



Hedonic methods for baskets of goods[☆]

Rachel Griffith ^{a,b}, Lars Nesheim ^{c,d,*}

^a Institute for Fiscal Studies, United Kingdom

^b University of Manchester, United Kingdom

^c CeMMAP, Institute for Fiscal Studies, United Kingdom

^d UCL, United Kingdom



HIGHLIGHTS

- We apply hedonic price methods to large complex baskets of goods.
- We combine hedonic price methods with revealed preference.
- We estimate bounds on willingness-to-pay for organic using scanner data with thousands of goods.
- Median lower bound is 0.2% of expenditure but 12.5% have a lower bound larger than 1%.
- Median upper bound is 31.5% with most households having a bound between 20% and 40%.

ARTICLE INFO

Article history:

Received 28 February 2013

Accepted 24 April 2013

Available online 3 May 2013

JEL classification:

C8

D1

D4

L1

L8

Keywords:

Hedonic prices

Scanner data

Willingness to pay

Organic

ABSTRACT

Existing hedonic methods cannot be easily adapted to estimate willingness to pay for product characteristics when willingness to pay depends on a very large basket of goods. We show how to marry these methods with revealed preference arguments to estimate bounds on willingness to pay using data on purchases of seemingly impossibly high dimensional baskets of goods. This allows us to use observed purchase prices and quantities on a large basket of products to learn about individual household's willingness to pay for characteristics, while maintaining a high degree of flexibility and also avoiding the biases that arise from inappropriate aggregation.

We illustrate the approach using scanner data on food purchases to estimate bounds on willingness to pay for the organic characteristic.

© 2013 The Authors. Published by Elsevier B.V. All rights reserved.

1. Introduction

Researchers, policy makers and firms often want to estimate consumers' willingness to pay for a characteristic of a good. For example, there is much interest in estimating willingness to pay for organic products (for example, see Blow et al., 2008). For small scale problems, hedonic or discrete choice methods can provide estimates. However, these methods are not tractable when the number of relevant products is large or the characteristic space is large.

We consider the (common) situation in which a consumer buys a large basket of goods, each good having many characteristics. We propose a method to marry hedonic price methods to revealed preference methods for analysing these large and complex baskets of goods.

It has long been understood that analogues of classic revealed preference arguments apply to hedonic prices (see for example Scotchmer (1985), Kanemoto (1988), Pollak (1989), and Pakes (2003)). These papers show that hedonic prices can be used to bound willingness to pay and willingness to accept. We build on Scotchmer (1985) and Pollak (1989) to develop the argument when consumers buy a basket of goods. The idea is simple. The fact that a consumer paid some premium to purchase a basket of goods implies that the consumer must have been willing to pay at least as much as that premium.

We combine ideas from the hedonic pricing literature (Nesheim, 2008; Bajari and Benkard, 2005) with revealed preference ideas

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Correspondence to: IFS, 7 Ridgmount Street, London WC1E 7AE, United Kingdom. Tel.: +44 2076795826.

E-mail addresses: rgriffith@ifs.org.uk (R. Griffith), l.nesheim@ucl.ac.uk (L. Nesheim).

from Blow et al. (2008) to analyse willingness to pay when consumers purchase continuous quantities of a high dimensional basket of goods. A major benefit of our approach is that we can exploit rich data without introducing aggregation bias and without making unnecessary separability assumptions. Under very minimal assumptions we are able to estimate bounds on willingness to pay; with more restrictive assumptions (but ones that are common in the literature) we can obtain point estimates of households' willingness to pay.

We illustrate our approach by estimating bounds on willingness to pay for organic foods using data on the shopping baskets of a large number of households. These estimates can inform regulation over the licensing and labelling of organic foods, increase government knowledge about consumer valuations of agricultural and environmental policies, and help give firms a better understanding of the potential profitability of new product lines.

2. Theoretical background

To develop intuition, we first describe bounds on willingness to pay in the single product case. Then we extend the analysis to the choice of a basket of products.

2.1. Demand for a single product

Let $z \in Z \subseteq \mathbf{R}_n$ be the vector of all product characteristics that affect consumer choice. Let $z(1)$, the first coordinate of z , be the characteristic of interest. In our example, $z(1) = 1$ if a product is organic and $z(1) = 0$ otherwise. The product price is given by $p = h(z)$ where $h(z)$ is the equilibrium hedonic price.

Consider a consumer with characteristics x_h who buys a single unit of an organic product with product characteristics z^o and price p^o and elects not to buy a non-organic product with characteristics z^n and price p^n . Assume that the two products are identical in all dimensions other than organic. Let the consumer's indirect utility function be $v(x_h, z, p)$, where v is increasing in $z(1)$, continuously differentiable in p and strictly decreasing in p . If the consumer chooses the organic product, then revealed preference dictates

$$v(x_h, z^o, p^o) \geq v(x_h, z^n, p^n), \quad (1)$$

the consumer obtains weakly greater utility from the organic product. By the mean value theorem, there exists some $p^* \in [p^o, p^n]$ such that

$$v(x_h, z^o, p^o) + \frac{\partial v(x_h, z^o, p^*)}{\partial p} (p^o - p^*) \geq v(x_h, z^n, p^n),$$

where $-\frac{\partial v(x_h, z^o, p^*)}{\partial p} > 0$ is the marginal utility of income. After rearranging, we have

$$\frac{v(x_h, z^o, p^o) - v(x_h, z^n, p^n)}{-\frac{\partial v(x_h, z^o, p^*)}{\partial p}} \geq p^o - p^n. \quad (2)$$

The left side of this expression is the willingness to pay for the organic characteristic. The right side is the organic price premium. For all consumers who buy organic, the price premium defines a lower bound on the willingness to pay for organic. For all consumers who do not buy organic, the price premium provides an upper bound on the willingness to pay for organic.

2.2. Demand for a basket

Let B_g be the set of products in category g and let $B = \bigcup_g B_g$ be the set of all products. For each product $b \in B_g$, let $z_b \in Z_g$ be its vector of characteristics. Define $\bar{z} = \{z_b\}_{b \in B}$ as the vector of characteristics of all products.

Let $\bar{v} = v(x_h, \bar{z}, \bar{p})$ be the maximum utility obtainable given market prices \bar{p} and product characteristics \bar{z} . Each consumer chooses a vector of quantities of each product, \bar{q} , to minimise costs of attaining the fixed utility level \bar{v} . The consumer's total expenditure is

$$\begin{aligned} e_h &= c(\bar{p}, \bar{z}, x_h, \bar{v}) \\ &= \min_{\bar{q}} \{ \bar{p} \cdot \bar{q} \text{ subject to } u(x_h, \bar{z}, \bar{q}) \geq \bar{v} \}. \end{aligned}$$

In general, the basket purchased will include both organic and non-organic products and the fraction organic will vary across consumers.

Denote what the consumer would have paid to obtain the same utility level if all products were converted to non-organic products with non-organic prices,

$$e_h^n = c(\bar{p}^n, \bar{z}^n, x_h, \bar{v}),$$

where $\bar{p}^n = \{p_b^n\}_{b \in B}$ and $\bar{z}^n = \{z_b^n\}_{b \in B}$ are the vectors of prices and characteristics in the counterfactual world where all products are converted to non-organic varieties. For household x_h , the willingness to pay for organic is the difference between these expenditures,

$$WTP_h = e_h - e_h^n.$$

It is the negative of compensating variation.

If we assume that the utility function is known, then we can calculate a point estimate of willingness to pay using the price premia. More generally, if the utility function is not known, we cannot calculate willingness to pay. Nevertheless, revealed preference gives a lower bound,

$$WTP_h^n = e_h - e_h^n \geq (\bar{p} - \bar{p}^n) \cdot \bar{q}_h. \quad (3)$$

By choosing to purchase \bar{q}_h , the consumer has revealed that they are willing to pay at least $(\bar{p} - \bar{p}^n) \cdot \bar{q}_h$ to purchase organic. This follows immediately from cost minimisation since

$$\bar{p}^n \cdot \bar{q}_h \geq e_h^n.$$

We can also compute various upper bounds for willingness to pay by considering counterfactual bundles in which some non-organic products are converted to organic. For example, let $\bar{z}^o = \{z_b^o\}_{b \in B}$ be the "all-organic" counterfactual characteristics vector with $z_b^o(1) = 1$ and $z_b^o(j) = z_b(j)$ for $j > 1$ and for all $b \in B$. Let $p_b^o = h_g(z_b^o)$ for all $b \in B$ and for all g . For this counterfactual bundle, we can compute upper bounds on willingness to pay for each consumer using,

$$w_h^o = (\bar{p}^o - \bar{p}) \cdot \bar{q}_h.$$

In summary, for each consumer we can calculate lower and upper bounds on willingness to pay for organic using

$$w_h^n = (\bar{p} - \bar{p}^n) \cdot \bar{q}_h \leq e_h - e_h^n \quad (4)$$

$$w_h^o = (\bar{p}^o - \bar{p}) \cdot \bar{q}_h \geq e_h^o - e_h. \quad (5)$$

2.3. Data

We use data from the 2004 Kantar Worldpanel for the UK to estimate (4) and (5). Households record purchases of all food, toiletries and cleaning products that are brought into the home using hand-held scanners. Prices are recorded from till receipts collected from the households. We use information on prices, quantities and characteristics of food items purchased for home consumption by 16,881 households. The sample contains data on more than

11 million purchases. Leicester and Oldfield (2009), Griffith and O'Connell (2009) and Griffith and Nesheim (2010) provide further description of the data.

We use data on 75 food categories where organic is a relevant characteristic. Total expenditure in our sample of households is £12.8m (grossed up using sampling weights it is £19.7bn). Summary statistics for the 75 food categories are provided in Griffith and Nesheim (2010). Overall, 2.1% of expenditure is on organic products, but there is substantial variation from 0.4% of "Fresh Bacon Rashers" to 28.6% of "Chilled Meat and Vegetable Extract". Estimating willingness to pay for organic requires the analysis of a very large basket of goods.

Organic purchases also vary across households. Just under 20% of households never buy any organic products and over a quarter buy only a very small amount (less than 0.5% of total expenditure). However, 37% of households spend more than 1% of their budget on organic products and 7% of households spend over 5% of their budget on organic products. These numbers illustrate the tremendous heterogeneity in demand for organic products, and that organic is an important expenditure category for a significant fraction of the population.

2.4. Hedonic model

To evaluate willingness to pay we estimate a hedonic price model for each of the 75 food categories. In this example we assume a linear form, but the hedonic model could take a more flexible functional form. In fact because we have very rich data on characteristics of individual products, and because most characteristics are discrete, this is in fact quite flexible. Let (b, r, s, t) index items, regions, store types and time. For each product category, we estimate a hedonic regression of the form

$$\ln(p_{brst}) = \alpha_1 \delta_t + \alpha_2 \kappa_{bt} + \alpha_3 \phi_r + f(z_{bs}) + \varepsilon_{brst} \quad (6)$$

where δ_t is the vector of month dummies, κ_{bt} is a vector of indicators for special offers (ticket price reduction, multi-pack purchase and extra free), and ϕ_r is a vector of regional dummies. The coefficients of interest are those on the organic coefficient, which might vary with other product characteristics, $f(z_{bs})$. The residual ε_{brst} captures unobserved product characteristics that are mean independent of the observed characteristic.

2.4.1. Hedonic price estimates for milk

To illustrate the approach, we first present results for a single category, milk in Table 1. The first column includes only month and region effects along with the organic characteristic interacted with fat content. The adjusted R^2 for this regression is only 0.065. The interactions between organic and fat content are not significant (either individually or jointly)—firms in the UK do not charge differential premia on organic depending on the fat content. In columns 2 and 3, we drop these interactions and in column 3 we add in the full set of characteristics including package size and type, variety of milk, store fascia in which purchased and whether on special offer. Many of these are statistically significant, and the estimated organic premium declines significantly. The adjusted R^2 increases to 0.726. The additional characteristics explain a substantial proportion of the variation in prices. In the final column we include interactions between the organic characteristic and store fascia. Across all stores, the average price premium for organic milk is 15% and ranges from 0% at Asda to 13% at Tesco to 30% at Waitrose. Since the market share of organic milk is 2.2%, these results imply that roughly 2.2% of households have a willingness to pay a premium for organic milk of at least 15%, while 97.8% of households are willing to pay no more than 15%. However, those who buy organic milk at Waitrose reveal a lower bound on willingness to pay of 30%.

2.4.2. Hedonic price estimates for all food categories

We estimate 75 separate hedonic regressions of the form of (6). Each regression includes a set of characteristics that is common to all categories, as well as a set of category specific characteristics. The category specific characteristics vary in number and type. For example, there are over 200 flavours of soup and over 250 flavours of yoghurt. Eggs, on the other hand, have relatively few characteristics—whether they are barn reared or free range, eggs size and whether they are branded. The key point is that, as with milk, many of the characteristics are correlated with the organic characteristic, so failing to control for them will lead to biased estimates of the organic price premia. Complete results can be found in Griffith and Nesheim (2010).

The organic coefficients vary a great deal both across product categories and across stores. Of the 595 potential organic-fascia coefficients we are able to identify 518 (some stores never sell an organic version of some products), 462 are positive and 338 are significantly so (at the 5% level). The unweighted mean of the price premia is 0.40 (organic products are roughly 40% more expensive) and the median is 0.38. Asda and Safeway have low prices on average and have the smallest mean and median price premia as well as the most categories (8) with price premia less than or equal to zero. The other stores have higher average and median organic price premia, fewer categories (4 or 5) with non-positive coefficients. In all cases, the range of positive price premia is from zero to nearly 125%. Marks and Spencer has the highest mean and median markup, followed closely by Sainsbury's, Waitrose and Tesco.

The adjusted R^2 are high (with a few exceptions) suggesting that we have captured most of the product characteristics that affect pricing. We have detailed information on all product characteristics judged to be important by market research firms, including characteristics that vary over time (such as being on special offer) and space (such as being sold in a different store). As indicated by the adjusted R^2 , measured characteristics explain most of the variation of prices in our data.

2.5. Bounds on willingness to pay for baskets

We aggregate these individual price premia to obtain estimates of (4) and (5) for each household. The dimensions of the vectors \bar{p} and \bar{q}_h are each over 4 million (47,854 barcodes by 12 months by 7 stores). Each element of $(\bar{p} - \bar{p}^n)$ is computed using our estimated hedonic coefficients from (6). Households spend different amounts on food, so we present the estimates as a share of expenditure

$$sw_h^n = (e_h - e_h^n) / e_h, \quad (7)$$

$$sw_h^o = (e_h^o - e_h) / e_h. \quad (8)$$

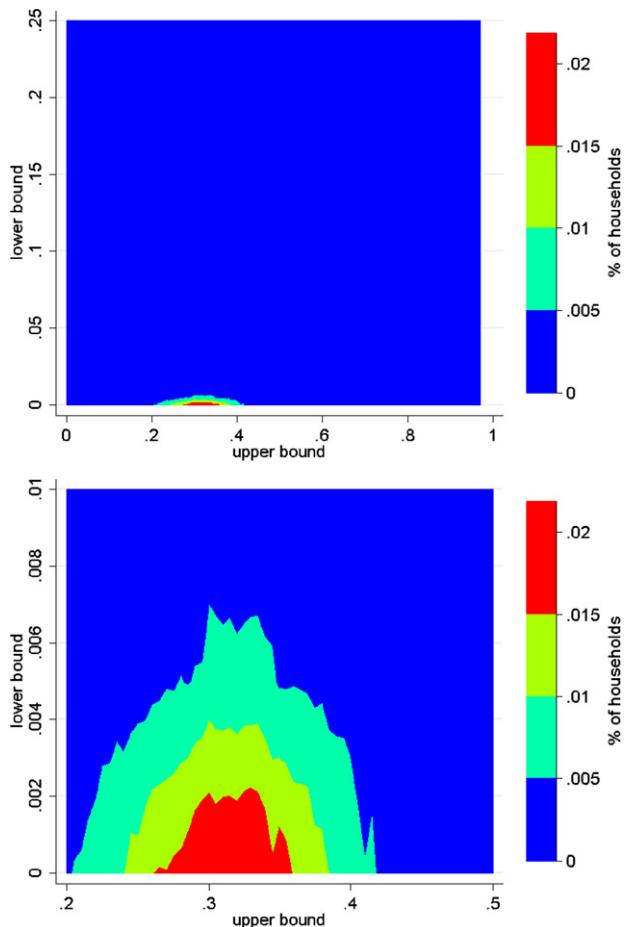
For a very small number of households (18) we estimate a negative lower bound on their willingness to pay for organic. For a much larger number (4121 (22.7%)) we find that they have zero willingness to pay for organic, either because they purchase no organic products, or the net price premia across all products purchased cancels out. The median lower bound on the willingness to pay is 0.2%, with just over 12.5% having a lower bound on their willingness to pay that is 1% or more. The median upper bound is 31.5%, with most households having a value between 20% and 40%.

The figures show a contour plot of the joint distribution of (7) and (8) across households. The first figure includes all households, the second figure focuses in on that part of the plot where most households are located. From the figures, one can see that the largest mass of households is located in the region where the lower bound ranges from zero to 0.2% of expenditure and the upper bound ranges from 35% to 45%.

Table 1
Hedonic regression for milk.

Dependent variable: ln (price)	(1)	(2)	(3)	(4)
Organic	0.296 (0.071)	0.256 (0.038)	0.150 (0.023)	0.166 (0.039)
Organic*semi-skimmed	-0.084 (0.083)			
Organic*skimmed	-0.031 (0.096)			
Organic*Asda				-0.170 (0.092)
Organic*M&S				0.061 (0.039)
Organic*Safeway				0.030 (0.035)
Organic*Sainsbury				0.030 (0.037)
Organic*Tesco				-0.034 (0.035)
Organic*Waitrose				0.136 (0.051)
Fat content	Yes	Yes	Yes	Yes
Other characteristics			Yes	Yes
Adjusted R ²	0.065	0.065	0.726	0.728

Notes: Regressions include 745,006 observations on 16,881 household purchases of milk over the calendar year of 2004. All regressions include month and region effects. Fat content is full fat, semi-skimmed or skimmed. Other characteristics include pack size, container, brand, type of milk, treatment, store, and whether on promotion. Standard errors in () are clustered at the barcode level and allow for general correlation.



3. Summary and conclusions

Rich data on spending behaviour are now widely available in a number of countries. These data offer great potential to learn about willingness to pay for many different characteristics. However, their use has been in part thwarted by the sheer scale of the data. Existing revealed preference approaches to estimating willingness to pay cannot deal with the large dimensionality of these data.

Methods such as Blow et al. (2008) illustrate how assumptions about separability and the absence of time varying demand shocks, combined with panel data, can be used to obtain point estimates of willingness to pay, at least for a fraction of the population. We extend the ideas developed in Blow et al. (2008) by incorporating market pricing equilibrium conditions, which help to reduce the

dimensionality of the problem, but allow us to retain much of the flexibility of their approach. We use standard assumptions about market pricing equilibrium and consumer revealed preference behaviour to compute consumer specific bounds on willingness to pay. We show how to aggregate estimates of willingness to pay for individual products across a basket of products in a manner that is consistent with consumer theory. These bounds are Laspeyres style price indexes for differentiated products. In order to recover point estimates of the willingness to pay we would need to make further assumptions about the structure of consumers' preferences. While this is certainly feasible for individual product categories, further work needs to be done to develop a tractable method to analyse the entire food basket.

We illustrate the application of these methods using rich data on households' purchases of food to estimate lower and upper bounds on willingness to pay for the organic characteristic in food. Our results suggest that there is a large amount of heterogeneity in willingness to pay for organic products.

Acknowledgements

The authors would like to thank James Banks, Richard Blundell, Martin Browning, Ian Crawford, Andrew Leicester, Aviv Nevo, Ariel Pakes and Carol Propper for many helpful comments. Financial support from the ERC through grant No. 249529 (MAPFAN) and the ESRC through the ESRC Centre for Microdata methods and Practice (CeMMAP) grant number ES/I034021/1 is gratefully acknowledged.

References

- Bajari, Patrick, Benkard, C. Lanier, 2005. Hedonic price indexes with unobserved product characteristics, and application to personal computers. *Journal of Business & Economic Statistics* 23 (1), 61–76.
- Blow, Laura, Browning, Martin, Crawford, Ian, 2008. Revealed preference analysis of characteristics models. *Review of Economic Studies* 75 (2), 371–389.
- Griffith, R., Nesheim, L., 2010. Estimating households' willingness to pay. CeMMAP Working Papers, CWP24/10.
- Griffith, R., O'Connell, M., 2009. The use of scanner data for research into nutrition. *Fiscal Studies* 30 (3/4), 339–365.
- Kanemoto, Y., 1988. Hedonic prices and the benefits of public projects. *Econometrica* 56, 981–989.
- Leicester, A., Oldfield, Z., 2009. Using scanner technology to collect expenditure data. *Fiscal Studies* 30 (3/4), 339–365.
- Nesheim, L., 2008. Hedonic pricing functions. In: New Palgrave Dictionary of Economics, second ed. Palgrave Macmillan, New York.
- Pakes, Ariel, 2003. A reconsideration of hedonic price indexes with an application to PCs. *The American Economic Review* 93 (5), 1578–1596.
- Pollak, Robert, 1989. The Theory of the Cost of Living Index. Oxford University Press, New York.
- Scotchmer, Susanne, 1985. Hedonic prices and cost/benefit analysis. *Journal of Economic Theory* 37, 55–75.