Welfare Analysis with Hedonic Budget Sets
(Preliminary Draft, Please do not Post Online)

Debopam Bhattacharya* — University of Cambridge
Ekaterina Oparina — LSE

Qianya Xu — University of Cambridge

September 8, 2024

Abstract

We analyze an empirical demand model where heterogeneous consumers maximize utility under a nonlinear, hedonic price function. We develop methods for empirical welfare analysis of interventions that change the price function. A key finding is Roy's identity for nonlinear budget sets, which produces a partial differential equation system. Under scalar unobserved heterogeneity and single-crossing, the coefficient functions in the PDEs are identifiable, and lead to point-identification of welfare effects. The key is that indirect utility is strictly increasing in income. The methods are used to study the effect of altering the relation between property-rent and neighborhood school-quality in the UK.

1 Introduction

Hedonic modelling has been a staple of demand analysis since Rosen, 1974. It is useful for analyzing markets for differentiated goods with a large number of available varieties, but

*Financial support from Economic and Social Research Council (ESRC)

where each variety can be viewed as a distinct bundle of a limited number of attributes. Examples include cars, houses, neighborhoods, colleges, hotels, restaurants etc. They have also been used to model labour markets where each worker may be viewed as a distinct bundle of a set of human capital attributes. An important characteristic of such markets is that in equilibrium, the marginal price of an attribute typically varies with quantity, i.e. budget sets are continuous but nonlinear. For example, the price of a residential property typically increases nonlinearly with its floor-area. The present paper develops methods for empirical welfare analysis of policy interventions in such settings.

As a motivating example, consider the well-studied relationship between housing costs and neighborhood school quality (cf. Sheppard 1999, Gibbons et al 2013). To enable children to attend nearby schools, public policy often mandates 'catchment area' rules, which restrict school-access solely to neighborhood children. This, however, means that the presence of a good school makes its adjoining residential neighborhood attractive, and raises house price and rent. This leads to wealthier families moving in from worse school districts, aggravating existing socioeconomic segregation. One potential way to interrupt this vicious cycle is to relax catchment area restrictions. This would lead housing costs to become less entangled with school quality, thereby changing choices in equilibrium. However, the overall welfare effects are likely to be heterogeneous, depending on both household preferences over consumption and neighborhood school-quality and on their income. The question is how can we calculate these welfare effects using the type of data usually available in surveys matched with school league tables.

In the present paper, we present an economic model of choice for a heterogeneous population of consumers, each facing a nonlinear, hedonic budget set. We then derive the analog of Roy's Identity for this setting, which yields a set of linear partial differential equations. Under a restriction on the dimension of heterogeneity, viz. that it is a scalar, and a single-crossing condition on preferences, the PDEs can be 'solved' for the indirect utility function, separately for each quantile of heterogeneity. We show that for calculating welfare effects of changing the budget set, a *general* solution to these PDEs is sufficient, and it is not necessary to obtain a particular solution using boundary conditions. This sufficiency is a consequence of the indirect utility function being increasing in income, an implication of preferences being nonsatiated in consumption.

Our work is thematically related to Heckman et al 2010, who show how to identify consumer preferences (over a single continuous attribute and consumption) nonparametrically in hedonic budget settings, where unobserved heterogeneity is a scalar, and preferences are quasilinear in consumption and satisfy a single-crossing condition. In fact, we borrow the Heckman et al 2010 set-up, except that utilities are not assumed to be quasilinear in consumption, and our focus is on welfare effects which are obtained *without* identifying the underlying preferences, the focus of Heckman et al 2010. Our paper is also related to Hausman and Newey 2016, who show that in a standard demand setting with one continuous inside good, linear budget frontiers and general heterogeneity, welfare distributions resulting from a price change are not point-identified. In contrast, we allow hedonic budget frontiers to be nonlinear, and show how to nonparametrically point-identify welfare by restricting heterogeneity to be one-dimensional and imposing a single-crossing condition, analogous to Heckman et al 2010. We also show how to include additional attributes into the analysis. As such, the key theoretical contributions of our analysis are the derivation of Roy's identity for nonlinear budget-frontiers, using the single-crossing condition on preferences to convert the coefficient functions in the resulting system of PDEs into estimable objects, and finally showing how to use monotonicity (in income) of the indirect utility function to obtain the necessary welfare effects from this system of PDEs.

We end this introduction by emphasizing that the key purpose of this paper is to derive welfare measures, *assuming that the budget frontiers are identifiable.* This means that the relation between price and the attribute of interest (e.g. school quality and property rent) is assumed to be obtainable from the data, as is also assumed in Heckman et al 2010. Therefore, we do not discuss – and indeed, do not contribute to solving – well-known issues of omitted variable problems that jeopardize identification of this relationship in the first place, cf. Black 1999.

2 Set-up

Denote the key product characteristic by s, e.g. quality of the nearest school in our motivating example, the hedonic price schedule describing the relation between housing cost and s is given by $p(s) \equiv p(s, \theta)$ where θ is a finite-dimensional parameter. Consumption is given by $c = y - p(s, \theta)$ where y is individual income. Individual preferences are described by the utility function $U(s, y - p(s, \theta), \varepsilon)$ where ε is unobserved heterogeneity. A household maximizes its utility by choosing s optimally. For the purpose of this paper, viz. identification of welfare effects, we assume that the function $p(s) \equiv p(s, \theta)$ is known to the analyst.[1]

The policy intervention we wish to investigate is one that changes the hedonic price frontier. In our empirical application, the leading case of interest is where school choice becomes less restrictive, which would weaken the relationship between rent and school-quality.

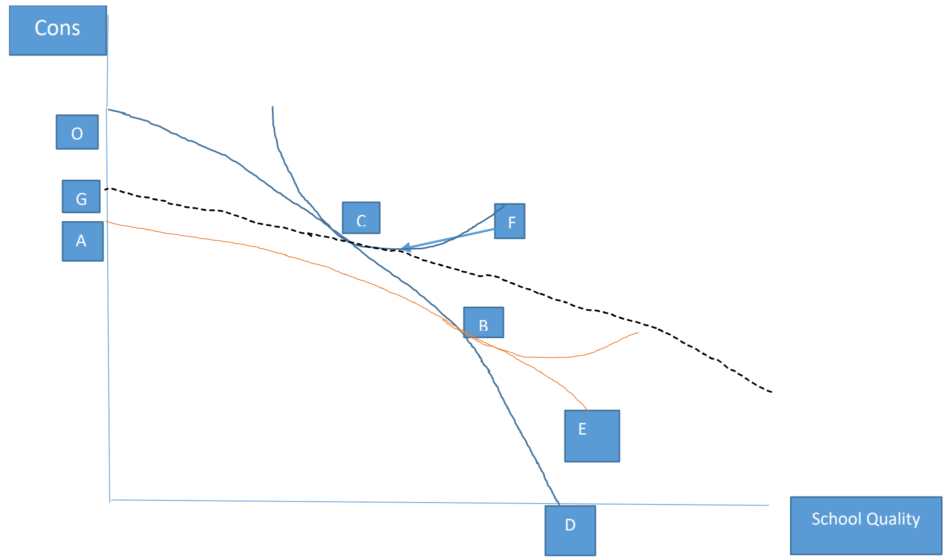
The pre- and post-intervention situations are depicted through the following graph where, for ease of exposition, ε is held fixed. In the graph, original hedonic price schedule is given by OCD. Utility is maximized at C. Due to a policy intervention (e.g. expanded school choice in our example), hedonic price schedule changes to ABE. Utility is now maximized at B.

We wish to compute the welfare effect of this intervention via the Compensating Variation, which calculates how much would a household need to be compensated, so that its maximized utility with the additional income in the post-intervention situation equals its maximized utility in the pre-intervention period with the original income.

To see this graphically, consider the curve depicted by the dashed line which is ABE translated vertically up and is tangent to the original indifference curve at F. Then AG equals compensating variation, which is the income supplement for the individual facing schedule OCD so that she can reach utility equal to the higher original indifference curve

Given the position of the indifference curves, the CV is positive, indicating that the consumer is losing as a result of the change, and hence needs to be compensated by a

[1] Indeed, θ will typically be estimated from the data, but at a parametric rate, and these estimated θs will be used subsequently as regressors, leading to standard measurement error issues. However, variance of the measurement error in θ is of order $O(n^{-1})$, where n is the number of observations in each market used to estimate θ in that market. Hence replacing θ by its estimate will lead to a very small attenuation bias when n is large. For related discussions, see Heckman et al 2010, Section 5.



positive income transfer to restore her utility to its pre-intervention state. However, if the indifference curve were tangent to OBD in the region BD, then the shift of the budget line to ABE would lead to tangency in the region BE, and thus represent a *gain* in utility. Such a consumer would benefit from the change, and the CV will be negative. This reasoning illustrates that welfare effects of a shift in the budget frontier can be heterogeneous in both magnitude and sign.

3 Parametric Analysis

Let s denote the key characteristic of interest, and x denote the other attributes. For example, in our empirical exercise, s denotes school quality in a neighborhood and x denote other attributes of residential property such as property size, neighborhood crime rate etc. Suppose price function and utility are given by

$$\begin{aligned} p(x,s) &= \theta_0 + \theta_1' x + \theta_2 \ln s \\ U(s,x,y-p(s,x),\varepsilon) &\equiv \varepsilon \ln s + \sum_j \gamma_j \ln(x_j) + \ln(y - p(x,s)), \end{aligned}$$

where ε is a random random coefficient on $\ln(s)$, satisfying $\varepsilon > 0$ w.p.1, i.e. better quality school is preferred by everyone. The γ_js are random coefficients on the non-school attributes of the property. We estimate θs via OLS or fixed-effect regression of *price* on $1, x, \ln s$ in a single market.

Now, first order condition for $\max_s \left\{\varepsilon \ln s_i + \sum_j \gamma_{ji} \ln(x_{ji}) + \ln(y_i - p(x_i, s_i))\right\}$ is given by

$$\frac{\varepsilon_i}{s_i} = \frac{\frac{\theta_2}{s_i}}{y - (\theta_0 + \theta_1' x_i + \theta_2 \ln s_i)}$$

which simplifies to

$$\varepsilon_i = \frac{\theta_2}{y_i - \theta_0 - \theta_1' x_i - \theta_2 \ln s_i} \tag{1}$$

First order condition for $\max_{x_i} \left\{\varepsilon \ln s_i + \sum_j \gamma_j \ln(x_{ij}) + \ln(y - p(x_i, s_i))\right\}$ is given by

$$\gamma_{ij} = \frac{\theta_{1j} x_{ij}}{y - \theta_0 - \theta_1' x_i - \theta_2 \ln s_i} \tag{2}$$

Note that (1) and (2), evaluated at their realized values $(x, s) = (x^*, s^*)$ in the data, give values of $\left(\varepsilon_i, \gamma_{ji}\right)$ $j = 1, 2, ...J$ and $i = 1, ...n$.

To calculate welfare effects of a change in the price function, we first perform the opposite exercise, viz. express the optimal s_i and x_{ji} as functions of the parameters θ, γ and y. From (1) and (2), we get

$$x_{ij} = \frac{\theta_2 \gamma_{ij}}{\theta_{1j} \varepsilon_i}. \tag{3}$$

Replacing in (2), we get

$$\begin{aligned} \frac{\theta_2}{\varepsilon_i} &= y_i - \theta_0 - \theta_1' x_i - \theta_2 \ln s_i, \text{ i.e.} \\ s_i &= \exp\left(\frac{y_i - \theta_0}{\theta_2} - \frac{\sum_{j=1}^J \gamma_{ij}}{\varepsilon_i} - \frac{1}{\varepsilon_i}\right). \end{aligned} \tag{4}$$

The corresponding indirect utility is given by

$$V(\theta, \gamma, \varepsilon; y) = \varepsilon_i \ln s_i^* + \sum_j \gamma_j \ln\left(x_j^*\right) + \ln\left(y_i - (\theta_0 + \theta_1' x_i^* + \theta_2 \ln s_i^*)\right) \tag{5}$$

Therefore, for a change in price function from $p(x, s, a) = a_0 + a_1' x + a_2 \ln s$ to

$$\tilde{p}(x, s, b) = b_0 + b_1' x + b_2 \ln s,$$

the compensating variation C_i solves

$$V(a, \gamma_i, \varepsilon_i; y_i) = V(b, \gamma_i, \varepsilon_i; y_i + C_i)$$

whence (5) implies

$$C_i = \begin{array}{l} \exp\left\{ \begin{array}{c} \varepsilon_i \ln\left(s_i^*/s_i\right) + \sum_j \gamma_{ij} \ln\left(x_{ji}^*/x_{ji}\right) \\ + \ln\left(y_i - \left(a_0 + a_1' x_i^* + a_2 \ln s_i^*\right)\right) \end{array} \right\} \\ - \left(y_i - \left(b_0 + b_1' x_i + b_2 \ln s_i\right)\right). \end{array}$$

4 Nonparametric Analysis

First, consider the case where s is the only attribute of interest, price function is given by $p(s,\theta)$, and utility of an ε-type consumer is given by $U(s, y - p(s,\theta), \varepsilon)$ where y represents disposable income, $y - p(s,\theta)$ is consumption of the non-S numeraire, and ε is unobserved heterogeneity. Now, utility maximization implies that at the optimal choice $s^* \equiv s^*(y,\theta,\varepsilon)$, we must have that

$$U_s(s, y - p(s;\theta), \varepsilon) - U_c(s, y - p(s;\theta), \varepsilon) \frac{\partial}{\partial s} p(s;\theta) \bigg|_{s=s^*(y,\theta,\varepsilon)} = 0 \tag{6}$$

A sufficient second order condition for s^* to be the unique interior maximum is that

$$\begin{array}{l} U_{ss}(s, y - p(s;\theta), \varepsilon) - 2\frac{\partial}{\partial s} p(s;\theta) \times U_{cs}(s, y - p(s;\theta), \varepsilon) \\ + \left(\frac{\partial p(s;\theta)}{\partial s}\right)^2 \times U_{cc}(s, y - p(s;\theta), \varepsilon) \end{array} < 0 \tag{7}$$

for all $s, y, \theta, \varepsilon$. Intuitively, this says that the hedonic budget frontier should be 'less convex' to the origin than the indifference curves.[2] In particular, (7) holds if the budget frontier is strictly concave and indifference curves are strictly convex to the origin, as in our graph above.

Finally, the indirect utility function is given by

$$V(y,\theta,\varepsilon) = U(s^*(y,\theta,\varepsilon), y - p(s^*(y,\theta,\varepsilon);\theta), \varepsilon). \tag{8}$$

[2]The slope of the indifference curves in the $S - C$ axes are given by $-\frac{U_S}{U_C}$, whereas the budget curve has slope $-p'$. Then (7) is equivalent to the difference between $-\frac{U_S}{U_C} - (-p') = p' - \frac{U_S}{U_C}$ being strictly negative, i.e. the indifference curves are more convex than the budget frontier.

Then for fixed θ, ε, we have the "envelope theorem" condition that

$$
\begin{aligned}
& \frac{\partial V(y,\theta,\varepsilon)}{\partial y} \\
= \ & U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \\
& + \nabla U_s\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right)' \frac{\partial s^*\left(y,\theta,\varepsilon\right)}{\partial y} \\
& - U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \times \nabla_s p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)' \frac{\partial s^*\left(y,\theta,\varepsilon\right)}{\partial y} \\
= \ & \begin{bmatrix} \nabla U_s\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \\ -U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \end{bmatrix}' \nabla_s p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right) \\
& + U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \\
= \ & U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right), \text{ by (6)}
\end{aligned} \tag{9}
$$

Therefore, $V(\cdot,\theta,\varepsilon)$ is strictly increasing if utility is non-satiated in consumption of the numeraire.

Further, letting $p_j\left(s^*\left(y,\theta,\varepsilon\right);\theta\right) = \left.\frac{\partial p(s;\theta)}{\partial \theta_j}\right|_{s=s^*(y,\theta,\varepsilon)}$, we have that

$$
\begin{aligned}
& \frac{\partial V(y,\theta,\varepsilon)}{\partial \theta} \\
= \ & U_s\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right)' \frac{\partial s^*\left(y,\theta,\varepsilon\right)}{\partial \theta} \\
& - U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \times p_\theta\left(s^*\left(y,\theta,\varepsilon\right);\theta\right) \\
& - U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \times \nabla p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)' \frac{\partial s^*\left(y,\theta,\varepsilon\right)}{\partial \theta} \\
= \ & -U_c\left(s^*\left(y,\theta,\varepsilon\right), y - p\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)\right) \times p_\theta\left(s^*\left(y,\theta,\varepsilon\right);\theta\right), \text{ by (6)}
\end{aligned} \tag{10}
$$

From (9) and (10), it follows that for each $j = 1, 2, ..., \dim(\theta)$, it must hold that

$$
-\frac{\frac{\partial V(y,\theta,\varepsilon)}{\partial \theta}}{\frac{\partial V(y,\theta,\varepsilon)}{\partial y}} = p_\theta\left(s^*\left(y,\theta,\varepsilon\right);\theta\right) \tag{11}
$$

which can be interpreted as Roy's identity for the Hedonic budget set. The compensating variation $C \equiv C(y,\varepsilon)$ solves

$$
V(y + C, b, \varepsilon) = V(y, a, \varepsilon) \tag{12}
$$

where a and b denote the values of the hedonic price parameter θ before and after the policy intervention, respectively. This is well-defined since $\frac{\partial V(y,\theta,\varepsilon)}{\partial y} > 0$ with probability 1, by (9).

To find the distribution of C, suppose, initially, that we know the value of ε, then we can learn $p_\theta\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)$ from the hedonic price schedule in the data. Now, equation (11) can be rewritten as a system of linear, first-order partial differential equations of order 1

$$\frac{\partial V\left(y,\theta,\varepsilon\right)}{\partial\theta_j}+\frac{\partial V\left(y,\theta,\varepsilon\right)}{\partial y}\times p_{\theta_j}\left(s^*\left(y,\theta,\varepsilon\right);\theta\right)=0,\ j=1,...J \tag{13}$$

Therefore, the goal is to solve for (12), where $V\left(\cdot\right)$ satisfies (13). Note that we do not need to explicitly solve for $V\left(\cdot\right)$ using boundary conditions. We simply need to solve (12) for C. So, identifying $V\left(\cdot\right)$ up to any increasing transformation would suffice.

The key difficulty in calculating welfare effects nonparametrically is that ε is unobserved. To bypass this problem, we proceed as follows. First, assume s is a scalar. If ε is a scalar and $s^*\left(y,\theta,\varepsilon\right)$ is strictly monotone and invertible in ε then we can interpret the observed τth quantiles of $s^*\left(y,\theta,\varepsilon\right)$ as the demand of the individual who is located at the τth quantile of the distribution of ε. This is identified by the τth quantile of demand for those at income y on the budget set $p\left(s;\theta\right)$, i.e.

$$s^*\left(y,\theta,F_\varepsilon^{-1}\left(\tau\right)\right)=F^{-1}_{s^*\left(y,\theta,\varepsilon\right)}\left(\tau\right),$$

where $\tau\in[0,1]$, and $F^{-1}_{s^*\left(y,\theta,\varepsilon\right)}\left(\tau\right)$ equals the τth quantile of demand for those with income y and facing a budget frontier characterized by θ. Further, letting the indirect utility function

$$V\left(y,\theta,\varepsilon\right)=\max_{s,c}U\left(s,c,\varepsilon\right)\ \text{ s.t. }\ c=y-p\left(s,\theta\right)$$

by the envelope theorem, we have that

$$\frac{\partial}{\partial y}V\left(y,\theta,\varepsilon\right)=\frac{\partial U}{\partial c}\left(s^*\left(y,\theta,\varepsilon\right),y-p\left(s^*\left(y,\theta,\varepsilon\right),\theta\right)\right)>0 \tag{14}$$

when utility is strictly increasing in consumption – a credible assumption.

Now, differentiating the LHS of (6), we get that

$$\left.\begin{array}{l}\left\{U_{s^*s^*}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)-U_{s^*c}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{\partial p}{\partial s}\right\}\frac{ds^*}{d\varepsilon}\\+U_{s^*\varepsilon}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)-U_c\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{\partial^2p\left(s^*;\theta\right)}{\partial s}\frac{ds^*}{d\varepsilon}\\-\frac{\partial}{\partial s^*}p\left(s^*;\theta\right)\times U_{cs^*}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{ds^*}{d\varepsilon}\\+\left\{\frac{\partial}{\partial s^*}p\left(s^*;\theta\right)\right\}^2\times U_{cc}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{ds^*}{d\varepsilon}\\-\frac{\partial}{\partial s^*}p\left(s^*;\theta\right)\times U_{c\varepsilon}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\end{array}\right|=0,$$

implying

$$\frac{ds^*}{d\varepsilon}=-\frac{\begin{array}{l}U_{s^*\varepsilon}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\\ -\frac{\partial}{\partial s^*}p\left(s^*;\theta\right)\times U_{c\varepsilon}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\end{array}}{\begin{array}{l}U_{s^*s^*}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\\ -U_c\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{\partial^2 p(s^*;\theta)}{\partial s}\\ \left\{\frac{\partial}{\partial s^*}p\left(s^*;\theta\right)\right\}^2\times U_{cc}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{\partial p}{\partial s^*}\\ -2U_{s^*c}\left(s^*,y-p\left(s^*;\theta\right),\varepsilon\right)\frac{\partial p}{\partial s^*}\end{array}}$$

The denominator of this expression is negative by (7). The numerator equals

$$\left.\frac{\partial^2}{\partial s\partial\varepsilon}U\left(s,y-p\left(s;\theta\right),\varepsilon\right)\right|_{s=s^*(y,\theta,\varepsilon)}$$

represents how the marginal utility with respect to s changes with ε. Under the assumption that the marginal utility is strictly increasing in ε, which is akin to indifference curves of different types of consumers (i.e. different εs) crossing only once, we will have $\frac{ds^*}{d\varepsilon}>0$. Heckman et al 2010 derived the analogous result for the case where utility is quasilinear in consumption.

The monotonicity of s^* w.r.t. ε will be used below for identifying the distribution of welfare effects.

Now, letting

$$Q_\tau\left(y,\theta\right)\equiv V\left(y,\theta,F_\varepsilon^{-1}\left(\tau\right)\right),$$

we have from (14) that

$$\frac{\partial}{\partial y}Q_\tau\left(y,\theta\right)>0,$$

and from (6) that for all $\tau\in[0,1]$,

$$\frac{\partial Q_\tau\left(y,\theta\right)}{\partial\theta}+\left.\frac{\partial p\left(s,\theta\right)}{\partial\theta}\right|_{s=q(y,\theta,\tau)}\times\frac{\partial Q_\tau\left(y,\theta\right)}{\partial y}=0. \tag{15}$$

Equation (15) is a system of linear PDEs which can in principle be solved (see below for examples) for $Q\left(y,\theta,\tau\right)$, which gives the marginal CDF of indirect utility, as τ varies over $[0,1]$.

Now, consider a change in the value of θ from a to b. Then the compensating variation at income y, defined as the income supplement required to maintain the utility of the consumer at the τth quantile of ε, is given by the solution C to

$$Q_\tau\left(y+C,a\right) = Q_\tau\left(y,b\right), \text{ i.e. } C\left(y,\tau\right) = Q_\tau^{-1}\left(Q_\tau\left(y,b\right),a\right) - y,$$

and the equivalent variation given by

$$Q_\tau\left(y,a\right) = Q_\tau\left(y-E,b\right), \text{ i.e. } E\left(y,\varepsilon\right) = y - Q_\tau^{-1}\left(Q_\tau\left(y,a\right),b\right).$$

Example: To see how we compute these expressions from data, note that the function $p\left(s,\theta\right)$ and the demand $s^*\left(y,\theta,\tau\right)$ are identified from the observed data. Suppose the hedonic price function is given by

$$p\left(s,\theta\right) = \theta_1 + \theta_2 \ln\left(S\right).$$

Our goal is to find C which solves

$$Q_\tau(y+C,b) = Q_\tau\left(y,a\right),$$

where $Q_\tau\left(\cdot,\theta\right)$ is strictly increasing, and satisfies the system

$$\begin{aligned} \tfrac{\partial Q_\tau(y,\theta_1,\theta_2)}{\partial\theta_1} + \tfrac{\partial Q_\tau(y,\theta_1,\theta_2)}{\partial y} &= 0 \\ \tfrac{\partial Q_\tau(y,\theta_1,\theta_2)}{\partial\theta_2} + \ln q_\tau\left(y,\theta\right) \times \tfrac{\partial Q_\tau(y,\theta_1,\theta_2)}{\partial y} &= 0, \end{aligned} \tag{16}$$

where the function $q_\tau\left(y,\theta\right)$ represents the τth quantile of demand for the attribute as a function of income y and the parameters θ of the hedonic price function. This can be identified from the data by running a quantile regression of the demanded attribute on individual income and the market level θ, when there are a large number of markets, each with its own θ.

To solve for C, we go through the following steps. Start with (y_0,a_1,a_2). To go from a to b, at each step, we fix all but one component of ■ and then travel along a characteristic curve from one of the PDEs for Q_τ; this is the curve along which Q_τ remains constant as y and the non-fixed component of θ varies (cf. Courant and Hilbert 1952). Specifically, first find C_1 such that

$$Q_\tau(y_0+C_1,b_1,a_2) = Q_\tau(y_0,a_1,a_2), \tag{17}$$

i.e. for fixed a_2, the function $Q_\tau(\cdot,\cdot,a_2)$ remains constant along the curve (traced out by varying y and θ_1) that joins the points (y_0+C_1,b_1) and (y_0,a_1). In the next step, analogously, find C_2 such that

$$Q_\tau(y_0+C_2,b_1,b_2)=Q_\tau(y_0+C_1,b_1,a_2). \tag{18}$$

Then we shall have that

$$Q_\tau(y_0+C_2,b_1,b_2)=Q_\tau(y_0+C_1,b_1,a_2)=Q_\tau(y_0,a_1,a_2).$$

The solution $C=C_2$ would be unique, since $Q_\tau(\cdot,\theta)$ is strictly increasing for fixed values of θ (see (14)); indeed, if $Q_\tau(y_0+C_2,b_1,b_2)=Q_\tau(y_0+\tilde{C}_2,b_1,b_2)$, then $C_2=\tilde{C}_2$.

To calculate C_1, note from Eq. (16) that Q_τ is constant w.r.t. y and θ_1 on characteristic curves given by $\frac{dy}{d\theta_1}=1$. Fix $\theta_2=a_2$. Solve the ODE $\frac{dy}{d\theta_1}=1$ for the characteristic curve passing through $y=y_0,\theta_1=a_1$. This gives $y-\theta_1=y_0-a_1$, whence (17) implies $y_0+C_1-b_1=y_0-a_1$, i.e., $C_1=b_1-a_1$.

In the next step, we have to find C_2 such that

$$Q_\tau(y_0+C_2,b_1,b_2)=Q_\tau(y_0+C_1,b_1,a_2). \tag{19}$$

To do so, fix $\theta_1=b_1$ and solve the ODE

$$\frac{dy}{d\theta_2}-\frac{r_1}{\theta_2}y=r_0-r_1\frac{\theta_1}{\theta_2}.$$

The standard method of integrating factors gives the solution

$$\theta_2^{-r_1}\left(y-\frac{\theta_1}{r_1}-\frac{\theta_2}{1-r_1}r_0\right)=K$$

where K is a constant. Therefore,

$$H_2(y,\theta_1,\theta_2)\equiv\theta_2^{-r_1}\left(y-\frac{\theta_1}{r_1}-\frac{\theta_2}{1-r_1}r_0\right)$$

remains constant as (y,θ_2) vary with θ_1 held fixed. Then (19) is equivalent to

$$b_2^{-r_1}\left(y_0+c_2-\frac{b_1}{r_1}-\frac{b_2}{1-r_1}r_0\right)=a_2^{-r_1}\left(y_0+c_1-\frac{b_1}{r_1}-\frac{a_2}{1-r_1}r_0\right),$$

implying

$$\begin{aligned}c_2 &= \left(\frac{b_2}{a_2}\right)^{r_1}\left(y_0+b_1-a_1-\frac{b_1}{r_1}-\frac{a_2}{1-r_1}r_0\right)\\ &\quad-\left(y_0-\frac{b_1}{r_1}-\frac{b_2}{1-r_1}r_0\right)\end{aligned}$$

4.1 Multiple Attributes

To incorporate additional hedonic attributes into the above analysis, consider the utility function

$$U(s,x,c,\varepsilon) = U_1(s,\varepsilon) + U_2(x) + U_3(c), \text{ with } c = y - P(s,x),$$

where s is the key attribute of interest, x represents the other attributes, y is income, and the hedonic price function is given by

$$P(s,x) = p_1(s,\theta) + p_2(x,\delta)$$

We want to measure the distribution of the compensating variation C that solves

$$V(y,\theta,\delta,\varepsilon) = V(y+C,\beta,\delta,\varepsilon),$$

where

$$V(y,\theta,\delta,\varepsilon) = \max_{s,x} U(s,x,y-p_1(s,\theta)-p_2(x,\delta),\varepsilon)$$

Now, the first order conditions for maximization are given by

$$\begin{aligned} \frac{\partial U_1(s,\varepsilon)}{\partial s} - U_3'(y-p_1(s,\theta)-p_2(x,\delta)) \times \frac{\partial p_1(s,\theta)}{\partial s} &= 0 \qquad (20)\\ \nabla_x U_2(x) - U_3'(y-p_1(s,\theta)-p_2(x,\delta)) \times \nabla_x p_2(x,\delta) &= 0 \qquad (21) \end{aligned}$$

while the second order condition for an interior maximum is that the matrix

$$H = \begin{bmatrix} \begin{array}{c} \frac{\partial^2 U_1(s,\varepsilon)}{\partial s^2} - U_3'' \times \left(\frac{\partial p_1(s,\theta)}{\partial s}\right)^2 \\ +U_3' \times \left(\frac{\partial p_1(s,\theta)}{\partial s}\right)^2 \end{array} & \nabla_x p_2(x,\delta) \times U_3'' \times \frac{\partial p_1(s,\theta)}{\partial s} \\ \nabla_x p_2(x,\delta) \times U_3'' \times \frac{\partial p_1(s,\theta)}{\partial s} & \begin{array}{c} \nabla_{xx} U_2 - U_3' \times \nabla_{xx} p_2(x,\delta) \\ -U_3'' \times \nabla_x p_2(x,\delta) \times \nabla_x p_2(x,\delta)' \end{array} \end{bmatrix} \qquad (22)$$

is negative definite.

Now, evaluating the first-order conditions (20)-(21) at the optimal choice and differentiating w.r.t. ε, we have that

$$\frac{\partial U_1(s^*,\varepsilon)}{\partial s} - U_3'(y-p_1(s^*,\theta)-p_2(x,\delta)) \times \frac{\partial p_1(s^*,\theta)}{\partial s} = 0$$

$$\begin{bmatrix} \frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s \partial \varepsilon} + \frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s^2}\frac{\partial s^*}{\partial \varepsilon} \\ +U_3''(y-p_1(s^*,\theta)-p_2(x,\delta)) \times \left\{\frac{\partial p_1(s^*,\theta)}{\partial s}\right\}^2 \times \frac{\partial s^*}{\partial \varepsilon} \\ +U_3''(y-p_1(s^*,\theta)-p_2(x,\delta)) \times \frac{\partial p_1(x^*,\theta)}{\partial s} \times \nabla_x p_2(x,\delta)\frac{\partial x^*}{\partial \varepsilon} \\ -U_3'(y-p_1(s^*,\theta)-p_2(x,\delta)) \times \frac{\partial^2 p_1(s^*,\theta)}{\partial s^2}\frac{\partial s^*}{\partial \varepsilon} \end{bmatrix} = 0 \qquad (23)$$

Similarly

$$\nabla_{xx}U_2(x) \times \frac{\partial x^*}{\partial \varepsilon} - U_3''\left\{\nabla_x p_2(x,\delta)\nabla_x p_2(x,\delta)'\right\}\frac{\partial x^*}{\partial \varepsilon} - U_3''\left\{\nabla_x p_2(x,\delta)\right\}\frac{\partial p_1(s^*,\theta)}{\partial s}\frac{\partial s^*}{\partial \varepsilon} = 0 \tag{24}$$

Equations (23)-(24) can be written in matrix notation as

$$H \times \begin{bmatrix} \frac{\partial s^*}{\partial \varepsilon} \\ \frac{\partial x^*}{\partial \varepsilon} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s \partial \varepsilon} \\ 0 \end{bmatrix}$$

where H is defined in (22). Therefore,

$$\begin{bmatrix} \frac{\partial s^*}{\partial \varepsilon} \\ \frac{\partial x^*}{\partial \varepsilon} \end{bmatrix} = H^{-1}\begin{bmatrix} -\frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s \partial \varepsilon} \\ 0 \end{bmatrix}$$

implying

$$\frac{\partial s^*}{\partial \varepsilon} = -H^{11} \times \frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s \partial \varepsilon}$$

where H^{11} is the $(1,1)$th entry of the matrix H^{-1}. Now since H is negative definite, so is its inverse. Therefore H^{11} must be strictly negative. Therefore, if $\frac{\partial^2 U_1(s^*,\varepsilon)}{\partial s \partial \varepsilon} < 0$, then it follows that $\frac{\partial s^*}{\partial \varepsilon} > 0$. That is, for given y, θ, δ, we have that $s^*(y,\theta,\delta,\varepsilon)$ is strictly increasing in ε. Therefore, we have that for each $\tau \in [0,1]$,

$$s^*\left(y,\theta,\delta,F_\varepsilon^{-1}(\tau)\right) = F^{-1}_{s^*(y,\theta,\delta,\varepsilon)}(\tau),$$

i.e. the value of $s^*(y,\theta,\delta,\varepsilon)$ at the τth quantile of ε equals the τth quantile of s^* for fixed values of y, θ, δ.

For measuring the welfare effect of a change in θ, holding δ fixed, we follow the essentially the same steps as outlined above. In particular, we have that

$$V(y,\theta,\delta,\varepsilon) = U_1(s^*,\varepsilon) + U_2(x^*) + U_3(y - p_1(s^*,\theta) - p_2(x^*,\delta))$$

so that, by the envelope theorem, one gets

$$-\frac{\frac{\partial V(y,\theta,\delta,\varepsilon)}{\partial \theta}}{\frac{\partial V(y,\theta,\delta,\varepsilon)}{\partial y}} = \left.\frac{U_3' \times \frac{\partial p_1(s,\theta)}{\partial \theta}}{U_3'}\right|_{s=s^*(y,\theta,\delta),x=x^*(y,\theta,\delta)} = \left.\frac{\partial p_1(s,\theta)}{\partial \theta}\right|_{s=s^*(y,\theta,\delta)}.$$

This last simplification, i.e. that the RHS depends only on $s^*(y,\theta,\delta)$ and not on $x^*(y,\theta,\delta)$, results from the additive separability of the hedonic price function.

Evaluating this at $\varepsilon = F_{\varepsilon}^{-1}(\tau)$, we get (15) replaced by

$$\frac{\partial Q_{\tau}(y,\theta,\delta)}{\partial\theta} + \left.\frac{\partial p_1(s,\theta)}{\partial\theta}\right|_{s=F^{-1}_{s^*(y,\theta,\delta)}(\tau)} \times \frac{\partial Q_{\tau}(y,\theta,\delta)}{\partial y} = 0, \tag{25}$$

where $F^{-1}_{s^*(y,\theta,\delta)}(\tau)$ is the τth quantile of the optimal (i.e. chosen) s across individuals with income y in markets characterized by (θ,δ). Therefore, we can apply the method outlined in the previous subsection, holding δ fixed, and obtain the value of the compensating variation for each type defined by a quantile of ε.

4.2 Computation Steps

The computation of the above can be done through the following steps:

1. Construct the scalar index X which equals the first principal component of all non-S attributes (STATA command pca)

2. Divide locations into M markets. For each market, estimate the price function by regressing price of unit on $\ln S$ and X; call the coefficients $\alpha_{1m}, \alpha_{2m}, \delta_m$, $m = 1, ...M$

$$p_{mi}(S,X) = \alpha_{1m(i)} + \alpha_{2m(i)} \ln S_i + \delta_{m(i)} X_{m(i)}$$

3. Run a linear median regression (qreg in STATA), using all observations, of $\ln S_i$ on $\frac{y_i - \alpha_{1m(i)}}{\alpha_{2m(i)}}$ and $\delta_{m(i)}$

$$\begin{aligned} med\left(\ln S_i | y_i, \alpha_{1m(i)}, \alpha_{2m(i)}, \delta_{m(i)}\right) &= r_0 + r_1 \frac{y_i - \alpha_{1m(i)}}{\alpha_{2m(i)}} + r_2 \delta_{m(i)} + \varepsilon_i \\ q(\alpha_{1m}, \alpha_{2m}, \delta_m, y) &= \exp\left(r_0 + r_1 \frac{y_i - \alpha_{1m(i)}}{\alpha_{2m(i)}} + r_2 \delta_{m(i)}\right) \end{aligned}$$

   where $m(i)$ is the market in which i lives

4. Fix a value of $y = y_0$, $\delta = \delta_0$ (say, median values of y and δ in the data)

5. Then consider the change in α_1, α_2 from (a_1, a_2) to (b_1, b_2) (say, from the bottom quartile to top quartile

6. Calculate compensating variation as

$$C = \begin{array}{l} \left(\frac{b_2}{a_2}\right)^{r_1} \left(y_0 + b_1 - a_1 - \frac{b_1}{r_1} - \frac{a_2}{1-r_1} r_0\right) \\ -y_0 + \frac{b_1}{r_1} + \frac{b_2}{1-r_1} r_0 \end{array} \tag{26}$$

7. Replace 0.5 in Step 2 to other quantile, e.g. $\tau = 0.25, 0.75$ etc. and repeat steps 3-6.

5 Empirical Application

5.1 Data

The primary dataset used in our analysis comes from the restricted-access version of Wave 2015 of English Housing Survey (DCLG, 2018), which is a nationally representative survey providing data on the housing stock, conditions, and household characteristics. We use the data on rented properties. For each property, we observe the annual rent as well as a range of property characteristics, including floor area, number of floors, dwelling type (terrace, detached, flat, etc.), age, number of bathrooms, bedrooms, and living rooms, and the deprivation decile of the Lower Layer Super Output Area. We also observe housholds characteristics, including the tenure type (private, local authority, housing association), whether the household receives housing benefits, and net household income.

To proxy for school quality, we use the average point score per pupil for secondary schools. The point score comes from the Key Stage 4 data in the School Performance Tables, commonly known as *league tables*.[3] The data is publicly available via the UK government's official website, GOV.UK. We exclude independent, i.e. private, schools and schools for children with special educational needs, i.e. special schools, because they follow different addmission procedures and cater to a distinct population.

Each property is matched to the nearest school based on postcode proximity. The matching process was carried out using QGIS. We start with a total of 4,686 property-school matched observations. We then exclude 265 cases where household incomes are negative

[3] Key Stage 4 represents the two years of education for students aged 14 to 16, corresponding to Years 10 and 11 in the English education system.

after accounting for rent. To remove the outliers, we further drop the households with rent-to-income ratios below the 5th percentile and those within the top or bottom 5 percent of the income distribution, leading to excluding 584 observations. The final sample includes 3,837 properties (or households) observations, each matched to the nearest school.

Table 5 of the Appendix presents the descriptive statistics for the dataset. On average, households in our sample have a post-tax income of approximately £19,000, with around £6,000 allocated to rent. Around 29 percent of households rent privately, while the remainder are social renters, either from local authorities or housing associations. Approximately half of the respondents receive housing benefits, and 53 percent reside in areas within the three most deprived deciles.

5.2 Results

5.2.1 Parametric Results

In this section, we report the results obtained by applying the methods outlined in Section 3. To allow for the comparability of the results with the nonparametric estimates, we reduce the dimensionality of the property and household characteristics by using its first principal component.[4]

Table 1 presents the results of the hedonic regression for property rental prices. The rents are positively and significantly associated with the quality of the nearest school: an increase in one standard deviation in school quality is associated with approximately extra £396 or 7 percent of annual rent. The literature on the relationships between school quality in rental prices is scare, making it hard to compare our result to earlier findings.[5] Our estimate is at the upper bound of what is typically found in the larger literature that focuses on purchase

[4] Appendix Table 6 reports the correlations between the original variables, while Appendix Table 7 provides the loadings on the first principal component.

[5] To the best of our knowledge, Bayer, Ferreira, and McMillan (2007) is the only paper that includes an analysis of rental prices. They use a dataset that combines rental and purchase prices to study the association with elementary school quality. They found that households are willing to pay less than 1 percent more in house prices when the average school performance increases by 5 percent. We estimate a moderately higher relationships of 1.7 percent, focusing solely on secondary schools and renters.

property prices (see a survey by Machin (2011)).[6]

Table 1: Hedonic Regression Results.

	Annual rent
Logarithm of total average point score per pupil	1980.60***
	(173.40)
First principal component of property characteristics	254.23***
	(17.54)
Constant	-6860.85***
	(1022.64)
R-squared	0.08
Number of observations	3,837

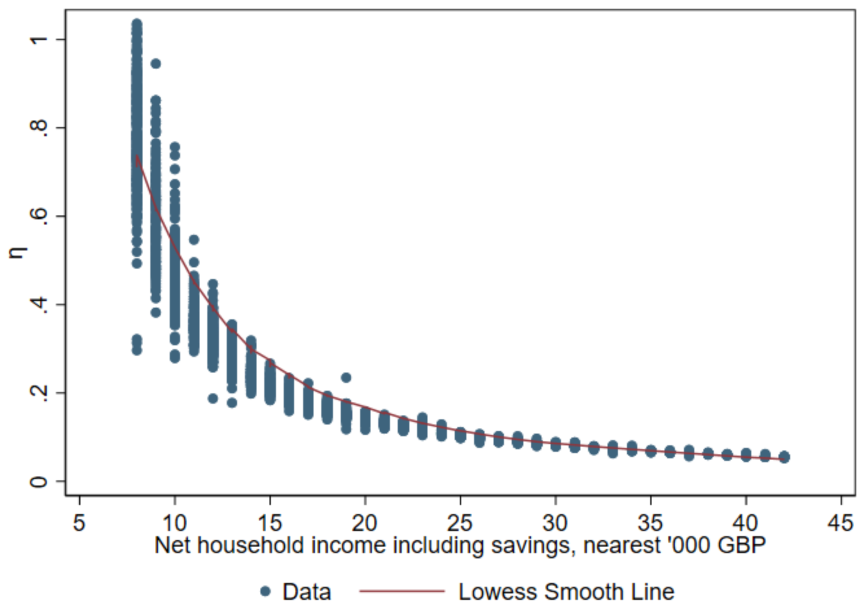
Note: Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

The approach in Section 3 allows estimating the household preference for school quality represented by the coefficient on the logarithm of school quality, η, in the utility function. Since the coefficient on the logarithm of consumption in the utility function is normalised to one, we interpret the coefficient on the logarithm of school quality as the ratio between the two. The model produces the average estimate of η at 0.24 with the standard deviation of 0.20. Then, to increase school quality by 1 percent, households, on average, are willing to sacrifice 0.20 percent of their consumption.

We estimate that, in our dataset, relatively poor households value school quality more than relatively rich households. This is in contrast with the earlier literature (see e.g. Bayer, P., Ferreira, F., & McMillan, R. (2007) who find that higher households income are associated with higher demand for better schools). Figure 1 presents the estimates of the preference for school quality, η, plotted against household income.

We now introduce a policy change that increases the sensitivity of rental prices to school quality by 5%. An example of such policy would be introducing more strict distance criteria

[6] We use the model with the first principal component instead of the full set of controls to maintain comparability between the parametric and nonparametic cases. The coefficient on the school quality is significantly lower in a regression with a full set of controls, as reported in the Appendix Table 8.



2.png

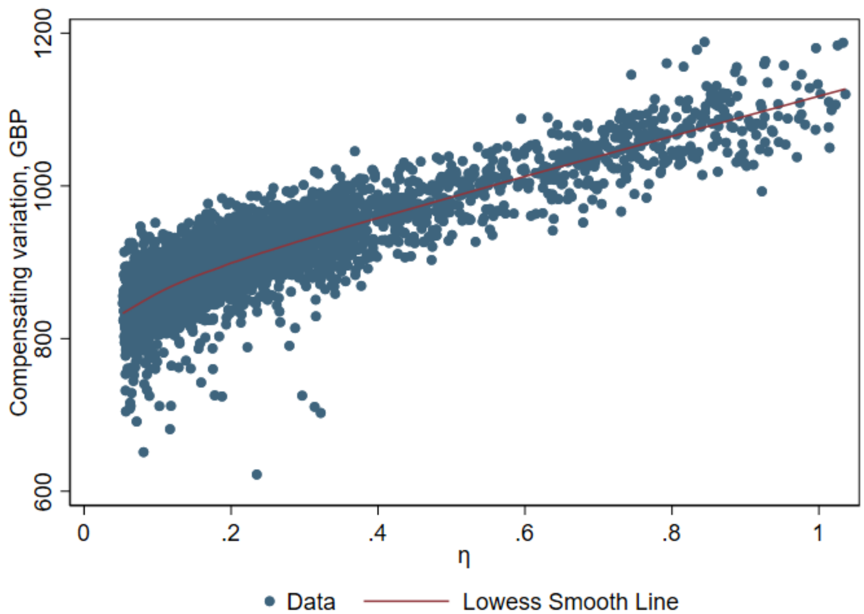
Figure 1: Relationship between Preferences for School Quality and Household Income.

for school admission. In particular, we change price function from $p(x, s, a) = a_0 + a_1' x + a_2 \ln s$ to $\tilde{p}(x, s, b) = b_0 + b_1' x + b_2 \ln s$, where $b_2 = 1.05 * a_2$ and $b_0 = 0.95 * a_0$. The results of the policy change are presented in Table 2. We observe that households reallocate to schools of lower quality: the average logarithm of school quality post-policy is lower than the average log of school quality in the current allocation.[7] From a welfare perspective, all households experience a welfare loss, with an average compensating variation of approximately £908.

Table 2: Effect of the Policy on Choose Choice and Welfare.

	Mean	SD	p5	p25	p50	p75	p95
Logarithm of school quality, pre policy	5.88	0.20					
Logarithm of school quality, post policy	5.09	0.27	4.62	4.92	5.11	5.28	5.48
Compensating variation, GBP	907.33	68.74	816.78	861.26	897.51	940.76	1046.54

[7] We limit potential choices of school quality to those currently existing in the dataset.



1.png

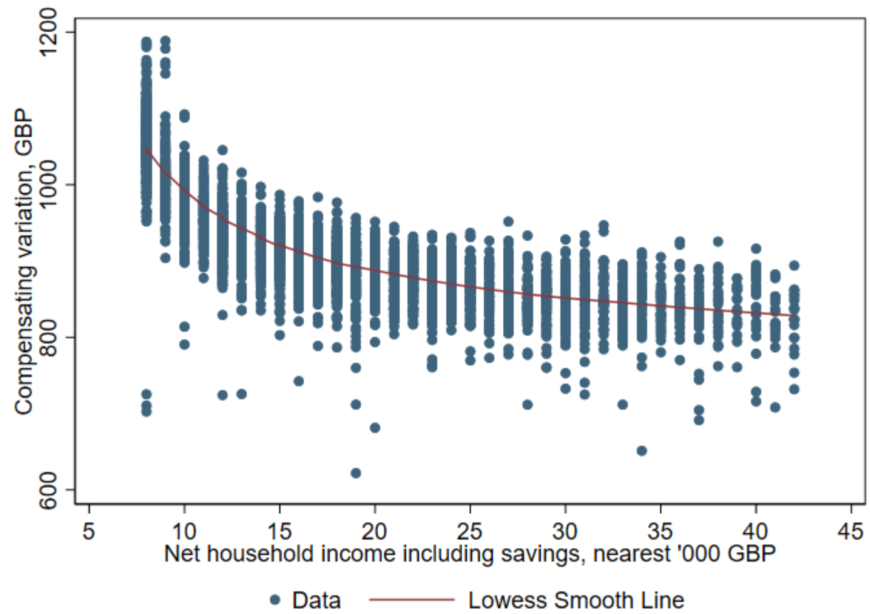
Figure 2: Relationship between Compensating Variation and Preferences for School Quality.

Unsurprisingly, the households that value school quality more lose more welfare that those who value school quality less. Figure 2 shows a positive relationship between the preference for school quality and compensating variation.

Relatively poor households, whose welfare is more sensitive to local school quality, lose more than relatively rich households as a result of the policy that increased the sensitivity of rental prices to local school quality. Figure 3 shows negatively sloping relationships between compensating variation and household income.

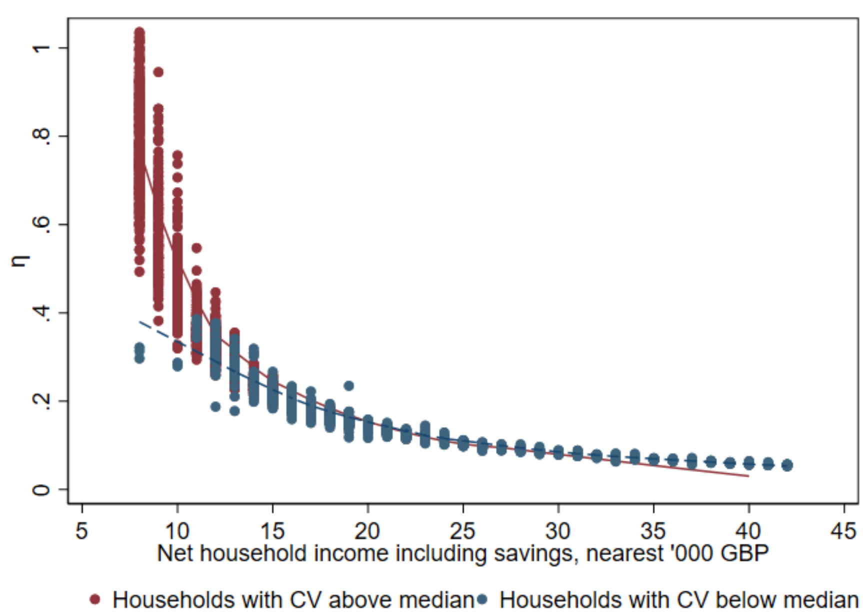
Table 3 provides the descriptive statistics of the households above and below the median of compensating variation. It shows a clear divide between the households that lose less and those who lose more both in household income, £25,000 a year vs. £14,000 a year, and in the preference for school quality, 0.12 vs. 0.36. Figure 4 illustrates this divide further by replicating Figure 1 for the two groups of households separately.

We propose the following explanation for this result: By construction, increasing the sensitivity of rental prices to school quality raises rents around better schools and lowers rents around worse schools. As a consequence of the price increase, households, on average,



3.png

Figure 3: Relationship between Compensating Variation and Household Income.



2a.png

Figure 4: Relationship between Preferences for School Quality and Household Income by the level of Compensating Variation.

Table 3: Characteristics of the Households below and above the Median of Compensating Variation.

	HH below Median CV		HH above Median CV	
	Mean	SD	Mean	SD
Compensating variation, GBP	856.12	32.17	958.54	56.32
Net household income, GBP	24773.08	7533.02	13581.27	4865.31
η	0.13	0.06	0.36	0.22
Observations	1899		1899	

Note: η is estimated preference for school quality from individual utility function.

reallocate to areas with worse schools, as shown in Table 2. All households incur welfare losses from moving to worse schools, but relatively poorer households suffer more because they place a higher value on school quality. This disparity may be further amplified, as relatively richer households are more likely to benefit from the reduction in rental prices around worse schools, given their lower preference for school quality.

5.2.2 Nonparametric results

In this section, we report the results obtained by applying the methods outlined in Section 4 with a single attribute, viz. school-quality. For the nonparametic estimation we split the dataset into 9 markets, represented by English regions: North, East Yorkshire and the Humber, North West, East Midlands, West Midlands, South West, East England, and South East London. Appendix Table 9 presents the results of hedonic regressions for rental prices estimated for different regions. The model is described in Step 2 of Section 4.2. The estimates for the relationship between rental prices and school quality vary from 183.04 in North East to 2344.44 in London. Appendix Table 10 reports the results of a linear quantile regression for the logarithm of the school quality presented in Step 3 of Section 4.2.

We estimate the welfare effects of changing the relationships between the school quality and rental prices from the 25th percentile of the distribution across the markets (1215.39 in North West) to the 75th percentile (1557.56 in East Midlands). We adopt respective coefficients for the constant term. The resulting estimates of compensating variation for different percentiles of the distribution of preference for school quality are presented in Table 4.

The non-parametric results are in line with the findings of the parametric case. The policy results in a universal welfare loss. As with the parametric results, households that value school quality more experience a greater welfare loss than those that place less value on it. The magnitude of the estimates is similar to those found in the previous section.

Table 4: The estimates of compensating variation for different percentiles of the distribution on the preference for school quality.

Percentile	Compensating variation
10th percentile	884.55
25th percentile	919.88
50th percentile	956.22
75th percentile	1001.82
90th percentile	1039.53

Appendix

A Supplementary Material for Parametric Estimates

Table 5: Descriptive statistics for matched property-school observations.

Variable	Mean	SD
Total annual rent	5819.23	2259.61
Net household income, GBP	19063.53	8488.34
Total average point score per pupil	364.02	67.94
Logarithm of total average point score per pupil	5.88	0.20
Floor area, sqm	66.31	21.04
Number of floors: 1	0.10	0.30
Number of floors: 2	0.68	0.47
Number of floors: 3	0.13	0.34
Number of floors: 4	0.04	0.20
Number of floors: 5 or more	0.05	0.21
Dwelling type: end terrace	0.12	0.33
Dwelling type: mid terrace	0.22	0.41
Dwelling type: semi detached	0.21	0.40
Dwelling type: detached	0.02	0.12
Dwelling type: bungalow	0.10	0.30
Dwelling type: converted flat	0.04	0.20
Dwelling type: purpose built flat, low rise	0.27	0.44
Dwelling type: purpose built flat, high rise	0.03	0.16
Dwelling age: pre 1850	0.01	0.10
Dwelling age: 1850 to 1899	0.05	0.22
Dwelling age: 1900 to 1918	0.06	0.24
Observations	3837	

Table 5: Descriptive statistics for matched property-school observations.

Variable	Mean	SD
Dwelling age: 1919 to 1944	0.13	0.34
Dwelling age: 1945 to 1964	0.27	0.44
Dwelling age: 1965 to 1974	0.17	0.37
Dwelling age: 1975 to 1980	0.07	0.26
Dwelling age: 1981 to 1990	0.09	0.28
Dwelling age: 1991 to 2002	0.08	0.27
Dwelling age: post 2002	0.07	0.25
Number of bedrooms: 1	0.22	0.42
Number of bedrooms: 2	0.36	0.48
Number of bedrooms: 3	0.37	0.48
Number of bedrooms: 4	0.04	0.19
Number of bedrooms: 5 or more	0.01	0.07
Number of living rooms: 0	0.01	0.08
Number of living rooms: 1	0.88	0.32
Number of living rooms: 2	0.10	0.30
Number of living rooms: 3	0.00	0.06
Number of living rooms: 4 or more	0.00	0.04
Number of bathrooms: 0	0.00	0.02
Number of bathrooms: 1	0.95	0.21
Number of bathrooms: 2	0.04	0.20
Number of bathrooms: 3	0.00	0.07
Number of bathrooms: 4	0.00	0.03
Tenure type: Private rented	0.29	0.45
Tenure type: Local Authority	0.30	0.46
Tenure type: Housing Association	0.41	0.49
Observations	3837	

Table 5: Descriptive statistics for matched property-school observations.

Variable	Mean	SD
Housing benefits: Yes	0.51	0.50
Housing benefits: No	0.49	0.50
Deprivation decile: 1 - Most deprived	0.22	0.41
Deprivation decile: 2	0.17	0.38
Deprivation decile: 3	0.14	0.35
Deprivation decile: 4	0.11	0.32
Deprivation decile: 5	0.09	0.29
Deprivation decile: 6	0.08	0.28
Deprivation decile: 7	0.07	0.25
Deprivation decile: 8	0.05	0.22
Deprivation decile: 9	0.04	0.20
Deprivation decile: 10	0.02	0.15
Region: North East	0.06	0.24
Region: Yorkshire and the Humber	0.12	0.32
Region: North West	0.14	0.35
Region: East Midlands	0.09	0.28
Region: West Midlands	0.11	0.32
Region: South West	0.10	0.30
Region: East England	0.12	0.33
Region: South East	0.14	0.35
Region: London	0.11	0.32
Observations	3837	

Table 6: Correlation between floor area and other property characteristics.

Variable	Floor area, sqm
Floor area squared	0.931
Number of floors: 2	0.179
Number of floors: 3	0.00710
Number of floors: 4	-0.0350
Number of floors: 5 or more	-0.0558
Dwelling type: mid terrace	0.198
Dwelling type: semi detached	0.226
Dwelling type: detached	0.263
Dwelling type: bungalow	-0.219
Dwelling type: converted flat	-0.114
Dwelling type: purpose built flat, low rise	-0.348
Dwelling type: purpose built flat, high rise	-0.0397
Dwelling age: 1850 to 1899	0.0332
Dwelling age: 1900 to 1918	0.0688
Dwelling age: 1919 to 1944	0.051
Dwelling age: 1945 to 1964	0.0362
Dwelling age: 1965 to 1974	-0.0586
Dwelling age: 1975 to 1980	-0.0624
Dwelling age: 1981 to 1990	-0.119
Dwelling age: 1991 to 2002	-0.0239
Dwelling age: post 2002	0.0749
Number of bedrooms: 2	-0.138
Number of bedrooms: 3	0.421
Number of bedrooms: 4	0.344
Number of bedrooms: 5 or more	0.256
Observations	3837

Table 6: Correlation between floor area and other property characteristics.

Variable	Floor area, sqm
Number of living rooms: 1	-0.266
Number of living rooms: 2	0.273
Number of living rooms: 3	0.147
Number of living rooms: 4 or more	0.0135
Number of bathrooms: 1	-0.290
Number of bathrooms: 2	0.246
Number of bathrooms: 3	0.182
Number of bathrooms: 5	0.0169
Deprivation decile: 1 most deprived = 2	-0.0258
Deprivation decile: 1 most deprived = 3	-0.0123
Deprivation decile: 1 most deprived = 4	-0.00969
Deprivation decile: 1 most deprived = 5	0.0206
Deprivation decile: 1 most deprived = 6	0.0193
Deprivation decile: 1 most deprived = 7	-0.0233
Deprivation decile: 1 most deprived = 8	-0.0178
Deprivation decile: 1 most deprived = 9	0.0408
Deprivation decile: 1 most deprived = 10	-0.00831
Tenure type: Local Authority	-0.0503
Tenure type: Housing Association	-0.0445
Housing benefits: No	0.0679
Observations	3837

Table 7: Principal component loadings and unexplained variance for property characteristics.

Variable	**Loadings on PC1**	**Unexplained Variance**
Floor area, sqm	0.4037	0.3517
Floor area squared	0.3762	0.4372
Number of floors: 2	0.2282	0.7929
Number of floors: 3	-0.0840	0.9720
Number of floors: 4	-0.0705	0.9802
Number of floors: 5 or more	-0.1049	0.9562
Dwelling type: mid terrace	0.1297	0.9331
Dwelling type: semi detached	0.1946	0.8494
Dwelling type: detached	0.1766	0.8760
Dwelling type: bungalow	-0.1372	0.9251
Dwelling type: converted flat	-0.0434	0.9925
Dwelling type: purpose built flat, low rise	-0.2510	0.7494
Dwelling type: purpose built flat, high rise	-0.0817	0.9734
Dwelling age: 1850 to 1899	0.0342	0.9954
Dwelling age: 1900 to 1918	0.0627	0.9843
Dwelling age: 1919 to 1944	0.0885	0.9689
Dwelling age: 1945 to 1964	0.0198	0.9984
Dwelling age: 1965 to 1974	-0.0871	0.9698
Dwelling age: 1975 to 1980	-0.0581	0.9866
Dwelling age: 1981 to 1990	-0.0633	0.9841
Dwelling age: 1991 to 2002	-0.0086	0.9997
Dwelling age: post 2002	0.0162	0.9990
Number of bedrooms: 2	-0.1683	0.8873
Number of bedrooms: 3	0.2803	0.6875
Number of bedrooms: 4	0.1675	0.8885
Observations	3837	

Table 7: Principal component loadings and unexplained variance for property characteristics.

Variable	**Loadings on PC1**	**Unexplained Variance**
Number of bedrooms: 5 or more	0.1207	0.9421
Number of living rooms: 1	-0.2717	0.7063
Number of living rooms: 2	0.2673	0.7158
Number of living rooms: 3	0.1010	0.9594
Number of living rooms: 4 or more	0.0154	0.9991
Number of bathrooms: 1	-0.2189	0.8094
Number of bathrooms: 2	0.1944	0.8497
Number of bathrooms: 3	0.1072	0.9543
Number of bathrooms: 4	0.0145	0.9992
Deprivation decile, 1 most deprived: 2	-0.0426	0.9928
Deprivation decile: 3	-0.0148	0.9991
Deprivation decile: 4	0.0030	1.0000
Deprivation decile: 5	0.0131	0.9993
Deprivation decile: 6	0.0207	0.9983
Deprivation decile: 7	0.0115	0.9995
Deprivation decile: 8	0.0051	0.9999
Deprivation decile: 9	0.0337	0.9955
Deprivation decile: 10	0.0027	1.0000
Tenure type: Local Authority	-0.0627	0.9843
Tenure type: Housing Association	-0.0348	0.9952
Housing benefits: No	0.0611	0.9852
Observations	3837	

Table 8: Regression of Rental Price on School Quality and Housing Characteristics

Variable	Annual rent	
	Est.	SE
Logarithm of total average point score per pupil	438.154***	(132.737)
Floor area, sqm	-1.718	(4.627)
Floor area squared	0.021	(0.022)
Number of floors: 2	-549.665	(782.566)
Number of floors: 3	-546.859	(786.789)
Number of floors: 4	-284.931	(794.709)
Number of floors: 5 or more	-414.423	(805.263)
Dwelling type: mid terrace	10.714	(91.238)
Dwelling type: semi detached	19.180	(94.343)
Dwelling type: detached	328.627	(232.894)
Dwelling type: bungalow	-549.596	(785.803)
Dwelling type: converted flat	225.234	(171.536)
Dwelling type: purpose built flat, low rise	82.798	(110.693)
Dwelling type: purpose built flat, high rise	180.914	(260.403)
Dwelling age: 1850 to 1899	333.936	(266.189)
Dwelling age: 1900 to 1918	599.068**	(264.359)
Dwelling age: 1919 to 1944	759.981***	(258.878)
Dwelling age: 1945 to 1964	689.686***	(255.606)
Dwelling age: 1965 to 1974	733.169***	(258.054)
Dwelling age: 1975 to 1980	735.382***	(267.546)
Dwelling age: 1981 to 1990	885.409***	(264.608)
Dwelling age: 1991 to 2002	782.403***	(265.320)
Dwelling age: post 2002	1034.321***	(268.680)
Number of bedrooms: 2	495.065***	(85.073)
Observations	3837	

Table 8: Regression of Rental Price on School Quality and Housing Characteristics

Variable	Annual rent	
	Est.	SE
Number of bedrooms: 3	948.227***	(113.827)
Number of bedrooms: 4	1398.225***	(183.901)
Number of bedrooms: 5 or more	695.818*	(390.612)
Number of living rooms: 1	30.286	(332.877)
Number of living rooms: 2	144.038	(345.274)
Number of living rooms: 3	545.582	(547.391)
Number of living rooms: 4 or more	-262.646	(680.054)
Number of bathrooms: 1	3526.067**	(1602.798)
Number of bathrooms: 2	4261.406***	(1607.767)
Number of bathrooms: 3	5153.728***	(1650.977)
Number of bathrooms: 5	8418.313***	(1841.039)
Tenure type: Local Authority	-2688.617***	(78.268)
Tenure type: Housing Association	-2079.075***	(69.769)
Housing benefits: No	-77.874	(52.872)
Deprivation decile: 2	-73.997	(83.082)
Deprivation decile: 3	1.986	(89.146)
Deprivation decile: 4	156.026	(96.019)
Deprivation decile: 5	391.328***	(104.397)
Deprivation decile: 6	564.317***	(107.837)
Deprivation decile: 7	532.712***	(116.460)
Deprivation decile: 8	141.440	(131.126)
Deprivation decile: 9	608.439***	(144.938)
Deprivation decile: 10	495.030***	(181.481)
Region: Yorkshire and the Humber	-53.918	(124.884)
Observations	3837	

Table 8: Regression of Rental Price on School Quality and Housing Characteristics

Variable	Annual rent	
	Est.	SE
Region: North West	138.634	(122.000)
Region: East Midlands	110.975	(133.278)
Region: West Midlands	426.660***	(125.830)
Region: South West	478.500***	(132.135)
Region: East England	1064.652***	(125.899)
Region: South East	1511.217***	(124.711)
Region: London	3075.788***	(134.679)
Constant	-458.425	(1969.894)
Adjusted R-squared	0.53	
Observations	3837	

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

B Supplementary Material for Nonparametric Estimates

Table 9: Hedonic Regression Results by Region.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	North East	Yorkshire and the Humber	North West	E. Midlands	W. Midlands	South West	East England	South East	Lond
School quality	183.04	1444.45***	1215.39***	1557.56***	870.89***	1609.16***	1412.95***	1518.81***	2344.4
	(199.86)	(308.84)	(353.23)	(475.76)	(335.34)	(548.30)	(425.77)	(486.54)	(1034.
First PC	163.50***	330.41***	242.35***	283.76***	255.02***	300.50***	354.92***	287.16***	584.06
	(33.49)	(30.10)	(30.51)	(40.57)	(32.40)	(43.58)	(42.12)	(50.28)	(85.9
Constant	3502.68***	-3604.40**	-2028.26	-4064.48	99.83	-3680.16	-2083.54	-2252.04	-5639.
	(1162.66)	(1805.68)	(2073.23)	(2776.58)	(1966.96)	(3253.55)	(2502.25)	(2875.15)	(6144.
R2	0.10	0.23	0.12	0.16	0.14	0.13	0.16	0.07	0.10
Observations	244	455	543	328	439	376	472	547	433

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 10: The results of quantile regression for 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles

	10th perc.	25th perc.	50th perc.	75th perc.	90th perc.
r_1	-0.00045	0.00043*	0.00073***	0.00019	-0.00036
	(0.00035)	(0.00022)	(0.00021)	(0.00022)	(0.00041)
r_2	0.00023***	0.00026***	0.00025***	0.00018***	0.00017**
	(0.00006)	(0.00004)	(0.00003)	(0.00003)	(0.00007)
r_0	5.59476***	5.68747***	5.78996***	5.92980***	6.04663***
	(0.02192)	(0.01398)	(0.01295)	(0.01364)	(0.02594)
Observations	3837	3837	3837	3837	3837

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

References

1. Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. Journal of political economy, 115(4), 588-638.

2. Black, S.E., 1999. Do better schools matter? Parental valuation of elementary education. The quarterly journal of economics, 114(2), pp.577-599.

3. Department for Communities and Local Government. (2018). English Housing Survey, 2014-2016: Secure Access. [data collection]. 3rd Edition. UK Data Service. SN: 8121, http://doi.org/10.5255/UKDA-SN-8121-3

4. Gibbons, S., Machin, S. and Silva, O., 2013. Valuing school quality using boundary discontinuities. Journal of Urban Economics, 75, pp.15-28.

5. Hausman, J.A. and Newey, W.K., 2016. Individual heterogeneity and average welfare. Econometrica, 84(3), pp.1225-1248.

6. Heckman, J.J., Matzkin, R.L. and Nesheim, L., 2010. Nonparametric Estimation of Nonadditive Hedonic Models, Econometrica, 78(5), pp.1569-1591.

7. Hausman, J.A. and Newey, W.K., 2016. Individual heterogeneity and average welfare. Econometrica, 84(3), pp.1225-1248.

8. Machin, S. (2011). Houses and schools: Valuation of school quality through the housing market. Labour Economics, 18(6), 723-729.

9. Sheppard, S., 1999. Hedonic analysis of housing markets. Handbook of regional and urban economics, 3, pp.1595-1635.