PANEL DATA HEDONICS:  
ROSEN'S FIRST STAGE AS A "SUFFICIENT STATISTIC"

H. Spencer Banzhaf*
Georgia State University, PERC, NBER

April 2017

I thank David Albouy, Olivier Beaumais, Kelly Bishop, Jared Carbone, Kerry Smith, Chris Timmins, Reed Walker and Jeffrey Zabel for helpful comments and suggestions on previous drafts. I am especially grateful to Paul Ferraro and Nick Kuminoff for very detailed comments and discussions. I also thank seminar participants at the Univ. of Illinois, Univ. of South Carolina, and Yale Univ. and participants at an Urban Economics Association session at the 2013 NARSC meetings (Atlanta, GA), the 2013 National Tax Association meetings (Tampa, FL), the 2014 Workshop on Nonmarket Valuation (Aix-en-Provence, France), and the 2015 NBER Summer Institute.

* Professor, Dept. of Economics, Georgia State University, PO Box 3992, Atlanta, GA, 30302. hsbanzhaf@gsu.edu.
Panel Data Hedonics:
Rosen's First Stage as a "Sufficient Statistic"

Abstract
Traditional cross-sectional estimates of hedonic price functions can recover marginal willingness to pay for characteristics, but face endogeneity problems for estimating the demand functions required for non-marginal welfare measures. I show that when panel data on household demands are available, one can construct a second-order approximation to non-marginal welfare measures using only the first-stage marginal prices. Under a single-crossing restriction, the approach remains valid for repeated cross sections of product prices. A variant also remains valid when individual demands shift over time.

1. Introduction
For decades, the hedonic model has been the starting point for understanding people's values for differentiated products (Palmquist 2005). Its applications include willingness to pay in higher housing prices for local public goods and spatial amenities, compensating wage differentials for attributes like job safety, pricing of quality-differentiated consumer products like computers and cars, and quality-adjustments in national accounting.

Part of the hedonic model's appeal has always been the simple relationship between hedonic prices and consumer demand: the derivative of a hedonic price function with respect to a characteristic is equal to a household's marginal willingness to pay for the characteristic. This aspect of the model is highly appealing because price functions can feasibly be estimated with simple, transparent research designs, yet they also have a clear welfare interpretation.

Unfortunately, the marginal willingness to pay potentially observed from only the hedonic price gradient is generally viewed as inadequate information for welfare evaluations of large policy shocks. Accordingly, since Rosen (1974), economists have sought to identify households' willingness-to-pay functions for amenities in a second stage. But recovering these willingness-to-pay functions from a single cross section has proved to be a challenge. Because only one point on each individual's demand function is observed (where marginal WTP is equal to the derivative of the hedonic price function), the only variation in the data comes from the way different households sort across choice alternatives in equilibrium. The standard solution is to model heterogeneity in individual demands. Unfortunately, the unobserved components in demand (e.g. tastes) systematically vary both with levels of amenities and their marginal prices.
This correlation gives rise to a well-known endogeneity problem (Bartik 1987, Bishop and Timmins 2017, Epple 1987, Heckman, Matzkin, and Nesheim 2010a).

Proposed solutions to this problem combine, in one way or another, the economic logic of sorting along with some structure imposed on heterogeneity in tastes. In the hedonic model, Ekeland, Heckman, and Nesheim (2004) consider the case of additive hedonic models, noting that nonlinearities in the equilibrium price function justify using nonlinear functions of observed demand shifters as instruments for the observed quantities demanded. Heckman, Matzkin, and Nesheim (2005, 2010a) and Bishop and Timmins (2017) discuss strategies for imposing functional form restrictions that allow one to map quantities of characteristics demanded into demands. Departing somewhat from the continuous world of the hedonic model, structural models of discrete choices follow a similar basic strategy (e.g. Bayer, Ferreira, and McMillan 2007, Kuminoff 2012, Sieg et al. 2004; see Kuminoff, Smith, and Timmins 2013 for a review). In general, because the economics of the models imply a particular mapping from households' preferences to the way they sort in equilibrium, the logic can be inverted to recover preferences from observed sorting, conditional on the assumed structure. Typically, this involves imposing distributional assumptions about unobserved tastes. For example, they may be assumed to have an extreme value distribution (e.g. Bayer, Ferreira, and McMillan 2007) or a log-normal distribution (e.g. Sieg et al. 2004); similarly, WTP functions may be assumed to have errors following some known distribution (e.g. normal in Bishop and Timmins 2017). Alternatively, one can relax these distributional assumptions but forego point identification of the underlying parameters (Kuminoff 2012). This literature has provided a tremendous advance on our ability to model general equilibrium counterfactuals as well as non-marginal welfare effects. However, these advantages come at the price of imposing additional structure and complication—thus losing some of the simple reduced-form appeal of the hedonic model.

As Bajari and Benkard (2005) and Kuminoff and Pope (2012) have pointed out, the problem becomes considerably easier when individuals are observed in multiple settings, as then individuals' willingness-to-pay functions can be fitted to two or more points. To my knowledge, however, the literature has not noticed that when households are observed two or more times, their observed choices, together with knowledge of the first stage hedonic price functions, are sufficient to estimate welfare measures that are proportional to second-order approximations to a change in utility for any constant demand function—without explicitly modeling heterogeneity at
all and without imposing any distributional assumptions on unobserved demand parameters.¹

With the advent of "big data," such panel data are becoming increasingly available, even in the context of housing markets. For example, in the United States, researchers are beginning to make use of data available under the Home Mortgage Disclosure Act (HMDA) to match households to the houses they live in over time (e.g. Bayer, McMillan, and Rueben 2011, Bayer et al. 2012, Bishop and Timmins 2017, Depro and Timmins 2012). In principle, such data may be even easier to come by in other contexts, such as automobile or computer purchases.

In this paper, I show how such data provide us with an opportunity to reinterpret the hedonic model in the spirit of calls from Chetty (2009) and Heckman (2010) to seek compromises that combine the clarity of reduced form econometric models with the ability of structural models to speak to welfare effects. Chetty (2009) recommends economists look for simple "sufficient statistics" that can be used to quantify non-marginal welfare measures. Heckman (2010) similarly urges us to follow "Marschak's maxim" and solve well-posed economic problems with minimal assumptions. In this spirit, I show that, with multiple time periods, it is possible to combine the economic logic of the hedonic model with estimation of only Rosen's first stage hedonic price function to identify non-marginal welfare effects under minimal assumptions. This is in contrast to the standard view that knowledge of the hedonic price function alone is insufficient to analyze welfare effects of large policy shocks with general equilibrium effects. The argument parallels Harberger's (1971) argument for thinking of consumer surplus in terms of an index number, averaging ex ante and ex post marginal values.

In particular, I first consider the situation where it is possible to estimate a hedonic price function in a single cross-section, but where we seek more welfare information than marginal values. This has been the traditional hedonic approach for decades and continues to be invoked in many models (e.g. Bishop and Timmins 2017, Ekeland, Heckman, and Nesheim 2004, Heckman, Matzkin, and Nesheim 2005, 2010a). I assume the observed policy is the only change to the economic environment shifting implicit prices or equilibrium levels of characteristics; thus

¹ Kuminoff and Pope (2012) suggest using exogenous shifts in the supply of amenities to derive within-market instruments for Rosen’s "second stage." Although their suggestion is based on the same basic insight of this paper (that exogenous supply shocks can trace out a demand curve), in contrast I am suggesting that a similar procedure replace the second stage entirely, to identify a sufficient statistic for welfare measurement without estimating the deep structural parameters.
the approach taken here is most relevant for a narrow window surrounding a relatively sudden policy shock. In this case, the first-stage hedonic price functions can be used to derive a "sufficient statistic" for non-marginal welfare changes, in the sense of Chetty (2009). This approach provides a hybrid between simpler reduced-form and structural approaches to hedonic estimation. It simplifies the estimation problem by side-stepping the difficult endogeneity problem of Rosen's second stage or associated structural models, while using the full economics of the hedonic model to make inference about sorting and welfare effects.

I also consider the minimal assumptions about household demands for this approach. In the case where panel data on households and their choices are available and where demands are constant, the sufficient statistic approach is feasible under very general conditions and heterogeneity does not need to be modeled explicitly. In Sections 4 and 5, I show that when only repeated cross-sections of hedonic prices are available or where demands are shifting, the sufficient statistic approach remains feasible under additional restrictions to heterogeneity, namely a single-crossing restriction commonly invoked in the existing literature.

Finally, I demonstrate these results using simulations of hedonic equilibria. The results of the simulations are consistent with the empirical approach outlined. The sufficient statistic approach is a compromise between two Hicksian welfare measures (compensating variation [CV] and equivalent variation [EV]). I also find that, in practice, this approach may be a reasonable approximation even when the assumptions justifying it do not hold.

To fix ideas, I specifically discuss the example of housing markets with spatially varying amenities and I primarily will discuss connections to that literature. However, the implications of this paper are not limited to that setting and apply equally to labor markets or to other contexts with differentiated commodities.

2. Model Basics

Consider a closed city (or region) with a constant set of households. Let \( \mathcal{H} \) denote the set of houses with typical element \( h \) and let \( \mathcal{I} \) denote the set of households with typical element \( i \). Equilibrium in each time period consists of a one-to-one correspondence of households to houses.

\[ \text{Equation} \]

---

\[ ^2 \text{The area modeled does not literally need to be one city (or housing market). Nor need it coincide with the area affected by the policy of interest. However, as always, economists modeling demand must make judgments about the set of relevant substitutes.} \]
(all households occupy a house and all houses are occupied by a household). Households rent their houses from absentee landlords.\(^3\)

Houses are differentiated by price \(p\), the continuous amenity of interest \(g\), and a vector of continuous housing characteristics \(x\) with characteristics indexed by \(r=\{1,\ldots,R\}\) (lot size, dwelling size, and so forth). (The variable \(g\) may be thought of as an index of public goods, or alternatively other public goods of secondary interest may be thought to be embedded in \(x\).) Notationally, it will sometimes be more convenient to work with a more parsimonious notation with the vector \(z'=\begin{bmatrix}g, x'\end{bmatrix}\) and with the elements of \(z\) indexed by \(j\).

At any point in time \(t\), households differ by their income \(y\) and by their current-period preferences, which can be represented by a twice differentiable quasi-concave conditional indirect utility function \(v^t_i(y^t_i-p_h, g_h, x_h)\), with \(\partial v^t_i/\partial y^t_i > 0\) and \(\partial v^t_i/\partial g \neq 0\) everywhere \(\forall i\). Note that \(-\partial v^t_i/\partial p = \partial v^t_i/\partial y^t_i \equiv \lambda^t_i\).

On the supply side of the market, the profit function for house \(h\) is \(\pi_h = p_h - c_h(x_h)\), where the cost function \(c_h(\ )\) is twice differentiable. For convenience, I assume \(c_h(\ )\) is constant over time, although this assumption could be relaxed.

Consider two time periods, with \(t=0\) in the initial situation and \(t=1\) in a later situation. Let \(F(\ )\) be the distribution function of \(g\) at time \(t\). Prices of houses are determined by the amenities and the equilibrium price function: \(p^t_h = p^t(g^t_h, x^t_h)\). The time superscript on the hedonic price function indicates that equilibrium hedonic prices may shift over time. In principle, these shifts may happen from changes in the distribution of \(g\) or changes in household demands, but not other changes in the economic environment.

In the initial situation, the household maximizing utility over a continuous choice set defined by the continuously differentiable hedonic function \(p^0=g^0(x^0)\). A policy then exogenously shocks the distribution of the amenity \(g\) available in the city. Consequently, the equilibrium price function adjusts to \(p^1=g^1(x^1)\), with the set of other available characteristics \(x\) possibly changing endogenously.

\(^3\) I impose this restriction here to facilitate the exposition. Strictly speaking, this assumption is only required for the models described in Sections 2.3 and 2.4. The basic model in Section 2.2 does not require this assumption.
I make the standard hedonic assumption that households are in a static equilibrium in each
time period, either because they choose a new product each period (as would be fitting for appli-
cations to consumer products) or because they can costlessly re-optimize. Maximizing utility in
period $t$, the household satisfies the first-order condition:

\[
\frac{\partial v_i^t}{\partial z_j} = - \frac{\partial v_i^t}{\partial p} \frac{\partial p^t}{\partial z_j} = \lambda_i^t \frac{\partial p^t}{\partial z_j}.
\]

Equation (1) represents the standard tangency condition, in which the derivative of the hedonic
function with respect to an amenity is equal to marginal willingness to pay for the amenity at the
optimal point.

Similarly, the landlord's first-order condition for profit maximization is

\[
\frac{\partial c_k}{\partial x_r} = \frac{\partial p^t}{\partial x_r}.
\]

The endogenous amenities $x$ are supplied according to similar tangency condition, with marginal
cost of supply equal to the marginal revenue.

The basic problem is to make inferences about non-marginal welfare effects from these
primitive conditions.

3. Non-Marginal Values when Demands are Constant and a Panel Of Households Is Avail-
able: The Hedonic Harberger Triangle

As a starting point, consider the arguably restrictive case covered by the following three assump-
tions.

ASSUMPTION A1 (Panel of Households). Panel data on household choices are available, so that $p$
and $\partial p^t / \partial z_j$ can be evaluated for each household in each time period at their choice of $z$.

---

4 This assumption continues to underlie the vast majority of work on hedonic markets (e.g. Bajari and
Benkard 2005, Bishop and Timmins 2015, Ekeland, Heckman, and Nesheim 2004, Heckman, Matzkin,
and Nesheim 2010) as well as structural sorting models of locational choice (e.g. Bayer, Ferreira, and
McMillan 2007, Kuminoff 2012, and Sieg et al. 2004). However, recent work is beginning to consider
dynamic optimization in the context of transaction costs, which may be substantial in applications to
housing (Bayer et al. 2016, Bishop 2012, Kennan and Walker 2011). The labor literature has a longer tra-
dition of considering such dynamic optimization (e.g. Keane and Wolpin 1997).
ASSUMPTION A2 (Constant Demands). Households' preferences and incomes are constant over the time period considered: \( v_{i}^{t}(y_{i}^{t}-p_{h}, g_{h}, x_{h}) = v_{i}(y_{i}^{t}-p_{h}, g_{h}, x_{h}) \) \( \forall t \), so that the optimal vector \( x \) and \( g \) are unchanging functionals of the hedonic price function.

ASSUMPTION A3 (Constant Economic Environment). The relevant policy shock to the distribution of \( g \) is the only change in the economic environment shifting slopes of the hedonic price function and the equilibrium levels of \( x \).

In this subsection, I show that, under Assumptions A1-A3, a second-order approximation of the general equilibrium welfare effects of a change in amenities can be constructed using only estimated marginal prices. The assumptions essentially guarantee that observed shifts in conditions are what is to be evaluated (A3), that these shifts trace out a demand curve (A2), and that points on the demand curve are observed (A1). Given Assumption A3, the approach taken in this section may be most relevant for sudden changes in conditions, such as discovery of a cancer cluster (Davis 2004) or the release of school report cards (Figlio and Lucas 2004). In subsequent sections, Assumptions A1 and A2 will be relaxed in turn.

For consumers, the Marshallian measure of the benefits of the (non-marginal) change in the distribution of \( g \) from \( F^{0} \) to \( F^{1} \) is given by:

\[
WTP_{i} = -dp_{i} + \int_{t=0}^{1} \frac{\partial f}{\partial g} [g_{i}^{*}(t), x_{i}^{*}(t)]dt,
\]

where \( x_{i}^{*}(t) \) and \( g_{i}^{*}(t) \) are the household's optimal levels of the respective amenities at notional time \( t \) given the prevailing price function; where \( g_{i}^{*}(t), x_{i}^{*}(t), \) and the price function continuously adjust between \( t=0 \) and \( t=1 \); and where \( dp_{i} \) is the price change experienced by the household (Scotchmer 1986, Bartik 1988). Equation (3) is impossible to observe literally, but it is instructive. It reduces the problem of measuring non-marginal willingness to pay to an index number problem, that is, to an average of marginal willingness to pay along the path between \([p^{0}, g^{0}, x^{0}]\) and \([p^{1}, g^{1}, x^{1}]\).

Although Equation (3) is based on a Marshallian construct, Arnold Harberger famously suggested that a linear approximation to (3) could be interpreted as a valid approximation to an exact welfare measure.\(^5\) More recently, Chetty (2009) has suggested that Harberger's approach

---

\(^5\) See Banzhaf (2010) for a discussion of this approach to welfare measurement in a historical context.
can be thought of as setting a paradigm for sufficient-statistic welfare measurement. Following Harberger (1971), consider a second order approximation to a change in utility for an individual in the hedonic model given Assumptions A2 and A3:

\[
 dv_i \approx \frac{\partial v_i}{\partial p} dp_i + \sum_j \frac{\partial v_i}{\partial z_j} dz_{j,i} + \frac{1}{2} \frac{\partial^2 v_i}{\partial p^2} dp_i^2 + \sum_j \frac{\partial^2 v_i}{\partial p \partial z_j} dp_i dz_{j,i} \\
+ \frac{1}{2} \sum_j \sum_{j'} \frac{\partial^2 v_i}{\partial z_j \partial z_{j'}} dz_{j,i} dz_{j',i},
\]

(4)

where \( dp_i = p^1(g^1_i, x^1) - p^0(g^0_i, x^0) \) is the household's change in expenditure and \( dz_{j,i} \) is the change in amenity \( j \) experienced by household \( i \) after all adjustments. These changes stem from a number of sources. At the household's initial optimal location, \( g \) may change directly from the policy and the price of the home may capitalize this change. Additionally, \( p \) changes as the hedonic function shifts. Finally, \( p, g, \) and \( x \) may all change from any readjustments by the household as it re-optimizes, and \( x \) also may change from any supply-side investments as landlords re-optimize. Whatever the source of the changes, welfare effects are evaluated taking all of them into account.

Equation (4) leads to the following lemma.

**Lemma 1.** Given Assumptions A1-A3, a second order approximation to the change in welfare for each consumer, \( dw_i \), from an exogenous change in the distribution of \( g \), can be constructed from observed prices and estimated marginal prices as follows:

\[
dw_i = \frac{dv_i}{\frac{1}{2}(\lambda^0_i + \lambda^1_i)} \approx -dp_i + \sum_j \frac{1}{2} \left( \frac{\partial p^0}{\partial z_j} |_{z^0_i} + \frac{\partial p^1}{\partial z_j} |_{z^1_i} \right) dz_{j,i}.
\]

(5)

**Proof:** See the appendix.

This expression is proportional to the utility change \( dv_i \), which is converted to the measuring rod of money using the average marginal utility of income, averaged between the starting point and ending point. Lemma 1 states that this change in welfare for a consumer is given by the change in rents \( dp_i \), plus the change in housing attributes and public goods experienced by the household after all adjustments, multiplied by the average marginal willingness to pay, again averaged between the starting point and ending point. The expression might be thought of as a "hedonic Harberger triangle" (or trapezoid). It is a compromise between two Hicksian measures,
CV and EV.

On the supply side of the market, landlords are directly better off by the change in rents \( dp \). This change in rents stems from shifts in the price function and from exogenous changes in \( g \), but also potentially from adjustments to \( \mathbf{x} \) that are costly to supply. Consequently, the cost of producing the change in \( \mathbf{x} \) must be netted out of the change in profits. The change in profits from any change in \( g \), the price function \( p(\cdot) \), or endogenous adjustments to \( \mathbf{x} \) is \( d\pi = dp - dc \). We can in turn take a second-order approximation to \( dc \) as follows.

\[
(6) \quad dc_h \approx \sum_r \frac{\partial c_h}{\partial x_r} dx_{r,h} + \frac{1}{2} \sum_r \sum_r \frac{\partial^2 c_h}{\partial x_r \partial x_r} dx_{r,h} dx_{r',h}
\]

This fact along with the first-order conditions leads to the following lemma.

**LEMMA 2.** A second order approximation to the change in profits for each landlord, \( d\pi_h \), can be constructed from observed prices and estimated marginal prices as follows:

\[
(7) \quad d\pi_h \approx dp_h - \sum_r \frac{1}{2} \left( \frac{\partial p^0}{\partial x_r} \bigg|_{z^0_h} + \frac{\partial p^1}{\partial x_r} \bigg|_{z^1_h} \right) dx_{r,h}.
\]

**Proof:** See the appendix.

LEMMA 2 says that the change in profits is just the change in price, net of an adjustment accounting for changes in costs due to endogenous changes in \( \mathbf{x} \), which can be approximated from marginal prices.

Let the change in aggregate welfare \( W \) be given by aggregating over the changes in consumer surplus and profits:

\[
(8) \quad dW \equiv \int_J dw_i di + \int_{J_h} d\pi_h dh.
\]

By lemmas 1 and 2, we can integrate over Expressions (5) and (7) and substitute them into the respective terms in Equation (8). Additionally, we can combine these into one integral, but doing so requires some additional notation because Expression (5) is evaluated at the choices made by a single household \( i \) (regardless of location) whereas Expression (7) is evaluated at a particular house \( h \) (regardless of who lives there). Let \( \mathbf{z}^t_{i(r)} \) represent the characteristics, in time
\( t \), of a house actually occupied by household \( i \) in time \( \tau \). Similarly, let \( \frac{\partial p^e}{\partial z_j} |_{z_{i(\tau)}^e} \) be the partial derivative of the period \( t \) hedonic price function with respect to attribute \( j \), evaluated at the period \( t \) attributes of the house actually occupied by household \( i \) in time \( \tau \). By definition, \( z_{i(\tau)}^e = z_i^e \).

However, this more general notation allows us to keep track of a household's former house or future house, even when it is not currently living there.

Equation (8) together with this notation lead to the following proposition.

**Proposition 1.** Given Assumptions A1-A3, a second-order approximation to the change in aggregate welfare from an exogenous change in the distribution of \( g \), when prices, households, and landlords adjust to the change endogenously, is

\[
dW \approx \int_I \frac{1}{2} \left( \left( \frac{\partial p^0}{\partial g} |_{z_{i(0)}^0} + \frac{\partial p^1}{\partial g} |_{z_{i(1)}^1} \right) dg_i 
+ \sum_r \left( \frac{\partial p^0}{\partial x_r} |_{x_{i(0)}^0} + \frac{\partial p^1}{\partial x_r} |_{x_{i(1)}^1} \right) (x_{r,i(1)}^1 - x_{r,i(0)}^1) 
+ \sum_r \left( \frac{\partial p^1}{\partial x_r} |_{x_{i(1)}^1} - \frac{\partial p^1}{\partial x_r} |_{x_{i(0)}^0} \right) (x_{r,i(0)}^1 - x_{r,i(0)}^0) \right) di.
\]

**Proof:** The proposition follows from Lemmas 1 and 2, simply integrating (5) over \( I \) and (7) over \( H \), adding the two together, fixing our indices so \( h = i(0) \) (which we can do given the bijective mapping between them), and re-arranging terms.6

Despite the intimidating subscript notation, Expression (9) is actually quite simple. It consists of three terms. The first is the change in \( g \) experienced by a household (across houses if it moves) evaluated at the partial derivative of the hedonic price function, averaged over those two points. The second term is the change in \( x \) experienced by the household, accounting for only the fact that it may move (i.e. netting out any supply adjustments in \( x \) at a fixed location), but with these changes again evaluated by the respective average partial derivatives experienced by the household (across houses). The third and final term is the supply response at the house-

\[ \text{The first pair of terms in parentheses in Expression (9) comes from Expression (5). Adding the other terms of Expressions (5) and (7), we have} \]

\[ \frac{1}{2} \sum_r \left( \frac{\partial p^0}{\partial x_r} |_{z_{i(0)}^0} + \frac{\partial p^1}{\partial x_r} |_{z_{i(1)}^1} \right) dx_{r,i} = \frac{1}{2} \sum_r \left( \frac{\partial p^0}{\partial x_r} |_{z_{i(0)}^0} + \frac{\partial p^1}{\partial x_r} |_{z_{i(0)}^0} \right) dx_{r,i(0)}. \]

\[ \text{We can write} \]

\[ dx_{r,i} = (x_{r,i(1)}^1 - x_{r,i(0)}^1) + (x_{r,i(0)}^0 - x_{r,i(0)}^0) = (x_{r,i(1)}^1 - x_{r,i(0)}^1) + dx_{r,i(0)}. \]

\[ \text{The rest follows by regrouping terms.} \]
hold's initial house, multiplied by the difference in the ex post partial derivatives between the final location and the initial location. The changes in prices $dp$ which appear in Expressions (5) and (7) cancel out as transfers between households and landlords.

Proposition 1 is the key result of this section. It says that, under our assumption that we observe a panel of housing attributes and our initial assumption that we can track households across locations, a second-order approximation to welfare can be estimated with only the derivatives of the first-stage hedonic price function at the two points in time. The approach essentially relies on the fact that the policy shock induces a new equilibrium, which allows demands to be approximated by tracing out the optimized points.

Expression (9) can be simplified if we are willing to assume that the supply side of the market is highly competitive and that profits (and rents to fixed factors) from adjustments to $x$ are approximately zero. Denote this as Assumption A4:

**ASSUMPTION A4 (Zero profits).** The change in profits due to adjustments in $x$ are approximately zero:

$$\int_h \left[ p^1(g^1_h, x^1_h) - p^1(g^1_h, x^0_h) \right] dh \approx \int_h \left[ c(g^1_h, x^1_h) - c(g^1_h, x^0_h) \right] dh.$$

Thus, the change in profits consists only of the change in rents to land and to the sunk capital $x^0$:

$$\int_h d\pi dh \approx \int_h \left( p^1(g^1, x^0) - p^0 \right) dh = \int_h \left[ dp - \left( p^1 - p^1(g^1, x^0) \right) \right] dh.$$

In this case, we would have the following variant of Proposition 1:

**PROPOSITION 2.** Given Assumptions A1-A4, a second-order approximation to the change in aggregate welfare from an exogenous change in the distribution of $g$, when prices, households, and landlords adjust to the change endogenously, is:

$$dW \approx \int_J \sum_i \frac{1}{2} \left( \frac{\partial p^0}{\partial z_j} \bigg|_{z^0_i} + \frac{\partial p^1}{\partial z_j} \bigg|_{z^1_i} \right) dz_{j, i} di \int_h \left( p^1_h - p^1_h(x^0, g^1_h) \right) dh.$$

Proposition 2 states that we can measure benefits by tracking the average change in amenities experienced by each household, weighted by the average of the derivatives of the two hedonic price functions evaluated at the households' choice, netting out any aggregate price changes that reflect real costs of adjustments in $x$. For simplicity, I will use Expression (10) as the measure of benefits for the remainder of this paper. However, this choice plays no substantive part of the results later derived, and parallels could be derived based on Expression (9) as well.
Note finally that in the simple case where there are no supply adjustments, then \( x_{i(0),r}^1 = x_{i(0),r}^0 \) and the entire expression collapses to

\[
(11) \quad dW \approx \int \sum_j \left( \frac{\partial p_0^j}{\partial z_j} |_{z_i^0} + \frac{\partial p_1^j}{\partial z_j} |_{z_i^1} \right) dz_{j,i} \, di.
\]

Expression (11) is just the individual household measure from Expression (5) summed over households, but with the \( dp \) terms cancelling as transfers between residents and landlords.

Figure 1 illustrates measure (11), or equivalently the first terms in Expressions (9) and (10). The figure shows the derivatives with respect to \( g \) of both the before-policy and after-policy hedonic functions; these derivatives are positive and continuous but unrestricted as to slope or curvature. The line \( \text{bid}(g) \) represents the linearized approximation to the Marshallian bid function. The points \( g^0 \) and \( g^1 \) represent the levels of the public good selected by the consumer in each scenario. Although it is a Marshallian measure, the area under the linearized bid function represents a second-order approximation to the welfare change associated with this change in \( g \). The figure illustrates this measure only in the dimension of \( g \), but note Expressions (10) and (11) require summing over all attributes \( j \). Even if there are no adjustments to \( x \) in the housing stock as a result of the policy, the welfare measure for this change in \( g \) still requires taking these terms into account, weighted by the average marginal WTP of the household, as shown in Expressions (10) and (11). The measures are no different if people are owner-occupiers. In that case, the wealth effects still cancel and the \( dz \) terms incorporate the wealth effects on demand for attributes, as would be appropriate. Regardless, a valid welfare measure can be obtained simply by adding up experiences changes in characteristics, weighted by the average marginal values.

4. When Only a Panel of Houses is Available

The previous subsection involves Assumption A1, that panel data on households' choices are available. As noted above, such data are becoming available with new big data sets. Nevertheless, they have not been available traditionally and will not always be. This subsection considers what we can learn if we relax this assumption.

First, to see the role of Assumption A1, note that Propositions 1 and 2 introduce expres-
sions involving $\sum_j \frac{1}{2} \left( \frac{\partial p^0_j}{\partial z_{j}} |_{z^0_i} + \frac{\partial p^1_j}{\partial z_{j}} |_{z^1_i} \right) dZ_{j,i}$ for each household $i$. This expression involves observing (i) in which house each household lives in both time periods; (ii) a panel of house attributes $(g, x)$ or, if we are willing to assume $dx=0$, a panel of $g$ together with data on $x$ at one time period; and (iii) either a panel of house prices (as in a repeat sales model) or repeated cross sections of housing prices sufficient for predicting $p_h$ at each location. Of these, the first is the most difficult to observe: in many cases analysts may observe housing characteristics and a sample of prices, but have no information whatsoever on who is occupying those houses. Unfortunately, in general such data are not sufficient to use the strategy outlined in the previous sub-section. Expression (10) requires information on the household's change in amenities, $d_z$. In general, information on the change in amenities at a fixed house $h$, $d_z h$, are not sufficient. The problem is depicted in Figure 2. The figure shows two points respectively chosen by two households in the first period, two in the second, and two possible pairs of demands for the two households. However, which household sorts into which house is unobserved. Consequently, the two households' demand curves may be the two solid lines or alternatively the two dashed lines. But the sums of the areas under those respective pairs of demand curves are not the same.

However, information on only the changes in $g$ and other characteristics of the houses, which is available in a great many hedonic applications (perhaps the majority), is sufficient under at least two special cases. The first is trivial: if households do not relocate in equilibrium then one can substitute the change in amenities at a fixed house $h$, $d_z h$, for the change in amenities consumed by a household, $d_z i$, in Expression (10). In this case, aggregating over houses amounts to the same thing as aggregating over households. However, for the large changes envisioned here, this patterns is unlikely to hold.

Nevertheless, data on $d_z h$ can still be used under a second, more interesting, case: a "single-crossing" restriction on any two households' indifference curves. This amounts to a restriction on preferences such that households' Marshallian bid functions for some observed amenity do not cross. That is, households can be ordered by their marginal willingness to pay for the amenity, and the ordering will be the same evaluated at any level of the amenity and under any equilibrium price function. Because households always sort in the same order, if we have single crossing on $g$ then in Figure 2 we can infer that the sorting is that of the dashed lines and rule out the solid lines. Essentially, the logic of single crossing provides a way to impute.
households' pattern of sorting, even when their actual locations are not observed. This property is formalized in Assumption A5:

**Assumption A5 (Single Crossing).** Let $V_{ji}(y_i, p(), z_j)$ be the indirect function conditional only on attribute $z_j$ given the price function $p()$, with the other attributes optimally chosen subject to $z_i, y,$ and $p()$ to determine the utility level. Let $J$ be a simply ordered set and let the distribution of demands be such that, for some amenity $z_j$, $v_i^j / \partial z_j \neq 0 \forall i$ and

$$\frac{\partial v_{ji}(y_i, p(), z_j)}{\partial z_j} / \frac{\partial v_{ji}(y_i, p(), z_j)}{\partial y}$$

is everywhere non-decreasing in $i$.

Assumption A5 requires single crossing in only one dimension of the characteristic space. The public good of interest $g$ may be a natural choice for that attribute, but that choice is not necessary. $z_j$ could be any characteristic or any scaler-valued index of characteristics. Even though we are modeling multidimensional characteristics, induced assortative matching along any one dimension is enough to impute households' choices. The single-crossing assumption, sometimes called the Spence-Mirrlees condition when written this way, guarantees that the level of $z_j$ chosen by households is always increasing in $i$. See, e.g., Milgrom and Shannon (1994) and Athey, Milgrom, and Roberts (1998 Ch. 3) for proofs. To see this intuitively, note that we could write A5 alternatively as follows. For any two households $i, i'$ if

$$\frac{\partial v_{ji}(y_i, p(), z_j)}{\partial z_j} / \frac{\partial v_{ji}(y_i, p(), z_j)}{\partial y} \geq \frac{\partial v_{ji'}(y_{i'}, p(), z_j)}{\partial z_j} / \frac{\partial v_{ji'}(y_{i'}, p(), z_j)}{\partial y}$$

for some $\hat{z}_j, \hat{p}()$, then

$$\frac{\partial v_{ji}(y_i, p(), z_j)}{\partial z_j} / \frac{\partial v_{ji}(y_i, p(), z_j)}{\partial y} \geq \frac{\partial v_{ji'}(y_{i'}, p(), z_j)}{\partial z_j} / \frac{\partial v_{ji'}(y_{i'}, p(), z_j)}{\partial y}$$

for all $z_j, p()$.

---

7 As Chiappori, McCann, and Nesheim (2010) discuss, extensions of the single-crossing property to the multi-attribute case lose the interpretation of inducing assortative matching. Although we are working with a multi-attribute model, Assumption 5 involves single-crossing of the WTP functions in the dimension of only one attribute.

8 As discussed by Athey, Milgrom, and Roberts (1998), Edlin and Shannon (1998), and Milgrom and Shannon (1994), a more general version of single crossing using monotone comparative statistics is sufficient to guarantee sorting by $i$. However, in the hedonic context we are already assuming the regularity conditions, associated with the hedonic tangency conditions, which as they show imply the Spence-Mirrlees condition specified here. (In particular, the assumptions that $v()$ is continuously differentiable, $\partial v / \partial y > 0$, and $\partial v / \partial z_j \neq 0$.) If we were to relax the tangency conditions and focus on welfare bounds associated with inequality conditions instead of equalities, the required restriction would be their single crossing condition (see e.g. Edlin and Shannon 1998, condition 1).

9 See, e.g., Theorem 3.2 in Athey, Milgrom, and Roberts (1998). Their proof applies directly if we assume the utility-maximizing choice for each household is a singleton. Moreover, it is easy to see that, with single crossing, a household can only be indifferent between two points on an equilibrium hedonic price function if at least one other household is too (otherwise, another household would bid up the price of the house it prefers and the first household would no longer be indifferent). But such ties are precisely the case where mistaking the sorting is immaterial.
Thus, if a household selects more $z_j$ than another household in the baseline scenario, it will do so in the ex post scenario as well. Intuitively, given that the Marshallian demand functions do not cross, this obviously must be so if the implicit price of $z_j$ is increasing in $z_j$ (i.e. if the hedonic price function is convex in $z_j$). However, even if the implicit price is decreasing in $z_j$ over portions of the range, the second-order condition for utility maximization requires that it cut the demand curves from below. In other words, the economics of the model require that the slope of the price function be greater than the slope of the demand curve in the neighborhood of the optimal choice. Thus, households will always "sort" across $z_j$ in the same order, even if they are changing consumption of other attributes or the numeraire.

All this suggests a simple approach for identifying Expression (10) with panel data on houses. Let $\mathcal{H}$ and $\mathcal{J}$ now be finite countable sets indexed by $h=\{1,\ldots,H\}$ and $i=\{1,\ldots,I\}$ with $H=I$. These can be viewed as finite samples of data drawn from the underlying distribution. Let $F_t^z(\cdot)$ be the distribution function of some continuously distributed amenity $z_j$ in period $t$. Given that $z_j$ is continuously distributed, for each observed percentile $\theta \in \{1/H, 2/H, \ldots, 1\}$ of the distribution of $z_j$, there will be a unique vector $z(\theta)$ in period $t$. Let $z_k^t(\theta)$ be the value of the $k^{th}$ attribute of this vector. Note for $k=j$, $z_j^t(\theta)=(F_t^z)^{-1}(\theta)$. Then we can now state the following proposition.

**Proposition 3.** Under Assumptions A2-A5, aggregate welfare for a change in the distribution of $g$, when prices, households, and landlords adjust to the change endogenously, can be computed from observed prices and estimated derivatives as follows:

$$
\begin{align*}
\text{Proposition 3 states that one can rank the houses by } z_j \text{ in both scenarios, find the change}
\end{align*}
$$
in each attribute at a constant percentile of the \( z_j \) distribution, and evaluate the derivatives of the hedonic price function at the \( z \) falling at the same percentile of the \( z_j \) distribution. The final term is the same price adjustments as in Expression (10). In practice, note that as long as one knows \( g \) and the other attributes at all locations and time periods, this estimate can be implemented with only a repeated cross section of housing prices (i.e. without a full panel): all that is required is that \( p^1 \) and \( \frac{\partial p^1}{\partial z_j} \) can be predicted for each house from the hedonic pricing model.

This single-crossing condition is routinely imposed in the literature on non-linear pricing (e.g. Athey, Milgrom, and Roberts 1998, Wilson 1993), including models of monopoly screening as well as locational sorting. Although imposing this property is undoubtedly a restriction relative to the more general treatment of heterogeneity in Section 3, it is actually less restrictive in this respect than many structural models, which employ the same single crossing property plus additional functional form restrictions or parametric restrictions on the distribution of unobservable demand shifters. For example, consider the common class of models which allow households \( i \) to differ in two dimensions, income \( y \) and a parameter \( \alpha \) reflecting tastes for \( g \). Many hedonic and sorting applications, including Bajari and Benkard (2005), Bishop and Timmins (2017), and Heckman, Matzkin, and Nesheim (2010b) impose the additional restriction that willingness to pay is strictly increasing in \( \alpha \) and that preferences are quasi-linear. These models implicitly impose Assumption A5: households are totally ordered by \( \alpha \), with increasing \( \alpha \) implying increasing \( g \). The same is true, after taking expectations over the additive errors, of many logit models such as Bayer, Ferreira, and McMillan (2007).

This is not to say that all papers impose these conditions. Other models in this class impose only a partial ordering on \( i \) rather than a total ordering. For example, Epple, Peress, and Sieg (2010), Kuminoff (2012), and Sieg et al. (2004) order households by \( \alpha \) conditional on \( y \), and vice versa. However, pairs of households differing in \( \alpha \) and \( y \) need not be ordered: one household may choose higher \( g \) than the other household in one equilibrium but not necessarily in another equilibrium. In this sense, Assumption A5 is stronger than the related single crossing assumptions imposed in those papers. However, in other respects the approach of this section still imposes weaker assumptions about heterogeneity. Epple, Peress, and Sieg (2010), Kuminoff (2012), and Sieg et al. (2004) essentially compensate for their weaker single crossing assumptions by imposing additional functional form restrictions on \( \nu() \) and (in the case of Sieg et al.)
parametric assumptions on the joint distribution of \((\alpha, y)\).

In practice, Expression (12) may be a reasonable approximation to Expression (10) even when households are not totally ordered, in violation of Assumption A5. In the simulations reported in Section 6, the matching-by-percentiles approach of Expression (12) gives results quite close to the matching-by-households approach of Expression (10), even when households are not totally ordered. The reason appears to be that even if households do not sort on \(g\) in exactly the same order in different scenarios, their rank orderings are still highly correlated. Consequently, violations of Assumption A5 are only local, over a range in which the hedonic price function is approximately linear (so marginal prices are approximately constant) and the households are similar enough that imputing one's marginal value to another creates only small errors. In this sense, estimates based on Expression (12) can be thought of as \textit{approximations} to Expression (10), even when Assumption A5 does not strictly hold.

Moreover, there may be compromises between Propositions 2 and 3. The model of Section 3 and Proposition 2 considered the case where individual households were observed in multiple time periods. So far, this Section and Proposition 3 have considered the polar opposite case where no information was available on the identity of which households sort into which houses. In between these two cases are a variety of intermediate ones where partial information is available on the households locating at a house. Not surprisingly, such partial information would allow us to partially relax Assumption A5.

For example, suppose we can track where \textit{types} of households live but not the individual household. Perhaps we can observe the race of a household occupying a given house, or its income, or some other characteristic or combination of characteristics. Then we would only require that the single crossing property hold within observable type.

In this context, the required single-crossing condition may be modified as follows.

**Assumption A5' (Single crossing within type).** Let \(\mathcal{T}\) be a set of observable types which partitions the set of households \(\mathcal{K}\). Let \(\tau \subset \mathcal{T}\) be the set of households of a specific type with measure \(\mu_\tau, \sum_\tau \mu_\tau = 1\), and let each \(\tau\) be a simply ordered subset of the partially ordered set \(\mathcal{I}\). For each \(\tau\), let the distribution of demands be such that, for some amenity \(z_j\),

\[
\frac{\partial v_{i,j}(y_i,p(z_j))}{\partial z_j} = \frac{\partial v_{i,j}(y_i,p(z_j))}{\partial y}
\]

is everywhere nondecreasing in \(i \in \tau\).

This property guarantees that, after conditioning on the observed type, households sort
along \( z_j \) in the same order. Thus, if a household selects more \( g \) than another household of the same type in the baseline scenario, it will do so in the ex post scenario as well. This approach allows households to be partially ordered overall but totally ordered within type.

In this case, Expression (12) can be modified as follows. Let \( I_\tau \) be the number of observed households of type \( \tau \) and let \( \theta_\tau = \{1/I_\tau, 2/I_\tau, \ldots, 1\} \) be the observed percentiles of the distribution of \( z_j \) among type \( \tau \). Then by the same argument as given in Proposition 3, aggregate welfare is

\[
dW \approx \sum_{\tau \in \mathcal{I}} \sum_{\theta_\tau} \frac{1}{I_\tau} \sum_k \frac{1}{2} \left( \frac{\partial p^0}{\partial z_k} |z^0(\theta_\tau) + \frac{\partial p^1}{\partial z_k} |z^1(\theta_\tau) \right) \left[ z^1_k(\theta_\tau) - z^0_k(\theta_\tau) \right]
\]

(13)

That is, for each type, one can take the set of houses occupied by that type and rank them by \( z_j \) in both scenarios, find the change in \( z_k \) at a constant percentile of the distribution, and evaluate the derivatives of the hedonic price function at the same percentile of the distribution. Then, one can take the weighted sum over types.

5. When Household Demands Shift

The models of Sections 3 and 4 both have relied on Assumption A2, that households' preferences and incomes are constant over the time period considered, so that the optimal vector \( z \) is an unchanging function of the hedonic price function. As depicted in Figures 1 and 2, this assumption allows us to identify two points on a single demand vector for \((g, x)\) as a function of the hedonic prices. Assumption A2 may be most fitting for hedonic applications to consumer goods such as computers (Bajari and Benkard 2005) or groceries (Griffith and Nesheim 2013). It may be less fitting for applications to housing if income or taste changes are important in that context or—to put it differently—when there are important changes in parameters affecting current-period utility, such as family status. In such cases, even when we observe points for two equilibria for a single household (Section 3), these may be two points on two different demand curves. Likewise, even if household preferences exhibit the single crossing property (Section 4), if incomes or preferences are shifting across equilibria then there is no reason to believe households are still
In this subsection, I consider relaxing Assumption A2 as well as Assumption A1. Changes in the willingness to pay for $g$ of course will have implications for equilibrium hedonic prices. But more importantly, they raise fundamental questions about what perspective to take when evaluating changes in $g$. I take the approach of Fisher and Shell (1972) and Pollak (1989), who suggest that welfare comparisons can be made in such situations from the perspective of one or the other preference relationships. This basic approach also is implicit in simulation exercises such as those in Bayer et al. (2011), Sieg et al. (2004), or elsewhere, where general equilibrium welfare effects are estimated for policy shocks under an assumption of constant preferences, even though preferences may well change over the time periods envisioned. Essentially, these simulations evaluate the policy shocks from the perspective of ex ante preferences.

From this perspective, we can again use the single crossing property along with a weaker version of Assumption A2:

ASSUMPTION A2'. The distribution of pairs $(v^i(), y^i)$ is constant over time. Consequently, the distribution of demand functionals $z(p^i( ))$ is constant.

Assumption A2' is implied by A2 but the reverse is not true. It says that although individual households' demand functions may change over time, the distribution of demand types remains constant.

With this weaker assumption we can now state Proposition 4.

**PROPOSITION 4.** Under Assumptions A2' and A3-A5, aggregate welfare for a change in the distribution of $g$, evaluated from the perspective of constant preferences for each household, can be computed from observed prices and estimated derivatives as given by Equation (12). Moreover, the aggregate evaluation is the same whether using ex ante or ex post preferences (though the distributional effects may be different).

**Proof:** See the appendix.

Proposition 4 says that we can still use the same measure for aggregate welfare as given in Proposition 3 under a weaker version of Assumption A2, in which individual household demands may shift due to changes in preferences or incomes, as long as the distribution of de-

---

10 See Stapleford (2011) for discussion in historical context.
mands does not change. However, the welfare estimates are now from a single period perspective, or alternatively are counterfactual estimates "as if" preferences had remained constant. If Assumption A2' is still considered too strong an assumption, alternative approaches may be possible introducing semiparametric controls for observed factors that change over time, such as family status (marital status, presence of children, etc.) or income. Extending the model in this direction would bring it close to an alternative model recently proposed by Bishop and Timmins (2017), which implicitly imposes single crossing plus constant distributions of demands, but imposes additional structure to condition on income and an annual additive shocks to WTP. In general, the strategy can be thought of as an across-time variant of the argument for using multiple markets under assumptions of constant distributions of tastes (Epple 1987, Heckman, Matzkin, and Nesheim 2010a).

6. Illustrative Simulations

In this section I illustrate the results from the previous two sections by simulating hedonic housing equilibria and shocking the equilibria with changes to $g$. In brief, 100 economies were simulated, each with 1000 households and 1000 houses. In the base model, households have Cobb-Douglas preferences over $g$ and a numeraire, with unobserved taste parameters on $g$ distributed triangular with nodes (0.1, 0.2, 0.6), and with unobserved income distributed log-normal with mean 11.1 and standard deviation 0.4 (and truncated at $30,000 and $180,000). Single crossing holds in this model across income conditional on tastes and across tastes conditional on income, but it does not hold between all pairs of individuals. Consequently, there is no \textit{a priori} reason that the approximation given by Equation (12) will be identical to that given by Equation (10). Below, I also consider an alternative model where tastes vary only by an observable, discrete type. Note $x$ is omitted from the simulations, which simply saves the need to condition on it in the analysis of the simulated data.

The public good $g$ is uniformly distributed on (1, 3) in the baseline scenario. In the ex post scenario, 50% of observations are "treated" by a policy. The probability of being treated is linearly decreasing over the support of $g^l$ from 0.75 at $g^l=1$ to 0.25 at $g^l=3$. If a house is treated, its level of $g$ improves such that $g^l = g^0 + (3 - g^0)/3 + 1$. Figure 3 illustrates the levels of $g$ in one representative simulation. The top panel shows the level of $g$ in the ex post scenario as a func-
tion of its level in the ex ante scenario. The bottom line, along a 45-degree ray, represents untreated houses, whereas the top line represents treated houses. The bottom panel of the figure shows the cumulative distribution functions (CDF) of \( g \) in each scenario. It shows the ex ante scenario is uniformly distributed, while the ex post scenario, which 1st-order stochastically dominates it, is not. Equilibrium rents range from 16% to 42% of income, with a mean of 27%, which approximates US expenditure shares for housing.

In the base model, equilibrium prices in each scenario were then perturbed by an error term, normally and independently distributed and calibrated such that the standard deviation of the error was equal to either 1% or 5% of the mean price. This error term can be interpreted as either measurement error in price (the dependent variable) or alternatively as an unobserved characteristic of the home that enters preferences as a perfect substitute for the numeraire good.

Hedonic price functions were fitted non-parametrically to this noisy data in each scenario with a local quadratic function, with bandwidths separately tuned in each simulation using leave-one-out cross validation. Local derivatives were taken analytically from the estimated quadratic function, and so are locally linear.\(^{11}\) Figure 4 illustrates the estimated price functions for one representative simulation in which the standard deviation of the error was set to 5% of the mean price. The upper left panel shows the price function fit to the data in the ex ante scenario; the upper right shows the respective relationship in the ex post scenario. The lower left shows the two price functions overlayed. Finally, the lower right shows the derivatives of the two respective price functions with respect to \( g \). While the first three panels suggest the relationship is fairly smooth and convex, the final panel does show that the second derivatives are not constant.

Figure 5 displays the relationship between these estimated derivatives and the households' marginal WTP for the public good. Recall that the first-order conditions for the household imply that these should be equal. The left panels in the figure represent the ex ante scenario and the right panels represent the ex post scenario. The top panels plot the slopes of the hedonic price function alongside the marginal WTP to pay of households occupying those houses. The

\(^{11}\) The optimal bandwidth for fitting prices was adjusted for fitting the first derivative. I also considered modeling the derivative directly from differenced data and tuning the bandwidth using leave-L-out cross validation as suggested by De Brabanter et al. (2013). However, that approach performed quietly poorly, especially near the endpoints.
two tend to run together except at very high levels of \( g \), where the estimated derivatives have difficulty keeping up with the rapid escalation in marginal WTP. The bottom panels plot the estimated slope against true marginal WTP. As suggested by the top panels, the estimated slopes fit marginal WTP well except for those with the highest WTP.

Finally, after computing these price functions and their derivatives, the welfare measures defined by expressions (10) and (12) were then computed. Additionally, the exact CV and EV for each household were calculated. Table 1 reports the results. The first two columns report the results from the base model described above, one column for each standard deviation of the error term. The first two rows show the "true" welfare measures of average CV and EV (averaging over households within a simulation), showing the median and the 5th and 95th percentiles of these averages across the 100 simulations. The median average CV is $6452 in the first model and the median average EV is $7249.

The third row in Table 1 shows the value of the Harberger approach when the full information on household sorting is available, as computed by Expression (10) (and expressed as a household average). The median estimate of the average value using this approach is $6926 in the first model. As expected, this value lies between the CV and EV measures. The following row places this estimate on to the unit interval between CV and EV. The median placement is 60.4% of the way from CV to EV, just above the midpoint. Similarly, 90 percent of the observations range between 52.1% and 68.1%. All the estimates lie within the interval bracketed by CV and EV. This illustrates that the "sufficient statistic" approach truly is sufficient in this case. Using only information from the first-stage hedonic price function, we can compute a welfare estimate that approximates non-marginal Hicksian welfare measures.

The fifth row shows the value of the Harberger approach when information on household sorting is not available, but such sorting is "imputed" by assuming households sort in the same order of \( g \) across the ex ante and ex post scenarios, as described by Expression (12). This imputation is guaranteed to be correct when households are simply ordered and heterogeneity exhibits the single crossing property. However, as noted above, the single crossing property does not strictly hold in this simulation. Nevertheless, the estimates using Expression (12) are virtually identical to those using Expression (10).

To explore the reasons for this result, Figure 6 illustrates the sorting patterns observed in
one illustrative simulation. The first panel shows the marginal WTP for $g$ of each household evaluated at the "average house," against the level of $g$ they actually choose in the ex ante scenario. That is, it shows an index of the household's demand-type for $g$ against its optimal $g$. If Marshallian bid functions never crossed, this figure would show an increasing function. While it is not strictly increasing, it is nearly so, indicating the any crossings in the bid functions are only very local. The second panel in the figure shows the households level of $g$ in the ex post scenario versus the ex ante scenario. It illustrates that households sort approximately in the same order, though not exactly. Finally, the third panel shows the level of $g$ predicted for a household in the ex ante scenario if it sorted in the same order as in the ex post scenario, against its actual level of $g$ in the ex ante scenario. It shows that the errors from imposing the single-crossing assumption in this scenario are small and local. Consequently, the hedonic price function is approximate constant over the range of these errors, and the over-all bias is minimal. Thus, even when household sorting is not observed, and even when the single crossing property does not hold locally, approximations based on this property potentially can be quite accurate.

In the second column, I consider the sensitivity of these results to changes in the variance of the additive errors, increasing them by a factor of 5. Though the estimates change somewhat, the pattern of results hold, with the Harberger approach still falling between the true CV and EV values.

In the next column, I consider an additional model where households differ only by five discrete types, as well as by unobserved continuously distributed income. Whereas in the base model households' tastes for $g$ were distributed triangular with nodes (0.1, 0.2, 0.6), in this model those values are rounded to the nearest value in the set {0.15, 0.25, 0.35, 0.45, 0.55}. The household's type is assumed to be observed, so when sorting is not observed it can now be imputed using Equation (13)—i.e. assuming single crossing within type—rather than Equation (12), which assumed single crossing globally. The Harberger approach performs just as well using single-crossing by type as when household locations are observed in the panel, although this is not surprising given the success of Equation (12) in the first model.

7. Conclusions

For decades, economists have used the hedonic model to estimate demands for the implicit characteristics of differentiated commodities, including the demands for otherwise unpriced local
public goods and amenities. The traditional cross-sectional approach to hedonic estimation has recovered marginal willingness to pay for amenities when unobservables are conditionally independent of the amenities, but has faltered over a difficult endogeneity problem when attempting non-marginal welfare measures. In this paper, I show that when marginal prices can be reliably estimated, and when panel data on household sorting is available, one can construct an approximation—using only the first-stage marginal prices—which is a "sufficient statistic" for non-marginal welfare measures. With this approximation, Rosen's second-stage estimation can be replaced with a simple average of first-stage parameters. Moreover, even when panel data on household sorting are unavailable, and only repeated cross sections of housing prices (together with a panel on house characteristics) are available, the sufficient statistic approach remains valid under a single crossing restriction. In practice, this approximation appears to perform well in these simulations even when this restriction does not strictly hold.
References


Kuminoff, N.V. 2012. "Recovering Preferences from a Dual-Market Locational Equilibrium,
mimeo.


Figure 1. Willingness to pay for non-marginal changes in public goods
Figure 2. Household sorting is unobserved
Figure 3. Depiction of Distributions of Public Good for One Representative Policy

3A. Ex post public good vs. ex ante public good (each circle is one house).

3B. CDF of ex post public good (solid) and ex ante public good (dashed)
Figure 4. Depiction of Results for One Representative Simulation: Price Fitting

Predicted and actual ex ante prices against g

Predicted and actual ex post prices against g

Predicted ex ante (solid) and ex post (dashed) prices against g

Predicted derivatives in ex ante (solid) and ex post (dashed) scenarios

- 1st scenario fitted prices
- 2nd scenario fitted prices

- Ex ante derivative
- Ex post derivative
Figure 5. Depiction of Results for One Representative Simulation: Estimated Hedonic Derivative and Marginal WTP
Figure 6. Sorting Behavior
<table>
<thead>
<tr>
<th>Table 1. Results of Simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistic</strong></td>
</tr>
<tr>
<td>(Averaged over households)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1. Average CV</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2. Average EV</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3. Value of Eq. (10), as avg.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4. Eq.(10) as pct. distance from CV to EV</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>5. Value of Eq. (12) [Eq. (13) when type observed], as avg.</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
APPENDIX. PROOFS OF LEMMAS AND PROPOSITIONS.

Proof of Lemma 1.
Since the marginal utilities are functions of $p$ and $z$, we can write Equation (4) in terms of changes in marginal utilities, $d \left( \frac{\partial v_i}{\partial p} \right)$ and $d \left( \frac{\partial v_i}{\partial z} \right)$. With this notation, Equation (4) simplifies to:

\[
d v_i \approx \frac{\partial v_i}{\partial p} d p_i + \sum_j \frac{\partial v_i}{\partial z_j} d z_{j,i} + \frac{1}{2} \left( d \left( \frac{\partial v_i}{\partial p} \right) d p_i + \sum_j d \left( \frac{\partial v_i}{\partial z_j} \right) d z_{j,i} \right).
\]

Taking the total derivative of the household's first-order condition given by (1), we have

\[
d \left( \frac{\partial v_i}{\partial z_j} \right) = d \lambda_i \frac{\partial p^0}{\partial z_j} \left| z_i^0 \right. + \lambda_i^0 d \left( \frac{\partial p}{\partial z_j} \right) + d \lambda_i d \left( \frac{\partial p}{\partial z_j} \right),
\]

where $d \left( \frac{\partial p}{\partial z_j} \right) = \frac{\partial p^1}{\partial z_j} \left| z_i^1 \right. - \frac{\partial p^0}{\partial z_j} \left| z_i^0 \right.$; that is, it is the change in the derivative of the price function from both the change in the price function itself and from the change in the point where it is evaluated. Inserting this expression along with $\frac{\partial v}{\partial p} = -\lambda$ into (14), we have:

\[
d v_i \approx -d p_i \left( \lambda_i^0 + \frac{1}{2} d \lambda_i \right) + \sum_j \left( \frac{\partial p^0}{\partial z_j} \right. \left| z_i^0 \right. d z_{j,i} \left( \lambda_i^0 + \frac{1}{2} d \lambda_i \right) + \frac{1}{2} \sum_j d \left( \frac{\partial p}{\partial z_j} \right) d z_{j,i} d \lambda_i.
\]

Adding and subtracting $\frac{1}{4} \sum_j d \left( \frac{\partial p}{\partial z_j} \right) d z_{j,i} d \lambda_i$, ignoring the remaining third-order terms (as we are taking a second-order approximation), and re-arranging, we have the desired expression. Note to compute this expression, we need Assumption A3, that the price derivatives and $dz$ terms can be computed for each household across time.

Proof of Lemma 2.
The proof follows a similar outline as that for Lemma 1. Since the marginal costs are functions
of \( x \), we can write Equation (6) in terms of changes in marginal costs \( d \left( \frac{\partial c_h}{\partial x_r} \right) \). With this notation, Equation (6) can be re-written as:

\[
d c_h \approx \sum_r \left( \frac{\partial c_h}{\partial x_r} + \frac{1}{2} \left( \frac{\partial c_h}{\partial x_r} \right) \right) dx_{r,h}.
\]

(17)

Substituting the first-order condition (2), we now have:

\[
d c_h \approx \sum_r \left( \frac{\partial p^0}{\partial x_r} \bigg|_{x_h^0} + \frac{1}{2} \left( \frac{\partial p}{\partial x_r} \bigg|_{x_h^1} \right) \right) dx_{r,h}
\]

(18)

= \sum_r \frac{1}{2} \left( \frac{\partial p^0}{\partial x_r} \bigg|_{x_h^0} + \frac{\partial p^1}{\partial x_r} \bigg|_{x_h^1} \right) dx_{r,h}

Substituting this expression into the equation \( d\pi = dp - dc \) completes the proof.

**Proof of Proposition 4.**

Consider the initial equilibrium described by the hedonic price function \( p^0(g, x; F^0) \). Now consider a change in the distribution of \( g \) and allow preferences for individual households to change, but such that they are still distributed the same (Assumption A2'). Denote the new equilibrium price function by \( p^1(g, x; F^1) \). Now consider a counter-factual scenario where the distribution of \( g \) is still given by \( F^1 \) but all households have their original demand functions. By Assumption A2', the equilibrium price function would still be \( p^1(g, x; F^1) \) since the distribution of demands would be no different in this counterfactual \( g \) and \( x \) are unchanged (though the assignment of households to houses may change). But by Assumption A5 and the argument in Proposition 3, any household choosing \( z_j \) in the initial equilibrium such that \( F^0(j) = \theta \) will choose \( z_j \) in the counterfactual equilibrium such that \( F^1(j) = \theta \). Moreover, these choices are consistent with the price equilibrium. The rest of the argument follows from Proposition 3.

Finally, note that the same argument could be made, *mutatis mutandis*, starting with \( p^1(\ ) \) and going back to \( p^0(\ ) \) under ex post preferences. Thus, the aggregate welfare evaluation is invariant to the perspective taken.