Nonlinear Hedonics and the Search for School District Quality*

Abbigail J. Chiodo
Rubén Hernández-Murillo
and
Michael T. Owyang†

Research Department
Federal Reserve Bank of Saint Louis
411 Locust Street
Saint Louis, MO 63102

January 29, 2004

Abstract

In an influential paper studying school districts in the Boston area, Black (1999), using a linear boundary fixed effects model, showed that hedonic pricing functions can overestimate the premium on school quality. We extend her analysis using data from the St. Louis area and show the premium for good schools is non-linear and, consequently, underestimated by Black's methodology. Moreover, we demonstrate that, unlike in the linear model, there is no penalty for houses in worse-than-average school districts. These houses are priced based on comparables alone. We motivate our results by outlining a simple housing search model. [JEL: C21, I20, R21]

Keywords: education capitalization, house prices, hedonics, search

*We thank Leora Friedberg and Howard Wall for discussions and suggestions. Jeremy Bixby, Kristie M. Engemann, Mark Opitz, and Deborah Roisman provided research assistance. The authors are grateful to Tara Hansen and Paul Keppel at First American Real Estate for providing house pricing data and technical support. The views herein are the authors' alone and do not represent the views of the Federal Reserve Bank of Saint Louis or the Federal Reserve system. All remaining errors are the sole responsibility of the authors.

†Corresponding author: `owyang@stls.frb.org`.

1 Introduction

Since the pioneering work of Tiebout (1956), economists have recognized that the quality of public services, especially schools, influence house prices. Many empirical studies have attempted to discern the extent to which the quality of public education affects house prices. Initially, researchers estimated hedonic pricing equations (Rosen, 1974). In a simple hedonic pricing model, a house's value depends on its comparable neighborhood and school district characteristics. A house's comparable characteristics include aspects such as the number of bedrooms, square feet, etc. Neighborhood characteristics typically include the distance to the nearest major downtown area, racial composition, and median household income. Education quality may be proxied by variables such as per-pupil spending, pupil/teacher ratio, and property taxes, which are usually available at the school district level, or it may be measured directly by state or local standardized tests scores, which are usually available at the school level.[1]

In an influential study, Black (1999) argues that past research estimating hedonic pricing functions may introduce an upward bias due to neighborhood quality effects that are unaccounted for in the data. Specifically, she notes that better schools may be associated with better neighborhoods, which could independently contribute to higher house prices. Black circumvents this problem by estimating a linear hedonic pricing function using data only from houses which border the school attendance zone boundaries.[2] She rationalizes that, while test scores make a discrete jump at attendance boundaries, changes in neighborhoods are more smooth.

Black's linear specification presupposes that the marginal valuation of worse-than-average

[1] A number of studies have demonstrated the influence of school quality on performance (e.g., Card and Krueger, 1992; Betts, 1995). Hanushek (1996), however, argues that input proxies of school district quality may be inappropriate.

[2] A school's attendance zone delimits the geographic area around the school the residents' children would have to attend. In this text, we often refer interchangeably to a school's attendance zone as the school, but this term should not be confused with *school district,* which is an administration unit in the public-school system often comprising several schools. Although schools within the same school district will often have similar test scores on average, we measure scores at the individual school level.

schools is equal to the valuation of better-than-average schools and results in a constant premium on school quality. Moreover, if school quality is normalized (i.e., expressed in terms of deviations from the mean), the linear capitalization term implies a penalty (increasing as quality decreases) for houses in attendance zones of schools performing below average. Thus, a linear model implies there exists a substantive pecuniary penalty for a really bad school compared to just a bad school. In this paper, we formulate a simple housing search model that yields a theoretical nonlinear pricing function.

The nonlinearity in our model reflects two aspects of the market for public education via housing. First, alternative schooling arrangements (e.g., private school, home schooling, magnet schools, etc) can provide home buyers with high quality education even if they choose to live in below average school districts. The existence of these options underlies our belief that an increasing penalty for below average quality school attendance zones may be theoretically unappealing. Second, if buyers have positive valuations for education, they may concentrate their efforts among the highest quality attendance zones, yielding an increasing market tightness as school quality increases. Thus, buyers may face increased competition for the highest quality schools and a rapidly increasing premium for houses in those attendance zones.

Motivated by our theoretical specification, we extend Black's analysis and examine the relationship between school quality and house prices in the St. Louis, Missouri, metropolitan area.[3] A previous study by Ridker and Henning (1967) found no evidence of education capitalization in St. Louis house prices. While their main concern was to determine the negative effect of air pollution on housing prices, they included a dummy variable which indicated residents' attitudes about the quality of the schools (above average, average, and below average).[4]

[3] In this paper, we limit ourselves to school districts in St. Louis County, Missouri, as some variables were not available for St. Louis City and a large number of schools in the city have programs that allow children to cross school attendance boundaries.

[4] Ridker and Henning acknowledge that their study may suffer from a small sample bias that could explain this seemingly contradictory finding.

Our goal is to determine the degree of education capitalization in the St. Louis MSA. We first measure education capitalization employing Black's methodology of considering only houses near attendance zone boundaries to control for neighborhood quality. This allows us to determine the extent to which Black's results extend to the St. Louis metro area. Then, we advance Black's methodology by considering the possibility that education capitalization affects house prices nonlinearly, as indicated by our theoretical framework.

This paper is organized as follows. Section 2 provides a review of some of the extensive literature on education capitalization and neighborhood quality effects. Section 3 offers a motivating model of housing search. Section 4 describes the econometric methodology, focusing on the differences between Black's linear boundary fixed effects model and the nonlinear boundary fixed effects model we propose. Section 5 reviews the data used to calculate the school premium for real estate. Section 6 presents the results of each of the econometric specifications. First, we compare the linear model's estimate of the school premium to that implied by the nonlinear model. Then, we conduct a number of specification tests to ascertain which model is preferred. Section 7 concludes.

2 Recent Empirical Literature

Tiebout (1956) argued that people self-sort into a community which meets their public spending preferences. Thus, unlike the federal government, local governments do not tend to adapt their policies to current residents, but, rather, maintain set policy patterns. Based on these policy patterns set by various communities, prospective residents may select the area which best meets their needs and desires. Property values will, therefore, be determined by the desired level of public services, particularly school quality, that are capitalized into real estate prices.

Empirical tests of this degree of capitalization of services into house prices has met disparate levels of success. Although studies using hedonic pricing functions to estimate the value of school quality have shown a positive correlation between school quality and house prices, these results may be biased upwards due to unobservable neighborhood characteristics (see, for example, Bogart and Cromwell, 1997; and Hayes and Taylor, 1996). While these studies focus on small areas (a county in Ohio or the city of Dallas, respectively), it is debatable whether or not these areas are small enough to eliminate differences across neighborhoods.[5] Thus, we forgo further discussion of the previous empirical literature on pure hedonic models of capitalization and, instead, focus on the recent studies attempting to correct for neighborhood effects.

Black (1999) employed a model augmented with attendance zone boundary dummies, which were included to disentangle education quality from neighborhood quality. Essentially, these dummy variables account for fixed effects at attendance zone boundaries. We will henceforth refer to the Black model as the linear boundary fixed effects model (LBFM). These dummies take advantage of the discrete break in school quality across the attendance zone boundary and rely on the hypothesis that neighborhoods change more smoothly. Examining house price and elementary school test score data for large Boston suburbs, Black finds that controlling for neighborhood effects decreases the capitalization effect by about half. She discovers that parents are willing to pay 2.5 percent more for a 5 percent increase in test scores. For her sample, this computes to a \$3948 premium at a median house price of \$188,000 (this is approximately half her estimate from the hedonic model of \$9280 at the same house price).

Recently, researchers have used Black's methodology to control for neighborhood effects. Figlio and Lucas (2000) find that both test scores and school "letter grades" are capitalized into housing prices in Gainsville, Florida.[6] They report that a letter grade assignment of "A" has a market value of about seven percent over a "B" grade. They also find that average test score data significantly affects house prices, in line with Black's (1999) results.

[5] Bogart and Cromwell (2000) extended their previous analysis by examining the effect of changes in school district quality caused by redistricting. They find that redistricting reduces house values by nearly 10 percent.

[6] Wiemer and Wolkoff (2001) perform a similar analysis for Monroe County, New York.

Further, they test the short-run effects of changes in school accountability ratings and find that such changes can significantly affect house prices. Downes and Zabel (2002) employ an augmented dataset that includes information taken from the American Housing Survey to further correct for neighborhood effects. Their intention is to uncover which elements parents take into account in their decision process. Consistent with other studies, they find that test scores are an important factor to prospective homeowners.

Kane, Staiger, and Samms (2003) use data from Mecklenburg county in North Carolina and find that school test scores are strongly related to higher house prices. They employ a version of Black's identification that includes neighborhood fixed effects at a school district level. They find that the impact of a school-level standard deviation is approximately 4 to 5 percentage points. They also test for stability of the education and real estate relationship by examining the short term effect of changes in school accountability ratings. However, unlike Figlio and Lucas (2000), they find no evidence that revisions to state accountability ratings had any effect on house prices.

Finally, Bayer, Ferreira, and McMillan (2003) use Black's method for the San Francisco Bay Area to find that on average people are willing to pay an additional one percent in house price when the school quality is increased by five percent. They also include interaction effects between model variables to estimate the heterogeneous preferences. Their estimates also indicate significant heterogeneity in preferences for school quality, as increased income and education considerably increase a household's willingness to pay more for better school performance.

3 A Housing Search Model

Previously, models of education capitalization have been extrapolations of the Tiebout sorting model briefly discussed above or heterogeneous utility models of the form proposed by Bayer, Ferreira, and McMillan (2003). While we acknowledge the explanatory power of these types of models, we are interested in investigating a different aspect of the education capitalization issue.[7]

The majority of the econometric specifications tested in the empirical literature include linear capitalization terms and, thus, a linear demand for education. We propose an alternative formulation motivated loosely from the search-theoretic framework of Mortensen and Pissarides (1994) and Den Haan, Ramey, and Watson (2000). While expansion in some cities is a possibility, for most MSAs the housing stock is, at least in the short-run, fixed. Thus, prospective homeowners that have preferences for purchasing properties in high quality school attendance zones may pay a higher percentage premium than those seeking housing in lower quality school attendance zones. Moreover, our conjecture is that there does not exist a penalty for low quality school and that house prices in attendance zones of schools below a certain quality (perhaps the mean) are based solely on comparables.[8]

3.1 Setup

In our model, buyers and sellers are matched each period. The value of a matched buyer and seller is determined by a stochastic *housing productivity* X drawn from a cumulative distribution $F(x|\mu)$, conditional on *school quality* μ and the buyer's *preference for education* $\alpha \geq 0$.[9] We interpret the stochastic housing productivity as the potential (matched) buyer's valuation over comparables. Buyers have heterogeneous preferences for education quality μ reflected in α.

School quality is *ex ante* observable but the stochastic productivity draw is unobserved until a match is formed. Thus, buyers with different education valuations can concentrate their efforts in particular school attendance zones but cannot observe comparables without

[7]We employ some characteristics of both the sorting model and the heterogenous utility model in our specification.

[8]The model we present in this section is highly stylized. It is intended to demonstrate that, under some conditions, the education premium is kinked. Of course, a more realistic model of housing search can be developed. We direct the reader to Wheaton (1990) for an example of housing search with mismatched agents.

[9]In the theoretical model, school quality μ reflects the deviation from the mean **and is normalized such that** $-1 \leq \mu \leq 1$.

participating in the matching market. Buyers and sellers pay a cumulative per period cost of searching c, regardless of the school attendance zone in which they search. Buyers of type α choose $\gamma(\mu,\alpha)$, the percentage of time spent searching in school attendance zone quality μ to maximize their expected returns from search.

Agents are matched via a technology that depends on $V(\mu)$, the ratio of sellers in each school attendance zone, U, the ratio of buyers in the market, and $\gamma(\mu,\alpha)$. The quality-dependent search technology defines the likelihood that the a prospective buyer is matched with a house as

$$\lambda(\mu,\alpha) = F\left(V(\mu), U\gamma(\mu,\alpha)\right), \tag{1}$$

where $\sum_{\mu}\gamma(\mu,\alpha) = 1$. The matching technology (1) has the property that an increase in the number of vacancies increases the likelihood of a match. Conversely, an increase in the number of searchers in a given school attendance zone (weakly) decreases the likelihood of a match. This can be produced either by an increase in the overall number of searchers or an increase in search time in a single attendance zone by individual searchers.

The buyer's contemporaneous valuation W of a house in an attendance zone with quality μ with drawn productivity X is

$$W = (1+\mu)^{\alpha} X, \tag{2}$$

where α determines the buyer's preferences for education quality. Education and house productivity are, therefore, complementary. Agents with no preference for schooling ($\alpha = 0$), therefore, have valuation that reflects only their utility from the house's comparables. The value of searching in attendance zone μ (either because of entering or reentering the market) is given by

$$\begin{aligned}\Lambda(\mu,\alpha) = \quad & \beta\{\lambda(\mu,\alpha)\textstyle\int_{R(\mu,\alpha)}^{\infty}\left[(1+\mu)^{\alpha}x + \Pi(x,\mu,\alpha)\right]dF(x|\mu) \\ & +\lambda(\mu,\alpha)\left[\sum_{\mu}\Lambda(\mu,\alpha) - c\right]\textstyle\int_{0}^{R(\mu,\alpha)} dF(x|\mu) \\ & +\left[1-\lambda(\mu,\alpha)\right]\left[\sum_{\mu}\Lambda(\mu,\alpha) - c\right]\},\end{aligned} \tag{3}$$

which depends on the expected future value of a match $\Pi(x, \mu, \alpha)$, the combined value of the search costs c, and the acceptance threshold $R(\mu, \alpha)$.[10] The acceptance threshold reflects the value at which all matches that yield house productivity above $R(\mu, \alpha)$ are accepted and all those below are not. The total value of remaining in the searching pool is simply the sum of (3) over school qualities, μ.

The Bellman equation that defines the total value of the house conditional on its school quality is

$$\Pi(X, \mu, \alpha) = \beta \left\{ (1 - \sigma) \left[(1 + \mu)^{\alpha} X + \Pi(X, \mu, \alpha) \right] + \sigma \left[\sum_{\mu} \Lambda(\mu, \alpha) - c \right] \right\}, \tag{4}$$

where σ is the probability that the new owner becomes mismatched in the future.[11] Our model differs from a standard labor-search model via the school quality μ that reflects the education quality.[12] The buyer and seller agree on a sale if the house productivity is sufficiently high. The threshold at which this occurs is determined by (3) and (4) in the following manner

$$R(\mu, \alpha) + \Pi(R(\mu, \alpha), \mu, \alpha) = \sum_{\mu} \Lambda(\mu, \alpha) - c, \tag{5}$$

which states that at the acceptance threshold $R(\mu, \alpha)$ agents are just indifferent between completing the transaction and reentering the search pool.

We assume the agents Nash bargain over the match surplus, with the seller having bar-

[10] The value of searching is equivalent to the total outside option in standard labor-search models. Many search-theoretic models include a contemporaneous outside option for agents in the search pool. This is usually interpreted as unemployment benefits in labor search models. Without loss of generality, we forgo inclusion of that term here.

[11] A mismatch could be produced by an exogenous life event (e.g., having children) which renders the current house insufficient for future consumption. We assume that becoming mismatched implies the owner must sell. A more realistic model might assume that σ forces a redraw of the house productivity. A draw that is sufficiently low will induce a sale.

[12] This model is similar in flavor to but still fundamentally different from the segmented labor market model of Mortenson and Pissarides (1999) in which workers are differentiated based on skill levels.

gaining weight θ. The match surplus is

$$S(X,\mu,\alpha) = (1+\mu)^{\alpha} X + \Pi(X,\mu,\alpha) - \left[\sum_{\mu} \Lambda(\mu,\alpha) - c\right] \tag{6}$$

and the school-quality conditional sale price is

$$P(X,\mu,\alpha) = \theta\left[(1+\mu)^{\alpha} X + \Pi(X,\mu,\alpha)\right]. \tag{7}$$

3.2 The Education Premium

Our primary interest is the premium associated with an increase in the school quality μ. Therefore, we leave detailed discussion of the model dynamics to future research. For the sake of discussion, we assume that the house productivity densities are invariant to school quality ($F(x|\mu) = F(x)\ \forall\ \mu$) and we restrict the model to two levels of education valuations ($\alpha = 0$ and $\alpha > 1$), indicating that some agents have no preference for education.[13] The model, then, collapses into two segmented markets: buyers who value education ($\alpha > 1$) search only in above-mean school attendance zones, and buyers who do not value education ($\alpha = 0$) search in below-mean school attendance zones. The former occurs because education-conscious buyers assign lower value to equivalent houses in below-mean school attendance zones. The latter occurs because sellers have higher outside options in rejecting offers from $\alpha = 0$ consumers to wait for education-conscious buyers.[14] Thus,

$$\begin{aligned} \gamma(\alpha &> 1, \mu \leq 0) = 0 \\ \gamma(\alpha &= 0, \mu > 0) = 0. \end{aligned}$$

[13]Conditioning productivity on school quality allows for potential neighborhood effects that are often discussed in the empirical literature. However, under certain parameterizations, neighborhood effects can induce some education-indifferent consumers to concentrate search in high quality school attendance zones. This leads to a lower but, potentially, unidentifiable education premia. For clarity, we exclude this possibility without affecting the qualitative nature of the theoretical predictions.

[14]This assumes reasonable discount rates and search costs.

In this case, the price paid for houses in below-mean ($\mu \leq 0$) schools is

$$P(X, \mu, \alpha|_{\alpha=0}) = \theta \left[X + \Pi(X, \mu\,, \alpha|_{\alpha=0}) \right], \tag{8}$$

which now depends on the school quality only through the proportion of education indifferent searchers.[15] Similarly, the price paid for a house in an above-mean ($\mu > 0$) school attendance zone is

$$P(X, \mu, \alpha|_{\alpha>1}) = \theta \left[(1+\mu)^{\alpha} X + \Pi(X, \mu, \alpha|_{\alpha>1}) \right]. \tag{9}$$

[Figure 1 about here.]

The education premium $\psi(X, \mu, \alpha)$ is the increase in price between two houses with the same productivity draw that results from a change in school quality. We calculate $\psi(X, \mu, \alpha)$ as

$$\psi(X, \mu, \alpha) = \frac{\partial P(X, \mu, \alpha)}{\partial \mu}. \tag{10}$$

The restrictions outlined above indicate a clear partition for the range of the education premium, yielding the following theoretical premium

$$\psi(X, \mu, \alpha) = \begin{cases} \theta\alpha\,(1+\mu)^{\alpha-1} X + \frac{\partial \Pi(X,\mu,\alpha)}{\partial \mu} & if \quad \mu > 0,\ \alpha > 1 \\ 0 & if \quad \mu \leq 0,\ \alpha = 0 \end{cases}. \tag{11}$$

Since $\Pi(X, \mu, \alpha > 1)$ is increasing in attendance zone quality, the premium for agents that prefer education is also strictly increasing.

The education premium for a fixed value of X is illustrated in Figure 1. For houses in better-than-average school attendance zones, increasing the house quality X increases the magnitude of the nominal premium through the first term in (11). Note, in particular, that the theoretical education premium is exactly zero for $\mu \leq 0$ but strictly increasing for

[15] We assert without proving formally that for $\mu \leq 0$ and for $F(x|\mu) = F(x)$ for all μ, the steady-state search ratio $\gamma(\mu, \alpha = 0) = \gamma$ is equal for all μ.

$\mu > 0$. Essentially, the heterogeneity in preferences for education causes a jump in the pricing function at the mean school quality.

In addition to the jump at the mean attendance zone, the model implies a strong theoretical prediction about the nature of the education premium for houses associated with above mean quality schools. Heterogeneity of preferences leads to strategic interaction between agents that prefer quality education. This obtains because as more agents search in high quality attendance zones, the rate at which matches occur decreases. We reflect this in a decrease in the μ-specific match rate $\gamma(\mu, \alpha > 1)$. However, increasing value of the school quality also increases the percentage of time that agents search in that district. If the increase in search rate dominates the strategic interaction, the education premium increases at an increasing rate. This is reflected as an exponentially increasing premium in Figure 1. In the next section, we examine whether these theoretical predictions are borne out in the data.

4 Econometric Methodology

Our intention is to estimate the value added to house prices of higher school performance. Specifically, we are interested in estimating the dollar value difference in home price a household pays for a quantified increase in school quality. Below, we discuss three alternative specifications which include two different identification techniques employed to disentangle neighborhood quality from school quality.

4.1 Pure Hedonic Pricing

As a benchmark, we introduce a hedonic pricing function in which the sale price is described as a function of the characteristics of the house, its location, and the value associated with each characteristic. A basic hedonic function can be described as follows:

$$\ln(p_{iaj}) = \kappa + X'_{iaj}\beta + Z'_j\delta + Y'_{aj}\psi^H + \epsilon_{iaj}, \tag{12}$$

where p_{iaj} is the price of house i in attendance zone a in school district j (within each school district there are several individual schools which are assigned students by attendance zones). The vector X_{iaj} represents the comparable aspects of house i (i.e., the number of bedrooms, bathrooms, etc.) and Z_j is a vector of neighborhood and school district characteristics. The value Y_{aj} is the quality of the school in attendance district a in school district j (measured by average test scores). The quantity of interest ψ^H is the education capitalization premium calculated with the hedonic pricing model.

While not all of these variables may be important to a buyer himself (someone without children, for example, will not use the school district), a buyer usually takes into account all variables which will be important to the resale value of the house. Therefore the house price will reflect all relevant attributes (i.e., good qualities of the home are capitalized into the house value even if they are not directly consumable to the current tenant). One potential problem is that the comparable house characteristics do not indicate the quality of the house (updates, condition, landscaping, layout, etc.), the quality of the surrounding neighborhood, and various other factors. The hedonic pricing function attempts to correct for some of these factors by incorporating the Z_j vector.

4.2 Linear Boundary Fixed Effects

The methodology of adding the neighborhood vector Z_j outlined in the previous section may reduce but not entirely account for all of the variation that can be introduced on a neighborhood level. Suppose that the neighborhood quality gradient is large in absolute value. This implies that houses a few blocks away from each other can vary a great deal in atmosphere, and, therefore, in price. This variation can come from distance to amenities, mass transit, and thoroughfares (i.e., highway access), proximity to commercial and industrial zoning, single-family housing density, etc. Since the vector Z_j is defined over a coarsely defined area, it may be unable to account for the neighborhood variation that confounds the estimate of the capitalization premium. As Black points out, many of these factors (though admittedly not all) can be corrected for by analyzing houses that are geographically close.

Essentially, we identify education effects by weeding out the neighborhood-level variation. While the hedonic pricing function can accomplish some of this at a school district level, Black's refinement is to consider only those houses that are geographically close to the school district boundary. Therefore, we replace Z_j with a full set of *pairwise boundary dummies* which indicate whether or not houses share an attendance zone boundary. This yields

$$\ln(p_{iab}) = \kappa + X_{iab}'\beta + K_b'\phi + Y_a'\psi^L + \epsilon_{iab}, \tag{13}$$

where K_b is the vector of boundary dummies and ψ^L is the education premium calculated with the LBFM. Equation (13), then, is equivalent to calculating differences in house prices on opposite sides of attendance boundaries while controlling for house characteristics and relating it to test score information. The boundary dummies allow us to account for any unobserved neighborhood characteristics of houses on either side of an attendance boundary because two homes next to each other generally have the same atmosphere.

4.3 Nonlinear Boundary Fixed Effects

As an alternative to the LBFM outlined in the previous section, we consider the possibility that the capitalization premium is not a constant percentage over the range of school qualities.[16] This is accomplished by testing whether the education capitalization term enters

[16] The model in section 3 makes two predictions. First, there is a kink in the premium. Second, heterogeneity in α may lead to a rapidly increasing premium for high quality schools. Based on the illustrative version of the model we present in section 3, one might expect that estimating a kinked linear premium is warranted. However, the nonlinear model we discuss in this section is more general and serves as a good approximation to a kinked-linear model.

nonlinearly. Consider the following pricing equation:

$$\ln(p_{iab}) = \kappa + X'_{iab}\beta + K'_b\phi + f(Y_a) + \epsilon_{iab}, \tag{14}$$

where $f(Y_a)$ represents a potentially nonlinear function of school quality. Suppose the function $f(Y_a)$ is composed of a linear and higher order polynomial terms in school quality. That is,

$$f(Y_a) = \psi_1 Y_a + \psi_2 Y_a^2 + \psi_3 Y_a^3 \tag{15}$$

where ψ_i are scalar parameters.[17] We then rewrite (14) as

$$\ln(p_{iab}) = \kappa + X'_{iab}\beta + K'_b\phi + \psi_1 Y_a + \psi_2 Y_a^2 + \psi_3 Y_a^3 + \epsilon_{iab}. \tag{16}$$

The specification (16) has a number of advantages over the linear form (13). First, the rate at which the nominal premium increases is not fixed. This allows us to capture the effects of increased search in high quality school attendance zones. Second, the linear model penalizes houses in low quality school attendance zones by valuing them below what would be predicted by their comparables. Moreover, the penalty increases as the school quality worsens. This is theoretically unappealing in our framework since agents can choose not to search in these school attendance zones. Our model implies houses in below threshold attendance zones receive no premium (and thus no penalty), a possibility that is explicitly excluded in the linear model.

5 Data

In this analysis we restrict our attention to single family residences and elementary school attendance zones. Each house represents an observation and its described by variables

[17] Note that our specification requires that we depart from measuring school quality in terms of test score points (as in Black, 1999, and others). We adopt the normalized measure of school quality discussed in the following section.

reflecting its physical characteristics, the quality of the elementary school that children in the household would have to attend, and the characteristics of the neighborhood the house is in.

5.1 Educational Data

We use test scores to assess the quality of schools. For this, we obtained tests scores data from the Missouri Assessment Program (MAP) for the years 1998 through 2001. Our sample consists of 121 elementary schools in 15 school districts in St. Louis county.

The MAP test includes a *Mathematics* section for grades 4, 8, and 10, a *Communication Arts* section (which includes a *Reading* portion) for grades 3, 7, and 11, a *Science* section for grades 3, 7, and 10, and a *Social Studies* section for grades 4, 8, and 11. We focus on the Math section, but results for the Science and Reading sections are available from the authors upon request.

Our dataset provides an overall score for each school and year in the sample, called the MAP Index. The MAP Index is the weighted sum of the percentages of students in each of five performance bins (Advanced, Proficient, Nearing Proficient, Progressing, and Step 1).[18] In our analysis, we determined the MAP index score for each school by averaging the Math scores available in the period 1998-2001 (although not all schools had a score available for each year in the period). We then normalized each score in standard deviations from the sample mean.[19]

[Table 1 about here.]

[18] MAP Index = (percent in Step1)*1 + (percent in Progressing)*1.5 + (percent in Nearing Proficient)*2 + (percent in Proficient)*2.5 + (percent in Advanced)*3

[19] For the linear model we could have used the raw scores as the measure of quality to compare with other studies. The nonlinear models, however, can only be reasonably estimated with the standardized school quality measures.

5.2 Discussion of summary statistics

5.2.1 Education Variables

Summary statistics for test scores and other education variables are described in Table 1. The Math test score index, for example, has a mean value of 211.44 points. Essentially all the schools in our sample have a Math score that lies within 2 standard deviations of the mean. This table also reports the following school district variables: student/teacher ratio, expenditures per pupil as of 1998, the fraction of students eligible for free or reduced-price lunch, and the property tax rate as of 1998.

5.2.2 Real Estate Prices and Housing Comparables

We obtained house price and house characteristics data from First American Real Estate, Inc. The observations selected correspond to single family residences sold during the 1998-2001 period. Summary statistics for these variables are reported in Table 2. We deflated house prices to 1998 dollars with a median house price index for the entire St. Louis metropolitan area.[20] The resulting adjusted house price has a mean of $152,473 and a standard deviation of $166,067; the median house price is $115,324. House characteristics include the total number of rooms, the number of bedrooms, the number of bathrooms, the lot size, the internal square footage, the age of the structure, and the number of stories in the house. We also include the distance to downtown Clayton, MO, a major business area in the St. Louis metro area, as an additional characteristic. After eliminating from the original dataset observations with missing or outlier house prices (i.e., outside a bound of 3.5 standard deviations from the mean unadjusted house price), our sample includes 38,676 single family residences.

For the boundary analysis, we determined the attendance zone for 121 schools in St. Louis county. For every pair of adjacent schools, we coded a boundary dummy, as in Black (1999). Finally, we assigned each house within a 0.1 mile buffer to the nearest (and therefore unique) pairwise boundary dummy. In this restricted sample there are 10,190 single family

[20] These data were obtained from the National Association of Realtors.

residences.[21]

[Table 2 about here.]

6 Empirical Results

6.1 Benchmark Linear Results

In order to provide a benchmark for comparison, we estimate the linear boundary fixed effects model (13) and present the results in the first column of Table 3. We find that housing comparables enter the house pricing equation with the expected sign. Increases in living area, lot size, and the number of rooms increases the price of a house on average. Conversely, the age of the building—perhaps an indicator of condition—adversely affects the price.

The key result is the coefficient on the MAP score reported in the first line of Table 3. We find that, in the LBFM, school quality is a statistically significant contributor to house prices and enters with the expected sign. Moreover, our estimate of the education premium implies that a 5 percent increase in the average test scores in a school attendance zone leads to an increase of 3.5 percent or $5,383 in house prices at the sample median price. This value is only slightly higher than that estimated by Black.[22] We attribute moderate cross study variation to differences in housing demand and preferences for education across MSAs.

6.2 The Nonlinear Model

[Table 3 about here.]

In addition to the LBFM, we estimate two specifications of the nonlinear boundary-fixed effects model (14) intended to account for an increasing marginal valuation of school quality.

[21] Black considers a number of different boundary width ranges and finds no significant differences. Our sample does not permit wider boundaries as these would encompass some attendance zones almost entirely.

[22] Black reports a 2.1 percent, or $3,948 at her sample mean, increase in house prices for a 5 percent increase in test scores.

We first consider only quadratic terms for school quality (QBFM); then, we examine the model with cubic terms (CBFM). Results are reported in the second and third columns of Table 3. In particular, note that the addition of the second order polynomial of school quality has only a small impact on the linear term. Thus, around the mean, there is little difference between the premium predicted by both models. However, when considering school qualities more than one-half of one standard deviation from the mean, the difference becomes more noticeable. These differences are summarized in Figure 2, where we plot the estimated premium against school quality for the linear and both nonlinear models, including the theoretical premium for comparison.

[Figure 2 about here.]

Figure 2 shows that the linear model predicts that houses in below-average quality school attendance zones must be sold at a discounted price (i.e., below the price that would be warranted by its comparables). However, this is not the case in the two models that include higher order polynomial terms. Inclusion of the polynomial terms allows the econometric model to approximate, in accordance with the model presented earlier, the nonlinearity in the premium around the mean school quality. Both the square and cubic term models predict that houses in below-average school zones are not penalized but, as predicted by the model, sold at prices driven entirely by comparables.

[Table 4 about here.]

Moreover, the LBFM underestimates the premium for high-quality schools. Table 4 summarizes the implied school premium for all three models. For a 5 percent increase in school quality from the mean, the quadratic and cubic BFMs predict a premium about 42.5 and 29.8 percent higher than the linear model, respectively. At the mean house price, this is equivalent to $1,733.5 and $1,213.8 respectively.

For above-average schools, the discrepancy is more apparent. The linear model presupposes a constant premium of $4,078.4 at the median house price per 5 percent (of the mean score) increase in school quality. The nonlinear models, however, allow for an accelerating premium. The premium for a 5 percent increase in quality at one standard deviation above the mean school is $7,352.8 for the quadratic model and $7,611.2 for the cubic model.

Finally, Table 4 reports the magnitude of a change in quality for below-average schools. Symmetry suggests that the linear model predicts a penalty of $4,078.4 at the median house price per 5 percent decrease in school quality. However, the theoretical model predicts a zero penalty. Consistent with this prediction, the quadratic and cubic models imply a $597.0 and $255.9 penalty for a 5 percent decrease in quality from one standard deviation below mean schools.

6.3 Specification Testing

[Table 5 about here.]

We consider a number of specification tests to determine whether the nonlinear model is preferred.[23] As a prelude, we consider the adjusted R^2 for the three specifications in Table 5. The explanatory power as computed by the adjusted R^2 of each of the specifications is nearly identical. We note that, while the cubic model generates the model with the highest likelihood, the estimated parameter on the cubic term in the cubic BFM is not significant at the 10 percent level.[24]

Therefore, we consider a battery of likelihood ratio tests to determine the optimal specification. Table 5 also includes these results. The tests we conduct assess the likelihood that the zero restrictions on terms in the school premium (15) can be rejected. Tests of all three models resoundingly reject the null hypothesis of a model with no education premium. We find that the restriction of no included quadratic or cubic term ($\psi_2 = \psi_3 = 0$) is rejected

[23] We compare the two nonlinear specifications (quadratic and cubic) to the LBFM and a model with no education effects.

[24] Surprisingly, for the science test scores (not reported here), tests of the linear model compared to the null of a model with no school district quality premium, the restricted model cannot be rejected at any reasonable significance level. This implies that the linear model may, in fact, not be preferable to a model with no premium.

at the 1 percent level. However, the restriction of no cubic term ($\psi_3 = 0$) is rejected at only the 60 percent level. Thus, we find evidence that the preferred model for the education premium is quadratic.

7 Conclusion

Current empirical models of the value-added of education for house prices are computed with linear contributions from school quality. Although the magnitude of the influence is arguable, these models show that the quality of primary school education is positively correlated with house prices. Recent use of linear boundary fixed effects models have shown that hedonic pricing models tend to overpredict the premium. However, even these augmented models have similar implications on the shape of the premium over school quality. Although the nominal value of the premium rises with school quality, the rate at which the premium increases is constant–even for houses in very good school attendance zones where the vacancy rate is low and demand is high. Moreover, the linear model implies an increasing penalty for increasingly bad schools, implying that homeowners are influenced by small changes in school quality even if their school is well below the MSA median school quality.

We propose an alternative formulation that accounts for these two possibilities: that very high quality schools demand an increasing premium and that below-mean schools are ex post identical as far as calculation of the premium is concerned. We show that not only does the nonlinear formulation match theoretical predictions from a housing search model, but that it is preferred by the data over the baseline LBFM and a model with comparables only. This leads to an underestimation of the education capitalization premium with the LBFM. We thus conclude that, at least for the St. Louis MSA, the theoretical prediction that housing prices for worse-than-average schools can be thought of as driven only by comparables while houses in better-than-average schools garner an increasing premium.

References

[1] Bayer, Patrick; Ferreira, Fernando, and McMillan, Robert. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," Unpublished Manuscript, July 2003, Yale University.

[2] Betts, Julian R. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics*, May 1995, 77(2), pp. 231-250.

[3] Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics,* May 1999, *114*(2), pp. 577-599.

[4] Bogart, William T. and Cromwell, Brian A. "How Much More is a Good School District Worth?" *National Tax Journal,* June 1997, *50*(2), pp. 215-232.

[5] Bogart, William T. and Cromwell, Brian A. "How Much is a Neighborhood School Worth?" *Journal of Urban Economics*, March 2000, 47(2), pp. 280-305.

[6] Card, David and Krueger, Alan B. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy,* February 1992, 100(1), pp. 1-40.

[7] den Haan, Wouter J.; Ramey, Garey, and Watson, Joel. "Job Destruction and the Experiences of Displaced Workers." *Carnegie-Rochester Conference Series on Public Policy,* June 2000, 52(0), pp. 87-128.

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

[9] Figlio, David N. and Lucas, Maurice E. "What's in a Grade? School Report Cards and House Prices," Working Paper Number w8019, National Bureau of Economic Research, November 2000.

[10] Hanushek, Eric A. "Measuring Investment in Education," *Journal of Economic Perspectives*, Fall 1996, 10(4), pp. 9-30.

[11] Hayes, Kathy J. and Taylor, Lori L. "Neighborhood School Characteristics: What Signals Quality to Homebuyers?" *Federal Reserve Bank of Dallas Economic Review,* Fourth Quarter 1996, *0*(0), pp. 2-9.

[12] Kane, Thomas J.; Staiger, Douglas O., and Samms, Gavin. "School Accountability Ratings and Housing Values," *Brookings-Wharton Papers on Urban Affairs,* Forthcoming, 2003.

[13] Mortensen, Dale T. and Pissarides, Christopher A. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies,* July 1994, 61(3), pp. 397-415.

[14] Mortensen, Dale T. and Pissarides, Christopher A. "Unemployment Responses to 'Skill-Biased' Technology Shocks: The Role of Labour Market Policy," *Economic Journal,* April 1999, 109(455), pp. 242-265.

[15] Ridker, Ronald G. and Henning, John A. "The Determinants of Residential Property Values with Special Reference to Air Pollution," *Review of Economics and Statistics,* May 1967, *49*(2), pp. 246-257.

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

[17] Tiebout, Charles M. "A Pure Theory of Local Expenditures," *Journal of Political Economy,* October 1956, *64*(5), pp. 416-424.

[18] Weimer, David L. and Wolkoff, Michael J. "School Performance and Housing Values: Using Non-Contiguous District and Incorporation Boundaries to Identify School Effects," *National Tax Journal*, June 2001, 54(2), pp. 231-253.

[19] Wheaton, William C. "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy*, December 1990, 98(6), pp. 1270-1292.

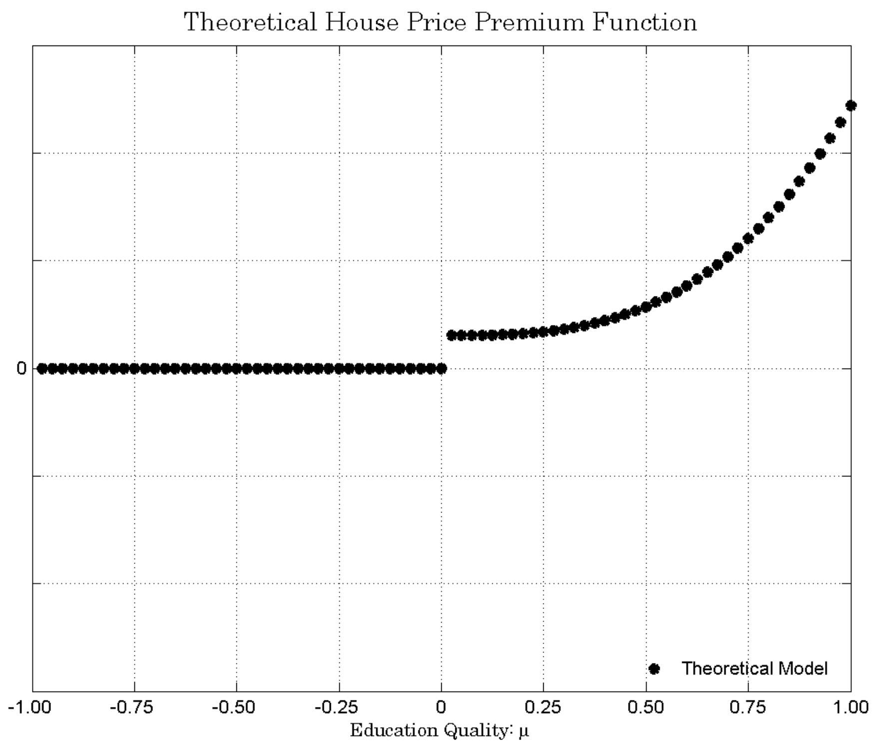

Figure 1: Theoretical House Price Premium Function

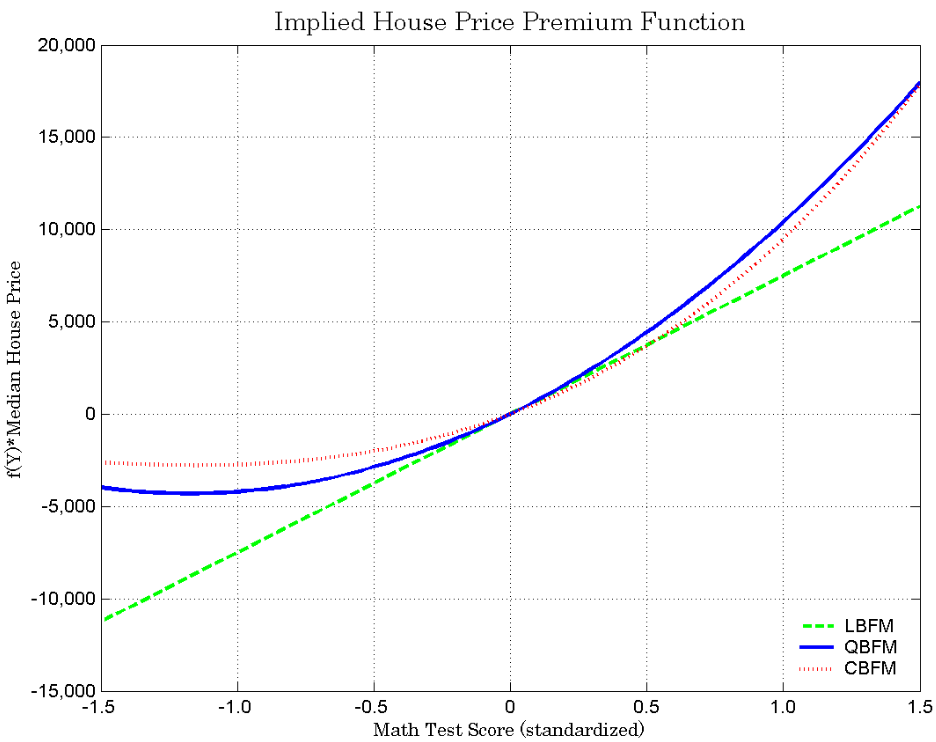

Figure 2: Implied House Price Premium Function

Table 1: Summary Statistics: Test Scores and School District variables

Variable	Mean	Std. Dev.	Min.	Max.
Math MAP score	211.44	19.44	168.15	250.18
Science MAP score	211.68	22.43	100.00	242.60
Reading MAP score	205.38	30.29	100.00	250.45
Student/Teacher Ratio	16.12	1.87	11.77	18.82
Total Expenditures per Pupil (1998: $1,000s)	8.07	3.08	5.89	18.52
Students with Free/Reduced-Price Lunch	0.29	0.24	0.00	0.74
Property Tax Rate (1998: $1/$1,000)	4.20	0.88	2.66	5.74

MAP scores are measured at 121 schools; other variables are measured at 15 school Districts.

Table 2: Summary Statistics: House and Neighborhood Characteristics

	Full Sample: N=38676		Boundary Sample: N=10190	
House Variables	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Sale Price (1998)	152,472.46	166,067.13	146,256.43	180,642.80
log of Sale price (1998)	11.65	0.73	11.59	0.74
Number of Bedrooms	2.96	0.84	2.90	0.84
Number of Bathrooms	2.01	0.95	1.95	0.93
Number of Bathrooms (square)	4.97	5.05	4.66	5.04
Age of Building	38.91	20.63	40.72	21.27
Age of Building (square)	1,939.39	1,922.69	2,110.15	2,028.41
Lot Area (sq.ft)	14,752.24	38,342.73	13,610.24	39,203.83
Living Area (sq.ft)	1,160.22	435.07	1,128.93	424.78
Total Number of Rooms	6.38	1.60	6.26	1.57
Distance to Clayton (miles)	6.91	3.66	6.42	3.68
Distance to Clayton (squared mi.)	61.19	61.68	54.76	57.91
Number of Stories	1.24	0.42	1.23	0.41
	Census Block Groups: N=674			
Neighborhood Variables	**Mean**	**Std. Dev.**		
Hispanic population	0.01	0.02		
Non-Hispanic population: black	0.22	0.31		
Population with less than high school completed	0.14	0.10		
Population with a bachelor's degree	0.20	0.12		
Population with a graduate degree	0.13	0.11		
Population between 0 and 9 years old	0.13	0.05		
Population 65 and over	0.15	0.09		
Female-headed households with children	0.12	0.11		
Median Household Income (1999: $1,000s)	54.39	27.26		

Table 3: Math Scores: Regressions with Boundary Dummies

	LBFM	QBFM	CBFM
Math MAP std. score	0.06503	0.06338	0.04820
	(2.569)**	(2.891)***	(1.688)*
Math MAP std. score (square)		0.02693	0.02932
		(2.476)**	(2.454)**
Math MAP std. score (cube)			0.00484
			(0.680)
Number of Bedrooms	0.03732	0.03737	0.03754
	(3.887)***	(3.900)***	(3.910)***
Number of Bathrooms	0.10978	0.10929	0.10935
	(5.901)***	(5.930)***	(5.941)***
Number of Bathrooms (square)	-0.00539	-0.00543	-0.00545
	(1.724)*	(1.743)*	(1.750)*
Age of Building	-0.00368	-0.00371	-0.00371
	(2.465)**	(2.490)**	(2.497)**
Age of Building (square)	0.00004	0.00004	0.00004
	(2.722)***	(2.743)***	(2.751)***
Lot Area (sq.ft)	0.00000	0.00000	0.00000
	(2.433)**	(2.436)**	(2.433)**
Living Area (sq.ft)	0.00035	0.00035	0.00035
	(15.476)***	(15.333)***	(15.334)***
Total Number of Rooms	0.05885	0.05863	0.05856
	(7.295)***	(7.249)***	(7.232)***
Number of Stories	0.27659	0.27644	0.27643
	(9.416)***	(9.415)***	(9.423)***
Constant	8.54855	8.45106	8.44670
	(79.388)***	(83.626)***	(83.506)***
Observations	10190	10190	10190
Adjusted R-squared	0.762	0.763	0.763

Standard errors clustered by school attendance zone.
Robust $t-$statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.
Coefficients for boundary dummies not shown.

Table 4: Implied House Price Premia from Math Test Scores

Mean house price	\$152,472.5		
Median house price	\$115,323.8		
Mean of Math Score	211.44		
St. Dev of Math Score	19.44		
ΔY_a (5% of mean score)	10.57		
	LBFM	QBFM	CBFM
	Standardized Scores		
Linear Coefficient	0.06503	0.06338	0.04820
Quadratic Coefficient		0.02693	0.02932
Cubic Coefficient			0.00484
Increase score by ΔY_a at $Y = 0.0$ to:	222.01	222.01	222.01
Percent change in house price	3.54	5.04	4.59
Dollar value at mean house price	5,392.20	7,684.12	6,997.03
Dollar value at median house price	4,078.44	5,811.94	5,292.26
Increase score by ΔY_a at $Y = +1.0$ to:	241.45	241.45	241.45
Percent change in house price	3.54	6.38	6.60
Dollar value at mean house price	5,392.20	9,721.39	10,063.01
Dollar value at median house price	4,078.44	7,352.85	7,611.24
Decrease score by ΔY_a at $Y = -1.0$ to:	181.43	181.43	181.43
Percent change in house price	-3.54	-0.52	-0.22
Dollar value at mean house price	-5,392.20	-789.39	-338.31
Dollar value at median house price	-4,078.44	-597.06	-255.88

The percent change in house price is computed as $f'(Y)\Delta Y_a$.

Table 5: Specification Tests on Boundary Fixed Effects. Math Scores

	LBFM	**QBFM**	**CBFM**
Model $f(Y_a)$	$\psi_1 Y_a$	$\psi_1 Y_a + \psi_2 Y_a^2$	$\psi_1 Y_a + \psi_2 Y_a^2 + \psi_3 Y_a^3$
Log-Likelihood	-4002.32	-3994.90	-3994.55
Adjusted R-squared	0.762	0.763	0.763
Null Hypothesis	$\psi_1 = 0$	$\psi_1 = \psi_2 = 0$	$\psi_1 = \psi_2 = \psi_3 = 0$
LR test	22.929	37.783	38.480
Prob $> \chi^2$	(0.000)	(0.000)	(0.000)
Null Hypothesis		$\psi_2 = 0$	$\psi_2 = \psi_3 = 0$
LR test		14.854	15.551
Prob $> \chi^2$		(0.000)	(0.000)
Null Hypothesis			$\psi_3 = 0$
LR test			0.697
Prob $> \chi^2$			(0.404)