Toward More General Hedonic Estimation: Clarifying the Roles of Alternative Experimental Designs with an Application to a Housing Attribute

Michael Eriksen
Texas Tech University

Thomas J. Kniesner
Claremont Graduate University, Syracuse University, and IZA

Chris Rohlfs
Morgan Stanley

Ryan Sullivan
Naval Postgraduate School

May 22, 2015

Abstract

Our research develops a more general hedonic model in which an exogenous shock to a single product attribute can affect other attributes, the markets for the product’s complements and substitutes, and aggregate quantity produced. The factors are shown to be empirically relevant and to cause bias in traditional approaches. Experimental estimators of attribute demand are introduced that address biases, are transparent, and are straightforward to implement. We apply one of the estimators developed to measure the marginal value placed by householders on subsidized carbon monoxide detectors.

Corresponding Author: Michael Eriksen, Rawls College of Business, Texas Tech University, Lubbock, TX 79409. mike.eriksen@ttu.edu.

Special thanks to Gus Bartuska and Qu Feng for expert research assistance and thanks also to Dan Black, Gregorio Caetano, Bill Horrace, Boyan Jovanovic, Jeff Kubik, Derek Laing, Derek Neal, Jan Ondrich, Stuart Rosenthal, Kevin Tsui, Andy Vogel, Pete Wilcoxen, Paul Wilson, and seminar participants at Clemson, Syracuse University Camp Econometrics, NYU, and Rochester and Syracuse Universities for many helpful comments. The views expressed here are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. government. Distribution for this article is unlimited.

Keywords: hedonic, identification, field experiment, marginal willingness to pay, heterogeneous goods, endogenous attributes

JEL Classifications: D12, C35, C31, D61, C9
Hedonic estimation and the measurement of marginal willingness to pay (MWTP) for product attributes are vital tools for quantifying the benefits of public policies that improve safety, environmental, school, or health care quality (Black 1999; Chay and Greenstone 2005; Cutler, Rosen, and Vijan 2006; Viscusi 1993, 1996). Hedonic methods are used to understand the demand for heterogeneous goods such as automobiles, computers, food, housing, and jobs (Bajari and Benkard 2005; Hamermesh 1999; Kiesel and Villas-Boas 2007; Raff and Trajtenberg 1995; Sheppard 1999). They are also used to calculate the Consumer Price Index and one fifth of expenditures in the Gross Domestic Product (Landefield and Grimm 2000; Moulton 2001). For the purposes of measurement and policy evaluation it is desirable to have robust hedonic estimators whose empirical results are correct generally. Our research demonstrates the identifiability of MWTP without the strong econometric restrictions often applied in earlier applications and presents straightforward estimators of MWTP and related measures for use in experimental empirical settings.

A cursory reading of the hedonics literature might yield the impression that MWTP cannot be identified without imposing highly restrictive assumptions about utility and markets, even when a natural experiment is available. Models adopted often assume that unobserved product attributes either are uncorrelated with observed ones or do not exist (Berry, Levinsohn, and Pakes (BLP) 1995; Epple 1987; Rosen 1974). In addition, the adopted models generally assume that the product of interest has no complements or substitutes, so that a location-specific attribute like weather cannot affect the labor market and housing market simultaneously. Finally, adopted models also typically specify aggregate quantity consumed as exogenous and unresponsive to price changes.¹

There is a widespread belief in the literature that the above restrictions are appropriate and necessary to estimate MWTP. Earlier applied studies of heterogeneous goods generally employ slight modifications of the hedonic frameworks, or measure reduced-form price effects without estimating MWTP directly. More recent empirical work in hedonic estimation focuses on quasi-experiments, and some innovative studies have incorporated quasi-experimental variation into existing hedonic models (Bayer, Ferreira, and McMillan 2007; Berry and Haile 2010; Boes and Nüesch 2011; Chay and Greenstone 2005; Klaiber and Smith 2009; Kuminoff and Pope 2010, 2012; Lewbel 2000; Parmeter and Pope 2013; Pope 2008a, 2008b).² To our knowledge, no previous hedonic frameworks simultaneously

¹ Rosen (1974) and Eppe (1987) additionally require that markets are sufficiently thick so that every conceivable product is available and that supply is competitive. BLP (1995) additionally requires specific functional forms for utility and firm costs, plus a specific distribution for heterogeneity in preferences.

² Some recent theoretical studies relax the functional form assumptions from earlier models but leave the frameworks largely intact elsewhere (Athey and Imbens 2007; Ekeland, Heckman, and Nesheim 2004; Heckman, Matzkin, and Nesheim 2010).
allow for unobserved product attributes that are affected by exogenous shocks, complementarity with the
good of interest, and aggregate quantities that vary.³

In what follows we first provide an intuitive discussion of the types of biases that endogenous
omitted attributes, complement and substitute goods, and aggregate quantity effects generate in traditional
hedonic approaches, especially with regards to housing. Next, we present experimental estimators to
address the biases. Of the estimators presented, we start with estimators with the least restrictive modeling
assumptions, but have the most demanding data requirements. The modeling assumptions become more
restrictive and the data requirements less demanding with successive estimators. We then focus on
developing nonparametric experimental estimators that identify the entire distribution across consumers
of the demand for a given product attribute. In particular, we present experimental estimators of the
aggregate demand for a product attribute among a population of consumers. The experimental estimators
we develop avoid the effects of endogenous omitted attributes and complement and substitute goods by
offering products and subsidies to consumers.

It is important to emphasize that the estimators we develop here rely upon straightforward,
transparent identification conditions that are relatively easy to implement in future research. Variations on
the estimators have been previously applied in recent studies to estimate the value of freedom from jail,
the demand for avoiding the Vietnam draft, the value of a statistical life, and the demand for class size
reductions in elementary school (Abrams and Rohlf 2011; Rohlf 2012; Rohlf, Sullivan, and Kniesner
2015; Rohlf and Zilora 2013). The new class of experimental hedonic estimators, however, has not been
applied within a housing context where researchers often adopt hedonic estimators with the strongest
econometric restrictions. As a final exercise, we illustrate how one of the proposed estimators could be
used to value a housing attribute by conducting a field experiment that randomly subsidized the price of
carbon monoxide detectors.

II. DISCUSSION OF POSSIBLE BIAS IN HEDONIC MODELS

To illustrate the sources of biases that our research addresses, let $Price_{ht}$ be the average price of
house $h$ in year $t$. Let $z_{1ht}$ be an observable attribute about house $h$, such as local school quality. Next let
the value of $z_{1ht}$ be determined by a quasi-experiment so that it varies exogenously across locations and
over time. Let $z_{2ht}$ be an attribute about house $h$ that is difficult to measure, such as the pleasantness of
neighbors in the area. Finally, let $Price_{ht}$ be a linear function of the two attributes and an error term
denoting unobserved attributes:

³ Roback (1982) allows for one type of complementarity (housing and jobs), and Sieg, Smith, Banzhof, and Walsh
(2002) include area-specific dummy variables to proxy for the areas’ job quality and public goods. BLP (1995)
allow for market shares (but not aggregate quantity produced) to vary. No single framework has addressed more
than one of the biases simultaneously.
\( (1) \) \( Price_{ht} = \beta_0 + \beta_1 z_{1ht} + \beta_2 z_{2ht} + \epsilon_{ht} \)

The aim of a hedonic price regression in this case is to identify \( \beta_1 \), the effect of attribute \( z_{1ht} \) on housing prices, holding all other attributes constant. Once identified, the hedonic price effect is used in a second-stage procedure to estimate MWTP for \( z_{1ht} \) (Epple, 1987; Rosen, 1974).\(^4\) For the second stage to produce accurate estimates, the estimates from the first-stage hedonic regression (1) must be consistent.

In the situation we are discussing suppose the assignment of \( z_{1ht} \) is random and consequently uncorrelated with any predetermined factors that influence \( Price_{ht} \). An improvement in school quality, however, may cause affluent and educated people to move into the area. If being affluent and educated is correlated with being a pleasant neighbor, then the school quality improvement will directly affect the pleasantness of neighbors. If \( z_{2ht} \) is not included as a control in the regression, OLS estimates of (1) would only measure the reduced-form effect of the shock to school quality on \( Price_{ht} \) and not the structural parameter \( \beta_1 \). The reduced-form effect includes the direct effect of \( z_{1ht} \) and the indirect effects of \( z_{1ht} \) through the mechanism of \( z_{2ht} \). Hence, OLS produces a biased estimate of \( \beta_1 \), and the magnitude of the bias is \( \beta_2 \ast cov(z_{1ht}, z_{2ht}) \). Even if data were available on \( z_{2ht} \), the values of the other \((z_{1ht})\) attribute were not assigned experimentally and are likely to be correlated with \( \epsilon_{ht} \). Consequently, without an instrument for \( z_{2ht} \), an OLS regression of \( Price_{ht} \) on \( z_{1ht} \) and \( z_{2ht} \) will produce biased estimates of \( \beta_2 \), and by failing to adequately control for \( z_{2ht} \), the regression will also produce biased estimates of the effect of \( z_{1ht} \).

However, the pleasantness of neighbors is not the only characteristic of additional concern. For example, many location-specific attributes are also influenced by the composition of local residents and businesses. Such neighborhood attributes will vary in response to any exogenous shock that causes consumers or firms to move. Similar biases arise in the markets for labor and schooling, where some workers' and students' behaviors affect the quality of the environment experienced by other workers and

---

\(^4\) The procedure proposed by BLP (1995) is different from that described here, but BLP require consistent estimation of the effect of the attribute on the decision to purchase the product.

\(^5\) If \( z_{2ht} \) is observable and data are available for some period preceding the shock to school quality, then one might consider instrumenting for \( z_{2ht} \) with the pre-treatment levels of the attribute. In general, however, the geographic variation in \( z_{2ht} \) will still be correlated with unobservable location-specific determinants of home value. Suppose, for example, that an additional omitted attribute \( z_{3ht} \) is the natural beauty of the area, which is time-invariant, positively correlated with \( z_{2ht} \), and positively valued by consumers. Instrumenting for \( z_{2ht} \) with the pre-treatment value will produce an upward-biased estimate of \( \beta_2 \) that captures the effects of both pleasantness of neighbors (which changes in response to a shock to school quality) and natural beauty (which is time-invariant, does not respond to the shock, and should not appear in the set of controls). Supposing that \( cov(z_{1ht}, z_{2ht}), \beta_1 \), and \( \beta_2 \) are all positive, then the upward bias in the estimation of \( \beta_2 \) will generate a downward bias in the estimation of \( \beta_1 \) so that the researcher attributes too much of \( z_{1ht} \)'s effects on housing prices to the increase in the pleasantness of neighbors.
students. In general, \( z_{2ht} \) should be treated as an endogenous variable that may change in response to a shift in \( z_{1ht} \).

A similar form of bias arises within a hedonic setting if goods that are complements or substitutes with housing are ignored. One key interaction is between the markets for housing and labor. Consumers often decide where to live based upon job availability, and the types of jobs that are available in an area affect housing prices and the types of people who live there. Additionally, consumers who are considering buying a home in an area may also consider the local price level and the quality and variety of local goods. Hence, local job characteristics and local prices are location-specific attributes that should be included in (1). In addition to being difficult to measure in an exhaustive way, locational factors are probably correlated with \( \varepsilon_{ht} \), and adequately controlling for them would require finding yet another credible instrument for each one.

Figure 1: Illustration of Bias Due to Aggregate Quantity Changes

---

Another problem particular to labor market studies is that individual wages are determined by workplace amenities and worker productivity, both of which are endogenous and difficult to measure. Note that even with an unconditionally random treatment, including endogenous outcomes as controls can produce conditional endogeneity bias of \( \beta_i \) and in turn the MWTP.
Variation in aggregate quantity is a third potential source of bias. As illustrated in Figure 1, the lower demand curve in the figure plots consumer willingness to pay for living in that area before a school quality improvement. The higher demand curve shows willingness to pay for living in that area after the improvement. The benefit of the intervention is illustrated by the vertical difference in the two demand curves. The intervention’s effect on prices, which is shown on the vertical axis, will in general be smaller than the benefit of the intervention. The price difference will only equal the benefit in the special case that supply is perfectly inelastic. The price difference may equal MWTP in the short run, when quantity is fixed, but in the long run quantity will increase, and the price difference in Figure 1 will shrink.

In the models of Rosen (1974) and Epple (1987), all of the benefits of a product improvement are capitalized into the price as supply is assumed perfectly inelastic. In a more general framework in which supply is not perfectly inelastic, some of the benefits of the improvement will be capitalized into the price, and the remainder will be experienced as consumer surplus. If the affected product has complements or substitutes, the intervention could affect prices or production quantities for other goods. Suppose, for example, that the supply of computers is perfectly competitive but the supply of operating systems is controlled by a monopoly with inelastic supply. A technological innovation that increased computing speed could have no effect on the price of computers, but the producer of operating systems could raise its prices and keep output constant.

BLP (1995) develop a discrete choice model in which the dependent variable is the decision to purchase a specific good, and the price of the good is an endogenous regressor. The quantity effects described above are partially considered in their model, as an exogenous shock to product attributes can affect prices and market shares – for example, the fraction of sales accounted for by a single product – but not aggregate quantity produced. One limitation of BLP’s approach is that, in addition to needing an instrument for the product attribute being studied, it is necessary to have an instrument that shifts prices but is unrelated to unobserved factors influencing demand. The types of quasi-experiments that generate valid instruments are rare, and it would be unusual to have a single dataset in which separate plausible instruments existed for both an important product attribute and the price of the good. Thus, researchers using the BLP approach may have to use instruments for the price that arguably do not satisfy the necessary exogeneity conditions.

We propose to address the problem of endogenous prices experimentally. Specifically, the researcher offers randomized products and subsidies to individual consumers to measure their response. Offers made to such atomless consumers do not generate any general equilibrium responses in prices. Table 1 below summarizes the experimental estimation strategies that we develop here. The estimators illustrate what specific aspects of the demand for product attributes can be identified in different situations resulting from the increase in the attribute.
Table 1: Description of MWTP Estimators Developed

<table>
<thead>
<tr>
<th>(1) Estimator</th>
<th>(2) Research design</th>
<th>(3) Key Assumptions</th>
<th>(4) Identifies . . .</th>
<th>(5) Applications described in text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Idealized Experiment</td>
<td>Offer consumers the option to “treat” all of their goods in the market of interest with an additional unit of the attribute $z_k$. Randomize the price for the treated option across consumers.</td>
<td><em>Local non-satiation</em>, price-taking consumers.</td>
<td>Distribution across consumers of MWTP for an attribute $z_k$.</td>
<td>Home improvements, product upgrades, and attributes that are artificially tied to houses or jobs, such as school district access or health care coverage.</td>
</tr>
<tr>
<td>2. Restricted Offer Experiment</td>
<td>Pay consumers to restrict consumption in the market of interest to “treated” or “untreated” versions of the same good. Randomize price for treated version across consumers.</td>
<td>Same as idealized experiment.</td>
<td>Distribution across consumers of “offer-restricted MWTP” for an attribute $z_k$.</td>
<td>The value of doctor visits and medical treatment, the value of internet bandwidth, the discount rate.</td>
</tr>
<tr>
<td>3. Randomized Offer Experiment</td>
<td>Offer some consumers “untreated” goods other consumers “treated” goods, both at a subsidized rate, where the subsidy is randomly assigned across consumers.</td>
<td>Same as idealized experiment.</td>
<td>“Marginal Surplus” (the vertical difference between the treated and untreated demand curves) at all points along the demand curve.</td>
<td>The value of different characteristics of small business loans or solicitations for charitable donations.</td>
</tr>
</tbody>
</table>
III. IDENTIFICATION OF MWTP DENSITIES

Previous hedonic and discrete choice models by Rosen (1974); Epple (1987); and BLP (1995) are special cases of the framework described here. The models consider for simplicity a homogeneous consumption good. In the housing market, a minor extension to the conventional models would allow consumption to consist of multiple homogeneous goods such as the comparison of those purchased in Atlanta and as compared to in Boston. A home buyer would certainly take the price level into account when choosing where to live, and a change in weather or school quality in one of the cities could affect local prices for the homogeneous good as well as local housing prices. The full market response to the weather or school quality shock includes both the effect on housing prices and the effect on local prices for everything else. Local price effects, which are assumed not to exist in previous hedonic and discrete choice models, are easy to accommodate in the current setup we propose here.

Complementarity also occurs between jobs and housing in the same location. A firm hiring in Syracuse, New York (which receives 115 inches of snow per year) must pay a higher wage to obtain the same level of talent than does a similar firm hiring in San Francisco (which has year-round pleasant weather). In the current model, one of the goods could be hours of work in a specific job in Syracuse, and another good could be hours of work in a specific job in San Francisco. Each consumer would be endowed with quantities of time that could be sold to the employer or consumed as leisure. The shape of the utility function would be such that selling hours to the firm in San Francisco (and failing to consume them as leisure) would greatly increase the utility benefit of housing in San Francisco. The two situations would also be strong substitutes in the utility function, so that selling hours of work in San Francisco would sharply increase the utility cost of selling hours of work in Syracuse, and working jobs in both cities would be rare.

In addition to ruling out substitutes and complements, previous hedonic models require that each consumer purchases exactly one unit of the good whose characteristics are under study. In the market for housing, this restriction rules out any effects of prices or location-specific attributes on the decision to buy a home or the total number of home buyers. In addition to allowing for endogenous homeownership, relaxing the restriction helps to describe markets in which consumers often buy more than one good whose characteristics are important, such as automobiles and computers, or markets in which quantity consumed various continuously, as in foods, music, and vacations.

In Rosen’s (1974) original study including the MWTP each consumer selects an interior solution for each attribute, and the MWTP for more of an attribute $z_k$ is exactly equal to the marginal cost to producers of adding an additional unit of $z_k$ to a good. In that framework, all owners of Ford Mustangs place the same value on additional units of safety, horsepower, and fuel efficiency. In the current setting, prices are not necessarily continuous functions of product attributes, markets are thin, and two buyers of
Ford Mustangs may have very different utility functions. Each might select a Mustang with factory specifications because it is one of the few products available at a preferred price.

Because we use a flexible functional form for utility and because markets may be thin, consumption bundles will not necessarily represent interior solutions. Consequently, it is more natural to consider discrete changes in attributes. Such discrete changes are also better than marginal changes are at reflecting the type of variation that is generated through experiments and quasi-experiments. Moreover, a primary goal of our study is to establish conditions for identifying the MWTP density. In many cases, ethical or cost considerations will prevent researchers from conducting experiments to accomplish this goal. Thus, we consider experiments that estimate alternative measures of attribute demand that are similar to MWTP.

We begin our discussion of identification with an idealized experiment where the researcher applies a so-called treatment technology and charges randomly assigned prices for that technology to different consumers. Specifically, a treatment technology, $T_{Z_\tau k}$, converts any bundle $q_z$ of goods in $Z$ (the set of all conceivable goods in the market of interest) into a new bundle $T_{Z_\tau k}(q_z)$ of goods in $Z$ where every unit of every good in the treated segment $Z_T$ is replaced with the equivalent good plus an additional unit of $z_k$. The treatment technology effectively increases each $z_k$ by one unit for every $z$ in $Z_T$ that $i$ chooses to consume.

Knowing that the treatment technology will be applied to a bundle of goods alters a consumer’s optimal choices. Let $(q_{xi}^*, q_{zi}^*)$ denote $i$’s optimal consumption bundle at equilibrium prices in the absence of any intervention, where $x$ denotes goods outside the market of interest. Let $(q_{ki}^*, q_{zi}^*)$ denote the consumption bundle that $i$ would choose to purchase at equilibrium prices, given the knowledge that the bundle of goods $q_{zi}^*$ will be converted into $T_{Z_k}(q_{zi}^*)$, where the treatment is applied to the entire set $Z$. If, for example, $z_k$ is local school quality, then $i$ might select a home in a relatively low quality school district with the understanding that the district will be improved by the treatment technology. The bundle $q_{zi}^*$ includes the home in the low quality district, and $T_{Z_k}(q_{zi}^*)$ represents the same bundle after the school quality improvement.

Given this, we define the benefit to $i$ of a one-unit increase in attribute $z_k$:

MARGINAL WILLINGNESS TO PAY. Consumer $i$’s Marginal Willingness to Pay (MWTP) for $z_k$ is scalar-valued, is denoted $MWTP_{ki}$, and equals $\theta_i(T_{Z_k}(q_{zi}^*), p_x^*, w_i^*, u_i^*) - p_z^* \cdot q_{zi}^*$. $MWTP_{ki}$ is a dollar-denominated measure of the consumer surplus that $i$ experiences due to the treatment technology -- from consuming $T_{Z_k}(q_{zi}^*)$ at the price $p_z^* \cdot q_{zi}^*$. Because consumer surplus is defined relative to the benchmark utility level $u_i^*$, and wealth level $w_i^*$ the formula gives the change in surplus that $i$
would experience from switching from the optimal untreated bundle, which provides zero surplus, to the optimal treated bundle.

For $MWTP_{ki}$ to accurately measure the benefit, it is essential that the final term, $p^*_z \cdot q^k_{zi}$, be subtracted off the reservation price $\theta_\bullet$, so that the expression returns a surplus and not a reservation price. If $i$ lives in a school district with high housing prices and $z_k$ is school district quality, being offered the treatment technology could induce $i$ to move to a more affordable area. After applying the treatment technology, the new area could be as desirable as or even less desirable than the old one. However, the treatment technology had a positive benefit by helping $i$ to save money.

A. Experimental Estimator 1: An Idealized Experiment

To formalize the concept of an idealized experiment we will use the following definition here:

**IDEALIZED EXPERIMENT.** To conduct the idealized experiment the researcher draws a sample of $N$ consumers from the population, where the draws are independent. Each consumer has the option to have the treatment technology for attribute $z_k$ applied to every good consumed. To receive this treatment, the consumer must pay a treatment price $\theta$, where $\theta$ is randomly assigned across consumers.

The idealized experiment can be applied to measure the MWTP for home improvements, product upgrades, and attributes that are artificially tied to specific houses or jobs, as with school district access or health care coverage.

In the absence of the treatment, $i$ selects the bundle $q^*_z$ of all goods in the market of interest $Z$ and obtains zero surplus. If the treatment is provided at a price $\tau_i = MWTP_{ki}$, then $i$ is able to purchase the bundle $T_{z_k}(q^k_{zi})$ at a cost of $\theta_i(T_{z_k}(q^k_{zi}), p^*_z, w^*_i, u^*_i)$. At the treatment price $\tau_i = MWTP_{ki}$, consumer $i$ is indifferent between selecting and not selecting the treatment. At any treatment price greater than $MWTP_{ki}$, purchasing the treatment would give $i$ negative surplus, and at any treatment price less than $MWTP_{ki}$, purchasing the treatment would give $i$ positive surplus.

Identification of $MWTP_{ki}$ in the idealized experiment is a straightforward application of a nonparametric discrete choice estimator (Pagan and Ullah, 1999, pp. 272-299). Consumer $i$ selects the treatment option if and only if $\tau_i \leq MWTP_{ki}$. At a given treatment price $\tau$, the fraction of consumers who select the treatment option is $Pr_i(MWTP_{ki} \geq \tau)$, where the probability is taken over all consumers $i$. This probability can be rewritten in terms of the cumulative density function ($F$) as $1 - F_k^{MWTP}(\tau)$. Let $Treat_i$
be a binary indicator of whether \(i\) selects the treatment. A consistent kernel estimator for \(F_k^{MWTP}(\tau)\) can be constructed following Li and Racine (2007, pp. 182-183, 209-210; 2008):\(^7\)

**BANDWIDTH AND WEIGHTING KERNEL:** A bandwidth \(h\) is defined as a decreasing function of the sample size \(N\). For simplicity, let the value \(h(N)\) be denoted \(h\). This function satisfies the conditions that \(\lim_{N \to \infty} h = 0\) and \(\lim_{N \to \infty} N \cdot h = \infty\). A weighting kernel \(\omega\) is a symmetric, bounded pdf that integrates to one.

**ESTIMATOR IN IDEALIZED EXPERIMENT:** Given a sample size \(N\), bandwidth \(h\), and weighting kernel \(\omega\), for any MWTP value \(\tau\), our estimator \(\hat{F}_k^{MWTP}(\tau)\) equals

\[
\frac{\sum_{i=1}^{N} \omega\left(\frac{\tau - \tau_i}{h}\right)(1 - \text{treat}_i)}{\sum_{i=1}^{N} \omega\left(\frac{\tau - \tau_i}{h}\right)}.
\]

**B. Experimental Estimator 2: Offer-Restricted Environments**

A related strategy to the existing tradeoffs approach is to induce consumers to participate in an experiment that restricts their choices:

**OFFER-RESTRICTED EXPERIMENT:** A sample of \(N\) consumers is drawn from the population, and each one is offered a large dollar payment \(\delta\) to participate in the experiment. Each participant must select bundles in \(Z\) that either include only goods in \(Z_0\) or only goods in \(Z_k\). Hence, for each \(i\), participation in the experiment requires that \(q_{zi}(z) = 0\) for every \(z \in Z_0\) or \(q_{zi}(z) = 0\) for every \(z \not\in Z_k\). The payment \(\delta\) is sufficiently large that every consumer opts to participate. Each participant may choose between consuming goods in \(Z_0\) or goods in \(Z_k\). The researcher randomly assigns a treatment price \(\tau_i\) across participants in the experiment. Let the price schedule in the offer-restricted experiment be denoted \((p_{x}^{i}, p_{z}^{T})\), where \(p_{z}^{Z_0}(z) = p_{z}^{i}(z)\) for all \(z \in Z_0\) and \(p_{z}^{Z_k}(z) = p_{z}^{i}(z)\) for all \(z \not\in Z_k\).

**OFFER-RESTRICTED CONSUMPTION BUNDLE:** For a set of goods \(Z_T\) and a dollar payment \(\delta\), consumer \(i\)'s offer-restricted consumption bundle for a set \(Z_T\) and payment \(\delta\) is \(i\)'s optimal consumption bundle \((q_{x}^{Z_T}, q_{zi}^{Z_T})\) given the restriction that \(q_{zi}^{Z_T}(z) = 0\) for all \(z \not\in Z_T\) and supposing that \(i\) is given a dollar payment \(\delta\). Analytically, the consumption bundle is the solution to the following price and wealth constrained utility optimization problem:

\[
\max_{q_{x}^{i}, q_{zi}} u_i(q_{x}^{i}, q_{zi}) \text{ subject to } p_{x}^{i} \cdot q_{x}^{i} + p_{z}^{T} \cdot q_{zi} \leq w_i + \delta \text{ and } q_{zi}(z) = 0 \text{ for all } z \not\in Z_T.
\]

**OFFER-RESTRICTED MWTP:** Let \(Z_0, Z_k \subseteq Z\) be two sets of goods such that \(Z_k\) contains the treated variant of every good in \(Z_0\). Let \(\delta\) be a dollar payment offered to \(i\). Consumer \(i\)'s offer-restricted MWTP for \(z_k\) from goods in \(Z_0\) with payment \(\delta\) is scalar-valued and equals

\[
\theta_i\left(q_{zi}^{Z_k} p_{x}^{i}, w_i + \delta, u_i(q_{zi}^{Z_0}, Z_0) - p_{z}^{Z_k} \cdot q_{zi}^{Z_k}\right) - p_{z}^{Z_k} \cdot q_{zi}^{Z_k}.
\]

The pdf \(F_k^{Z_0, \delta}\) denotes the density of the offer-restricted MWTP, and the corresponding CDF is written as \(F_k^{Z_0, \delta}\).

---

\(^7\) The estimator is similar to that used in a variety of discrete choice settings including contingent valuation experiments in which survey participants report how they might act if presented with certain hypothetical tradeoffs (Creel and Loomis 1997, Crooker and Herriges 2004, Kristrom 1990).
In the offer-restricted experiment, consumers are invited to join the study and to restrict their choices in \( Z \) to come either entirely from \( Z_0 \) or entirely from \( Z_k \). In exchange for accepting the restriction, participants are given a payment \( \delta \). As with the idealized and existing tradeoff experiments, there is a treatment price \( \tau_i \) that is randomly assigned across consumers, and consumers must pay this price in order to consume goods from \( Z_k \). Because consumers are paid to participate in the experiment, the offer-restricted MWTP is defined using utility \( u_i(q_{x_i}^Z, q_{z_i}^Z) \) rather than \( u_i(q_{x_i}^Z, q_{z_i}^Z) \) as the benchmark utility level.

The estimation strategy in the offer-restricted experiment is the same as in the idealized and existing tradeoff experiments. A nonparametric kernel regression of \( 1 - \text{Treat}_i \) on \( \tau_i \) identifies the CDF of the offer-restricted MWTP. Often the researcher will not be able to offer a sufficiently high payment \( \delta \) to induce every consumer to participate in the study. In such cases, the study will produce internally valid estimates of \( F_{Z_k}^{Z_0} \) for a selected sample of consumers who are particularly receptive to cash incentives or the chance to receive the treatment.

One important example of an offer-restricted experiment is the RAND Health Insurance Experiment (Manning, et al., 1987). The researchers randomly assigned health insurance plans across participants, so that some consumers faced high prices for doctor and hospital visits and others faced low prices. The authors use the random variation in prices to estimate the willingness to pay for doctor visits and other types of medical care. Another example of an offer-restricted experiment is the Internet Demand Experiment (Edell, and Variaya, 1999; Varian, 2001). Consumers participating in that study agreed to have their internet service provided by the researchers. Every time consumers went online, they faced a menu of different amounts of bandwidth, each sold at a different randomly assigned price. A third example of offer-restricted experiments involves laboratory or field experiments to measure the discount rate (Harrison, Lau, and Williams, 2002; McClure, et al., 2004). In such studies, consumers are offered a cash amount to be paid now or a slightly larger amount to be distributed later (such as $100 now or $100 plus some additional amount in seven months). The additional amount that is offered later is randomly assigned across consumers.8

C. Experimental Estimator 3: Randomized Product Offers

In many cases, no specific consumer faces a choice between treated and untreated sets of goods, but researchers can learn about the demand for treatment by measuring the extent to which it affects total sales of the product of interest. Such estimators can be used to identify the marginal surplus, MS.

---

8 A special case of the restricted offer experiment is an existing tradeoff experiment, in which the researcher selects a random sample of consumers already making a purchasing decision and randomly assigns taxes or subsidies for selecting an option. The existing tradeoff experiment approach is developed in Appendix A.
One approach for estimating marginal surplus is to generate randomized product offers whose characteristics vary continuously across the consumers being studied:

**OFFER DENSITIES.** Let the offer density functions $g_0$ and $g_k$ both be pdfs that assign density levels to each good in $Z$. These density functions are constructed to satisfy $g_0(z) = g_k(z_1, ..., z_k + 1, ..., z_n)$ for all $z$, so that goods drawn from $g_k$ have on average one more unit of $z_k$ than do those drawn from $g_0$.

**RANDOMIZED OFFER EXPERIMENT.** A sample of $N$ consumers is drawn from the population, where $N$ is even. For the first $N/2$ consumers, a good $z_i$ is randomly selected for each consumer from the distribution $g_0$; for the remaining $N/2$, $z_i$ is selected according to $g_k$. Each consumer is offered a subsidy of $\delta$ per unit consumed of the offered good, where $\delta$ is constant across consumers. Additionally, the researcher randomly assigns a per unit tax $\tau_i$ across participants in the experiment, where $\tau_i < \delta$. Each good $z_i$ is offered at a subsidized price of $p^*_z(z_i) = \delta + \tau_i$.

In the randomized offer experiment, each consumer in the sample is offered a different good. In addition to randomly selecting the product offers, subsidies for the offered goods are randomly assigned across consumers.

Let $h$ be a bandwidth and $\omega$ be a symmetric weighting kernel. Our parameter of interest and our estimator are defined as follows:

**MS FOR THE AVERAGE OFFERED GOOD.** The marginal surplus (MS) for the average offered good at quantity level $Q$ is denoted $E_Z[MS^k_{2i}(Q, s^*)]$ and equals $\int Z MS_{2i}^k(Q, s^*) g_0(z) dz$.

**RANDOMIZED OFFER ESTIMATOR.** The estimator $\hat{E}_Z[MS^k_{2i}(Q, s^*)]$ of MS for the average offered good at quantity level $Q$ equals $\arg\min_{Q_0} Q - \frac{\sum_{i=N/2+1}^N \omega_i (\frac{i-1}{h}) Q_{zi}(\tau_i - \delta, s^*)}{\sum_{i=1}^{N/2} \omega_i (\frac{i-1}{h})} - \frac{\sum_{i=N/2+1}^N p_{zi}(z_i) - \sum_{i=1}^{N/2} p_{zi}(z_i)}{N/2}$.

The first argmin term in our random offer estimator estimates the $\tau$ value at which consumption of the offered good would equal $Q$ for the average good selected according to the $g_k$ pdf. The argmin measures the *inverse aggregate treated demand* at quantity $Q$. The second argmin term estimates the $\tau$ value at which consumption of the offered good would equal $Q$ for the average good selected according to the $g_0$ pdf. The second argmin measures the *inverse aggregate (untreated) demand* at quantity $Q$. To estimate the two argmins, the researcher first estimates two separate nonparametric kernel regressions of $Q_{zi}(\tau_i - \delta, s^*)$ on $\tau_i$, one for each of the two halves of the sample. Next, the researcher inverts the functions by conducting a grid search of $\tau$ values or using Newton’s method. The third term measures the average price difference between the offered goods in the two samples. In the definition of MS, the treatment technology is applied holding prices constant. In the experiment just described the prices are different for
the average “treated” good drawn from the $g_k$ pdf and the average “untreated” good drawn from the $g_0$ pdf. The third term in the formula for the estimator corrects for the price difference.

In one recent application of a randomized offer experiment, Bertrand, et al. (2010) offered subsidized loans to small business owners in South Africa. The researchers randomized multiple features of the loan, including response deadlines, advertising content, and interest rates. Other researchers have implemented randomized offer experiments to study the supply of charitable contributions (Karlan and List, 2007; Landry, et al., 2006). Specifically, subjects were approached and solicited for donations; the researchers randomized the physical characteristics of the solicitors and the extent to which the charity offered matching contributions or lottery incentives.9

IV. APPLICATION TO A HOUSING ATTRIBUTE

To illustrate the idealized experiment estimation strategy (Estimator 1 in Table 1), we consider the economic value that consumers marginally place on a specific housing attribute, carbon monoxide detectors. A typical hedonic study of the MWTP for carbon monoxide detectors would use a linear regression model as shown in equation (1) where carbon monoxide detectors are denoted as the $z_{1ht}$ in the equation, and $\beta_1$ is the marginal effect of that attribute on housing prices, holding all other attributes constant. As discussed previously, such an identification strategy often ignores supply side issues or has data limitations that lead to omitted variable or endogeneity bias in the regressions. For instance, a hedonic estimation strategy might exclude some of the other safety features of the house in the regressions such as smoke detectors or home security systems. If some or all of other relevant attributes are not included as controls and correlated with the presence of carbon monoxide detectors, then the omitted variables might lead to biased estimates of $\beta_1$.

A. Overview of Field Experiment

To get around such identification issues, we conducted a small-scale field experiment between August and November 2014 that illustrates an example of estimator 1. The primary purpose of the study was to obtain unbiased estimates of consumers’ MWTP for carbon monoxide detectors, without the estimation problems just discussed. The US Centers for Disease Control (CDC) reports more than 400 Americans die and 20,000 visit emergency rooms of hospitals from unintentional CO poisoning each year.10 The CDC recommends all households to have a CO detector installed to prevent CO poisoning.

---

9 A special case of the randomized offer experiment is the take-up estimator where the researcher compares total sales for matched pairs of products that are similar in all but one attribute. The take-up estimator is developed in Appendix B.

10 See the CDC website (www.cdc.gov/co) for more information.
especially those with fireplaces, gas furnaces, and gas stoves. In idealized experiment form, we offered carbon monoxide detectors that retail for $20 at Home Depot at randomly offered prices of $5, $10, $15, and $20 to homeowners in Lubbock, TX. Our example study provides a basis for work in the area and how researchers may want to set up larger-scale experiments in the future, particularly in the urban economics field of study.

Participants for the field experiment were recruited through mailing a survey to the first name listed on owner-occupied residential property tax records provided by the Lubbock Central Appraisal District (LCAD). The city of Lubbock is located in western Texas and had a population of 229,573 as of 2010 (United States Census 2010), with an estimated 48,301 owner-occupied properties. There were 41,390 addresses of property owners identified as owner-occupied by their request of a homestead exemption on their property taxes, of which 1,000 (2.4 percent) were randomly selected and mailed a letter inviting them to participate in the research study. Households selected were also sent a brief survey asking about their household composition and the current presence of safety features within the household. Individuals were offered $5 for completing and returning the survey in an included postage paid envelope.

Unsolicited mailed surveys are known to have a low response rate (Shih and Fan, 2009). To avoid potential biases due to expected low response rates correlated with the generosity of the randomly offered subsidy, we adopted a two-step design where only the households returning a survey and indicating they wished to participate in a second component of the study were randomized. More specifically, the subset of households returning a survey were called using a phone number they provided and asked additional questions about their housing attributes and any changes in regards to household safety features since they completed the original survey. All participants were told at the beginning of the phone call that they would receive $20 in total compensation for answering a couple of additional question about their housing attributes, or the original $5 they were entitled for returning at the mailed survey. At the conclusion of the call, individuals were then offered the chance to use the $20 they received from participating in the study to purchase a carbon monoxide detector for the randomly offered price of $5, $10, $15, or $20. For example, individuals randomly selected to be offered a carbon monoxide detector for $5 had the option to be sent either $20 in the mail, or a carbon monoxide detector plus $15 delivered to their door.

Of the 1,000 surveys mailed, we received responses from 18 percent of the individuals. Table 2 shows basic demographics and summary statistics for those who returned the initial survey. Of the participants who responded to the initial survey, 26 percent indicated they did not wish to be contacted, and we were unable to reach 19 percent after at least three attempts using the phone number they
provided. Our final sample was therefore composed of 98 participants who completed all phases of the study. Table 2 below presents summary statistics of our sample.

Table 2: Summary Statistics of Participants offered Carbon Monoxide (CO) Detector

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Randomly Offered Price of CO Detector</td>
<td>$5</td>
<td>$10</td>
<td>$15</td>
</tr>
<tr>
<td>Housing Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>House Value</td>
<td>179,077.80</td>
<td>195,685.20</td>
<td>175,428.60</td>
<td>185,300.00</td>
<td>156,522.70</td>
</tr>
<tr>
<td>Number of Bedrooms</td>
<td>3.22</td>
<td>3.15</td>
<td>3.24</td>
<td>3.25</td>
<td>3.27</td>
</tr>
<tr>
<td>Number of Baths</td>
<td>2.26</td>
<td>2.20</td>
<td>2.29</td>
<td>2.30</td>
<td>2.27</td>
</tr>
<tr>
<td>Square Footage</td>
<td>2,069.32</td>
<td>2,113.11</td>
<td>2,037.33</td>
<td>2,184.90</td>
<td>1,941.05</td>
</tr>
<tr>
<td>Age of Structure</td>
<td>32.83</td>
<td>33.22</td>
<td>34.38</td>
<td>28.95</td>
<td>34.41</td>
</tr>
<tr>
<td>Carbon Monoxide Risks:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fireplace</td>
<td>0.80</td>
<td>0.71</td>
<td>0.83</td>
<td>0.86</td>
<td>0.71</td>
</tr>
<tr>
<td>Gas Furnace</td>
<td>0.60</td>
<td>0.64</td>
<td>0.63</td>
<td>0.68</td>
<td>0.42</td>
</tr>
<tr>
<td>Gas Stove</td>
<td>0.17</td>
<td>0.18</td>
<td>0.21</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td>Presence of CO Detector Before Study</td>
<td>0.50</td>
<td>0.39</td>
<td>0.63</td>
<td>0.46</td>
<td>0.58</td>
</tr>
<tr>
<td>Number of CO Detectors</td>
<td>1.14</td>
<td>1.22</td>
<td>1.33</td>
<td>0.85</td>
<td>1.14</td>
</tr>
<tr>
<td>Purchased CO Detector</td>
<td>0.18</td>
<td>0.29</td>
<td>0.13</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td>Sample Size</td>
<td>98</td>
<td>28</td>
<td>24</td>
<td>22</td>
<td>24</td>
</tr>
</tbody>
</table>

The average participant in our sample reported a house value of $179,077 and lived in a home with 3.22 bedrooms and 2.26 bathrooms. Their home also had on average 2,069 square feet of livable space and was approximately 33 years old. Before the study, 50% of participants reported they already had a carbon monoxide installed, and on average 1.14 were installed. Participants were also asked if they had a fireplace, gas furnace, or gas stove, each of which are listed as potential sources of CO poisoning on the CDC website. 80% of participants reported the presence of a fireplace, 60% a gas furnace, and 16% a gas stove within their home.

As mentioned above, we did not make participants aware of their subsidy condition until the end of the phone survey to avoid selection biases due to differential response rates. To test whether we
implemented randomization correctly we conducted a t-test for each of the attributes listed in Table 2 to determine whether assignment of any of the subsidy conditions was greater than chance alone. Of the attributes, we only found that persons assigned to the no-subsidy condition (those offered a CO detector for $20) were statistically less likely (t-value = -2.01) to have a gas furnace than those assigned to the other three conditions.

B. Experimental Estimates

Of participants offered a carbon monoxide detector in lieu of their full $20 in compensation, 18 per cent accepted the offer. To obtain unbiased elasticity estimates on the attribute in question (CO detectors), we use the following linear probability regression equation:

\[
(2) \quad \text{Purchase CO}_i = \beta_0 + \beta_1 \log(\text{Randomized Price})_i + Z_i'\delta + \epsilon_i
\]

where \(\text{Purchase CO}_i\) is an indicator variable equal to one if individual \(i\) purchased a CO detector (and zero otherwise), \(Z\) is a vector of control variables indicating four carbon monoxide risks within the household: fire place, gas furnace, and gas stove, and \(\epsilon\) is a white noise error term. \(\log(\text{Randomized Price})\) is the variable of interest and signifies the log of the offered CO detector price to consumer \(i\). Thus, \(\beta_1\) reveals the estimated elasticity from the idealized experiment.

Table 3 below presents the results for equation (2). The first two columns are of the full unrestricted sample of 98 participants. Column (1) portrays the regression results without controls and column (2) includes the full set of controls as described in equation (2). For the full sample, the estimates indicate that a 10 percent increase in price leads to about a one percentage point decrease in the probability of purchasing a carbon monoxide detector when controls are not included and a 0.5 percentage point decrease when all of the controls are included. Both of the specifications, however, are imprecisely estimated with neither of the coefficients being significantly different from zero at any conventional level.
Table 3: Effect of Ln(Randomized Price) on CO Detector Purchase

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable is Indicator for Selects CO Detector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Observations</td>
<td>-0.105</td>
<td>-0.056</td>
<td>0.016</td>
<td>0.048</td>
<td>-0.209</td>
<td>-0.195</td>
</tr>
<tr>
<td>Previously Had a CO Detector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Controls for CO Risk?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.021</td>
<td>0.168</td>
<td>0.000</td>
<td>0.225</td>
<td>0.125</td>
<td>0.228</td>
</tr>
<tr>
<td>Mean Dep Var</td>
<td>0.184</td>
<td>0.271</td>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>98</td>
<td>48</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors robust to heteroskedasticity are listed below each estimate in parentheses. The controls for carbon monoxide risk are presence of a fireplace, gas furnace, or gas stove. Asterisks denote statistical significance at the following levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

The next four columns of Table 3 stratify the sample based upon whether or not the participant indicated they already had a carbon monoxide detector installed at the initial survey. Columns (3) and (4) present the estimates for observations that did not previously have a CO detector. Both estimates are positive and relatively small (0.016 and 0.048, respectively). In addition, neither is significantly different from zero at any of the conventional levels. Columns (5) and (6) show the elasticity estimates for individuals who already had a CO detector. The estimates are the most precisely estimated of the group, with both elasticities being significant at the five percent level. The CO adoption likelihood elasticity estimates are -0.209 and -0.195.

What our estimates indicate is that the subsidies acted as an effective price incentive to purchase additional detectors for those who already had them. Although the sample size somewhat limits the preciseness of some of the estimates depending upon the specification type, we can still get a handle on how consumers react to changes in the price of such housing attributes. The research design shown here provides a template for future work in this area and mitigates estimation issues previously discussed.

C. Traditional Hedonic Regression Estimates

To underscore the advantages of utilizing the experimental methods just shown, we now compare those estimates with estimates of a participant’s MWTP for a carbon monoxide detector using a traditional hedonic model. The following model displays the traditional hedonic approach:
where the dependent variable $\text{Home Value}_i$ is the self-reported home value of participant $i$. The variable, $\text{CO Detectors}_i$, denotes the number of carbon monoxide detectors in observation $i$'s home. Thus, $\beta_1$ is the coefficient of interest and can be interpreted as the MWTP for CO detectors. $X$ is a vector of control variables which can change depending on the regression specification. The control variables include the number of bedrooms, number of baths, square footage of livable space, quadratic of square footage, age of structure, quadratic of age of structure, natural log of household income, number of children, and indicator variables for the presence of a fireplace, gas furnace, and a gas stove. The summary statistics for each of these attributes are listed in Table 3. Also included in all of the regressions, but not listed in the table, are indicator variables for each of the 12 postal zip codes in Lubbock, TX. The estimates for equation (3) are presented in Table 4 below.

The first column of estimates omits the presence of carbon monoxide detectors in the specification and is similar in form to what is typically found in the earlier hedonic literature to derive estimates of housing attributes. We find evidence that values are positively correlated with the number of bathrooms, square footage, and age of the structure. We also find holding square footage of livable space constant, the number of bedrooms has a negative, although statistically imprecise, relationship. We include the number of carbon monoxide detectors reported by the participants in their home for estimates reported in the second column of Table 4. Holding the other attributes constant, we estimate the MWTP of a CO detector to be $15,313. Indicative of the demand for multiple CO detectors is positively correlated with the size of the home; we report smaller coefficients on bathroom and square footage for the specification that includes carbon monoxide detectors. In the third column, we include the three carbon monoxide risk factors (presence of a fireplace, gas furnace, and gas stove) thought to be correlated with the presence of a CO detector. The coefficients on these attributes are positive, but statistically indistinguishable from 0. The attributes’ presence in the specification also marginally decreases the MWTP estimate of a CO detector to $14,963. In the last column of Table 4, we report coefficients where the specification includes the natural log of household income and number of children.

11 All home values are reported in year 2014 nominal dollars. In addition to the self-reported property values, we also collected estimated property values from Zillow.com. The resulting (Zillow) hedonic estimates closely resemble the self-reported results and are available upon request.

12 Notably, the number of observations displayed in this specification in Table 4 is only 90 in comparison to the 98 displayed in the experimental results in Table 3. The decrease in number of observations is the result of having missing information on the control variables for eight of the households. The number decreases further when additional controls for household-specific attributes are added in column 4.
Table 4: Hedonic Estimate of MWTP for CO Detector

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CO Detectors</td>
<td>15,313.62**</td>
<td>14,963.14**</td>
<td>13,967.90**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6,844.04)</td>
<td>(6,624.38)</td>
<td>(5,690.41)</td>
<td></td>
</tr>
<tr>
<td>Number of Bedrooms</td>
<td>-27,485.50</td>
<td>-23,467.38</td>
<td>-19,815.19</td>
<td>-27,269.29</td>
</tr>
<tr>
<td></td>
<td>(23114.70)</td>
<td>(16,138.56)</td>
<td>(16,620.52)</td>
<td>(17,328.08)</td>
</tr>
<tr>
<td>Number of Baths</td>
<td>73,735.80***</td>
<td>58,368.99***</td>
<td>55,567.18***</td>
<td>65,952.46***</td>
</tr>
<tr>
<td></td>
<td>(20429.22)</td>
<td>(16,467.41)</td>
<td>(17,592.10)</td>
<td>(19,430.89)</td>
</tr>
<tr>
<td>Square Footage</td>
<td>97.95***</td>
<td>92.62***</td>
<td>89.56***</td>
<td>88.20***</td>
</tr>
<tr>
<td></td>
<td>(21.70)</td>
<td>(12.98)</td>
<td>(12.38)</td>
<td>(12.05)</td>
</tr>
<tr>
<td>(Square Footage)$^2$</td>
<td>-3,884.14***</td>
<td>-3,239.21***</td>
<td>-3,312.95***</td>
<td>-3,187.81***</td>
</tr>
<tr>
<td></td>
<td>(814.34)</td>
<td>(823.86)</td>
<td>(884.38)</td>
<td>(951.24)</td>
</tr>
<tr>
<td>Age of Structure</td>
<td>25.99***</td>
<td>20.92**</td>
<td>20.71**</td>
<td>22.71**</td>
</tr>
<tr>
<td></td>
<td>(9.11)</td>
<td>(8.75)</td>
<td>(9.79)</td>
<td>(10.58)</td>
</tr>
<tr>
<td>(Age of Structure)$^2$</td>
<td>-45,864.99</td>
<td>-24,523.09</td>
<td>-2,831.77</td>
<td>-92,732.04**</td>
</tr>
<tr>
<td></td>
<td>(41701.79)</td>
<td>(45,155.10)</td>
<td>(46,806.83)</td>
<td>(42,796.97)</td>
</tr>
<tr>
<td>Fireplace</td>
<td>.</td>
<td>.</td>
<td>1,934.57</td>
<td>-357.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(13,916.71)</td>
<td>(15,770.84)</td>
</tr>
<tr>
<td>Gas Furnace</td>
<td>.</td>
<td>.</td>
<td>9,265.79</td>
<td>2,743.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10,251.21)</td>
<td>(10,411.89)</td>
</tr>
<tr>
<td>Gas Stove</td>
<td>.</td>
<td>.</td>
<td>17,854.66</td>
<td>9,083.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(17,345.84)</td>
<td>(22,822.36)</td>
</tr>
<tr>
<td>ln(Household Income)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>21,842.26*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(12,255.44)</td>
</tr>
<tr>
<td>Number of Children</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>15,928.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(10,685.58)</td>
</tr>
<tr>
<td>Sample Size (n)</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.83</td>
<td>0.88</td>
<td>0.88</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Note: Dependent variable is reported house value. Each regression also includes a separate intercept (or fixed effect) for each of the 12 zip codes of properties in the sample. Standard errors are reported in parentheses robust to heteroscedasticity. Asterisks denote statistical significance at the following levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 
within the household. The household attributes in the last column of Table 4 do not typically have a role in a traditional hedonic regression, but are suspected to be correlated with the demand for CO detectors. Although the coefficients on the attributes are only marginally significant, the MWTP estimate of CO detectors decreases to $13,967 when they are included in the specification.

The results presented here differ dramatically from estimates from the idealized experiment with weaker imposed assumptions. Depending on the specification, we estimate using traditional hedonic methods, that a household’s MWTP for CO detectors was $13,967 to $15,313. As a point of reference, carbon monoxide detectors generally sell for around $20 at local retail stores. One possible explanation for such a discrepancy of estimates is the existence of further omitted variables which we were unable to observe in our study when utilizing the traditional hedonic regression approach.

The discrepancy between MWTP estimates underscores the importance of seeking and applying more robust MWTP estimators. That stated, experiments do have some limitations. Conducting larger scale experiments can be cost prohibitive and researchers are not always able to randomize goods or prices in a particular market. For example, it would be difficult and costly to randomize the number of bedrooms in houses for consumers. In addition, external validity and representativeness of estimates may have to be at least partially sacrificed for the sake of internal validity. Given their limitations, the experimental estimation strategies discussed in this paper may only apply to a certain subset of goods.

V. CONCLUSION

We have extended the theory of hedonic estimation to incorporate three important aspects of markets for heterogeneous goods. First, many important product attributes are endogenous and change in response to exogenous shocks. Second, many heterogeneous goods have complements and substitutes, and exogenous shocks to the market of interest may affect the markets for those other products. Third, aggregate quantity supplied may change in response to an exogenous shock. For all three reasons, the benefits of an exogenous shock to one product attribute will not entirely be capitalized into the price of that product, and traditional hedonic estimators will produce biased estimates.

We began with an idealized experiment to highlight what assumptions are in principle necessary to identify MWTP. The new modeling concept serves as a theoretical benchmark to clarify the tradeoffs that researchers face between generality of the model and availability of experimental or quasi-experimental data. In the idealized experiment the researcher has the ability to effectively “treat” goods that a consumer purchases with an additional unit of the attribute of interest $z_k$. A small set of consumers is selected from the population, and each one is offered the treatment option at a different, randomly assigned price. Because the intervention only affects the small number of consumers participating in the study, it avoids biases from market-level changes to other attributes, other goods’ prices, or aggregate
quantity. Through such an idealized experiment it is possible to identify the distribution across consumers of the MWTP for the attribute $z_k$. In keeping with the policy focus of hedonic estimation, we paid exclusive attention to identification of the demand for product attributes, and did not impose extra assumptions to identify utility or cost functions.

We then developed alternative practical estimators that identify the distributions across consumers of policy-relevant measures of product attribute demand that differ slightly from MWTP. In the “offer-restricted experiment” the researcher artificially restricts consumers’ options to the untreated and treated sets of goods and provides financial compensation to the participants to induce them to accept the restriction. The researcher then can identify the value of the treatment by randomly assigning subsidies for selecting the treatment option across consumers in the study.

The final estimator presented compares total sales of untreated goods with their treated variants. For the treated-untreated research designs, some consumers face decisions between untreated goods and their best outside options, and others face decisions between treated goods and their outside options. It is not possible to identify MWTP for any consumer, but it is possible to identify the effect of the attribute $z_k$ on aggregate surplus. In a “randomized offer experiment” the researcher offers each participant in the experiment a subsidized good. The level of $z_k$ in the offered good and the amount of the subsidy are randomized across consumers. In the randomized offer experiment the researcher compares the effect of $z_k$ on the demand for the offered good to the effect of the dollar subsidy to measure the subsidy amount that would increase demand by as much as the attribute does.

We concluded by presenting a small-scale field experiment that randomized the price of carbon monoxide detectors offered to households to illustrate an application of the idealized experiment estimator. We showed a large discrepancy in MWTP estimates between those derived using traditional hedonic regression and more robust experimental methods for the sample. Although use of the idealized experiment and other estimators outlined in the paper may not be ideal in all research settings due to data limitations, we strongly encourage their adoption to provide more robust and transparent MWTP estimates of heterogeneous goods.
REFERENCES


Appendix A  
**Experimental Estimator: Studying Existing Tradeoffs**

In some cases it is possible to learn about consumers’ valuations of an important attribute or public good by examining decisions that specific consumers already face:

**EXISTING TRADEOFF EXPERIMENT:** Let $Z_0$ and $Z_k$ be two sets of goods such that $Z_k$ contains the treated variant of every good in $Z_0$. A sample of $N$ consumers is drawn from a population $a$ of consumers. The consumers in $a$ all have utility functions such that their optimal bundles in $Z$ either include only goods in $Z_0$ or only goods in $Z_k$. Hence, for all $i \in a$, at every value of $\tau$, $Q_{Z_0}(\tau, s^*) = Q_{Z_k}(\tau, s^*)$ or $Q_{Z_k}(\tau, s^*) = Q_{Z_0}(\tau, s^*)$, where $s^*$ is the state of the economy as described by prices and wealth. For each consumer in the sample, prices between the two sets differ by a consumer-specific constant $treatment price \tau_i$, so that $p_z(z) = p_k(z_1, \ldots, z_k + 1, \ldots, z_n) - \tau_i$ for each $z \in Z_0$. There is random variation across consumers in $\tau_i$.

In the existing tradeoff experiment, the researcher first identifies a population of consumers who face an important binary decision. This decision is modeled as a choice between selecting one’s goods in $Z$ entirely from the set $Z_0$ versus entirely from the treated variant $Z_k$. The researcher selects a random sample of consumers already making this decision and randomly assigns taxes or subsidies for selecting the treated option. Our estimator in the existing tradeoff setting is identical to the estimator used in the idealized experiment. The distribution of MWTP is estimated with a nonparametric kernel regression of $1 - Treat_i$ on $\tau_i$. The existing tradeoff estimator identifies the cumulative density function $F_k^{MWTP}$ and consequently generates internally valid estimates of the distribution of MWTP for the subpopulation $a$. The degree to which these estimates generalize to the overall population depends on the degree to which $a$ is representative. If decisions are only observed at one treatment price $\tau$, $F_k^{MWTP}$ can be estimated at the MWTP value $\tau$. If the treatment price is assigned according to a discrete probability distribution, then $F_k^{MWTP}$ can be estimated at the different $\tau$ values assigned in the experiment.

As an example of an existing tradeoff experiment, Duflo, et al. (2006), consider a population of consumers who must decide whether to invest their tax returns into individual retirement accounts or take them as cash. The authors randomly assign a subsidy for selecting the retirement account, so that the time-deferred asset is subsidized for some consumers and not for others. In another example, Abrams and Rohlfs (2011) examine criminal defendants who are deciding whether to post bail or to remain in jail until trial. The authors focus on an experiment conducted in 1981 in Philadelphia that generated random variation across defendants in the bail levels that they faced. The market of interest $Z$ is one’s activities over the 90 days until one’s trial. Defendants who select the set (and do not post bail) $Z_0$ must remain in jail, so that they have relatively little freedom over those 90 days, and defendants who select $Z_k$ and post bail are allowed to be free from jail over those 90 days. This exogenous variation in bail makes it possible to measure the value that defendants placed on 90 days of freedom from jail.
An example of a quasi-random discrete probability distribution for the price of treatment appears in a study by Rohlfis, Sullivan, and Kniesner (2015) of consumer valuations of automobile air bags. In that case, the market of interest is new and used vehicles, $Z_0$ includes vehicles without air bags, and $Z_k$ includes the same vehicles with air bags. Due to government regulations, the supply of air bags increased dramatically over a short period, so that in the early periods of the data, the premium on an air bag reflected the valuation of an air bag for someone at the high end of the MWTP distribution. As the air bags became steadily more common, the premium on an air bag reflected the valuation of an air bag for a consumer at the median, and later at the low end of the MWTP distribution, so that the shifts trace out the shape of the MWTP distribution.
Appendix B
Experimental Estimator: Studying Take-up Rates

One common experimental strategy for measuring attribute demands is to examine total sales for matched pairs of products that are similar in all but one attribute. Using such an approach to estimate MS involves conducting two experiments:

TAKE-UP EXPERIMENT: Let $Z_0, Z_k$ be two sets of goods such that $Z_k$ contains the treated variant of every good in $Z_0$. In a take-up experiment, $N_{\text{takeup}}$ consumers are selected from the population. The first $N_{\text{takeup}}/2$ consumers are offered a subsidy of $\delta$ per unit of goods consumed from $Z_0$. The second $N_{\text{takeup}}/2$ of consumers are offered a subsidy of $\delta$ per unit of goods consumed from $Z_k$. Selection into the first or second half of the sample is randomized. Let $s^*$ denote the state of the economy in the take-up experiment.

PRICE EXPERIMENT: In the price experiment, $N_{\text{price}}$ consumers are selected. Each consumer is charged a tax $\tau$, which could be positive or negative, per unit of goods consumed from the set $Z_T$. Let $s^T$ denote the state of the economy in the price experiment.

There are many different types of studies that have the general structure of the take-up experiment, and take-up experiments are often easier and less costly to implement than the idealized, existing tradeoff, and offer-restricted experiments are.

This relative ease of implementation comes at a cost, however. It is not possible using the take-up experiment to measure the MWTP for any consumers, and we instead focus on estimating marginal surplus for the marginal unit:

MARGINAL SURPLUS FOR THE MARGINAL UNIT. For a set $Z_T$ of goods, a set $a$ of consumers, a subsidy of $\delta$ per unit of goods consumed in $Z_T$, and a state of the economy $s$, marginal surplus for $z_k$ for the marginal unit is defined as $MS^k_{Z_Ta}(\delta, s, s)$.

The marginal surplus for the marginal unit is the vertical difference in aggregate demand between $Z_T$ and its treated variant, evaluated at the quantity level $Q_{Z_Ta}(-\delta, s)$ that would be consumed in the absence of the treatment, supposing that all consumers in $a$ were given a subsidy of $\delta$ per unit purchased of goods in $Z_T$. This surplus can be viewed as the social benefit associated with applying the treatment technology to the last unit of $Z_T$ that is consumed.

Another limitation with the take-up framework is that to identify the parameter of interest, it is necessary to impose restrictions on the functional form for aggregate demand:

LINEAR DEMAND. Aggregate demand is linear in goods in $Z_T$ if and only if the “own price effect” of the tax $\tau$ on quantity demanded of $Z_T$ is constant with respect to the price. Hence, linear demand for goods in $Z_T$ implies that, for all $s$ and $\tau$, $\partial Q_{Z_Ta}(\tau, s) / \partial \tau = \beta_{Z_Ta}(s)$. The function $\beta_{Z_Ta}$ describes the own price effect on quantity demanded as it varies with the state of the economy. This function is constant with respect to $\tau$.
CROSS-SAMPLE VALIDITY OF OWN PRICE EFFECT. For two aggregate demand functions $Q_{Z,a}(\tau, s)$ and $Q'_{Z,a'}(\tau', s')$ cross-sample validity of the own price effect is satisfied if
\[
\frac{\partial Q_{Z,a}(\tau,s)}{\partial \tau} = \frac{\partial Q'_{Z,a'}(\tau', s')}{\partial \tau}.
\]

For two aggregate demand functions $Q_{Z,a}(\tau, s)$ and $Q'_{Z,a'}(\tau', s')$ to satisfy cross-sample validity of the own price effect, the effect of the per unit tax $\tau$ on quantity demanded of $Z_T$ by consumers in $a$ in state of the economy $s$ at tax level $\tau$ must equal the effect of the per unit tax on quantity demanded of $Z'_T$ by consumers in $a'$ in state of the economy $s'$ at tax level $\tau$. We require that that $Q_{Z_0}(\tau, s^*)$ is linear in $\tau$ and that cross-sample validity of the own price effect is satisfied between $Q_{Z_0}(\tau, s^*)$ and $Q_{Z_*}(\tau, s^*)$.

Given these two conditions, the own price effect estimated from the price experiment is a constant slope that can be applied to the results from the take-up experiment.\(^{14}\)

ESTIMATED PRICE EFFECT. The estimated price effect from the price experiment is the coefficient on $\tau$ from an ordinary least squares regression of $Q_{Z,i}(\tau_i, s^*)$ on $\tau_i$. For a given sample size $N_{price}$, let this coefficient be denoted $\hat{\beta}_T$.

TAKE-UP ESTIMATOR. Our take-up estimator of the marginal surplus for the marginal treated unit of $Z_0$ is denoted $\hat{MS}_{Z_0}^K(Q_{Z,k}(-\delta, s^*), s^*)$. This parameter is calculated as
\[
\hat{\beta}_T \sum_{i=1}^{N_{takeup}/2} Q_{Z,k}(-\delta,s^*) - \sum_{i=1}^{N_{takeup}/2} Q_{Z_0}(-\delta,s^*)
\]

The difference inside the brackets is estimated from the take-up experiment and measures the effect of the additional unit of $z_k$ on demand for goods in $Z_0$; however, the effect is measured in quantity units rather than dollars. To obtain a dollar measure, the researcher uses the price experiment to measure the effect of exogenous price shocks on quantity demanded. By cross-sample validity of the own price effect, $\hat{\beta}_T$ is an unbiased estimate of $\beta_{Z_0a}(s^*)$, the own price effect for the take-up experiment. Dividing the quantity effect by negative one times the own price effect converts this quantity effect into a dollar-denominated measure of the benefit of the additional unit of $z_k$.

Figure B-1 below illustrates this estimation strategy. The lower and higher curves plot demand for goods in $Z_0$ and $Z_k$. Prices in the take-up study are the same for goods in the two sets, and the

\[\text{Figure B-1 below illustrates this estimation strategy. The lower and higher curves plot demand for goods in } Z_0 \text{ and } Z_k. \text{ Prices in the take-up study are the same for goods in the two sets, and the} \]

\[\text{We require that the two functions examine the demand for equally-sized sets of goods among equally-sized sets of consumers. If the assumption is relaxed, then the estimated own price effect from one economy must be multiplied by a correction factor to accurately estimate the own price effect in the other economy.} \]

\[\text{In principle, linearity of demand is not required for identification. Aggregate demand could have curvature of unknown form that could be identified using nonparametric regressions on data from the price experiment. If, however, the price effect is estimated from an entirely different economy from that studied in the take-up experiment, it is unlikely that the curvature in demand would be similar across the two experiments. In the special case in which } Z_0 = Z_* \text{ and } s^* = s^*, \text{ so that a single economy is being studied, the curvature is likely to be similar across the two samples, and relaxing the linearity assumption would be appropriate.} \]
difference shown along the horizontal axis illustrates the effect of the additional unit of $z_k$ on quantity demanded. Multiplying this difference by negative one times the slope of the demand curve gives $MS$, the vertical increase in the demand curve. The ratio gives the amount of the dollar subsidy for goods in $Z_0$ that would be required to increase quantity demanded by as much as the treatment did.

One straightforward application of take-up experiments is audit studies. The classic audit study design involves two job applicants, one black and one white, who are given similar resumes and are trained to act similarly. In this setting, the consumer is the employer, and the product attribute is worker race. The difference in callback or hiring rates between the black and white applicants is used to measure the extent to which employers prefer workers of a certain race, holding other factors constant (Fix and Turner, 1998). Research designs similar to the take-up experiment have also been used to measure banks’ valuations of the race and sex of loan applicants and home sellers’ valuations of the race and sex of potential home buyers or renters (Hu, et al., 2010; Zhao, Ondrich, and Yinger, 2006; Hanson and Hawley, 2011). In some studies, applicants are not trained to act similarly, but researchers use propensity score matching to identify pairs of job or credit applicants or home buyers who are similar in attributes other than race or sex. “Correspondence studies” represent one variation on audit studies in which researchers generate resumes with randomized attributes including experience, education, and a signal of the applicant’s race or sexual orientation (Bertrand and Mullainathan, 2004; Weichselbaumer, 2003). Here $Z_0$ might represent the set of black workers in the data and $Z_k$ could represent the set of white workers in the data. To convert the hiring effects from quantity units into dollar units the researcher can use separate estimates of the hiring effect of subsidizing certain workers, as in Woodbury and Spiegelman’s (1987) experimental study that offered bonuses to employers for each worker hired from one of their treatment groups. The resulting estimator identifies the amount of the bonus for hiring a black worker at which the marginal employer would be indifferent between hiring a white or a black worker.

In addition to studying discrimination, the design of the take-up experiment can be applied to attrition rates in experiments. An influential study by Philipson and Hedges (1998) suggests that researchers can use differential attrition rates between treatment and control groups in experiments to measure the extent to which consumers value the treatment. Similarly, Rohlf’s and Zilora (2013) examine the effect of being in the treatment group in the Tennessee STAR class size experiment on the decision to remain in the study. There, children who were assigned to large classes left the public school system at higher rates than did children assigned to small classes. In this context, $Z_0$ consists of a single good, education in a large class in Tennessee Public Schools, and $Z_k$ consists of a single good, education in a small class in Tennessee Public Schools. The effect of being assigned to a small class on remaining in the school system measures the quantity difference shown on the horizontal axis in Figure B-1 below. The slope of the demand curve is estimated using data from other studies of the effect of receiving a private
school voucher on the decision to leave the public school system. Dividing the quantity effect by the voucher effect gives an estimate of the private school voucher amount that would be necessary to cause as many people to leave the school system as being assigned to a large class did. The ratio gives measures the dollar equivalent of the benefit of small classes.

Another potential application of the take-up estimator would use data from the Moving to Opportunity Experiment (Kling, Ludwig, and Katz, 2005). The population of interest there is individuals living in higher poverty neighborhoods. Consumers in the treatment group were offered Section 8 housing vouchers, which are long-term housing subsidies, as incentives to move to low-poverty neighborhoods. Consumers in the control group received a housing voucher without a requirement to move to low poverty neighborhood. Here $Z_0$ represents apartments in low poverty neighborhoods, and $Z_k$ represents the same apartments supplemented with Section 8 housing vouchers. The take-up rate in measures the extent to which offering the Section 8 vouchers caused consumers to move to low poverty areas. When combined with separate estimates of the slope of the demand for housing in a certain area, the effect of the voucher on moving measures the value that poor consumers placed on Section 8 vouchers. Estimates of the slope of the demand for housing in a specific area can be taken from studies of the migration effects of welfare benefits or state taxes (Gelbach, 2004; McKinnish, 2007). The resulting estimator measures the increase in welfare benefits or the reduction in taxes that the marginal person moving to a wealthier area would regard as equivalent to a Section 8 voucher.
Figure B-1: Identification of Marginal Surplus using the Take-up Estimator

Surplus equals zero

Demand for products in $Z_k, \tau_{Z_k}(Q, s^*)$

Demand for products in $Z_0, \tau_{Z_0}(Q, s^*)$

$MS^k_{Z_0}(Q_{Z_k}(-\delta, s^*), s^*)$

Observed change in quantity demanded

Aggregate consumption of offered products