Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay?[*]

By NICOLAI V. KUMINOFF AND JAREN C. POPE[1]

This paper develops a welfare theoretic framework for interpreting evidence on the impacts of public programs on housing markets. We extend Rosen's hedonic model to explain how housing prices capitalize exogenous shocks to public goods and externalities. The model predicts that trading between heterogeneous buyers and sellers will drive a wedge between these "capitalization effects" and welfare changes. We test this hypothesis in the context of changes in measures of school quality in five metropolitan areas. Results from boundary discontinuity designs suggest that capitalization effects understate parents' willingness to pay for public school improvements by as much as 75%.

---

1. Introduction

In his seminal 1974 paper, Sherwin Rosen explained how market transactions can reveal buyers' willingness to pay for the characteristics of a differentiated product. Rosen's static model is frequently used to assess the benefits of policies targeting public goods and externalities. The logic is simple. Homebuyers implicitly purchase the right to consume a bundle of local public goods when they buy a house. It follows that a hedonic price function for housing can be used to infer buyers' willingness to pay for polices that would alter the provision of public goods. Unfortunately, this is easier said than done.

One of the main complications with using housing markets to infer the willingness to pay for public goods is that the market clearing process can present endogeneity problems for estimation.[2] As heterogeneous households sort themselves across an urban area they also vote on the provision of local public goods, they interact with their neighbors, and their collective actions may increase congestion and degrade the natural environment. These mechanisms have the potential to confound reduced form estimators for the willingness to pay by inducing correlation between the public good of interest and latent attributes of households and neighborhoods. Goldstein and Pauly (1981) first called attention to this problem, labeling it "Tiebout bias" since it arises from the sorting mechanism in Tiebout's (1956) conceptual model of residential sorting and local public goods provision. Rubinfeld, Shapiro, and Roberts (1987), Epple and Sieg (1999), and Epple, Romer, and Sieg

---

[2] Other modeling challenges include forward looking behavior, the unknown form of the equilibrium price function, and identification problems with estimating demand curves for product characteristics. For discussion of these issues see Epple 1987, Cropper, Deck, and McConnell 1988, Ekeland, Heckman, and Nesheim 2004, Kuminoff, Parmeter, and Pope 2010, Bishop and Murphy 2011, and Kuminoff, Smith, and Timmins 2013.

(2001) were among the first to develop strategies for estimating households' preferences for public goods in a way that accounted for Tiebout sorting. Subsequent studies addressed peer effects in addition to modeling sorting based on public goods (e.g. Calabrese et al. 2006, Bayer and Timmins 2007, Ferreyra 2007, Walsh 2007).

A recent wave of empirical research has sought to estimate the willingness to pay for public goods without modeling sorting behavior by instead placing hedonic price functions within the econometric framework for program evaluation (Imbens and Wooldridge 2009). This approach treats Tiebout bias as an omitted variable problem to be addressed using instruments, panel data, and regression discontinuity designs. The most common strategy is to use a plausibly exogenous source of *temporal* variation in the quality of a public good to identify how the quality change was capitalized into housing prices. These "capitalization effects" are then interpreted as welfare measures.[3] Researchers have relied on this logic in order to draw strong conclusions about important problems such as the value of a statistical life (Davis 2004), the benefits of the Clean Air Act (Chay and Greenstone 2005), and homeowners' willingness to pay to reduce their exposure to crime risk (Linden and Rockoff 2008, Pope 2008). More generally, over the past decade the hedonic program evaluation framework has become a leading approach to measuring the public's willingness to pay for public goods, with numerous applications published in the top general interest and field journals in economics (see Parmeter and Pope 2012 for a survey).

---

[3] Consistent with recent program evaluation studies, we use the word "capitalization" to describe how shocks to amenities at a point in space cause prices to change over time. This description differs from an earlier literature that used "capitalization" to describe the equilibrium spatial relationship between prices and amenities at a point in time (see Kanemoto 1988 for an example and references). We define capitalization formally in section 2.

While it is routinely asserted that "capitalization effects" measure the willingness to pay for public goods, the recent program evaluation studies have not provided any evidence to support this claim. That is, none of the studies that have interpreted capitalization effects as welfare measures have developed models of the capitalization process to support their interpretations. In order to understand capitalization effects, one would need to use Rosen's (1974) model to conduct an appropriate comparative static analysis of how changes in public goods affect housing market equilibria.

The purpose of this paper is to investigate the validity of interpreting capitalization effects as welfare measures when the price functions that clear a market for a differentiated good arise from the equilibrium sorting process described by Rosen. In the first half of the paper we extend Rosen's (1974) conceptual model to express the capitalization effect for a public good as a general function of structural parameters describing household preferences, production technology, and market institutions. We find that the capitalization effect does not have a specific welfare interpretation in this environment. When there is an exogenous shock to the spatial distribution of a public good, the gradient of the hedonic price function will generally adjust in order to clear the housing market. This adjustment drives a wedge between the average capitalization effect and the average household's willingness to pay.

For example, an improvement in the quality of public education will change the shadow price of access to public schools. The shadow price adjusts because the demand for school quality is downward sloping and/or because the composition of households in a given neighborhood changes. An improvement in school quality may also change what people

3

are willing to pay for complementary housing attributes such as locations near public parks. The problem is that the capitalization effect conflates the public's willingness to pay for the improvement in school quality with changes in the shadow prices of school quality and other housing attributes. This type of conflating appears to be a general feature of the hedonic equilibrium model. It even occurs in simple specifications for consumer preferences such as the linear-quadratic-normal model considered by Epple (1987) and Ekeland, Heckman, and Nesheim (2004).

In the second half of the paper we investigate the empirical implications of "conflation bias" in the willingness to pay for public goods. We develop and demonstrate a methodology for testing whether capitalization effects reveal welfare measures. Given a parametric specification for the hedonic price function, we derive sufficient conditions for interpreting the marginal capitalization effect experienced by a household as a measure of that household's marginal willingness to pay. Importantly, these conditions can be tested within the hedonic program evaluation framework. Our main test relies on having a research design for identifying the gradient of the equilibrium price function both *before* and *after* the shock to public goods that defines the capitalization effect.

Our empirical demonstration of the methodology uses a boundary discontinuity design to estimate parents' valuation of public school quality before and after there were large changes in publicly reported measures of academic performance. This research design exploits a series of laws that create spatial discontinuities in the way that children are assigned to public schools. Children living in physically similar houses in the same neighborhood are sometimes assigned to different schools where students tend to score better or

worse on standardized exams. These assignment laws underlie our strategy for estimating the shadow price of school-level academic performance. We estimate shadow prices in 10 housing markets: five metropolitan areas (Los Angeles, Philadelphia, Detroit, Fairfax, and Portland) each observed at two points in time (2003 and 2007) that were chosen because they bracket substantial changes in the measures of test scores that were reported to parents and the general public. Prior studies such as Black (1999) and Bayer, Ferreira, and McMillan (2007) have used the same research design to estimate the shadow price of public school test scores in a single metro area at a single point in time. Their results provide a baseline for comparison. Our study is the first to provide evidence on variation in the shadow price of public school test scores across time and space.

We find that the average shadow price of a 1% increase in test scores increased by 28% between 2003 and 2007. This average reflects considerable heterogeneity across markets. Changes in the shadow prices of test scores and other housing attributes are conflated with homebuyers' valuation of school quality, causing our estimates for capitalization effects to understate hedonic measures of the willingness to pay by as much as 75%.

Overall, the evidence from our conceptual and empirical models suggests the bias in interpreting capitalization effects as measures of the willingness to pay for public goods is of first-order importance. Our work raises the bar for future research. In order to use capitalization effects to draw credible inferences about consumer welfare, the analyst must first demonstrate that the evolution of the price function supports their interpretation.

The next section provides context for our study and explains our research design. Sections 3 and 4 develop our conceptual and econometric models. Section 5 describes the ap-

plication to school quality, section 6 presents results, and section 7 concludes.

2. The Hedonic Method and Benefit Measurement

2.1. *Identifying Capitalization Effects for Endogenous Public Goods*

To illustrate the issues at stake, we begin with a standard reduced-form model of the relationship between housing prices and public goods. We define public goods broadly to include any nonmarket goods and services conveyed to homeowners through their choice of a neighborhood. Examples include local public goods (such as school quality), urban and environmental services (such as crime rates and air quality), and variables describing the demographic composition of the community (such as race and educational attainment). Virtually all reduced-form studies in the literature pose a version of the following model,

$$(1) \qquad\qquad p_1 = g_1\theta_1 + h_1\eta_1 + \varepsilon_1,$$

where p represents the price of housing, g is the public good of interest, h represents all other observable attributes of houses and neighborhoods, ε is an error term that arises, in part, due to unobserved attributes of neighborhoods, and their subscripts denote the time period. With $j = 1,...,J$ houses and $k = 1,...,K$ observable attributes, p_1 and g_1 are $Jx1$ vectors and h_1 is a JxK matrix. The elements of p_1, g_1, and h_1 are typically measured in levels or logs. Finally, θ_1 is the parameter of interest. We will discuss its interpretation in section 2.2.

Many public goods are endogenously determined through the housing market in ways that are likely to induce correlation between g_1, h_1, and ε_1, creating a problem for OLS

estimation of θ_1. For example, imagine trying to isolate the impact of registered sex offenders on nearby property values. If the sex offenders sorted themselves into subdivisions with higher preexisting crime rates, where housing was cheaper, then the OLS estimator for θ_1 will confound the sex offenders' impact on property values with the impact of preexisting crime. The two effects cannot be distinguished by controlling for crime because data on crime rates are generally unavailable below the level of a zip code. This type of confounding is widely believed to pervade the literature.

Recent studies have developed research designs that mitigate confounding (e.g. Black 1999, Davis 2004, Chay and Greenstone 2005, Linden and Rockoff 2008, Pope 2008, Greenstone and Gallagher 2008, Baum-Snow and Marion 2009, Cellini, Ferreira, and Rothstein 2010). With the exception of Black (1999), these studies exploit sources of *temporal* variation in *g*. This variation is used to estimate an econometric model specified in terms of fixed effects, first differences, or difference-in-differences. For example, suppose *p*, *g*, and *h* are observed again after the distribution of *g* has changed. *h* and ε may have changed as well. Differencing the data produces a panel model,

(2) $$\Delta p = \Delta g \phi + \Delta h \gamma + \Delta \varepsilon ,$$

where $\Delta p = p_2 - p_1$, for example. Equation (2) describes how prices adjusted to the change in *g*, controlling for concomitant changes in *h*. This is the standard first differences model used in the literature. Some studies also use instruments for Δg to address possible correlation between Δg and $\Delta \varepsilon$. Notice that the identified parameter, ϕ, is not necessarily the same as θ_1 in equation (1). We refer to ϕ as the "capitalization effect" because it

7

describes how the change in *g* was capitalized into housing prices. Our interest lies in analyzing how this effect can be interpreted when (1) is the correct model.

2.2. *The Welfare Interpretation of Capitalization Effects*

The interpretation of ϕ begins with the interpretation of θ_1. It is standard to translate θ_1 into a welfare measure by appealing to hedonic theory. First, the price function is assumed to be continuously differentiable. Differentiating with respect to *g* defines the marginal price function for *g*. The marginal price paid by the buyer of house *j* is defined by a $(\theta_1, p_{1j}, g_{1j})$ triplet.[4] Next, buyers and sellers are assumed to satisfy the smoothness conditions of Rosen's (1974) model, including: (i) differentiability of utility functions and cost functions; (ii) free mobility; (iii) the ability to consume and produce continuous quantities of *g* and *h;* (iv) perfect information about *p*, *g*, and *h*; and (v) no market power on the part of any buyer or seller. Under these conditions, Rosen demonstrates that the marginal price function for *g*, evaluated at p_{1j}, g_{1j}, will equal the buyer's willingness to pay for a marginal change in *g* (henceforth MWTP).

In contrast, Rosen (1974) does not interpret ϕ. He considers market equilibrium, not the adjustment process that would follow an exogenous change in product attributes. Studies that estimate capitalization effects have addressed this knowledge gap by assuming that the gradient of the price function is constant over the duration of the study period (i.e. $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$). This assumption is crucial. It allows household-specific measures

---

[4] The formula for the marginal price depends on whether *p* and *g* are measured in levels or logs. However, this distinction is not important for our analysis. Our conclusions hold regardless of how the analyst chooses to scale of elements of *p*, *g*, and *h*.

of MWTP to be defined by $\left(\phi, p_{1j}, g_{1j}\right)$ triplets in period 1 and by $\left(\phi, p_{2j}, g_{2j}\right)$ triplets in period 2. These definitions follow from the interpretation of θ_1 and simple algebra.[5] If we instead consider the fixed-effects or difference-in-difference or instrumental variables analogs to (2), we reach the same conclusion: The assumption of a time-constant gradient is crucial to the analyst's ability to translate the identified parameters of the econometric model into welfare measures.

Recent studies have used the time-constant gradient assumption (henceforth TCGA) to translate capitalization effects into welfare measures for changes in cancer risk (Davis 2004), crime risk (Linden and Rockoff 2008, Pope 2008), hazardous waste (Greenstone and Gallagher 2008), invasive species (Horsch and Lewis 2009), investment in education (Cellini, Ferreira, and Rothstein 2010), low income housing credits (Baum-Snow and Marion 2009), open space (Bin, Landry, and Meyer 2009), and particulate matter (Chay and Greenstone 2005) to list only a few.[6] In these studies, the gradient is assumed to be fixed for 10 to 20 years, spanning large changes in g, h, and potentially ε.[7]

Given the importance of developing credible estimates of MWTP for public goods, it is surprising how little is known about the evolution of hedonic price functions. None of the

---

[5] Write the period 2 price function as $p_2 = g_2\theta_2 + h_2\eta_2 + \varepsilon_2$. Subtracting the period 1 price function from the period 2 price function reduces to the capitalization model in (2) as long as $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$. Thus, $\phi = \theta_1 = \theta_2$.

[6] For example, Chay and Greenstone (2005, p.418) conclude that their analysis "demonstrates that quasi-experimental approaches can be effective in estimating parameters derived from economic models (e.g. MWTP)" and that welfare calculations based on their estimates for capitalization effects, "suggest that the mid-1970s TSPs nonattainment designation provided a $45 billion aggregate gain to homeowners in nonattainment counties."

[7] The length of the study period and the sizes of the changes in variable are typically dictated by the instruments needed to support the analyst's preferred identification strategy.

studies invoking TCGA have tested it or provided evidence to validate it. Nor can we find any prior studies that explain what (if anything) must by assumed about preferences in order to guarantee that a hedonic gradient will be invariant to the types of changes in public goods, wealth, and information that occur over 10-20 year periods.

2.3. *Related Evidence from Previous Studies*

Three sets of studies have considered issues that relate to our research question. First, theory papers by Lind (1973) and Starrett (1981) ask whether a policy that alters the distribution of a public good will produce changes in land values that reveal the social benefits of the policy. Their answer is 'no', not if heterogeneous households react to the policy by moving. Sieg et al. (2004) reach the same conclusion in a numerical simulation. One *might* expect their common finding—that price changes do not reveal welfare effects—to extend to our hedonic setting. However, this is an intuitive leap. The models developed by Lind, Starrett, and Sieg et al. relax some of the smoothness conditions that support equilibria with a one-to-one mapping between marginal prices and MWTP in Rosen's model. Therefore, their results do not have direct implications for the relationship between capitalization and MWTP in environments based on Rosen (1974).

Second, Palmquist (1988, 1992) considers how hedonic price functions could be used to measure welfare effects for changes in environmental quality. His 1988 paper explains how Hicksian welfare measures could, in principle, be constructed from data on an *individual's* choices before and after a quality change, if such data were available and if it were possible to identify price functions before and after the change. In the special case where the change is "localized", Palmquist (1992) conjectures that it might be possible to con-

struct welfare measures from the ex-ante price function.[8] Neither paper addresses the assumptions needed to support TCGA; nor do they consider whether it is possible to recover MWTP from data on price changes following a non-marginal change in quality.

Finally, a few empirical studies have reported evidence of temporal instability in the parameters used to characterize gradients of housing price functions. For example, Brookshire et al. (1985) found that a shock to information about earthquake risk changed the implicit price of earthquake risk over a 6-year period, and Beron, Murdoch, and Thayer (2001) reported annual changes in the implicit price of visibility in Los Angeles between 1980 and 1995. However, the evidence from these studies looks dubious when viewed through the lens of the modern program evaluation literature. The problem is that their research designs do not use modern tools for addressing omitted variables.

2.4. *Our Research Design*

A direct way to test the hypothesis of a time-constant gradient is to identify single-period price functions before and after a change in the distribution of public goods. While there is no methodological panacea for overcoming omitted variable bias in cross-section data, Sandra Black's (1999) boundary discontinuity design is generally viewed as a credible strategy for mitigating the problem. For example, Greenstone and Gallagher (2008 p.997) include it among their short list of papers "*demonstrating that it is possible to iden-*

---

[8] By "localized", Palmquist means that the quality change has no impact on the equilibrium price function. Subsequent to Palmquist's work, it has been recognized that localized changes can trigger tipping effects via Tiebout sorting that produce large changes in equilibria (Sethi and Somathathan 2004, Card, Mas, and Rothstein 2008, Banzhaf and Walsh 2013). For example, Banzhaf and Walsh (2008) find that increasing emissions of toxic air pollutants alters neighborhood demographics by increasing emigration in general and increasing emigration of wealthier households in particular.

tify research designs that mitigate the confounding that has historically undermined the credibility of conventional hedonic approaches to valuing nonmarket goods." [9]

Black (1999) used spatial discontinuities in the laws assigning children to public schools to identify the impact of standardized test scores on property values in the Boston suburbs in the early 1990s. While there have been many subsequent applications of her methodology, to our knowledge none have tracked how the implicit price of test scores on property values has evolved over time. Nor has any study compared multiple markets at the same point in time. We extend this literature by estimating 10 hedonic price functions, describing five metropolitan areas, each observed during two years: 2003 and 2007. This period brackets significant changes in the spatial distribution of publicly reported measures of school quality. Details of the data and application begin in section 5. Now we develop the conceptual model.

3. Hedonic Equilibria and the Capitalization of Market Shocks

This section reviews the primitives of Rosen's model in the context of a housing market, characterizes equilibrium, and defines restrictions on preferences and technology that guarantee the marginal price schedule will be unaffected by exogenous changes in non-market attributes of a private good.

3.1. *Demand, Supply, and Market Equilibrium*

Price-taking households are assumed to be free to choose a house with any combination of physical attributes (e.g. bedrooms, bathrooms, sqft) in the neighborhood that provides

---

[9] Similar opinions are expressed through the discussion of Black's work in quasi-experimental hedonic studies such as Cellini, Ferreira, and Rothstein (2010), Chay and Greenstone (2005), Linden and Rockoff (2008), and Pope (2008).

their desired bundle of public goods. The model is static so that consumers are assumed to be myopic with respect to the future evolution of prices and public goods.[10] The utility maximization problem is

(3) $\max\limits_{g,x,b} \ U(g,x,b;\alpha)$ subject to: $y = b + P(g,x;\omega),$

where x is a vector containing all attributes of houses and neighborhoods, other than g. Thus x includes the elements of h along with any omitted variables that enter the error term in the estimator for the hedonic price function in (1). A household chooses levels of attributes and the composite good (b) to maximize utility, given its preferences (α), income (y), and the after-tax price of housing, $P(g,x;\omega)$, which is expressed as a general parametric function of g, x, and a parameter vector, ω. The first order conditions are

(4a) $$\frac{\partial P(g,x;\omega)}{\partial g} = \frac{\partial U/\partial g}{\partial U/\partial b},$$

(4b) $$\frac{\partial P(g,x;\omega)}{\partial x_k} = \frac{\partial U/\partial x_k}{\partial U/\partial b} \quad \text{for } k = 1,...,K.$$

Equation (4a) implies that each household will choose a neighborhood that provides a quantity of g at which their willingness to pay for an additional unit equals its marginal implicit price. Equation (4b) states the analogous condition for physical housing attributes.

Let $C(g,m,x;\beta)$ denote a producer's cost function, where m is the number of houses the producer sells and β is a vector of parameters describing the producer's idiosyncratic

---

[10] Recent studies have begun to consider how forward looking behavior may affect the estimation of structural models of the sorting process (e.g. Bayer et al. 2011, Bishop and Murphy 2011). The implications for reduced form estimation of price functions have yet to be determined. This is an important area for further research.

costs. Following Rosen, we treat each producer as a price taker who specializes in produc-

ing exactly one housing type but is free to vary the number of units sold. For convenience,

g is treated as exogenous.[11] In this case, the profit maximization problem is

$$(5) \qquad \max_{x,m} \ \pi = m \cdot P(g,x;\omega) - C(g,m,x;\beta),$$

with the corresponding first order conditions

$$(6) \quad P(g,x;\omega) = \frac{\partial C(g,m,x;\beta)}{\partial m}, \qquad \frac{\partial P(g,x;\omega)}{\partial x_k} = \left(\frac{1}{m}\right)\frac{\partial C(g,m,x;\beta)}{\partial x_k} \ \text{ for } k = 1,...,K.$$

Producers choose m to set the offer price of the marginal house equal to its production cost,

and they choose x to set the marginal per unit cost of each attribute equal to its implicit

price.

The primitives of the model include the distribution of consumer types, $R(y,\alpha)$, the

distribution of producer types, $S(\beta)$, and the spatial distribution of the public good, $T(g)$.

Equilibrium occurs when the first order conditions in (4) and (6) are simultaneously satis-

fied for all consumers and producers. This system of differential equations implicitly de-

fines the equilibrium hedonic price function that clears the market. It will be useful to re-

write the price function to acknowledge its dependence on model primitives:

$$(7) \qquad P(g,x;\omega) \equiv P[g,x;\omega(a,b,c)],$$

where a, b, and c are parameter vectors describing the distributions of consumers, produc-

---

[11] The results of this section are not altered by making g endogenous or x exogenous. We need only assume that g may be influenced by

forces that are exogenous to the model.

ers, and the public good: $R(y,\alpha) \sim a$, $S(\beta) \sim b$, and $T(g) \sim c$.[12] Importantly, the reduced form parameters describing the shape of the price function are endogenously determined by the structural parameters. It follows that shocks to the distributions of income and preferences, technology, or public goods may change the shape of the price function which, in turn, will change the implicit price schedule for g.

3.2. *Interpreting Capitalization Effects as Welfare Measures*

Now we depart from Rosen to consider equilibria in the same geographic market, before and after an exogenous shock to $T(g)$. The change in the value of a particular house j depends on the difference in the pre and post-shock price functions,

$$(8) \quad P\big[g_{2j}, x_{2j}; \omega(a_2, b_2, c_2)\big] - P\big[g_{1j}, x_{1j}; \omega(a_1, b_1, c_1)\big],$$

where the subscripts denote pre and post-shock equilibria. To isolate the capitalization effect for house j, we condition on x and divide the change in P by the change in g,

$$(9) \quad \phi_j = \frac{P\big[g_{2j}; \omega(a_2, b_2, c_2) \big| x_{2j} = \bar{x}\big] - P\big[g_{1j}; \omega(a_1, b_1, c_1) \big| x_{1j} = \bar{x}\big]}{g_{2j} - g_{1j}}.$$

This difference quotient provides a general expression for the parameter estimated in the hedonic program evaluation literature.

Because ϕ_j depends on two (potentially different) price functions, it is not the measure

---

[12] The distribution of physical housing types (i.e. x-types) is also an equilibrium outcome of this model. Since its distribution just depends on model primitives, we suppress it to avoid clutter. Alternatively, (7) could be written as $P\big[g, x; \omega(a, b, c, d(a, b, c))\big]$, where the vector d describes the distribution of housing types.

of MWTP from Rosen (1974). To convert ϕ_j into MWTP, we must restrict preferences

and technology to assure that the capitalization effect 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 the distribution of g is small, then we need only restrict

$a_1 = a_2$ and $b_1 = b_2$. Under this condition, the difference quotient in (9) approaches the

partial derivative in (4a) as $g_{2j} - g_{1j}$ approaches zero for all j.[13] In the limit, pre-shock

MWTP equals post-shock MWTP which equals the capitalization effect. This is intuitive.

An infinitesimal change in a single attribute will not alter the shape of the hedonic price

function; equilibrium prices will simply increase by MWTP. This special case is con-

sistent with Palmquist's (1992) characterization of localized externalities.

However, as noted earlier, empirical studies typically analyze large shocks. In this case,

three restrictions are jointly sufficient to establish a welfare interpretation for the capitali-

zation effect. We state this formally as

ASSUMPTION 1.

  i. $a_1 = a_2$ and $b_1 = b_2$.

  ii. $\partial P(g, x; \omega)/\partial g = f(x, \omega)$.

  iii. $\partial \omega/\partial g = 0$.

Condition (*i*) restricts preferences, income, and technology to be constant over the duration

of the study. Condition (*ii*) implicitly restricts the shapes of supply and demand curves so

that the marginal price of g does not depend on its level. Condition (*iii*) further restricts

---

[13] Proof of this statement follows immediately from the definition of a derivative.

supply and demand such that changes in g do not affect the hedonic gradient. We discuss violations of each condition after proving the theorem.

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

Proof. Consider any house, j, with characteristics $x_j = \bar{x}$ for which g_j changes from g_{j1} to g_{j2}. Since $a_1 = a_2$, $b_1 = b_2$, and $\partial \omega / \partial g = 0$, we know that $\omega_1 = \omega_2$. Combining this result with the assumption that $\partial P(g, x; \omega) / \partial g = f(x, \omega)$ implies $f(\bar{x}, \omega_1) = f(\bar{x}, \omega_2)$. It follows from the Mean Value Theorem that $\phi_j = f(\bar{x}, \omega_1) = f(\bar{x}, \omega_2)$. 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.

Alternatively, if assumption 1 is violated, the Mean Value Theorem generally implies

$$\phi_j \neq \partial P(g_{1j}, x_{1j}; \omega_1) / \partial g \neq \partial P(g_{2j}, x_{2j}; \omega_2) / \partial g \,.$$

For example, suppose conditions (*ii*) and (*iii*) hold, but the shock to $T(g)$ coincides with a shock to income or information, changing what households are willing to pay for g. This example violates condition (*i*). Since the parameters defining the hedonic gradient depend on preferences and income, ω_2 may differ from ω_1, causing the hedonic gradient to adjust, driving a wedge between the capitalization effect and MWTP.

All else constant, the credibility of condition (*i*) declines as periods 1 and 2 grow further apart. The more time passes the greater the scope for preferences, income, and technology to change in ways that alter the market-clearing price function. Notice that the shock to $T(g)$ need not be large to produce a large wedge between capitalization and MWTP. Even

in the special case where the shock is marginal and localized, $a_1 \neq a_2$ or $b_1 \neq b_2$ can invalidate a welfare interpretation of a capitalization effect.

Now suppose conditions (*i*) and (*iii*) hold so that the price function is stable (i.e. $a_1 = a_2$, $b_1 = b_2$, and $\omega_1 = \omega_2$). Condition (*ii*) restricts the curvature of its gradient. This restriction avoids problems that can occur if the gradient depends on *g*. To see this, notice that movement along a nonlinear price function will generally change marginal prices. If an increase from g_{j1} to g_{j2} corresponds to a change in its price, then the capitalization effect cannot simultaneously equal ex ante MWTP and ex post MWTP, since the two measures of MWTP differ.[14] The strength of condition (*ii*) is underscored by Ekeland, Heckman, and Nesheim's (2004) finding that the hedonic gradient is generically nonlinear in *g*.

Finally, consider condition (*iii*). The only obvious restriction on market primitives that supports $\partial \omega / \partial g = 0$ is that the demand for *g* is perfectly elastic. If the demand is downward sloping, then a positive shock to *g* will decrease individual MWTP (changing ω). Utility should also be separable in *g* and *x*. Otherwise, a shock to the distribution of *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 security systems, fences, or proximity to city parks. These restrictions on own and cross-price elasticities also apply to elements of *x* that are subject to exogenous shocks. A

---

[14] Empirical capitalization studies typically assume $\partial \Delta P / \partial \Delta g = \phi$. If the standard model were generalized to allow for heterogeneous local average capitalization effects, then condition (*ii*) could also be relaxed. For example, if equation (2) were to include polynomial functions of Δg, then condition (*ii*) would adjust to match the highest order polynomial; e.g. $\partial^2 P(g, x; \omega) / \partial g^2 = f(x, \omega)$ in the case of a quadratic. We thank a referee for pointing this out.

change in the relative price of *any* attribute violates $\partial\omega/\partial g = 0$ and can drive a wedge between MWTP and the capitalization effect for *any other* attribute.

Conditions (*i*)-(*iii*) are obviously strong restrictions. They seem unlikely to be satisfied in most applications. If they are violated, then the hedonic gradient may be unstable, producing a wedge between the identified capitalization effect and the policy-relevant measure of MWTP.[15] We illustrate this with a brief example using a version of Tinbergen's (1959) linear-quadratic-normal model.

3.3. *Example: Linear-Quadratic-Normal Model*

Suppose the housing stock is fixed, utility is quadratic, and preferences and housing characteristics are normally distributed. These assumptions conveniently yield a closed-form linear expression for the equilibrium price function (Tinbergen 1959, Epple 1987). Specifically, define the utility from a house with attributes $k = [g, x]$ as

$$(10) \qquad U = -(k - \alpha)' \frac{\zeta}{2}(k - \alpha) + b,$$

where ζ is a positive definite diagonal scaling matrix. When k and α are both normally distributed such that $k \sim N(\mu_k, \sigma_k)$ and $\alpha \sim N(\mu_\alpha, \sigma_\alpha)$, the price function can be expressed as

---

[15] As noted earlier the three conditions are sufficient, but they are not strictly necessary. It is possible to construct examples where simultaneous violations of two or more conditions are exactly offsetting. This should not diminish their importance. To provide an analogy: selecting the right empirical specification and valid instruments is sufficient, but not necessary to identify causal parameters in applications of the instrumental variables model. It is possible to construct examples where the bias from invalid instruments is exactly offset by biases from measurement error. This certainly does not diminish the importance of omitted variable bias.

(11) $P(\cdot) = \psi' k + k' \dfrac{v}{2} k$, where $\psi = \zeta \left(\mu_\alpha - \sigma_\alpha^{0.5} \sigma_k^{0.5} \mu_k \right)$ and $v = -\zeta \left(I - \sigma_\alpha^{0.5} \sigma_k^{0.5} \right)$.

Notice that the reduced-form parameters describing the shape of the price function $\omega = [\psi, v]$ are themselves functions of the structural parameters describing the distributions of household preferences $(\mu_\alpha, \sigma_\alpha)$ and housing characteristics (μ_k, σ_k). The structure of this simple model clearly violates the last two conditions of Assumption 1.

Now consider a shock to g. Before the shock, $MWTP_1 = \psi_1 + v_1 k$. After the shock, $MWTP_2 = \psi_2 + v_2 k$. It follows from (11) that, in general, $\psi_1 \neq \psi_2$ and $v_1 \neq v_2$ so that $MWTP_1 \neq MWTP_2$. The rate at which the shock is capitalized into property values is

(12) $\phi = \dfrac{P_2\{\cdot\} - P_1\{\cdot\}}{g_2 - g_1} = \dfrac{\psi_2' k_2 + k_2' \dfrac{v_2}{2} k_2 - \psi_1' k_1 - k_1' \dfrac{v_1}{2} k_1}{g_2 - g_1}$.

Hence $MWTP_1$, $MWTP_2$, and approximations to MWTP based on ϕ will generally differ, with the signs and magnitudes of these differences depending on the values of the structural parameters and the change in g .[16]

While the parametric structure of the linear-quadratic-normal model helps to illustrate the mechanics underlying conflation bias, program evaluation studies in the hedonic literature aim to avoid making explicit parametric assumptions about consumer preferences by instead assuming a parametric form for the equilibrium price function. We follow this approach in the next section, deriving an expression for conflation bias in terms of the data and parameters of a standard reduced form model.

---

[16] We provide numerical examples in the supplemental appendix.

4. Sufficient Conditions for Capitalization Based Welfare Measurement

The linear price functions that describe market equilibria before and after an unexpected shock to the distribution of g are $p_1 = g_1\theta_1 + h_1\eta_1 + \varepsilon_1$ and $p_2 = g_2\theta_2 + h_2\eta_2 + \varepsilon_2$. Recall that h represents the subset of housing attributes observed by the analyst ($h \subset x$), while the econometric error term, ε, captures the effect of unobserved attributes. Parameter subscripts recognize that the shape of the function may have been altered by the shock to public goods and by concomitant changes in h, ε, $R(y,\alpha)$, and $S(\beta)$. Note that we do not take a stance on approximation error in the use of a linear functional form.[17] Since virtually all empirical studies use linear models, doing so here allows us to focus attention on the relationship between capitalization and MWTP. The results in this section should be viewed as a best-case scenario where the price function is specified correctly.

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

$$(13) \qquad \Delta p = (g_2\theta_2 - g_1\theta_1) + (h_2\eta_2 - h_1\eta_1) + \Delta\varepsilon.$$

In the special case where $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$, equation (13) reduces to the standard first-differenced estimator from (2). Alternatively, if $\theta_1 \neq \theta_2$ but we restrict $h_2\eta_2 = h_1\eta_1$ then (13) reduces to a simple Oaxaca decomposition: $\Delta p = \Delta g\theta_2 + g_1\Delta\theta + \Delta\varepsilon$.

More generally, we can apply the Frisch-Waugh Theorem to write the relationship between the estimator for the capitalization effect ($\hat{\phi}$) and the price function parameters de-

---

[17] For discussions of hedonic functional form see Cropper, Deck, and McConnell (1988), Ekeland, Heckman, and Nesheim (2004), and Kuminoff, Parmeter, and Pope (2010).

scribing MWTP (θ_1, θ_2) as:

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

where $r = \Delta g - \Delta h(\Delta h' \Delta h)^{-1}\Delta h'\Delta g$. Let z denote a valid instrument for Δg. The IV ana-log to (14) simply replaces the Δg's in r with $\Delta\hat{g} = z(z'z)^{-1}z'\Delta g$.

Equation (14) reports what we can expect to learn about MWTP from estimating (2) when (13) is the true model. The IV estimator for the capitalization effect, $\hat{\phi}_{IV}$, depends on all of the parameters of the price functions that precede and follow the shock, as well as correlations between levels and changes in housing characteristics. The first term to the right of the equality in (14) is a parameter defining MWTP in the new equilibrium.[18] The second term is a "price effect" arising from a change in the implicit price of g. The third term is a "substitution effect" arising from changes in the implicit prices of other housing attributes that affect utility and, in some sense, serve as substitutes for g. The last term reflects the bias from correlation between changes in observed and unobserved variables.

The implicit price and substitution effects arise when the hedonic price function acts as the market clearing mechanism that Rosen described, adjusting to clear the market following the change in $T(g)$ and any concomitant changes in market primitives. Summing the price and substitution effects defines the conflation bias in interpreting a capitalization effect as a parameter of a hedonic price function. The direction of the bias is indeterminate.

---

[18] Applying Rosen's FOC to the price function defines MWTP for the occupant of house j by a $(\theta_2, p_{2j}, g_{2j})$ triplet, with the exact formula determined by the scaling of the variables in the hedonic price function.

Using $\hat{\phi}_{IV}$ to predict MWTP at a house j may produce an estimate that falls outside the range of values for the true MWTP for the occupants of j in the pre-shock and post-shock equilibria.[19,20] To establish a mapping between capitalization effects and welfare measures, some additional restrictions will be needed.

At least two sets of conditions are sufficient to translate capitalization effects into MWTP. The first set of conditions follows directly from assumption 1. If assumption 1 is satisfied, the hedonic gradient must be time-constant. Adding the usual orthogonality restriction on the econometric error term gives us

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

Under these restrictions, equation (14) reduces to $\hat{\phi}_{IV} = \theta_1 = \theta_2$. In this case the capitalization model (2) can be used to develop an unbiased estimator of ex ante MWTP which equals ex post MWTP. If estimation of single-period price functions is possible, the time-constant gradient assumption can be tested.

The second set of conditions replaces TCGA with additional restrictions on the data. It can be seen from (14) that $\hat{\phi}_{IV} = \theta_2$ under the following conditions

---

[19] For example, consider a quality improvement that decreases MWTP without affecting the control variables or their marginal prices. In this case, (14) implies that $\hat{\phi} < \theta_2 < \theta_1$ if $\Delta g' g_1 > 0$. Alternatively, $\theta_2 < \theta_1 < \hat{\phi}$ if $\Delta g' \Delta g < -\Delta g' g_1$.

[20] From the perspective of welfare measurement, conflation bias is more problematic than the standard complications with interpreting local average treatment effects (LATE). In the presence of heterogeneous treatment effects, LATE can identify parameters that are "structural" in the sense that they are invariant to policy changes operating through z (see Heckman 2010). In contrast, capitalization effects for public goods are not policy invariant. The market clearing function of the hedonic gradient makes capitalization effects endogenous to changes in implicit prices of non-market goods that will, in turn, vary with the policy change operating through z.

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

If the instrument is randomized in the sense that it is orthogonal to the initial level of the public good, and to the initial levels of the control variables, and to changes in those variables, then the capitalization effect identifies MWTP in the post-shock equilibrium, even if the gradient changes.[21] This identification argument is implicit in regression discontinuity designs such as Greenstone and Gallagher (2008), where the instrument is a policy that induces a Δg "treatment" for observations that lie above a certain g_1 threshold. If treatment is randomized with respect to h_1 and Δh, then focusing on observations in a very small neighborhood around the threshold may approximately identify ex post MWTP.

However, the policy relevance of ex post MWTP still depends on the nature of the instrument and the evolution of the hedonic gradient. For example, consider a policy that produces a large improvement in g, driving MWTP to zero. Knowing θ_2 (but not θ_1) does not allow us to distinguish the hypothesis that people were made better off by the policy from the alternative hypothesis that people were indifferent to the improvement that occurred. Now imagine a second random event causes g to deteriorate, increasing MWTP. Data from periods two and three could be combined to recover θ_3. In principle, θ_2 and θ_3 could be used to develop a linear approximation to a MWTP function for g,

---

[21] The requirement that z be orthogonal to Δh arises because of the potential endogeneity of Δh in equation (2) when (13) is the true model (e.g. Δh may be correlated with the h_1 term that is omitted from (2)). It follows that if z is correlated with some of the elements of Δh then the IV capitalization model will generally provide an inconsistent estimator for ex post MWTP. Condition (16) can be relaxed to allow correlation between z and Δh if additional instruments are available for the endogenous elements of Δh, consistent with the usual logic for two stage least squares estimation with multiple endogenous variables.

providing a more credible foundation for policy analysis.[22]

In summary, equations (15)-(16) define sufficient conditions for translating capitalization effects into welfare measures.[23] Econometric consistency is established, in part, through an assumption of temporal stability in the hedonic gradient. If this assumption is valid, then the identified parameters can be translated into ex-ante MWTP, which equals ex-post MWTP. If the gradient changes but an instrument randomizes the public goods treatment, then a capitalization effect can be translated into ex-post MWTP. Yet, the scope for using ex-post MWTP in policy evaluations depends on the evolution of the gradient. Thus, studying the evolution of price functions is essential to understanding the mapping between capitalization effects and the willingness to pay for public goods.

5. Evidence on the Evolution of Hedonic Price Functions

The 40-year history of research on valuing school quality is a microcosm for the broader literature on using housing markets to value public goods.[24] Because a household's access to a public school has traditionally been determined by whether the household lives in the attendance zone for that school, property values should reflect what parents are willing to pay for their children to attend schools where students score higher on standardized tests.[25]

---

[22] The approach suggested here would serve as a quasi-experimental analog to Palmquist's (1988) proposal for using hedonic price functions to calculate welfare measures for quality changes.

[23] This is similar to Chetty's (2009) "sufficient statistics" for quasi-experimental welfare measurement

[24] Kain and Quigley (1975) is among the early contributions. Recent applications include Downes and Zabel (2002), Figlio and Lucas (2004), Reback (2005), and Bayer, Ferreira, and McMillan (2007).

[25] The American Housing Survey provides strong evidence that school quality affects where many movers decide to live. Between 14% and 18% of recent homebuyers surveyed between 2003 and 2007 specifically cited "good schools" as one of the reasons they chose to move into their neighborhood. A steady 7% cited good schools as the main reason.

Early studies appeared to confirm this intuition. Then researchers noted a potential source of confounding—schools with higher test scores tend to be located in more exclusive neighborhoods. Subsequent studies refined the research design to mitigate confounding from omitted neighborhood amenities. This work began with Black (1999). She argued that school quality shifts discretely as one crosses an attendance zone boundary, but other amenities do not. Therefore, the composite price effect of all unobserved amenities that are common to houses 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 similar houses 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 racial composition also tends 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 the demographic composition of the neighborhood in order to isolate the implicit value of academic performance.[26]

We use the boundary discontinuity design for valuing school quality to estimate single-year price functions for five metropolitan areas at five-year intervals. Then we calculate

---

[26] In addition to refining Black's (1999) reduced-form estimation strategy, Bayer, Ferreira, and McMillan (2007) also develop and estimate a random utility model of sorting behavior.

MWTP for school quality in each year, test TCGA, and compare estimates for MWTP to capitalization effects following changes in test scores that occurred over the first four years of the No Child Left Behind Act (NCLB). Throughout the application, we follow the data collection and econometric procedures outlined by Black (1999) and Bayer, Ferreiria, and McMillan (2007). Readers are referred to their papers for additional background. The remainder of this section briefly summarizes NCLB and the data sets we have assembled.

5.1. *No Child Left Behind*

The No Child Left Behind Act was one of the most sweeping reforms in the recent history of public education. Beginning in 2003, it required states 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 school is making "adequate yearly progress" toward the goal of having every student attain state-specific standards for minimum competency in reading and math by 2014. Schools that do not make adequate yearly progress face a series of repercussions.

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 tracking performance, publicizing results, and sanctioning poorly performing schools. Critics argue that NCLB creates perverse incentives to teach to the test, to lower standards, to expel poorly performing students, or even lie when reporting scores. Several authors have investigated these issues. The emerging consensus seems to be that NCLB has improved performance, despite its flaws. For example, Dee and Jacob (2011) identify the impact of NCLB on test scores from the National Assessment of Education Progress

(NAEP). A key feature of their research design is that changes in NAEP scores should be unaffected by the perverse incentives of NCLB. They find that NCLB produced large and broad gains in NAEP math scores of 4th and 8th graders, especially in the bottom decile of the achievement distribution.[27] These results suggest that the upward trend in NCLB scores is consistent with alternative metrics for judging school quality.

NCLB scores are the official source of public information about school quality, and they are easily observed. Every school is required to track the share of its students who achieve proficiency in each subject. Results are mailed to parents and posted on websites such as greatschools.org.

5.2. *Ten Boundary Discontinuity Designs*

We estimate housing price functions for the metropolitan areas of Portland OR, Fairfax County VA, Philadelphia PA, Detroit MI, and Los Angeles CA during the 2003 and 2007 school years. After an exhaustive search over prospective study areas, these five were chosen because they had: (i) a large number of boundary zones; (ii) a large number of housing transactions; and (iii) data on NCLB scores in 2003 and 2007.[28]

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 schools based on the attendance zones where their parents live. However, school-specific assignment is no longer the norm. Since the mid-

---

[27] Mean increases in the NAEP math test scores were approximately 1-8 points from the start of NCLB to 2007 for 4th and 8th grade math scores. In a related study, Neal and Schanzenbach (2011) find that NCLB increased reading and math scores for fifth graders in the middle of the achievement distribution in the Chicago Public School system.

[28] States were not required to start reporting test scores until 2006. Some states did not report scores in 2003.

1990s, there has been an explosion of state and local regulations mandating "open enroll-ment" at the school district level. In an open enrollment area, parents are free to send their children to any public school within the district. There is evidence that parents take ad-vantage of these laws by sending their children to schools outside the zone where their house is located (Reback 2005). Philadelphia, Detroit, and Los Angeles all have open en-rollment policies. For these areas, our identification strategy is based on the relationship between property values and test scores on opposite sides of district boundaries. [29]

Implementing the boundary discontinuity design at the district level requires taking a weighted average over the 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 can vary discretely across districts. District boundaries may also be more likely than attendance zone boundaries to overlap with features of the urban landscape. Therefore, we control for property tax rates and we use visual inspection to ex-clude boundaries that overlap with landscape features such as rivers and highways.

5.3. *Data and Summary Statistics*

We assembled data on test scores, neighborhood characteristics, and houses sold during the 2003 and 2007 school years. [30] The scores that we use are combined rates of math and

---

[29] Dhar and Ross (2009) discuss tradeoffs between studying attendance zone boundaries and district boundaries. For example, district boundaries may have the advantage of being perceived as more permanent by homebuyers, but they are also more likely to overlap with property tax rates and neighborhood demographics, underscoring the importance of controlling for these variables in a regression. While interesting, the distinction between attendance zone and district boundaries does not affect our main empirical findings.

[30] 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. These definitions reflect the fact that NCLB scores and school grades for the preceding 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—

reading proficiency reported by states under NCLB. We matched each housing sale with lagged scores for the relevant school or school district.[31]   Houses sold during the 2003 school year were matched with scores from the 2002 school year, for example. We will refer to the lagged scores as the "2003 score" and "2007 score" from here on.

Table 1 reports the 2003 NCLB scores and 2007-2003 differences for the 10th, 50th, and 90th percentiles of schools in each study area. In Fairfax, for example, math/reading scores in the bottom 10th percentile increased by an average of 11 points (or 14%) with a standard deviation of 8 points. The corresponding changes for the other four areas are all positive and typically large.[32] There are smaller gains (and even losses) at the middle and 90th percentiles. These statistics are consistent with Dee and Jacob's (2011) finding that NCLB had the biggest impact on schools that began the program with the lowest scores.

<< Insert Table 1 Around Here >>

The remaining components of the data were collected from various sources. Sale prices and physical attributes of every house sold during the 2003 and 2007 school years were purchased from *DataQuick*. Tax rates were calculated using assessment data from public records. Finally, each house was matched with data on the demographic composition of residents living in the Census block group.[33]

---

test scores—to influence home buying decisions.

[31] 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.

[32] Scores are not directly comparable across study areas because each state has its own testing system.

[33] Annual block group data were obtained from Geolytics. Their data are developed using information from the decennial Census,

Table 2 reports summary statistics for Fairfax County, VA. Columns 1-2 report means and standard deviations for every variable in the final data set. In 2003 the average house sold for approximately $567,000. 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.[34] This small change in the average masks considerable heterogeneity across schools (table 1). The average house 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 18 years of age, 85% of houses were owner occupied, 1% of houses were vacant, and 0.37 was the normalized measure of population density. The average ratio of assessed to taxed value called a "tax rate" was 112.

<< Insert Table 2 Around Here >>

Columns 3-5 summarize the subsample that we use in the boundary discontinuity regressions. Column 3 reports means over houses located within 0.2 miles of a boundary. While this cuts the sample in half, there are almost no changes in the characteristics of the average house (comparing columns 1 and 3). Column 4 reports the difference in mean characteristics of houses located on the "high score" and "low score" sides of a boundary, and column 5 reports T-statistics on the differences. Differences in scores are large and statistically significant whereas differences in housing characteristics tend to be small and insignificant. Like Bayer, Ferreira, and McMillan (2007), we find differences in the racial

---

annual Census surveys, postal records, and actuarial tables of births and deaths.

[34] The mean 2003 score is slightly different than the corresponding mean in Table 1. This is because Table 2 scores are weighted by enrollment whereas Table 1 is weighted by housing transactions.

composition of homeowners on the high and low-score sides of a boundary. This under-scores the importance of controlling for demographics during the estimation.

Columns 6-7 report means and standard deviations for the average house in each Census block group. Because there are too few repeated sales of individual houses to estimate a first differenced model using micro data, we use the block group data to estimate capitalization effects for changes in test scores between 2003 and 2007. Notice that aggregation does not substantially change the summary statistics relative to the micro data. Finally, columns 8-9 report correlations between the changes in test scores and levels and changes in all other variables.

The Fairfax county data illustrate several features that are also common to the data for Portland, Philadelphia, Detroit, and Los Angeles: (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 boundaries; (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. Summary statistics for each area are provided in the online appendix.

6. Results

6.1. *Single-Year Hedonic Regressions*

After pooling data from 2003 and 2007, we estimate the following specification for the hedonic price function:

$$(17) \qquad \ln(price_j) = g_j\theta + g_j D_j \theta_{07} + h_j\eta + h_j D_j \eta_{07} + BFE_{j,03} + BFE_{j,07} + \varepsilon_j,$$

where g_j denotes the log of the NCLB test score for the year prior to the sale of house j, D_j is an indicator that equals 1 if the sale occurred in 2007, h_j is a vector containing the physical attributes of house j along with its neighborhood demographics and tax rate, and $BFE_{j,03}$ and $BFE_{j,07}$ are year-specific boundary fixed effects. The boundary regions are 0.2 mile areas that overlap adjacent attendance zones (Fairfax, Portland) or districts (Philadelphia, Detroit, Los Angeles).[35] Notice that equation (17) allows the price function gradient to change between 2003 and 2007. In the special case where the gradient is time constant, $\theta_{07} = \eta_{07} = 0$. An F-test of this hypothesis provides a formal test of TCGA.

We begin by using the sample of houses that sold within 0.2 miles of a boundary. Panels A and B of table 3 report OLS estimates of θ and θ_{07} from regressions with and without boundary fixed effects. Since NCLB scores are measured in logs, their coefficients are elasticities. For example, the results in column 2 indicate that the prices of houses sold in Portland during 2003 were 0.456% higher in attendance zones where NCLB scores were 1% higher. The elasticity is very similar for school districts in Philadelphia (column 3). Notice that Philadelphia is one of four areas to have a significant increase in the price elasticity. It increased from 0.481 in 2003 to 0.710 in 2007 (0.481 + 0.229). Overall, panel A provides tentative evidence that (i) NCLB scores capture a dimension of school quality that matters for property values and; (ii) the functional relationship between NCLB scores and property values changed over the duration of our study.

<< Insert Table 3 Around Here >>

---

[35] Our main results are unaffected by using boundary regions of 0.35 or 0.15 miles instead.

The evidence in panel A is tentative because we have not controlled for correlation between school quality and unobserved amenities. Positive correlation seems likely. To see this, first note that household income is a strong predictor of a child's academic performance.[36] Now consider a household's location choice problem. If homebuyers appreciate low crime rates, access to parks, and scenic views, they will bid up prices in neighborhoods with those amenities. Wealthier parents who can afford to live in high-amenity neighborhoods will have children who perform better on standardized tests. Therefore, our inability to control for crime, parks, and views will produce an upward bias on the OLS estimator for the test score coefficient. Boundary fixed effects can mitigate this problem by absorbing the price effect of unobserved amenities in each boundary zone.

Panel B reports regression results after adding boundary fixed effects. Consistent with intuition, the coefficients of variation increase and the test score coefficients decrease.[37] Comparing panels A and B reveals that boundary fixed effects decrease most of the elasticities by more than 50%.

NCLB scores are not directly comparable across states because each state develops its

---

[36] 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.

[37] 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. Results are suppressed for brevity and will be provided upon request. 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.

own tests. Nevertheless, since the state-specific scores represent different proxy measures of the same underlying variable—school quality—they can be compared in terms of a common proportionate change. The 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 (1999) reports an increase of 0.42% for Boston in 1993-1995 and the results from Bayer, Ferreira, and McMillan (2007) indicate an increase of 0.12% for San Francisco in 1990.

In 2007 our range of point estimates for the test score elasticity is wider: 0.04 to 0.57. These estimates are calculated by summing the baseline coefficient for 2003 and the differential for 2007. The changes are large and significant for Fairfax, Portland, Detroit, and Los Angeles. Several factors may be contributing to these changes, including: (i) changes in NCLB scores; (ii) changes in wealth as housing values and assets grew during the housing boom; (iii) the information shock created by the format for tracking performance 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 and other potential contributing factors would require estimating a demand system for school quality and other attributes of houses and neighborhoods—a challenge that we leave for future research.[38] Regardless of what drives the temporal instability of the hedonic price function, the large changes in the test score coefficients signal that the hedonic gradient changed, violating TCGA. Moreover, changes in other coefficients are large enough to reject the

---

[38] See Epple (1987), Ekeland, Heckman, and Nesheim (2004), and Kuminoff, Smith, and Timmins (2013) for an overview of the challenges with hedonic demand estimation.

hypothesis of a time-constant gradient for every metro area (F-tests are reported in panel B). Philadelphia is the only area with a p-value near the 0.05 threshold. These results clearly indicate the presence of conflation bias.

6.2. *Capitalization Effects Measured Over 5-Year Intervals*

To assess the magnitude of conflation bias, we use capitalization effects to calculate another set of test score elasticities. We regress price changes on changes in test scores, treating the average house in each block group as an observation.[39] The control variables include changes in tax rates, changes in the characteristics of residents living in each block group, and changes in the physical characteristics of the average house sold within the block group. Differencing the data purges omitted characteristics of block groups that are constant between 2003 and 2007.

Panel C of table 3 reports results based on the full sample of block groups. Los Angeles is the only area where the capitalization effect implies an elasticity (0.17%) within the range defined by the parameters of the 2003 and 2007 price functions (0.14% to 0.22%). In Fairfax, Portland, Philadelphia, and Detroit, the capitalization effects are far below the lower bound of point estimates from single year price functions. The implied elasticity is at least positive and marginally significant in Philadelphia. In Fairfax and Portland, the elasticities are close to zero. In Detroit the estimated capitalization effect is negative and marginally significant. This could reflect specification error in the linear form of the esti-

---

[39] As noted earlier, there are too few repeat sales to support a micro data analysis. Our use of block group averages provides greater resolution than recent studies that defined the unit of observation as a census tract median or a county average (Chay and Greenstone 2005, Greenstone and Gallagher 2008, Baum-Snow and Marion 2009).

mating equation, but the hedonic estimates in column 9 seem plausible by contrast.[40]

One might worry that the results in panel C are confounded by omitted variables. As we noted earlier, schools with lower test scores in 2003 tended to experience larger increases in test scores. These increases may not be exogenous. If changes in unobserved attributes of block groups are negatively correlated with changes in scores, then our estimators for capitalization effects may be biased downward. Recall that we control for changes in property taxes and changes in observable characteristics of block group populations. This means that any confounding must come from changes in unobserved variables that co-vary with changes in NCLB scores across the block groups within a metro area (e.g. crime rates). While this is certainly possible, it seems unlikely that localized amenities (other than school quality) would change sharply as we cross an attendance zone boundary. Based on this logic, adding boundary fixed effects to the regression will mitigate potential confounding by absorbing the capitalization of changes in all unobserved variables that are common to both sides of a 0.2-mile boundary zone.

To implement our panel data version of the boundary discontinuity design we first drop all houses located more than 0.2 miles from a boundary. Then we aggregate the micro data on either side of each boundary. Finally, we add fixed effects for each boundary zone and

---

[40] As an exploratory exercise, we repeated the estimation using a Oaxaca decomposition that added baseline test scores as controls: $\Delta p = \Delta g \theta_2 + g_1 \Delta \theta + \Delta h \gamma + \Delta \varepsilon$. This specification replaces the assumption that $\Delta \theta = 0$ with the assumption that $E[\Delta \varepsilon \mid g_1] = 0$. Doing so improved our results for Fairfax and Detroit. The implied test score elasticities were positive and significant for Fairfax in 2007 and Detroit in 2003 and 2007, though they still understated our preferred estimates from panel B by 45% to 62%. However, results for the other three metro areas were statistically insignificant, making it difficult for us to draw general conclusions about the relative performances of the two models. Investigating Oaxaca-type extensions of the standard capitalization model would be an interesting area for further research.

estimate the resulting first-differenced model,

$$(18) \qquad \Delta \ln(price_j) = \Delta testscore_j \phi + \Delta h_j \gamma + BFE_j + \Delta \varepsilon_j.$$

This specification uses the same geography and the same fixed effects as the cross-section regressions. If changes in omitted variables are negatively correlated with changes in test scores, then we would expect the capitalization effects to increase. Results are reported in panel D. While standard errors on the capitalization-based elasticities have increased due to the decrease in sample size and the inclusion of fixed effects, the point estimates are very similar to our baseline results. The estimates in columns 16-20 all fall within 95% confidence intervals of the estimates from columns 11-15. Thus, we do not find strong evidence of confounding from time-varying omitted variables.

Overall, comparing the results in panels A and C of table 3 (without boundary fixed effects) and comparing the results in panels B and D (with boundary fixed effects) suggests that the price function gradients changed over time and that these changes created wedges between capitalization effects and hedonic price function parameters. We ran additional robustness checks to investigate the possibility that these wedges are influenced by aggregation bias (moving from micro data to block groups or boundary zones) and/or sample selection bias (moving from the full metro area to 0.2 mile boundary zones) and found evidence against both explanations. Details are provided in the online appendix.

6.3. *Implications for Welfare Measurement*

The results from our single-year regressions suggest that hedonic price functions adjusted to changes in housing market conditions. These changes matter for evaluating the bene-

fits of public education. Table 4 provides a summary comparison between hedonic and capitalization based estimates for the average resident's willingness to pay for a 1% increase in NCLB scores. Each column reports the MWTP predicted by a specific model, averaged over the samples from all five study regions. In columns 1-3 we do not control for omitted variables. The resulting predictions are fairly robust to how we define a data point (house, 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 predictions from models using boundary fixed effects to mitigate confounding (column 4).

<< Insert Table 4 Around Here >>

The boundary discontinuity design in column 4 is our preferred specification. It mitigates confounding; it controls for race-based sorting; and it predicts MWTP using data for a single school year, which seems consistent with the static description of equilibrium in Rosen's model. It implies the average household would have been willing to pay $536 (year 2000 dollars) for a 1% improvement in school quality in 2003. Looking across metro areas, average MWTP ranges from $422 for Detroit to $743 for Philadelphia. This range lies within the range of estimates for San Francisco ($372) and Boston ($917) reported by Black and Bayer, Ferreira, and McMillan.

There were several changes to housing markets between 2003 and 2007. Property values increased by 6% on average, test scores increased by 10% on average, and there were smaller changes in the demographic compositions of neighborhoods. There was also steady media coverage of the NCLB program and changes to the national economy that would have affected expectations about future wealth (e.g. rapid growth in stock market

indices and personal income). These changes were accompanied by changes in hedonic gradients which, in turn, increased our prediction for average MWTP to $688 in 2007.

In contrast, capitalization effects suggest much smaller measures of MWTP. Column 5 reports the average MWTP predicted by the simple first-differenced model ($134 in 2003, $152 in 2007).[41] These figures are about ¼ the size of estimates from single-year boundary discontinuity regressions! The difference only narrows slightly when boundary fixed effects are used to mitigate confounding by time-varying omitted variables (column 6). Overall, these results seem to confirm the predictions from our theoretical and econometric models. Shocks to the spatial distribution of public goods and changes in market fundamentals cause the hedonic price function to adjust, driving a wedge between capitalization effects and welfare measures.

7. Conclusion

Rosen's (1974) static hedonic model provides a starting point for developing revealed preference estimates of the willingness to pay for public goods and externalities. A recent wave of empirical research has sought to improve the credibility of these estimates by refining conventional research designs to mitigate confounding by omitted variables. The leading strategy uses panel data to identify how exogenous shocks to public goods are capitalized into property values. Unfortunately, credible estimates for capitalization effects do not generally provide credible measures of consumer welfare. We have shown that moving to a capitalization framework changes the economic interpretation of the identified pa-

---

[41] These figures were calculated by combining results from columns 11-15 in table 3 with data on average property values and populations in tables 2 and A3-A6.

rameters. This change occurs when the price function adjusts to clear the market following shocks to the distributions of public goods, preferences, wealth, or technology. Capitalization effects conflate these temporal adjustments to the price function with the spatial price differentials of hedonic equilibria. Our application suggests the resulting bias can be serious, causing capitalization effects to understate the willingness to pay for improved school quality by as much as 75%.

Conflation bias is a potential problem for any panel data estimator that ignores temporal changes in the economic parameters of the underlying cross-section model. For example, Banzhaf and Walsh (2013) use a Oaxaca decomposition to characterize conflation bias in differences-in-differences estimates for the effect of changes in public goods on the racial composition of neighborhoods. Their simulation illustrates how a standard program evaluation model can fail to identify the correct sign of a policy-relevant parameter.

A key challenge for future research is to develop research designs that mitigate endogeneity problems without undermining the economic interpretations of the identified parameters. One strategy is to carefully model the mechanisms that cause variables to be endogenous within a well posed structural model of the sorting process (e.g. Epple, Romano, and Sieg 2006, Ferreyra 2007). Another option is to adapt tools from the program evaluation literature to refine the design of a structural sorting model (e.g. Bayer, Ferreira, and McMillan 2007, Galiani, Murphy, and Pantano 2012). A third possibility is to refine the econometric tools of the program evaluation literature to mitigate omitted variable bias in a way that maintains a consistent link to an equilibrium description of the structural model. Our application to valuing public school quality illustrates how this can be done. We sus-

pect that similar approaches can be adapted to provide more credible hedonic measures of the willingness to pay for other public goods and externalities.

AUTHOR AFFILIATION

Kuminoff: Arizona State University, U.S.A.

Pope: Brigham Young University, U.S.A.

REFERENCES

BANZHAF, H.S. AND R. P. WALSH, "Do People Vote with Their Feet? An Empirical Test of Tiebout's Mechanism," *American Economic Review* 98 (2008), 843-63.

_____, AND R. P. WALSH, "Segregation and Tiebout Sorting: The Link between Placed-based Investments and Neighborhood Tipping," *Journal of Urban Economics* 74 (2013), 83-98.

BAUM-SNOW, N., AND J. MARION, "The Effects of Low Income Housing Tax Credit Developments on Neighborhoods," *Journal of Public Economics* 93 (2009), 654-66.

BAYER, P., F. FERREIRA, AND R. MCMILLAN, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy* 115 (2007), 588-638.

_____, R. MCMILLAN, A. MURPHY, AND C. TIMMINS, "A Dynamic Model of Demand for Houses and Neighborhoods," NBER Working Paper #17250, 2011

_____, AND C. TIMMINS, "Estimating Equilibrium Models of Sorting across Locations," *The Economic Journal* 117 (2007), 353-74.

BERON, K., J. MURDOCH, AND M. THAYER, "The Benefits of Visibility Improvement: New Evi-

dence from the Los Angeles Metropolitan Area," *Journal of Real Estate Finance and Economics* 22 (2001), 319-337.

BIN, O., C. E. LANDRY, AND G. E. MEYER, "Riparian Buffers and Hedonic Prices: A Quasi-Experimental Analysis of Residential Property Values in the Neuse River Basin," *American Journal of Agricultural Economics* 91 (2009), 1067-79.

BISHOP, K. C. AND A. D. MURPHY, "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *American Economic Review* 101 (2011), 625-629.

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

BROOKSHIRE, D. S., M. A. THAYER, J. TSCHIRHART, AND W. D. SCHULZE, "A Test of the Expected Utility Model: Evidence from Earthquake Risks," *Journal of Political Economy* 93 (1985), 369-89.

CALABRESE, S., D. EPPLE, T. ROMER, AND H. SIEG, "Local Public Good Provision: Voting, Peer Effects, and Mobility," *Journal of Public Economics* 90 (2006), 959-981.

CARD, D., A. MAS, AND J. ROTHSTEIN, "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics* 123 (2008), 177-218.

CELLINI, S. R., F FERREIRA, AND J. ROTHSTEIN, "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *Quarterly Journal of Economics* 125 (2010), 215-261.

CHAY, K.Y. AND M. GREENSTONE, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy* 113 (2005), 376-424.

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

CROPPER, M. L., L. B. DECK, AND K. E. MCCONNELL, "On the Choice of Functional Form for

Hedonic Price Functions," *Review of Economics and Statistics* 70 (1988), 668-75.

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

DEE, T., AND B. JACOB, "The Impact of No Child Left Behind on Student Achievement," *Journal of Policy Analysis and Management* 30 (2011), 418-446.

DHAR, P., AND S. L. ROSS, "School Quality and Property Values: Re-examining the Boundary Approach," University of Connecticut working paper #2009-37, 2009.

DOWNES, T. A., AND J. E. ZABEL, "The Impact of School Characteristics on House Prices: Chicago 1987-1991" *Journal of Urban Economics* 52 (2002), 1-25.

EKELAND, I., J. J. HECKMAN, AND L. NESHEIM, "Identification and Estimation of Hedonic Models," *Journal of Political Economy* 112 (2004), S60-S109.

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

_____, R. ROMANO, AND H. SIEG, "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education," *Econometrica* 74 (2006), 885-928.

_____, T. ROMER, AND H. SIEG. "Interjurisdictional Sorting and Majority Rule: An Empirical Analysis." *Econometrica* 69 (2001), 1437-1465.

_____, AND H. SIEG. "Estimating Equilibrium Models of Local Jurisdiction." *Journal of Political Economy*, 107 (1999), 645-81.

FERREYRA, M. M, "Estimating the Effects of Private School Vouchers in Multi-District Economies," *American Economic Review* 97 (2007), 789-817.

FIGLIO, D. N., AND M. E. LUCAS, "What's in a Grade? School Report Cards and the Housing Market," *American Economic Review* 94 (2004), 591-604.

GALIANI, S., A. MURPHY, AND J. PANTANO, "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment," Working paper, 2012.

GOLDSTEIN, G. S. AND M. V. PAULY. "Tiebout Bias on the Demand for Local Public Goods." *Journal of Public Economics*, 16 (1981), 131-43.

GREENSTONE, M. AND J. GALLAGHER, "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *Quarterly Journal of Economics* 123 (2008), 951-1003.

HECKMAN, J.J, "Schools, Skills, and Synapses," *Economic Inquiry* 46 (2008), 289-324.

_____. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature* 48 (2010), 356-398.

HORSCH, E. J., AND D. J. LEWIS, "The Effects of Aquatic Invasive Species on Property Values: Evidence from a Quasi-Experiment," *Land Economics* 85 (2009), 391-409.

IMBENS, G. W., AND J. M. WOOLDRIDGE, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature* 47 (2009), 5-86.

KANEMOTO, Y., "Hedonic Prices and the Benefits of Public Policies," *Econometrica* 56 (1988), 981-989.

KAIN, J. F., AND J. M. QUIGLEY, "Housing Markets and Racial Discrimination." New York: National Bureau of Economic Research, 1975.

KUMINOFF, N. V., C. F. PARMETER, AND J. C. POPE, "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities?" *Journal of Environmental Economics and Management* 60 (2010), 145-160.

_____, V. K. SMITH, AND C. TIMMINS, "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets," *Journal of Economic Literature*, forthcoming.

LINDEN, L. AND J.E. ROCKOFF, "Estimates of the Impact of Crime Risk on Property Values from

Megan's Laws," *American Economic Review* 98 (2008), 1103-27.

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

NEAL, D. AND D. W. SCHANZENBACH, "Left Behind By Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics* 92 (2010), 263-283.

PALMQUIST, R. B., "Welfare Measurement for Environmental Improvements Using the Hedonic Model: The Case of Nonparametric Marginal Prices," *Journal of Environmental Economics and Management* 15 (1988), 297-312.

_____. "Valuing Localized Externalities," *Journal of Urban Economics* 31(1992), 59-68.

PARMETER, C. F. AND J. C. POPE. "Quasi-Experiments and Hedonic Property Value Methods," in J.A. List and M.K. Price, eds. *Handbook on Experimental Economics and the Environment* (Northampton, MA: Edward Elgar Publishing, 2013), 3-66.

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

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

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

RUBINFELD, D. L., P. SHAPIRO, AND J. ROBERTS. "Tiebout Bias and the Demand for Local Public Schooling." *Review of Economics and Statistics*, 69 (1987), 426-37.

SETHI, R. AND R. SOMANATHAN, "Inequality and Segregation," *Journal of Political Economy* 112 (2004), 1296-321.

SIEG, H., V. K. SMITH, H. S. BANZHAF, AND R. WALSH, "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods," *International Economic*

Review 45 (2004), 1047-77.

STARRETT, D. A., "Land Value Capitalization in Local Public Finance," *Journal of Political Economy* 89 (1981), 306-27.

TIEBOUT, C. M. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64 (1956), 416-24.

TINBERGEN, J., "On the Theory of Income Distribution," in L.H. Klaassen, L.M. Koych and H.J. Witteveen, eds., *Selected Papers of Jan Tinbergen*. (Amsterdam: North Holland, 1959), 243-263.

WALSH, R. L., "Endogenous Open Space Amenities in a Locational Equilibrium," *Journal of Urban Economics* 61 (2007), 319-44.

TABLE 1
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 2
SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND TEST SCORES IN FAIRFAX, VA

Fairfax County, 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)			
	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 houses 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 3
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 boundary sample without boundary fixed effects)				
	(1)	(2)	(3)	(4)	(5)
log (test score), 2003 coefficient	0.122 (0.027)	0.456 (0.020)	0.481 (0.045)	0.524 (0.036)	0.274 (0.012)
log (test score), 2007 differential	0.554 (0.056)	0.034 (0.032)	0.229 (0.067)	0.516 (0.086)	0.084 (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 boundary sample with boundary fixed effects)				
	(6)	(7)	(8)	(9)	(10)
log (test score), 2003 coefficient	0.116 (0.040)	0.200 (0.028)	0.272 (0.071)	0.208 (0.047)	0.140 (0.015)
log (test score), 2007 differential	0.293 (0.081)	-0.165 (0.048)	-0.120 (0.101)	0.357 (0.126)	0.075 (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 without boundary fixed effects)				
	(11)	(12)	(13)	(14)	(15)
change in log (test score)	-0.037 (0.073)	0.007 (0.096)	0.116 (0.068)	-0.289 (0.134)	0.174 (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
	D. Test Score Parameters from Capitalization Regressions (boundary zone data from 0.2 mile sample with boundary fixed effects)				
	(16)	(17)	(18)	(19)	(20)
change in log (test score)	0.008 (0.111)	-0.025 (0.091)	0.130 (0.180)	-0.445 (0.521)	0.231 (0.177)
R^2	0.83	0.82	0.91	0.87	0.83
Number of observations	404	603	176	213	251

NOTE.—All regressions use Eicker-White standard errors and include controls for property taxes, physical 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 house. All control variables are interacted with a dummy for sales made during the 2007-2008 school year. In cols. 11 through 20 the dependent variable is the change in the natural log of the average sale price in a census block group or a 0.2 mile boundary zone. See the text for additional details.

TABLE 4
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	boundary zone
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 3 and A1. For example, the estimates in col. 4 are based on the elasticities reported in cols. 6 through 10 of table 3.