp \Delta g$ and $\Delta g, \Delta h \perp \Delta\varepsilon$. (13)

In words: if the shock to g is orthogonal to its initial level, and to the initial levels of the control variables, and to changes in those variables, then the capitalization rate provides an unbiased estimate of MWTP in the post-shock equilibrium, even if the gradient changes between the two observation periods. The data may still contain capitalization bias. If $\theta_1 \neq \theta_2$ or $\eta_1 \neq \eta_2$ the price change for any given home may lie above or below the resident's ex post MWTP. However the positive and negative differentials for individual homes cancel out of the first-differenced estimate for ϕ due to the linearity of the price function and the orthogonality

of the shock.[20]

If instruments are available, sufficient conditions 1 and 2 can be relaxed. Some authors have sought to develop instruments for Δg out of concern for the potential correlation between changes in observed and unobserved variables. Notably, Chay and Greenstone (2005) and Greenstone and Gallagher (2008) use discontinuities in the structure of public policies to break potential correlation between Δg and $\Delta\varepsilon$. Equally important is the fact that these "policy discontinuity" instruments offer the potential to identify subsets of "treated" and "untreated" homes that are similar in many other respects. With this in mind, let z denote a set of valid instruments for Δg. The instrumental variables analog to the capitalization bias function in (11) simply replaces Δg with $\Delta\hat{g} = z(z'z)^{-1}z'\Delta g$. Likewise, (12) and (13) are replaced with (14) and (15).

SUFFICIENT CONDITION 1.a. $\theta_1 = \theta_2$, $\eta_1 = \eta_2$, and $z, \Delta h \perp \Delta\varepsilon$. (14)

SUFFICIENT CONDITION 2.a. $g_1, h_1, \Delta h \perp z$ and $z, \Delta h \perp \Delta\varepsilon$. (15)

Assuming valid instruments are available, it is straightforward to test whether they induce sufficient randomization to satisfy the orthogonality condition in the first part of (15). Also notice that z must not contain Δh. Adding the elements of Δh as control variables in a first-stage regression would violate the first orthogonality condition.

There is an important caveat to the randomization strategy justified by conditions 2 and 2.a. If the shape of the price function changes over the duration of the study, an accurate estimate of θ_2 may be of limited use for policy evaluation. Consider an extreme case where a large positive shock to the public good drives the MWTP to zero. The shock may have dramatically increased consumer welfare, but knowing θ_2 does not allow us to distinguish this outcome from the alternative hypothesis that people do not care about the change that occurred. More generally, a welfare approximation based on $\theta_2 \times \Delta g$ will understate the bene-

[20] Linearity and orthogonality are both necessary. In the context of assumption 1, we have strengthened condition *c* such that $\partial P(g, X; \Theta)/\partial g = f(\Theta)$. Under the original condition *c*, where the marginal price function may be nonlinear in *X*, orthogonality restrictions on the data may not be sufficient to identify θ_2.

fits from a ceteris paribus improvement during the study period and overstate the costs of a decline.

In summary, the relationship between capitalization and MWTP depends on the evolution of the price function gradient. If the gradient is found to be time-constant, (12) or (14) can be invoked to interpret capitalization rates as measures of ex-ante MWTP. If the gradient changes but the data satisfy the orthogonality conditions in (13) or (15), capitalization rates can be interpreted as measures of ex-post MWTP. Together, (12)-(15) define sufficient conditions for developing consistent welfare measures based on quasi-experimental estimates for capitalization effects. These conditions are analogous to Chetty's (2009) "sufficient statistics" for quasi-experimental welfare measurement.

To assess the practical importance of conceptual differences between capitalization and willingness to pay requires tracking how the hedonic gradient evolves over time. The difficulty lies in identifying single-period price functions. Perhaps the most credible identification strategies to date are the boundary discontinuity designs used to measure the willingness to pay for improving the quality of public schooling (Black 1999, Bayer, McMillan, and Reuben 2007). Therefore we focus the remainder of our attention on using this strategy to compare the capitalization of changes in school quality with the willingness to pay for improvements.

V. Capitalization of School Quality Changes and the MWTP for Improvements

Understanding the willingness to pay for school quality is crucial for determining the benefits of undertaking a wide range of academic reforms. Hedonic property value models offer the most intuitively appealing method. Because a household's access to a public school is determined by whether or not the household lives within the attendance zone for that school, property value differentials should reflect what parents are willing to pay for their children to attend schools where students score higher on standardized tests. A large empirical literature evolved around this idea, beginning with Oates (1969).[21]

The 40-year history of the literature on valuing school quality is a microcosm for the

[21] Kain and Quigley (1975) is another early example. Recent applications include Black (1999), Bogart and Cromwell (2000), Downes and Zabel (2002), Gibbons and Machin (2003), Reback (2005), and Bayer, Ferreira, and McMillan (2007). Figlio and Lucas (2004) is an example of a capitalization-based study.

broader literature on valuing public goods. Early studies used cross-section models with few or no controls for omitted variables. Then researchers noted a potential source of confounding—schools with higher test scores tended to be located in more exclusive neighborhoods. Subsequent studies sought to avert the potential bias by developing quasi-experimental identification strategies. This work began with Black (1999). She noticed that school quality shifts discretely as one crosses an attendance zone boundary, but other neighborhood characteristics do not (e.g. crime rates, air quality, access to the city center). Therefore, the composite price effect of all the unobserved amenities that are common to homes on both sides of a boundary can be absorbed by a fixed effect for the "boundary zone". By focusing on sales that occurred near a boundary and including fixed effects for each boundary zone, Black forced the identification to come from price differentials between structurally similar homes located on opposite sides of a boundary.

Bayer, Ferreira, and McMillan (2007) refined Black's approach to control for correlation between preferences for schools and preferences for the demographic characteristics of one's neighbors. The problem stems from sorting. If preferences for school quality are correlated with demographic characteristics, such as race or education, then similar "types" of households will tend to locate in the same attendance zones. This helps to explain why neighborhood demographics also tend to shift discretely as one crosses an attendance zone boundary. Since prospective homebuyers may care about the characteristics of their neighbors, one must control for changes in the demographic composition of the neighborhood in order to isolate the implicit value of academic performance.

We use the hedonic boundary discontinuity design developed by Black and refined by Bayer, Ferreira, and McMillan to identify single-period price functions in five metropolitan areas. Then we calculate the MWTP for school quality, test for time-constancy of the hedonic gradient, and compare our estimates for MWTP to capitalization rates for the changes in test scores that occurred during the first four years of the No Child Left Behind Act. The remainder of this section summarizes the Act, our data, and key features of the research design.

A. No Child Left Behind

President George W. Bush announced his "No Child Left Behind" framework for education reform three days after taking office, and within a year the NCLB act had been passed. NCLB was one of the most sweeping reforms in the recent history of public education in the United States. Since its enactment, states have been required to implement accountability systems that measure student performance in reading and math. Standardized testing is done in grades 3 through 8 and at least once during high school. State test scores are used to determine if each public school is making "Adequate Yearly Progress" (AYP) toward the goal of having 100% of its students attain state-specific standards for minimum competency in reading and mathematics by 2014. Schools that do not meet AYP face a series of repercussions.

Importantly, NCLB established a consistent set of metrics for comparing academic performance across schools and improved accessibility of the information. To obtain a ranking of schools in their area or to see specific test scores, parents need only visit one of several websites that collect the information and distribute it freely.[22] A low cost of obtaining information should strengthen the link between property values and the willingness to pay for higher academic performance.

While test scores have trended up since NCLB was enacted, its impact on the quality of education has been debated. Advocates argue that school quality will be improved by developing consistent metrics for tracking school performance, publicizing results, and sanctioning schools that fail to meet APY. Detractors argue that NCLB creates perverse incentives to "teach to the test", to lower state standards, to expel poorly performing students, or even to lie when reporting scores. Several authors have investigated these issues. Perhaps the most convincing analyses are those by Neal and Schanzenbach (forthcoming) and Dee and Jacob (2009). These papers also provide excellent reviews of the literature. Neal and Schanzenbach show that the introduction of NCLB increased reading and math scores for students in the middle of the achievement distribution for fifth graders in the Chicago Public School system. Dee and Jacob (2009) employ a comparative interrupted time series design to identify the impact of NCLB on a panel of state test scores from the National Assessment of Education Progress (NAEP). The key feature of their research design is that changes in NAEP scores

[22] See for example the popular websites: www.greatschools.org and www.schooldigger.com .

should be unaffected by the perverse incentives that critics of NCLB have emphasized. They found that NCLB did indeed cause large and broad gains in NAEP math achievement scores of 4th and 8th graders, especially in the bottom decile of the achievement distribution.[23] These results suggest that the upward trend in NCLB scores is consistent with alternative metrics for judging public school quality.

B. Ten Boundary Discontinuity Designs

There have been few applications of the boundary discontinuity methodology to study the willingness to pay for school quality, and the vintage of data used by Black and Bayer, Ferreira, and McMillan was early to mid 1990's. We significantly update and extend the literature by applying the methodology to 10 new markets: 5 geographic regions (Fairfax County VA, Portland OR, Philadelphia PA, Detroit MI, and Los Angeles CA) in 2 distinct periods (the 2003 and 2007 school years). After an exhaustive search over prospective study regions, these five areas were chosen because they each satisfied three key criteria: (i) a sufficient number of boundary zones to conduct the estimation;[24] (ii) a sufficient number of housing transactions available for estimation; and (iii) NCLB test scores were reported for the 2003 and 2007 school years.[25] Together, the five regions also provide considerable geographic diversity.

Black (1999) and Bayer, Ferreira, and McMillan (2007) used elementary school attendance zones as the basis for identification. We use this same approach in Fairfax and Portland, where children are still assigned to elementary schools based on the attendance zones where their parents live. However, this type of school-specific assignment is no longer the norm. Since the mid-1990s, there has been an explosion of state and local regulations that mandate open enrollment at the school district level. In an open enrollment area, parents are free to send their children to any public school that lies within the school district. There is evidence that parents take advantage of these laws by sending their children to schools outside the ele-

[23] Mean increases in the National Assessment of Educational Progress math test scores were approximately 1-8 points from the start of NCLB to 2007 for 4th and 8th grade math scores.
[24] Boundary discontinuity analysis is extremely data-intensive because it discards housing transactions that occur beyond small distances from the school district boundaries.
[25] States were not required to start reporting test scores until 2006 and so many states did not have test score data available early enough for the analysis.

mentary attendance zone where their home is located (Reback 2005, 2008). Philadelphia, Detroit, and Los Angeles all have open enrollment policies. For these areas, our identification strategy is based on the relationship between property values and average test scores on opposite sides of the school district boundary.

Implementing the boundary discontinuity design at the school district level requires taking a weighted average over the test scores in each district. This has the advantage of smoothing over idiosyncratic variability in annual school-specific scores. Yet, it also requires extra caution. Property tax rates often vary discretely across school districts, and district boundaries may be more likely than attendance zone boundaries to overlap with features of the landscape. Therefore, we are careful to control for property tax rates and to drop all district boundaries that overlap with discernable landscape features such as rivers and highways.

C. Data and Summary Statistics

We collected detailed information on test scores, neighborhood demographics, and homes that were sold during the 2003 and 2007 school years. The 2003 school year is defined as October 1, 2003 through September 30, 2004, and the 2007 school year is defined as October 1, 2007 through September 30, 2008.[26] The test scores that we use are combined rates of math and reading proficiency reported by states under NCLB. Scores are reported at the school and school district levels. We matched each housing sale with lagged test scores for the relevant school or school district. Homes that sold during the 2003 school year were matched with math/reading proficiency scores from the 2002 school year, for example. We will simply refer to the lagged scores as the "2003 score" and "2007 score" from here on.[27] Unlike the NAEP data used by Dee and Jacob (2009), these scores are not directly comparable across states. We use the state-specific NCLB scores because they capture variation within metro areas and they

[26] These definitions for the school year were chosen because the NCLB test scores and school grades for the preceding school year are typically announced at the end of August or the beginning of September. Thus we want to allow time for our proxy for school quality—test scores—to matter in the home buying decisions.

[27] The school quality information was obtained from www.schooldatadirect.org. The combined measure of reading and math is an overall measure (calculated by Standard & Poor's) that provides an average of the proficiency rates achieved across all reading and math tests, weighted by the number of tests taken for each school or school district.

contain the same information that is readily available to prospective homebuyers.

Table 2 reports the 2003 baseline NCLB test scores and 2007-2003 differences for the 10th, 50th, and 90th percentiles of schools within each study area. In Fairfax, for example, math/reading scores in the bottom 10th percentile of schools increased by an average of 11 points (or 14%) with a standard deviation of approximately 8 points. The corresponding changes for the other four metro areas are all positive and typically large. There are smaller gains (and even losses) at the middle and 90th percentiles. These statistics are consistent with Dee and Jacob's (2009) finding that NCLB had the biggest impact on schools that began the program with the lowest scores.

The remaining components of the data were collected from various sources. Sale prices and structural characteristics of every home sold during the 2003 and 2007 school years were purchased from a commercial vendor that assembles the data from public records maintained in the county/counties that comprise each study region. Tax rates were calculated using tax assessment data also available from public records. All other neighborhood characteristics were collected at the Census block group level, using annual data from Geolytics.[28] These block group data were spatially merged to each home.

Table 3 provides summary statistics for the complete set of data from our Fairfax County, VA sample. Columns 1-2 report means and standard deviations of all the variables used in the hedonic and capitalization regressions. In 2003 the average home sold for approximately $567,000 but by 2007 the price had dropped slightly to $563,000. Over this same period, the average test score rose from 83.56 to 84.36.[29] This seemingly small change masks considerable heterogeneity across individual schools (table 2). The average home was 34 years old, with 4 bedrooms, 3 baths, and 2,100 square feet of living area on a 0.4 acre lot. It was located in a block group where 23% of the neighborhood was nonwhite, 24% was under

[28] Geolytics combines demographic information from the decennial Census with postal records and actuarial tables of births and deaths to develop an annual series for neighborhood demographics of Census block groups.

[29] It should be noted that the mean for the 2003 score levels is slightly different than the 2003 score level reported in Table 2 and the corresponding appendix tables. This is because Table 3 scores are weighted by enrollment whereas Table 2 is weighted by housing transactions. In other words, the difference represents the fact that the spatial distribution of housing transactions is not the same as the spatial distribution of enrollments.

18 years of age, 85% of homes were owner occupied, 1% of homes were vacant, and 0.37 was the normalized measure of population density. The average ratio of assessed to taxed value called a "tax rate" in this area was 112.

Columns 3-5 summarize the sub sample used in the boundary discontinuity analysis. Column 3 reports means over sales of homes located within 0.2 miles of a boundary. While this cuts the sample in half, the characteristics of the average home are virtually the same as in the full sample (column 1). Column 4 reports the difference in mean characteristics of homes located on the "high score" and "low score" sides of a boundary, and column 5 reports T-statistics on the differences. Differences in test scores are large and statistically significant whereas differences in housing characteristics are mostly small and insignificant. Like Bayer, Ferreira, and McMillan (2007), we find significant differences in the racial composition of homeowners on the high and low-score sides of a boundary. This underscores the importance of controlling for demographic characteristics during the estimation.

Columns 6-7 report means and standard deviations for the average home in each Census block group. These are the data we use to estimate the capitalization rate for changes in test scores between 2003 and 2007.[30] Notice that aggregation does not substantially change the summary statistics relative to the micro data. Finally, columns 8-9 report correlations between the change in test scores and levels and changes in all other variables. The orthogonality condition in (13) is clearly violated.

The Fairfax county data illustrate several features that are common to the data sets for Portland, Philadelphia, Detroit, and Los Angeles. In particular: (i) variable means are very similar across the full micro, 0.2 mile micro, and block group samples in each metro area; (ii) test scores and racial composition both tend to change discretely across the boundary zones; (iii) changes in test scores are negatively correlated with the baseline level of test scores; and (iv) changes in test scores are generally correlated with levels and changes in other housing characteristics. Complete summary statistics for each metro area are reported in the appendix.

[30] There are insufficient repeated sales of individual homes to implement a micro data analysis as in Davis (2004). Relative to our block-group averages, other recent capitalization studies have used more aggregate data such as census tracts medians or county averages (e.g. Chay and Greenstone 2005, Greenstone and Gallagher 2008, Baum-Snow and Marion 2009).

VI. Results

A. Single-Period Hedonic Regressions

Our hedonic estimates of the MWTP for school quality are based on the following specification for the price function:

$$\ln(P) = testscore\,\theta_{03} + D \cdot testscore\,\theta_{07} + h\eta_{03} + D \cdot h\eta_{07} + BFE_{03} + BFE_{07} + \varepsilon\,. \quad (16)$$

$testscore$ denotes the log of math and reading proficiencies for the year prior to the housing sale, D is an indicator for sales that occurred in the 2007 school year, h includes all structural housing characteristics, neighborhood demographic variables, and the tax rate, and BFE_{03}, BFE_{07} are boundary fixed effects in 2003 and 2007. The boundary regions are 0.2 mile areas that overlap adjacent school attendance zones (Fairfax, Portland) or adjacent school districts (Philadelphia, Detroit, Los Angeles).[31] Under the null hypothesis that the hedonic gradient is constant over the duration of the study, $\theta_{07} = \eta_{07} = 0$.

We begin by using the sample of homes that sold within 0.2 miles of a boundary to assess the bias from omitted variables. Panels A and B of table 4 report the OLS estimates of θ_{03} and θ_{07} from regressions with and without boundary fixed effects. Since test scores are measured in logs, their coefficients are elasticities. For example, the results in column 2 indicate that the prices of homes sold in Portland during 2003 were approximately 0.456% higher in school attendance zones where math and reading proficiency was 1% higher. This price elasticity is virtually the same when we compare school districts in Philadelphia in column 3. Notice that Philadelphia is also one of four metro areas to have a significant increase in the price elasticity over the first four years of the NCLB program. It increased from 0.481 in 2003 to 0.710 in 2007 (i.e. 0.481 + 0.229). Overall, panel A provides tentative evidence that NCLB test scores matter for property values and that the functional relationship between them

[31] We found similar results for boundary regions of 0.35 and 0.15 miles.

changed over the duration of our study.

The evidence in panel A is tentative because we have not controlled for possible correlation between school quality and unobserved amenities. Positive correlation seems likely to arise from the sorting mechanism that underlies hedonic equilibrium. The intuition for this mechanism begins with the observation that household income is a strong predictor of a child's academic performance.[32] With this in mind, consider the household's location choice problem. If homebuyers appreciate low crime rates, access to parks, and scenic views, they will bid up housing prices in the neighborhoods that provide those (and other) amenities. Wealthier parents who can afford to live in the higher-amenity neighborhoods will have children who tend to perform better on standardized tests. Therefore, the inability to control for crime, parks, and views will produce an upward bias on the OLS estimator for the test score coefficient. The boundary fixed effects address this problem by absorbing the average price effect of unobserved amenities in the regions between adjacent school districts or adjacent attendance zones, allowing us to isolate the property value effect of higher test scores.[33]

Panel B reports the regression results after adding boundary fixed effects. In each metro area the coefficient of variation increases and the test score coefficients decrease, consistent with intuition.[34] A quick comparison between panels A and B confirms that omitted variables are a serious problem. They inflate most of the test score elasticities by more than 100%!

Raw test scores are not directly comparable across states because each state develops its own standardized tests. Nevertheless, since the state-specific scores represent different proxy measures of the same underlying variable—school quality—they can be compared in

---

[32] Correlation between household income and academic performance reflects a web of interaction between several underlying factors. Income is correlated with parental education and ability which, in turn, may help to explain the quality of the early parenting environment. Income is also correlated with the education and ability of the parents' of the child's peers, and so on. While positive correlation between income and test scores is sufficient to develop intuition for the endogeneity problem in our model, understanding the underlying causal mechanisms is critical to the development of effective education policies. See Heckman (2008) for a summary of the evidence.

[33] For more background on this identification strategy see Black (1999) and Bayer, Ferreira, and McMillan (2007).

[34] The impact on the test score coefficients of including the boundary fixed effects is quite similar (in percentage terms) to the results reported by Black (1999) and Bayer, Ferreira, and McMillan (2007). Coefficients on the control variables are generally consistent across metro areas with the usual signs and plausible magnitudes. Like Bayer, Ferreira, and McMillan we find that, more often than not, inclusion of the boundary fixed effects decreases the magnitudes of the coefficients on neighborhood demographics.

terms of a common proportionate change. The test score elasticities in columns 6-10 are remarkably similar across the five metro areas in 2003. They suggest a 1% increase in math and reading proficiency would increase property values by 0.12% to 0.27%. In comparison, Black's (1999) preferred specification indicates an increase of approximately 0.42% for Boston suburbs in 1993-1995 and the results from Bayer, Ferreira, and McMillan (2007) indicate an increase of approximately 0.12% for the San Francisco metro area in 1990.

In 2007 our range of point estimates for the test score elasticity is considerably wider: 0.04 to 0.57. The changes are large and significant for Fairfax, Portland, Detroit, and Los Angeles. Several factors may be contributing to the changes in elasticities between 2003 and 2007. These include: (i) changes in NCLB test scores; (ii) changes in wealth; (iii) the information shock created by the new format for reporting test scores under the NCLB program; (iv) changes in neighborhood demographics; (v) changes in other housing characteristics that serve as substitutes or complements for school quality; and (vi) changes in the stock of housing. Parsing out the relative importance of these effects would require estimating the demand curve for school quality. While demand estimation is beyond the scope of this study, we conjecture that changes in the hedonic gradient may provide the extra information needed to overcome past problems with identification. An explanation is saved for section 8. Until then, we continue to focus on the relationship between marginal effects in the capitalization and hedonic models.

B. Capitalization Regressions and Robustness Checks

Large changes in the hedonic test score coefficients provide the first signal that capitalization rates are unlikely to identify MWTP. A second indication is the fact that changes in other coefficients are large enough to reject the hypothesis of a time-constant gradient in each metro area (F-tests are reported in panel B). Since we lack a randomized instrument for the change in test scores, there is little hope for circumventing capitalization bias. Measures of correlation in tables 2 and 3 (and appendix tables 1-4) reveal that the orthogonality conditions in (13) are systematically violated. For example, the changes in NCLB test scores for schools in Fairfax are positively correlated with some neighborhood characteristics (e.g. percent nonwhite resi-

dents in 2003, percent renting in 2003, population density in 2003) and negatively correlated with others (e.g. NCLB score in 2003, tax rate, change in percent nonwhite). Thus, it comes as no surprise that the capitalization-based estimates for the test score elasticity in panel C of table 4 look very different from their hedonic counterparts in panel B.

The results in panel C were generated by OLS estimation of the first-differenced capitalization model using the full sample of block groups. Notice that Los Angeles is the only place where the capitalization rate (0.17%) lies within the range defined by the price function parameters from 2003 and 2007 (0.14% to 0.22%). In Fairfax, Portland, Philadelphia, and Detroit, our capitalization-based estimates for the test score elasticity are far below the lower bound of point estimates from the hedonic model. The capitalization rate is at least positive and marginally significant in Philadelphia. In Fairfax and Portland the downward bias is so large that capitalization rates would imply the willingness to pay for improved school quality is essentially zero. In Detroit the capitalization rate is negative and marginally significant. This could simply reflect approximation error in the linear form of the estimating equation, but the hedonic estimates in column 9 are quite reasonable by contrast.

We consider three alternative explanations for the large differences between our estimates for capitalization rates and hedonic parameters: sample selection, data aggregation, and unobserved shocks that may be correlated with the change in school quality. Table 5 reports the results from indirect tests of each hypothesis.

First consider the scope for sample selection bias. Houses located outside the 0.2 mile boundary zones are included in the capitalization model but excluded from the hedonic regressions. The excluded homes comprise a large share of total housing sales in each metro area, from 35% in Portland to 92% in Los Angeles. Differences between the capitalization and hedonic results could arise from differences in the distribution of properties located in the excluded and included areas. To test this possibility, we repeat estimation of the basic hedonic model (without boundary fixed effects) using all of the micro data that were used to construct the block group averages for the capitalization model. Results are reported in columns 1-5 of table 5. They essentially mirror the original hedonic estimates from columns 1-5 of table 4. Given the large sample sizes, it is remarkable that only two of the ten coefficients are statisti-

cally different (Fairfax and Detroit in 2003). From this we conclude that sample selection is unlikely to explain the differences between our baseline results from the hedonic and capitalization models.

A second possibility is that the capitalization results are driven by aggregation bias that arises from averaging the micro data over Census block groups. The issue is that the "average" home in a given block group need not correspond to any point on the hedonic price surface. It is difficult to predict the direction and magnitude of the resulting bias. Past studies that have used Census aggregates have assumed the bias is sufficiently small to ignore (Chay and Greenstone 2005, Greenstone and Gallagher 2008, Baum-Snow and Marion 2009). To test this assumption, we aggregate the micro data from panel A into block groups and repeat the estimation. Results are reported in panel B. Comparing the two panels reveals that aggregation does not affect the general pattern of results. The magnitudes of the coefficients do change a bit, but the differences are mostly insignificant.

Finally, our estimates for the capitalization rate could be confounded by omitted variables. If changes in unobserved amenities are negatively correlated with changes in school quality, the first-differenced estimator will be biased downward. To test this possibility we extend the boundary discontinuity identification strategy to a panel data setting. First, we drop all houses that do not fall within 0.2 miles of a boundary. Then we aggregate the micro data into "boundary neighborhoods" on either side of each boundary. Finally, we add fixed effects for each boundary and estimate the resulting first-differenced model,

$$\Delta \ln(P) = \Delta testscore\, \phi + \Delta h \gamma + \Delta BFE + \Delta \varepsilon . \tag{17}$$

These "boundary difference fixed effects" will absorb the capitalization of changes in unobserved amenities in each boundary region. Results are reported in panel C. While standard errors on the elasticities are quite large due to the decrease in sample size and the inclusion of fixed effects, the point estimates are remarkably similar to our baseline results. The point estimates in columns 11-15 of table 5 all fall within the 95% confidence intervals on the corresponding estimates from columns 11-15 of table 4. Thus, omitted variables do not provide much help in explaining why our baseline estimates for the capitalization rate are so much lower than the elasticities from the hedonic model. We are left to conclude that the differences

we observe are primarily due to changes in the gradient of the hedonic price function.

VII. Summary and Implications of Empirical Results

The results from our boundary discontinuity regressions demonstrate that hedonic gradients can change significantly over a short period of time. We are not the first to document this type of instability. Costa and Kahn (2003) found that the implicit price of living in a metropolitan area with a temperate climate doubled between 1970 and 1980, and then doubled again between 1980 and 1990. In an application that more closely resembles ours, Brookshire et al. (1985) found that a discrete shock to information about earthquake risk changed the hedonic gradient over a 6-year period. Other researchers have reported annual changes in the relative prices of structural housing characteristics (Meese and Wallace 1997, Murphy 2007). However, all of these studies are vulnerable to the usual concern about confounding from omitted variables. By using a sharp discontinuity to mitigate the omitted variable pitfall, we have provided the strongest evidence to date that hedonic gradients do change.

We also find that changes in the hedonic gradient matter for evaluating the benefits of public education. Table 6 provides a summary comparison between our hedonic and capitalization based estimates for the average resident's willingness to pay for a 1% increase in test scores. Each column reports the MWTP predicted by a specific econometric model, averaged over the samples from all five study regions. In columns 1-3 we do not control for omitted variables. The resulting predictions for MWTP are fairly robust to how we define a data point (house, Census block group) and how we define the extent of the market (full metro area, 0.2 mile boundary zone). However, these predictions are twice as large as the ones from the model with boundary fixed effects (column 4). This reinforces past evidence on the potential for omitted variables to confound the results from property value studies (Black 1999, Chay and Grenstone 2005, Bayer, Ferreira, and McMillan 2007).

The hedonic boundary discontinuity design in column 4 is our preferred specification. It controls for omitted variables; it controls for sorting across boundaries on the basis of race; and it provides a theoretically consistent prediction for MWTP at a single point in time. It implies the average household would have been willing to pay $536 for a 1% improvement in

school quality during the first year of the NCLB program.[35] Over the next four years, there were several changes. Property values increased by 6% on average, test scores increased by 10%, and there were smaller changes in the demographic composition of neighborhoods. There was also steady media coverage of the NCLB program and changes in the broader economy that would have affected expectations about permanent income (e.g. rapid growth in stock market indices and personal income per capita). These changes were accompanied by changes in hedonic gradients which, in turn, increased our prediction for average MWTP to $688 for the 2007 school year.

Relative to our hedonic model, capitalization rates severely understate the willingness to pay for academic performance. Column 5 reports the average MWTP predicted by our first-differenced capitalization model ($134 in 2003, $152 in 2007).[36] These figures are about ¼ the size of estimates from our hedonic boundary discontinuity regressions! The difference between hedonic MWTP and capitalization rates only narrows slightly when we add controls for time-varying omitted variables to the capitalization regressions (column 6). Placing these results in the context of our conceptual framework suggests that researchers must be cautious in using capitalization rates as the basis for evaluating the benefits of public programs.

VIII. Conclusions

The hedonic property value model and the land value capitalization model are typically viewed as separate frameworks. We have sought to connect them. By extending Rosen (1974) to describe how the equilibrium price function adjusts to changes in the supply of a public good, we were able to express market capitalization as a function of hedonic willingness to pay. This unified framework provides a welfare theoretic basis for interpreting evidence on capitalization rates for shocks to public goods.

Our conceptual model produced three insights into the relationship between capitaliza-

[35] This average reflects variation across metro areas, from a low of $422 in Detroit to a high of $743 in Philadelphia. Interestingly, the area with the highest average MWTP, Philadelphia, also ranked third among all U.S. cities in terms of the volume of Google searches on the phrase "No Child Left Behind" in 2004. The top two cities were Pittsburg and Washington D.C.

[36] These figures were calculated by combining results from columns 11-15 in table 4 with data on average property values and populations in tables 3 and A1-A4.

tion and MWTP. First, the scope for divergence between the two concepts grows with the size of the shock and the length of the study period. As both dimensions approach zero, the capitalization rate approaches MWTP. Second, if we want to guarantee that ex-ante MWTP is recoverable from the capitalization of a non-marginal shock, we must add further assumptions about preferences and technology to Rosen's model. These new assumptions have a testable implication. They imply the hedonic gradient will be constant over time. Finally, if the hedonic gradient changes over time we can still recover ex-post MWTP as long as the price function is linear in parameters and the shock (or an instrument for the shock) is randomized.

In the application to school quality, the average difference between capitalization and MWTP was quite large. To recover MWTP we developed the most comprehensive set of estimates to date on the contribution of academic performance to residential property values, using hedonic boundary discontinuity designs to control for omitted variables. By analyzing five metro areas at two points in time, we were able to generate ten separate estimates for the elasticity of property values with respect to test scores. We found that these hedonic gradients changed over time. As a result, our estimates for MWTP were three to four times as large as capitalization rates for changes in test scores.

More generally, our framework can guide future research on valuing public goods by illustrating how to overcome problems with capitalization-based welfare measurement. The simplest solution is to avoid interpreting capitalization rates as measures of willingness to pay unless the data make it possible to track small shocks over brief intervals. If the goal is to recover ex-post MWTP, a second solution is to find an instrument that randomizes the intensity of the public good "treatment". The instrument must be orthogonal to baseline levels of every control variable and to changes in those variables. This is a tall order which no instrument is likely to satisfy completely. But some natural experiments and policy discontinuities may come close. The validity of candidate instruments can be judged from the coefficients on the "price effect" and "substitution effect" terms in our expression for the capitalization bias (equation 11). For example, if $|r'g_1/r'r| < 0.01$ the bias in estimated MWTP that arises from a change in the implicit price of g will be less than 1% of the change that occurred.

Finally, we conjecture that the same market forces that drive a wedge between capitali-

zation and MWTP also have the potential to help us overcome the classic problem with identifying demand curves. The problem is that the equilibrium price function intersects each household's demand curve at exactly one point (Epple 1987; Bartik 1987). To identify the rest of the curve from market data, we must observe similar households making choices along a different price surface. Pooling data from different geographic markets, while possible, raises concerns with selection bias (Rubinfeld, Shapiro, and Roberts 1987). Our conceptual model suggests a different solution. We have demonstrated that large shocks to public goods can change the hedonic price surface in a single geographic market. Thus, it may be possible to identify demand curves from repeated cross-sections of households collected before and after a shock to the distribution of public goods supplied in a single metro area. We view this quasi-experimental approach to hedonic demand estimation as a promising direction for future research.

REFERENCES

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse. 1963. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association*, 58(304): 933-42.

Bajari, Patrick and C. Lanier Benkard. 2005. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *Journal of Political Economy*, 113(6): 1239-76.

Bartik, Timothy J. 1987. "The Estimation of Demand Parameters in Hedonic Price Models." *Journal of Political Economy* 95(1): 81-88.

Baum-Snow, Nathaniel, and Justin Marion. 2009. "The Effects of Low Income Housing Tax Credit Developments on Neighborhoods." *Journal of Public Economics* 93(5-6): 654-66.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588-638.

Bergstrom, Theodore C., and Robert P. Goodman. 1973. "Private Demands for Public Goods." *American Economic Review* 63(3): 280-96.

Bin, Okmyung, Craig E. Landry, and Gregory E. Meyer. 2009. "Riparian Buffers and Hedonic Prices: A Quasi-Experimental Analysis of Residential Property Values in the Neuse River Basin." *American Journal of Agricultural Economics* 91(4): 1067-79.

Bishop, Kelly, and Christopher Timmins. 2008. "Simple, Consistent Estimation of the Marginal Willingness to Pay Function: Recovering Rosen's Second Stage without Instrumental Variables." *Mimeo*.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 114(2): 577-99.

Bogart William T., and Brian A. Cromwell. 2000. "How much is a neighborhood school worth?" *Journal of Urban Economics* 47: 280-305.

Brookshire, David S., Mark A. Thayer, John Tschirhart, and William D. Schulze. 1985. "A Test of the Expected Utility Model: Evidence from Earthquake Risks." *Journal of Political Economy*, 93(2): 369-89.

Cellini, Stephanie Riegg, Fernando Ferreira, and Rothstein Jesse. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *Quarterly Journal of Economics*, forthcoming.

Chay, Kenneth Y., and Michael Greenstone. 1998. "Does Air Quality Matter? Evidence from the Housing Market." *NBER Working Paper #6826*.

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

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics* 1: 451-488.

Costa, Dora L and Matthew E. Kahn. 2003. "The Rising Price of Nonmarket Goods." *American Economic Review*, 93(2): 227-32.

Cropper, Maureen L., Leland B. Deck, and Kenenth E. McConnell. 1988. "On the Choice of Functional Form for Hedonic Price Functions." *Review of Economics and Statistics*, 70(4): 668-75.

Davis, Lucas. 2004. "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster." *American Economic Review*, 94(5): 1693-704.

Dee, Thomas, and Brian Jacob. 2009. "The Impact of No Child Left Behind on Student Achievement." *NBER Working Paper 15531*.

Downes, Thomas A., and Jeffrey E. Zabel. 2002. "The Impact of School Characteristics on House Prices: Chicago 1987-1991." *Journal of Urban Economics* 52: 1-25.

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

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy*, 95(1): 59-80.

Epple, Dennis and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdiction." *Journal of Political Economy*, 107(4): 645-81.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94(3): 591-604.

Gibbons, Steve., and Stephen Machin. 2003. "Valuing English Primary Schools." *Journal of Urban Economics* 53: 197-219.

Greenstone, Michael and Justin Gallagher. 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *Quarterly Journal of Economics*, 123(3): 951-1003.

Heckman, James J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46(3): 289-324.

Horsch, Eric J., and David J. Lewis. 2009. "The Effects of Aquatic Invasive Species on Property Values: Evidence from a Quasi-Experiment." *Land Economics* 85(3): 391-409.

Kain, John F., and John M. Quigley. 1975. "Housing Markets and Racial Discrimination." New York: National Bureau of Economic Research.

Klaiber, Allen H., and V. Kerry Smith. 2009. "Evaluating Rubin's Causal Model for Measuring the Capitalization of Environmental Amenities". *NBER Working Paper 14957*.

Kuminoff, Nicolai V. and Abdul S. Jarrah. 2010. "A New Approach to Computing Hedonic Equilibria and Investigating the Properties of Locational Sorting Models." *Journal of Urban Economics*, 67(3): 322-335.

Linden, Leigh and Jonah E. Rockoff. 2008. "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws." *American Economic Review*, 98(3): 1103-27.

Lind, Robert C. 1973. "Spatial Equilibrium, the Theory of Rents, and the Measurement of Benefits from Public Programs." *Quarterly Journal of Economics*, 87(2): 188-207.

Meese, Richard A. and Nancy E. Wallace. 1997. "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches." *Journal of Real Estate Economics and Finance.* 14: 51-73.

Murphy, Alvin. 2007. "A Dynamic Model of Housing Supply." *Mimeo*.

Neal, Derek and Diane W. Schanzenbach. forthcoming. "Left Behind By Design: Proficiency Counts and Test-Based Accountability." *The Review of Economics and Statistics*.

Oates, Wallace. 1969. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy*, 77: 957-971.

Palmquist, Raymond B. 1982. "Measuring Environmental Effects on Property Values without Hedonic Regressions." *Journal of Urban Economics*, 33(3): 333-47.

Palmquist, Raymond B. 2005. "Property Value Models," in *Handbook of Environmental Economics, Volume 2*. Karl-Göran Mäler and Jeffery Vincent eds. Amsterdam: North Holland Press.

Pope, Jaren C. 2008. "Fear of Crime and Housing Prices: Household Reactions to Sex Offender Registries." *Journal of Urban Economics*, 64(3): 601-14.

Reback, Randall. 2005. "House Prices and the Provision of Local Public Services: Capitalization under School Choice Programs." *Journal of Urban Economics* 57: 275-301.

Reback, Randall. 2008. "Demand (and supply) in an inter-district public school choice program." *Economics of Education Review* 27: 402-416.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1): 34-55.

Rubinfeld, Daniel L., Perry Shapiro, and Judith Roberts. 1987. "Tiebout Bias and the Demand for Local Public Schooling." *Review of Economics and Statistics* 69(3): 426-37.

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

Scotchmer, Suzanne. 1986. "The Short-Run and Long-Run Benefits of Environmental Improvement." *Journal of Public Economics*, 30(1): 61-81.

Simons, Robert. 2006. *When Bad Things Happen to Good Property*. Washington D.C.: Environmental Law Institute.

Starrett, David A. 1981. "Land Value Capitalization in Local Public Finance." *Journal of Political Economy*, 89(2): 306-27.

Tiebout, Charles. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64(5): 416-24.

Tinbergen, Jan. 1959. "On the Theory of Income Distribution," in *Selected Papers of Jan Tinbergen*. L.H. Klaassen, L.M. Koych and H.J. Witteveen eds. Amsterdam: North Holland.

TABLE 1
SUMMARY OF HEDONIC AND CAPITALIZATION-BASED ESTIMATES OF THE WILLINGNESS TO PAY FOR A SMALL IMPROVEMENT IN PUBLIC SCHOOL QUALITY

	STUDY REGION FOR ESTIMATES	TEST SCORE ELASTICITY		MEAN WILLINGNESS TO PAY FOR 1% INCREASE IN TEST SCORES	
		hedonic	capitalization	hedonic	capitalization
		(1)	(2)	(3)	(4)
PREVIOUS STUDIES					
Black (1999)	Boston, MA [1990]	0.42		917	
Bayer, Ferreira, McMillan (2007)	San Francisco, CA [1993-95]	0.12		372	
THIS STUDY	Fairfax, VA [2003]	0.12	-0.04	608	-194
	Portland, OR [2003]	0.20	0.01	447	16
	Philadelphia, PA [2003]	0.27	0.12	743	317
	Detroit, MI [2003]	0.21	-0.29	422	-587
	Los Angeles, CA [2003]	0.14	0.17	596	740
	All Five Regions [2003]			536	134
	All Five Regions [2007]			688	152

NOTE.—Children in Boston, San Francisco, Fairfax, and Portland were assigned to elementary schools based on the attendance zones where their parents lived. Children in Philadelphia, Detroit, and Los Angeles were assigned to school districts but free to choose between schools within a district. Each state develops its own standardized tests, which change over time. Assignment laws and test scores are discussed in section V. In cols. 1 and 3, the hedonic estimates are identified by boundary discontinuity designs that use fixed effects to control for omitted variables. In cols. 2 and 4, the capitalization estimates are identified by first-differenced regressions that control for changes in neighborhood demographics and purge omitted variables. All measures of willingness to pay are reported in constant year 2000 dollars.

TABLE 2
SUMMARY STATISTICS OF SCHOOL TEST SCORE DIFFERENCES

	FAIRFAX, VA		PORTLAND, OR		PHILADELPHIA, PA		DETROIT, MI		LOS ANGELES, CA	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
2002/2003 math-reading score	81.88	10.44	79.35	11.34	67.43	13.19	67.17	11.39	45.73	17.34
Changes in math-reading score										
10th decile	11.35	8.03	1.45	9.22	18.78	0.80	15.08	2.43	10.60	3.26
middle deciles	0.62	5.02	-4.02	6.61	10.45	4.85	11.15	2.78	9.24	2.01
90th decile	-0.44	2.79	-4.50	4.07	6.28	1.36	5.13	1.72	5.97	1.51
% change in 10th decile	13.87%		1.83%		27.85%		22.46%		23.18%	

NOTE.—Means and standard deviations for test scores are based on NCLB information aggregated and reported by www.schooldatadirect.org. The math reading score is an overall measure (calculated by Standard & Poor's) that provides an average of the proficiency rates achieved across all reading and math tests, weighted by the number of tests taken for each elementary school (Fairfax and Portland) or school district (Philly, Detroit and LA). Raw scores are not directly comparable across states because each state develops its own standardized tests.

TABLE 3

SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND TEST SCORES IN FAIRFAX, VA

	Full Sample (micro data: N = 10,255)		Sample: 0.20 Mile Boundary Zone (micro data: N = 5,843)			Full Sample (Census block group data: N = 438)			
Fairfax County, VA	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									-0.02
2003 price	567,322	247,727	546,575	-3,036	-0.40	571,742	226,270	0.06	
2007 price	562,683	305,748	542,998	14,512	1.36	599,474	268,952		
Average math/reading test result									
2003 score	83.56	9.54	83.01	8.91	38.67	82.86	9.25	-0.49	
2007 score	84.36	8.25	83.90	5.11	24.81	83.92	8.17		
Housing characteristics:									
square feet (100's)	21.12	9.93	20.66	0.07	0.26	21.32	7.06	0.01	-0.02
bathrooms	3.24	1.08	3.21	0.00	-0.10	3.24	0.72	0.00	-0.05
age	34.07	15.82	34.13	0.89	2.13	35.21	12.63	0.04	-0.02
lot acres	0.38	0.43	0.35	0.00	-0.51	0.43	0.42	0.07	0.04
bedrooms	3.94	0.77	3.93	0.03	1.56	3.92	0.36	-0.03	-0.08
Neighborhood characteristics:									
% block group nonwhite	0.23	0.11	0.23	-0.02	-6.89	0.24	0.12	0.16	-0.12
% block group under 18	0.24	0.04	0.24	0.00	-0.82	0.23	0.03	-0.04	0.07
% block group owner occupied	0.85	0.15	0.84	0.00	0.08	0.81	0.18	-0.18	0.00
% block group vacant	0.01	0.02	0.01	0.00	-2.06	0.02	0.02	0.10	0.04
block group pop density	0.37	0.22	0.40	0.00	0.77	0.39	0.26	0.06	-0.11
tax rate	111.85	49.52	111.45	-0.30	-0.28	117.30	38.00	-0.08	

NOTE.—This table reports summary statistics for the key variables included in the analysis for Fairfax, VA. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score homes on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

TABLE 4
TEST SCORE COEFFICIENTS FROM HEDONIC AND CAPITALIZATION REGRESSIONS

	FAIRFAX, VA	PORTLAND, OR	PHILADELPHIA, PA	DETROIT, MI	LOS ANGELES, CA
	A. Test Score Parameters from Hedonic Regressions (micro data from 0.2 mile sample without boundary fixed effects)				
	(1)	(2)	(3)	(4)	(5)
log (test score), 2003 coefficient	0.122	0.456	0.481	0.524	0.274
	(0.027)	(0.020)	(0.045)	(0.036)	(0.012)
log (test score), 2007 differential	0.554	0.034	0.229	0.516	0.084
	(0.056)	(0.032)	(0.067)	(0.086)	(0.023)
R^2	0.74	0.70	0.68	0.68	0.75
Number of observations	6,036	14,443	3,973	6,252	12,287
	B. Test Score Parameters from Hedonic Regressions (micro data from 0.2 mile sample with boundary fixed effects)				
	(6)	(7)	(8)	(9)	(10)
log (test score), 2003 coefficient	0.116	0.200	0.272	0.208	0.140
	(0.040)	(0.028)	(0.071)	(0.047)	(0.015)
log (test score), 2007 differential	0.293	-0.165	-0.120	0.357	0.075
	(0.081)	(0.048)	(0.101)	(0.126)	(0.028)
R^2	0.85	0.77	0.76	0.74	0.85
Number of observations	6,036	14,443	3,973	6,252	12,287
F-test on H_0: time-constant gradient	4.69	1.98	1.86	4.41	8.22
p-value on F-test	0.000	0.031	0.047	0.000	0.000
	C. Test Score Parameters from Capitalization Regressions (block group data from full sample)				
	(11)	(12)	(13)	(14)	(15)
change in log (test score)	-0.037	0.007	0.116	-0.289	0.174
	(0.073)	(0.096)	(0.068)	(0.134)	(0.033)
R^2	0.53	0.45	0.29	0.21	0.18
Number of observations	438	754	1,199	1,477	6,975

NOTE.—All regressions included controls for property taxes, structural housing characteristics (square feet, number of bathrooms, age, lot size, number of bedrooms) and neighborhood characteristics measured at the block group level (population density, percent nonwhite, percent under 18, percent owner occupied, and percent vacant). In cols. 1 through 10, the dependent variable is the natural log of the sale price of the home. All control variables are interacted with a dummy for sales made during the 2007-2008 school year. In cols. 11 through 15 the dependent variable is the change in the natural log of the average sale price in the census block group. All regressions use Eicker-White standard errors.

TABLE 5
ROBUSTNESS CHECKS ON TEST SCORE COEFFICIENTS

	FAIRFAX, VA	PORTLAND, OR	PHILADELPHIA, PA	DETROIT, MI	LOS ANGELES, CA
	A. Test Score Parameters from Hedonic Regressions (micro data from full sample without boundary fixed effects)				
	(1)	(2)	(3)	(4)	(5)
log (test score), 2003 coefficient	0.227	0.540	0.546	0.751	0.260
	(0.023)	(0.016)	(0.017)	(0.019)	(0.004)
log (test score), 2007 differential	0.550	0.024	0.396	0.565	0.041
	(0.044)	(0.024)	(0.028)	(0.042)	(0.008)
R^2	0.75	0.69	0.72	0.68	0.70
Number of observations	10,662	25,294	29,327	32,485	146,783
	B. Test Score Parameters from Hedonic Regressions (block group data from full sample without boundary fixed effects)				
	(6)	(7)	(8)	(9)	(10)
log (test score), 2003 coefficient	0.148	0.388	0.229	0.813	0.321
	(0.068)	(0.045)	(0.046)	(0.052)	(0.012)
log (test score), 2007 differential	0.475	0.086	0.367	0.717	0.082
	(0.121)	(0.070)	(0.083)	(0.108)	(0.022)
R^2	0.84	0.82	0.78	0.77	0.73
Number of observations	889	1,553	2,647	3,333	14,727
	C. Test Score Parameters from Capitalization Regressions (block group data from 0.2 mile sample with boundary fixed effects)				
	(11)	(12)	(13)	(14)	(15)
change in log (test score)	0.008	-0.025	0.130	-0.445	0.231
	(0.111)	(0.091)	(0.180)	(0.521)	(0.177)
R^2	0.83	0.82	0.91	0.87	0.83
Number of observations	422	603	176	213	251

NOTE.—All regressions included controls for property taxes, structural housing characteristics (square feet, number of bathrooms, age, lot size, number of bedrooms) and neighborhood characteristics measured at the block group level (population density, percent nonwhite, percent under 18, percent owner occupied, and percent vacant). In cols. 1 through 10, the dependent variable is the natural log of the sale price of the home. All control variables are interacted with a dummy for sales made during the 2007-2008 school year. In cols. 11 through 15 the dependent variable is the change in the natural log of the average sale price in the census block group. All regressions use Eicker-White standard errors.

TABLE 6
IMPACT OF IDENTIFICATION STRATEGY ON ESTIMATES FOR THE AVERAGE RESIDENT'S WILLINGNESS TO PAY FOR A 1% INCREASE IN TEST SCORES

	(1)	(2)	(3)	(4)	(5)	(6)
Estimates for willingness to pay:						
2003 school year	1,238	1,222	1,041	536	134	169
2007 school year	1,685	1,572	1,660	688	152	190
Identification strategy:						
Model	hedonic	hedonic	hedonic	hedonic	capitalization	capitalization
Sample	full	full	0.2 mile	0.2 mile	full	0.2 mile
Data point	block group	house	house	house	block group	block group
Sample size	23,149	244,551	42,991	42,991	10,843	1,665
Controls for omitted variables	none	none	none	boundary fixed effects	differencing	differencing + boundary fixed effects

NOTE.—All measures of willingness to pay are reported in constant year 2000 dollars. Each measure is averaged over the samples from our five study regions, using the elasticities reported in tables 4 and 5. For example, the estimates in col. 4 are based on the elasticities reported in cols. 6 through 10 of table 4.

APPENDIX TABLE 1
SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN PORTLAND, OR

	Full Sample (micro data: N = 25,294)		Sample: 0.20 Mile Boundary Zone (micro data: N = 16,539)			Full Sample (Census block group data: N = 754)			
Portland Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									0.00
2003 price	241,875	142,991	237,021	2,664	0.91	244,315	127,426	0.00	
2007 price	324,181	173,171	316,236	-4,500	-1.14	336,280	156,920		
Average math/reading test result									
2003 score	79.82	10.93	79.89	7.41	44.93	78.94	10.33	-0.28	
2007 score	76.00	10.83	75.96	4.79	28.77	75.24	10.69		
Housing characteristics:									
square feet (100's)	17.88	7.76	17.75	-0.06	-0.48	17.59	5.18	0.04	-0.06
bathrooms	2.22	0.93	2.22	0.04	2.41	2.09	0.59	0.01	-0.03
age	39.67	30.16	38.39	-0.56	-1.17	50.17	22.91	0.00	-0.01
lot acres	0.20	0.29	0.18	0.00	0.72	0.30	0.43	0.08	0.01
bedrooms	3.07	0.94	3.07	-0.01	-0.41	3.03	0.47	0.09	-0.12
Neighborhood characteristics:									
% block group nonwhite	0.17	0.10	0.17	-0.01	-3.64	0.16	0.11	-0.14	0.00
% block group under 18	0.23	0.04	0.24	0.00	-1.14	0.22	0.04	-0.03	-0.02
% block group owner occupied	0.66	0.19	0.66	0.00	1.60	0.61	0.22	0.03	-0.06
% block group vacant	0.05	0.03	0.05	0.00	-7.41	0.05	0.03	-0.05	0.00
block group pop density	0.53	0.29	0.55	-0.01	-1.35	0.56	0.34	-0.08	0.04
tax rate	54.62	8.02	54.74	-0.23	-1.86	54.70	9.01	0.13	

NOTE.—This table reports summary statistics for the key variables included in the analysis for Portland, OR. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score homes on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

APPENDIX TABLE 2
SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN PHILADELPHIA, PA

	Full Sample (micro data: N = 29,333)		Sample: 0.20 Mile Boundary Zone (micro data: N = 3,973)			Full Sample (Census block group data: N = 1,199)			
Philadelphia Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									-0.07
2003 price	295,845	188,924	285,243	32,498	4.73	273,104	148,829	-0.35	
2007 price	334,662	221,967	324,197	49,222	5.44	316,940	170,204		
Average math/reading test result									
2003 score	67.88	13.93	69.43	11.05	30.63	64.38	16.84	-0.74	
2007 score	78.61	10.90	79.51	7.20	25.67	75.99	13.23		
Housing characteristics:									
square feet (100's)	20.87	9.48	20.03	1.21	4.21	19.85	5.95	-0.27	0.00
bathrooms	2.37	1.00	2.28	0.09	3.02	2.15	0.73	-0.35	-0.02
age	42.03	27.85	46.32	3.50	4.23	49.54	21.16	0.03	0.00
lot acres	0.49	0.65	0.44	0.02	1.15	0.45	0.46	-0.15	-0.01
bedrooms	3.38	0.77	3.33	0.07	2.82	3.40	0.44	-0.04	-0.03
Neighborhood characteristics:									
% block group nonwhite	0.12	0.14	0.11	-0.01	-1.30	0.14	0.19	0.22	-0.20
% block group under 18	0.23	0.04	0.22	0.00	2.54	0.22	0.04	0.01	0.10
% block group owner occupied	0.78	0.18	0.79	0.02	3.13	0.74	0.21	-0.12	0.04
% block group vacant	0.03	0.03	0.03	0.00	-3.77	0.04	0.03	0.14	-0.02
block group pop density	0.34	0.39	0.36	-0.04	-3.83	0.46	0.53	0.28	-0.13
tax rate	29.05	14.28	28.38	2.74	6.55	25.47	14.65	-0.30	

NOTE.—This table reports summary statistics for the key variables included in the analysis for Philadelphia, PA. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score homes on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

APPENDIX TABLE 3
SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN DETROIT, MI

	Full Sample (micro data: N =32,486)		Sample: 0.20 Mile Boundary Zone (micro data: N = 6,285)			Full Sample (Census block group data: N = 1,477)			
Detroit Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									0.14
2003 price	219,857	131,658	214,048	14,186	2.92	214,626	123,303	-0.38	
2007 price	166,801	131,839	157,640	11,017	2.44	169,829	104,428		
Average math/reading test result									
2003 score	68.76	12.39	67.91	7.77	27.21	68.01	12.18	-0.60	
2007 score	79.28	10.52	78.51	7.72	32.09	78.80	10.38		
Housing characteristics:									
square feet (100's)	16.57	7.79	16.01	0.66	3.27	16.47	5.93	-0.32	0.02
bathrooms	2.06	1.00	2.00	0.08	3.12	2.03	0.73	-0.38	0.08
age	46.06	23.24	46.78	0.19	0.36	46.87	18.00	0.26	-0.07
lot acres	0.36	0.52	0.30	-0.02	-2.06	0.39	0.46	-0.11	0.01
bedrooms	3.15	0.73	3.11	0.03	1.66	3.15	0.45	-0.22	0.00
Neighborhood characteristics:									
% block group nonwhite	0.13	0.18	0.12	-0.03	-7.53	0.14	0.20	0.04	-0.30
% block group under 18	0.23	0.04	0.23	0.01	8.67	0.23	0.04	0.04	0.16
% block group owner occupied	0.80	0.18	0.82	0.02	3.79	0.78	0.20	-0.15	-0.01
% block group vacant	0.04	0.03	0.03	0.00	0.80	0.04	0.04	0.12	0.02
block group pop density	0.40	0.28	0.46	0.01	2.00	0.40	0.28	0.25	-0.07
tax rate	27.09	11.25	25.90	-0.61	-2.42	27.70	9.73	-0.11	

NOTE.— This table reports summary statistics for the key variables included in the analysis for Detroit, MI. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score homes on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

APPENDIX TABLE 4
SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN LOS ANGELES, CA

	Full Sample (micro data: N =146,788)		Sample: 0.20 Mile Boundary Zone (micro data: N = 12,287)			Full Sample (Census block group data: N = 6,975)			
Los Angeles Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									-0.04
2003 price	460,747	353,758	509,207	38,511	4.03	486,539	350,354	-0.15	
2007 price	486,752	463,063	563,566	55,621	3.72	551,685	458,579		
Average math/reading test result									
2003 score	39.75	13.75	41.86	13.92	49.94	39.11	14.14	-0.46	
2007 score	48.81	12.92	51.14	12.81	50.50	48.39	13.22		
Housing characteristics:									
square feet (100's)	17.06	7.67	17.11	0.84	5.81	16.16	5.63	-0.19	0.02
bathrooms	2.13	0.86	2.16	0.13	7.89	1.97	0.61	-0.20	0.02
age	43.06	22.98	44.87	-3.49	-8.98	53.11	19.18	0.21	0.00
lot acres	0.25	0.38	0.20	0.00	0.24	0.22	0.28	-0.13	0.01
bedrooms	3.18	0.87	3.21	0.07	4.66	3.07	0.54	-0.14	0.00
Neighborhood characteristics:									
% block group nonwhite	0.26	0.18	0.31	0.01	3.68	0.28	0.18	0.14	-0.25
% block group under 18	0.25	0.06	0.25	0.00	4.27	0.25	0.06	0.09	0.09
% block group owner occupied	0.68	0.21	0.70	0.01	3.99	0.60	0.25	-0.09	-0.01
% block group vacant	0.06	0.10	0.05	0.00	2.47	0.05	0.07	-0.06	0.00
block group pop density	0.72	0.60	0.82	-0.03	-2.74	0.96	0.75	0.17	0.05
tax rate	84.33	137.66	82.44	1.11	5.61	83.00	16.05	-0.04	

NOTE.— This table reports summary statistics for the key variables included in the analysis for Los Angeles, CA. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score homes on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.