

Household willingness to pay for organic products

Rachel Griffith
Lars Nesheim

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP18/08



An ESRC Research Centre

Household willingness to pay for organic products

Rachel Griffith

Institute for Fiscal Studies and UCL

Lars Nesheim

CEMMAP

Institute for Fiscal Studies and UCL

July 2008

Abstract

We use hedonic prices and purchase quantities to consider what can be learned about household willingness to pay for baskets of organic products and how this varies across households. We use rich scanner data on food purchases by a large number of households to compute household specific lower and upper bounds on willingness to pay for various baskets of organic products. These bounds provide information about willingness to pay for organic without imposing restrictive assumptions on preferences. We show that the reasons households are willing to pay vary, with quality being the most important, health concerns coming second, and environmental concerns lagging far behind. We also show how these methods can be used for example by stores to provide robust upper bounds on the revenue implication of introducing a new line of organic products.

JEL: D12, L11, L81, Q51, C81

Correspondence: rgriffith@ifs.org.uk, l.nesheim@ucl.ac.uk

Acknowledgement: 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 ESRC through the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (CPP) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) is gratefully acknowledged. All errors remain the responsibility of the authors.

1 Introduction

How much are households willing to pay for organic products? How does this vary across households? Does it reflect a desire to improve the environment, achieve better health or eat better quality food? Answer to these questions are important for a number of reasons including to inform regulation, licensing and labelling of organic foods, to increase government knowledge about household valuations of agricultural and environmental policies, and to understand the potential profitability of new products. The widespread availability of rich microdata sources that provide precise measurements of household level prices, quantities and utility relevant characteristics for individual products (defined at the barcode level) make it feasible to explore these questions in greater detail than was previously possible. In particular, they allow for the study of hedonic prices and hedonic demand across the entire basket of household food purchases at the household level.

These data also present research challenges due their size and dimensionality. One way to proceed is to focus on a small number of goods. A second is to aggregate goods, or to use economic theory to impose structure that makes the problem tractable. We take another approach, inspired by the revealed preference literature (e.g. Varian (1982), Blow, Browning and Crawford (2008)) and from the recent growing interest in partially identified models (e.g. Manski (2003), Chernozhukov, Hong, and Tamer (2007), and Pakes (2003)), to study what can be learned with minimal revealed preference assumptions and relying more on the quality of the data.

We find that hedonic prices estimated using scanner data allow us to compute household specific lower and upper bounds on willingness to pay for various baskets of organic products. They do not allow us to point identify structural demand parameters, but do provide informative bounds on willingness to pay for organic products taking the current equilibrium as a starting point. They do this without imposing overly restrictive assumptions on preferences, which are needed in order to point identify and estimate a structural demand system. Our estimates rely only on the assumptions that: households know the

distribution of prices in the market and willingly choose to purchase organic or not, household utility depends on observable characteristics (and possibly unobserved characteristics), the set of marketed products is rich enough to allow us to unbundle product characteristics. More precise estimates of willingness to pay require further assumptions about household preferences.

Our contribution in this paper is threefold. First, we explore how much can be learned empirically if we rely only on the assumptions of hedonic revealed preference and use detailed household level data on purchases, prices and characteristics. This type of data is now becoming widely available across many countries. We observe actual prices and quantities of individual items at the bar code level, along with characteristics of all food items purchased and brought into the household. Compared to other empirical papers in the literature our data are not aggregated in any way, and we observe the prices actually paid, including all discounts and special offers. Second, we use information on purchases of all types of organic foods from a nationally representative panel of households across the UK to estimate households' revealed preference for organic foods. In contrast to much of the hedonic literature, we compute willingness to pay not for a discrete item, but for baskets of products. We compute quantity weighted hedonic price premia (expenditure premia) and use these to bound households' willingness to pay. This is the first comprehensive look at product pricing and household purchase behaviour using detailed disaggregate data on the entire basket of food purchases, which account for on average 13% of total consumer expenditure.¹ Third, we combine information on revealed preferences with responses to attitudinal questions on households' preferences and beliefs to disentangle households' bounds on willingness to pay for organic into bounds on willingness to pay for the environment, health and quality.

Precise measurements of prices, quantities and characteristics allow us to accurately analyse the relationship between prices and characteristics and to explain most of the vari-

¹Work using similar data has looked either at the aggregate basket of groceries (e.g. Smith, 2004) or at single product categories, for example, breakfast cereals (Nevo, 2001), ketchup (Pesendorfer, 2002), yoghurt (Ackerberg, 2001) or carbonated soft drinks (Dube, 2005).

ation in prices. They also allow use to estimate heterogeneous bounds on willingness to pay for individual households. Willingness to pay, or at least bounds on willingness to pay, vary across products and across households. A large fraction of households never buy any organic products and have a willingness to pay bounded above by observed price premia. The fraction that buys organic varies significantly across product categories ranging from 0% to 13%. Of those buying organic products, a significant fraction are willing to pay quite a large amount. These variations across products and across households in bounds on willingness to pay are obscured by higher levels of aggregation. Additionally, we show how to aggregate estimates of the willingness to pay coefficients for different products in a manner that is consistent with consumer theory to produce bounds on willingness to pay for baskets of goods. These bounds are Laspeyres style price indexes for differentiated products.

We find that aggregate willingness to pay for organic foods in 2004 in England was at least £114m per annum, and that the most that households would have been willing to pay for a fully organic food market was £10.9bn. The latter number is an extreme counterfactual, and we also consider other "feasible" organic baskets. These yield lower estimates on the upper bound of £6.6bn and £5bn.

We use our results to undertake a number of counterfactual simulations. We consider the potential revenue implications for a supermarket chain of converting a product range to organic. We also show that households vary in the reasons they are willing to pay for organic products, with product quality and health being the most important factor, and environmental concerns lagging far behind. We estimate that the total lower bound on willingness to pay for health is around £17m, for the environment around £5m, and for quality around £20m. These results have implications for the regulation of organic labelling, and for the way that firms may want to advertise organic products.

In the household demand literature, our work is most directly related to Blow, Browning and Crawford (2007), who develop non-parametric revealed preference methods to estimate willingness to pay and apply it to organic milk sales in Denmark. In contrast to their work

we: look at the entire basket of food products for which organic is a relevant characteristic; analyse the data at a more disaggregated level (both in terms of time and product) by using actual transaction level unit prices rather than monthly average unit prices. In order to obtain more precise results they impose several theoretical restrictions, specifically, a known separability structure on preferences and no time varying preference shocks. In contrast, we impose fewer restrictions and rely more on the highly detailed disaggregate data.

In the hedonic literature², our work is most closely related to work on hedonic prices in retail settings such as Pakes (2003), Bajari and Benkard (2005a, 2005b), and Erickson and Pakes (2007). Bajari and Benkard (2005a, 2005b) study hedonic prices and willingness to pay for computers. They discuss nonparametric estimation of hedonic price functions and use a factor structure to model unobserved characteristics and recover prices of these characteristics. Bajari and Benkard (2005b) note that a hedonic price index can ignore pricing of unobserved characteristics if the relationship between observed and unobserved is stable. In our setting, the stable relationship assumption justifies our decision to ignore unobserved characteristics. Bajari and Benkard also discuss potential selection corrections for cases where selection may matter, as do Pakes (2003) and Erickson and Pakes (2007). In our setting, where technical change is slow, there is little product entry and exit, and where highly detailed information on characteristics and prices is available, we argue that these issues are not of first order importance. Unmeasured characteristics contribute little to the variation of prices in our data and, because of the stability of the market, it is quite plausible that the relationship of any unmeasured characteristics to measured is stable.

Our work is also related to the price index literature. The most closely related papers are Hausman (2003), Pakes (2003), and Triplett (2004). Triplett (2004) presents a comprehensive discussion of uses of hedonic methods in constructing price indexes. Effectively, what we have done is compute household specific hedonic price indexes and analyse the implications of these

²Our work is most closely related to work on hedonic prices in retail market settings. However, it is also related to work on hedonic prices in labor economics (Sattinger (1995) and Leeth and Ruser (2003)), in environmental economics (Freeman (1995), Smith and Huang (1995), Chay and Greenstone (2005), Sieg et al. (2005)) and urban and public economics (Epple and Sieg (1999), and Sieg et al. (2005)).

for consumer valuation of organic foods. Hausman (2003) discusses biases in the consumer price index (CPI). He notes that biases in the CPI are mainly due to substitution bias, new goods bias and store bias. He notes that statistical authorities often use hedonic regressions but is skeptical of the value of these regressions since there is no simple relationship between price and consumer valuation. Since price is determined by the interaction of demand, cost and competitive conditions, a hedonic price is not a compensating variation. While this is correct, we show that the hedonic price is a lower bound to compensating variation in general models as discussed in Scotchmer (1985), Pollak (1989) and Pakes (2003).

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 while hedonic prices do not identify structural supply or demand parameters, they can be used to bound willingness to pay and willingness to accept (at least given the existing equilibrium). Pakes (2003) develops the argument when a consumer makes a discrete choice. We build on Scotchmer (1985) and Pollak (1989) to develop the argument when consumer choice involves choice of a discrete option along with a continuous intensity of use for a basket of goods. Essentially, the fact that a household paid a premium to purchase a basket of goods implies that the household must have been willing to pay at least as much as the premium.³

A key issue in hedonic models is interpretation of the error term. This has been emphasised at least since Griliches (1961) and is discussed in detail in Epple (1987), Pakes (2003), Triplett (2004), Bajari and Benkard (2005a, 2005b) and Erickson and Pakes (2007). Despite the fact that our hedonic price regressions explain the large majority of price variation for nearly all products, it still may be important to account for the impact of unobserved product characteristics on price and on willingness to pay. We discuss this further in section 3.

³The bound is not structural except under very restrictive assumptions. It may change when market prices change. To estimate structural demand parameters or supply parameters one must use techniques such as those in Epple and Sieg (1999), Ekeland, Heckman, and Nesheim (2004), Smith (2004), Bajari and Benkard (2005), or Heckman, Matzkin and Nesheim (2008).

The rest of the paper is structured as follows. The next section describes a model of household behaviour. Section three discusses the market environment. Section four describes the data and our empirical implementation. Section five presents estimates of the lower and upper bounds on households' willingness to pay for organic. Section six considers the extent to which this willingness to pay reflects concerns about the environment, health or quality. A final section summarises and makes some concluding remarks.

2 Household demand

The model underlying our analysis is a simple rational choice model of a price taking household maximising utility. Each household chooses a shopping basket given the set of stores in the market and the prices of all products. The household's choices then reveal bounds on their willingness to pay. In particular, the willingness to pay for organic is the willingness to pay for all of the organic characteristics in the basket. We make this more precise below. First, to develop intuition, we describe the model in the case of a discrete choice of a single product. We then extend the analysis to model the choice of a basket of products, some fraction of which are organic.

2.1 Demand for a single product

First, we recall a simple single product discrete choice model. A household maximises utility by choosing one product from a finite number of options each described by a vector of characteristics and a price. For some product, let $z \in Z \subseteq \mathbf{R}_n$ be the vector of **all** product characteristics observed by the households and let $z(j)$ be the j 'th coordinate of z . The set Z is the set of feasible characteristics for the product. The product characteristics include the organic characteristic, package size, brand, store identity and location, flavour, variety and others (see Table 3 for details). Let $z(1) = 1$ if a product is organic and let $z(1) = 0$ otherwise. The product price is given by $p = h(z)$, where $h(z)$ is the equilibrium hedonic price defined for all $z \in Z$ including those not sold in equilibrium. We discuss the market

environment and determination of prices in section 3.

The vector z is the vector of product characteristics observed by the household. These are the characteristics that affect the household's choice. In the empirical section of the paper, we change the notation slightly to account for measured characteristics z (that are measured in the economic analyst's data) and unmeasured characteristics ε . In this section, to simplify the notation, we use the single letter z to indicate the full vector of measured and unmeasured characteristics.

Consider a household (indexed h) with characteristics x_h that buys a single unit of an organic product o with product characteristics z^o and price p^o and elects not to buy a non-organic product n with characteristics z^n and price p^n . The vector x_h includes household income, measured demographic characteristics and survey responses, and unmeasured heterogeneity and preference shocks such as the presence of household visitors, weather, travel cost shocks, random variation in who within the household does the shopping, etc. Note that x_h may also include unobserved shocks such as those arising in a logit discrete choice model.

Assume that o and n are identical in all dimensions other than organic. Let the household's indirect utility function be denoted $v(x_h, z, p)$, where v is increasing in z (1), differentiable in p and $\frac{\partial v}{\partial p} < 0$. If the household chooses the organic product, then standard revealed preference arguments imply

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

The household obtains higher utility from the organic product. By the mean value theorem, this inequality is equivalent to

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

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

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

The left side of this expression is the willingness to pay for organic for a single organic item. This expression is completely general. All else, including expenditures on all other goods and period specific preference shocks, is held fixed in (2). In case the products are not identical, inequality (1) would be replaced by $v(x_h, z^o, p^o) \geq v(x_h, \tilde{z}^n, p^n) + (v(x_h, z^n, p^n) - v(x_h, \tilde{z}^n, p^n))$ where $\tilde{z}^n(1) = 0$ and $\tilde{z}^n(j) = z^o(j)$ for $j \geq 1$ and inequality (2) would also include this additional term. The right side is the organic price premium. Revealed preference implies that the price premium is no bigger than willingness to pay. If the household bought the organic product, then they must have been willing to pay a premium of at least $p^o - p^n$. For all households that buy organic, the price premium defines a lower bound on their willingness to pay for organic. The complementary statement is also true. For all households that do not buy organic, the price premium provides an upper bound on their willingness to pay for organic.

2.2 Demand for a basket

When a household buys a basket of goods, which includes some organic and some non-organic products, the reasoning is similar. The willingness to pay for an organic basket will be defined in terms of the characteristics of all the goods in the basket and the total cost of the basket. Some baskets are more organic than others; they have a larger fraction of items that are organic. We define a non-organic basket to be one in which no products are organic. A fully organic basket is one in which all products are organic.

We first define some notation that is necessary both to define the notion of willingness to pay for a basket and for the empirical implementation in the next section. Households shop in a finite set of stores S and choose how much to spend on a finite set of products B . The set of products is partitioned into G categories, with each category labeled by an integer $g \in \{1, \dots, G\}$. The set B_g is the set of products in category g and $B = \bigcup_{g \in G} B_g$. Each product $b \in B_g$ sold in store $s \in S$ has a vector of characteristics $z_{bs} \in Z_g \subseteq \mathbf{R}_{n_g}$ that affect utility. The set Z_g is the set of feasible characteristics for product category g . As in the

single product example above, we assume that $z_{bs}(1) = 1$ if and only if (b, s) is organic. We define $\bar{z} = \{z_{bs}\}_{(b,s) \in B \times S}$ to be the vector of all characteristics of all products.

The price of each product $(b, s) \in B \times S$ is p_{bs} and the vector of all prices is $\bar{p} = \{p_{bs}\}_{(b,s) \in B \times S}$. For each category g , the price of product (b, s) is given by $p_{bs} = h_g(z_{bs})$ where h_g is the hedonic price function for category g . For each g , the function h_g is defined for all $z \in Z_g$ including those not sold in equilibrium.

Given market prices \bar{p} and product characteristics \bar{z} , each household chooses a vector of quantities of each product, $\bar{q}_h = \{q_{hbs}\}_{(b,s) \in B \times S}$, to maximise utility. That is,

$$v(x_h, \bar{z}, \bar{p}) = \max_{\bar{q}_h} \{u(x_h, \bar{z}, \bar{q}) \text{ subject to } \bar{p} \cdot \bar{q} = x_h(1)\}$$

where $x_h(1)$ is income and $v(x_h, \bar{z}, \bar{p})$ is the indirect utility obtained by household h . Let the attained utility level be \bar{v} . In general, the basket purchased will include both organic and non-organic products and the fraction organic will vary across households.

For each household, we observe the actual basket purchased and the price and characteristics of all items purchased. How do we define willingness to pay for organic? In the discrete choice case, willingness to pay is defined with respect to an alternative that is identical in every dimension except organic. When the household purchases a basket however, there are multiple dimensions of organic, one for each product in the basket. We calculate the lower bound on willingness to pay by comparing the household's actual expenditure to what would have been spent if all the products purchased were transformed into non-organic products. That is, how much would the household have spent if they had purchased the same bundle \bar{q}_h but with counterfactual vectors of characteristics and prices (\bar{z}^n, \bar{p}^n) where we use the n superscript to denote "non-organic".

Formally, let $\bar{z}^n = \{z_{bs}^n\}_{(b,s) \in B \times S}$ be a counterfactual vector of characteristics with $z_{bs}^n(1) = 0$ and $z_{bs}^n(j) = z_{bs}^0(j)$ for $j > 1$ and for all $(b, s) \in B \times S$. The vector \bar{z}^n is the vector of characteristics in the counterfactual world in which all organic products are transformed into non-organic products. If the hedonic price schedule is unchanged in this counterfactual world, then for all $(b, s) \in B \times S$ and for all g , prices are given by $p_{bs}^n = h_g(z_{bs}^n)$ where

$h_g : Z_g \rightarrow \mathbf{R}$ is the hedonic price function for product category g . In hedonic equilibrium, the price function prices all bundles, even those that are not marketed. Below we check in our data that the support of the marketed products is rich enough so that households can effectively unbundle product characteristics, as discussed in Heckman and Scheinkman (1987). We also discuss how to compute a counterfactual bundle by replacing each product with its nearest non-organic neighbour in the product space when the household can not unbundle.

We can also write the dual of the household maximisation problem. Given prices \bar{p} , characteristics \bar{z} , and utility level \bar{v} , household expenditure is

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

The value e_h is the cost minimising expenditure when characteristics are \bar{z} and prices are \bar{p} . The optimal consumer basket in this case is \bar{q}_h .

Consider the thought experiment of transforming all products to non-organic, maintaining the current price schedule, keeping everything else constant, and calculating household expenditure while holding utility constant. This expenditure would be:

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

where the vector $\bar{p}^n = \{p_{bs}^n\}_{(b,s) \in B \times S}$ is the vector of prices under the counterfactual assumption that every organic product price premium is set to zero. In this counterfactual thought experiment when characteristics are \bar{z}^n and prices are \bar{p}^n , the cost minimising basket is \bar{q}_h^n .

The amount

$$e_h - e_h^n = WTP_h^n = -CV_h^n$$

is the total organic expenditure premium the household would be willing to pay for the bundle \bar{q}_h . That is, it is the willingness to pay for the organic characteristics in the bundle \bar{q}_h . Its negative is the compensating variation; the amount that exactly compensates a household for a change from (\bar{p}, \bar{z}) to (\bar{p}^n, \bar{z}^n) .

If the utility function were known or \bar{q}_h^n were known, we could calculate this willingness to pay. However, as in the discrete choice case, revealed preference gives a simple bound

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

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

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

That is, with counterfactual prices \bar{p}^n , the cost of the original bundle is at least as large as the new cost minimising basket.

The expenditure premium is household specific. It depends on the bundle purchased. It is the quantity weighted sum of the individual item price premia. The price premium for an individual item or a specific basket is the same for all households. The expenditure premia we calculate are household specific because we use household specific baskets.

In a similar way, we can use our estimates to calculate bounds on willingness to pay for alternative counterfactual baskets. For example, we consider what would be the lower bound on a household's willingness to pay if a store converted a range of its products to organic (and the household continued to purchase the same basket.) This counterfactual would be of interest to a firm considering its product strategy or to competition authorities studying hypothetical market outcomes.

We can also compute an upper bound for willingness-to-pay for a completely organic bundle. For example, we consider the somewhat extreme counterfactual bundle in which all products are organic. Let $\bar{z}^o = \{z_{bs}^o\}_{(b,s) \in B \times S}$ be the "all-organic" counterfactual characteristics vector with $z_{bs}^o(1) = 1$ and $z_{bs}^o(j) = z_{bs}(j)$ for $j > 1$ and for all $(b, s) \in B \times S$ and let $p_{bs}^o = h_g(z_{bs}^o)$ for all $(b, s) \in B \times S$ and for all g . We compute upper bounds on willingness to pay for each household using,

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

This characteristics bundle is the maximally organic bundle; all products are transformed into organic products. No households actually purchase this bundle. We can also compute various "feasible" organic bundles, which consists of those products for which an equivalent to a non-organic product is available (i.e. all characteristics match except the organic one.).

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

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

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

For each household that purchases any organic items, equation (3) provides a lower bound on willingness to pay for the bundle of organic items actually purchased. For all households, equation (4) provides an upper bound for willingness to pay for organic for all non-organic items purchased. These lower and upper bounds are not strictly comparable; they are not bounds on the same goods. The former is a lower bound for willingness to pay for organic for the fraction of the basket that is organic. The latter is an upper bound for willingness to pay for the fraction of the basket that is non-organic.

This fact reveals a basic limitation of the data and of our minimal set of assumptions. Without more information or more structure, it is not possible to estimate an upper bound for the organic fraction of the basket. Nor is it possible to estimate a lower bound for the non-organic fraction of the basket. To estimate these missing bounds, we would need to add minimal assumptions on preference shocks and store switching behaviour and use the panel aspect of the data and cross-store price variation to identify willingness to pay.⁴ What we do in this paper is consider what we can learn without this additional structure. We use a minimal set of assumptions and see how well results accord with prior beliefs about reasonable values for willingness-to-pay. As one can see in the tables at the end of the paper

⁴In ongoing work, we are estimating a discrete-continuous demand model that imposes further structure on utility, exploits the panel nature of the dataset and exploits household level price variation across transactions induced by random shocks to the store choice process. Exploiting the repeated observations in the panel data is more complicated than in a simple discrete choice framework, for example, because the dimension of the vector \bar{q}_{ht} is very large.

and in section 4, an enormous amount of information about willingness to pay is revealed from this simple but thorough analysis of this sort of high quality disaggregate data that is now commonly available.

2.3 Assumptions

As we emphasised in the introduction, our objective is to consider what can be learned about willingness to pay while making minimal assumptions. What have we had to assume to get this far? And how does that compare with other methods in the literature?

The approach described above to identify a lower and upper bound on willingness to pay requires the following:

1. Utility depends on observable characteristics measured in our data, and possibly on unobserved characteristics that are mean independent of the observed characteristics.
2. Households maximise utility, have complete knowledge of the market environment and incur no search frictions.
3. The set of marketed products is rich enough to allow households to effectively bundle product characteristics.

All of these assumptions are common throughout the hedonic literature and the differentiated product demand literature. Much of this literature imposes additional restrictions on household preferences. Blow, Browning, and Crawford (2006) assume that preferences for milk are separable, that milk characteristics enter household utility through a known linear technology, and that there are no time varying preference shocks. The discrete choice demand literature in retail settings (see for example, Berry, Levinsohn and Pakes (1995), Bajari and Benkard (2005) and Smith (2004)) assumes that preferences have an index structure, mostly ignores quantity choices (excluding Smith (2004)), and usually assumes that households have vectors of preference shocks affecting utility for different options in their choice sets. These shocks are usually assumed to be i.i.d. across time. We view the set of

assumptions in this paper as the minimal set of assumptions that is consistent with previous literature.

Some of our assumptions could be relaxed (at least to some extent) but at a cost. If the mean independence assumption is invalid, one could in principle add product fixed effects or impose a factor structure on the error terms in the hedonic regressions. Adding search frictions to the analysis would add an element of complexity to the interpretation of the results. First of all, search frictions could explain why households switch between organic and non-organic products. Different draws from the price and product distribution would correspond to unobserved characteristics in the hedonic price equation and perhaps to time varying unobserved preference shocks. A model that addressed these issues would have to allow for correlation between these unobservables and both product and household observables. Finally, the product unbundling assumption could be relaxed by comparing actual purchases to substitute purchases that exist in the market and by re-defining counterfactual assumptions. We discuss this last point further in section 4.3 below.

2.4 Household heterogeneity

In addition to computing lower bounds on willingness to pay for each household, we can analyse how these lower bounds vary with demographic characteristics and survey responses to attitudinal questions that capture some aspects of the main shopper’s preferences and beliefs. Our estimated lower bounds are

$$w_h^n = (\bar{p} - \bar{p}^n) \cdot \bar{q}(x_h, \bar{\eta}_h)$$

where we have written $\bar{q}(x_h, \bar{\eta}_h)$ to emphasise that household demand depends on observable household characteristics x_h and unobservable household characteristics $\bar{\eta}_h$. In our data, observable household characteristics include total expenditure, household structure, social class, age, region, and observable measures of preference and belief heterogeneity such as survey responses to questions about attitudes towards organic products, toward the environment and toward health and product quality. We discuss these data in more detail in

section 4.

Without imposing more structure and using the panel nature of the data, we cannot estimate the demand functions \bar{q} . In particular, we expect that quantities demanded will depend on prices and that prices will depend on unobserved heterogeneity. However, we can estimate a reduced form willingness to pay regression of the form

$$\ln w_h^n = d(x_h) + \eta_h,$$

where x_h is a vector of household characteristics as described above and η_h is a scalar error term with $E(\eta_h | x_h) = 0$. These regressions allow us to study how important each observable household characteristic is in explaining variation in household level lower bounds on willingness to pay for organic products.

3 Market environment

The set of products marketed and the price premia for individual items are determined in the market equilibrium of the groceries market. This is a market in which a small number of firms sell a large number of differentiated products to a large number of households. Thus, the set of products and the price premia are determined in an oligopoly marketing and pricing game. The observed market outcomes reflect the technological feasibility of producing and selling various products, the costs of different firms and different stores, the number and types of firms in the market, and the distribution of demand across locations and products.

In other words, the expected market equilibrium prices depend on costs, variation in elasticities of demand, and competitive conditions in the local grocery market. For example, if the marginal production cost differences between organic and non-organic beef are larger than those between organic and non-organic chicken then, everything else equal, we expect the organic price premium to be higher for beef than for chicken. On the other hand, it might be the case that organic beef demand is more price elastic than organic chicken demand, or that entry into organic beef production is less elastic to profits. Each of these factors plays a role in determining the hedonic price of meat in the grocery sector.

In this environment, for each food category g , we estimate a hedonic price function of the form

$$p = h_g(z, \varepsilon)$$

where we assume that h_g is the true hedonic price function for food category g , z is a vector of measured characteristics, ε is an unmeasured characteristic with $E(\varepsilon | z) = 0$ and $h_g : \mathbf{R}_{n_g} \rightarrow \mathbf{R}$. In theory, we could estimate each h_g non-parametrically. However, despite the very large sample sizes we have, the large dimensionality of z makes this impractical. Note, however, that all elements of z are discrete. Instead, for each product category we estimate a log-linear model of the form

$$\ln(p_{bst}) = \alpha_t + \beta z_{bs} + \varepsilon_{bst}$$

where α_t is a time dummy, and where ε_{bst} includes an indicator of whether the product is on sale and a common regional component of prices.⁵ For any particular food category, the predicted organic price premium of a specific product (b, s) is given by

$$\Delta p_{bst}^n = \exp\left(\widehat{\alpha}_t + \widehat{\beta} z_{bs} + \widehat{\varepsilon}_{bst}\right) - \exp\left(\widehat{\alpha}_t + \widehat{\beta} z_{bs}^n + \widehat{\varepsilon}_{bst}\right).$$

One limitation of the data is that p_{bst} is only recorded if item b is purchased in store s at time t . Otherwise, it is not observed. Let $d_{bst} = 1$ indicate that we observe at least one occurrence of the price. We assume that

$$E(\varepsilon_{bst} | s, d_{bst} = 1) = E(\varepsilon_{bst} | s, d_{bst} = 0) = 0.$$

That is, we assume that the distribution of the unobserved characteristics is no different amongst the items that are not purchased. This is a weak assumption for several reasons. First, the weighted TNS sample is nationally representative of both all consumers and all expenditure items. By construction, the sample is meant to have the desired property. Second, the sample size is extremely large and very high frequency. If an item is ever purchased, the probability it appears in the sample is high. The sample includes every item

⁵The achieved R^2 are all very close to one. In ongoing work, we are estimating more flexible specifications.

purchased on a daily basis by a representative sample of 17,162 individuals over a the period of a year. Third, items that are never or rarely purchased are items for which a very small fraction of the market has willingness to pay larger than the price. On average these items could have high prices, low willingness to pay or both. In our application, there is no reason to expect that pricing amongst these items is systematically different from pricing amongst sampled items. Erickson and Pakes (2007) find that a similar assumption does not hold in monthly data for the television market. However, in contrast to their study, our data sample has a much higher frequency (daily), focuses on a very different market (groceries), and is a much larger sample of individual transactions. In the empirical section, we discuss the extent of potential bias arising from violation of this assumption.

For each food category $g \in G$, the estimated hedonic price function \hat{h}_g is an estimate of the market opportunity set facing households. For existing products, each provides a consistent estimate of the market equilibrium price. For products that do not exist but that are technologically feasible however, it provides at most a close approximation to equilibrium prices. For products that do not exist, equilibrium prices are bounded below by the constraint that no households wants to buy them and bounded above by the constraint that no firms want to sell them. For these goods, it may not be appropriate to extrapolate from prices of existing goods. Instead, it may be more appropriate to substitute an existing good that is nearby in characteristics space. We investigate this issue empirically in section 4.3.

It has long been recognised that interpretation of the error terms in hedonic regressions plays an important role in hedonic analysis (see Griliches (1961), Epple (1989), Pakes (2003), Triplett (2004), Bajari and Benkard (2005), and Erickson and Pakes (2007).). In the literature, three main sources for this error term have been considered (see Triplett (2004) for a detailed discussion.). First, it could be measurement error. In this case, if the measurement error is mean independent of the observed characteristics, then the estimated hedonic price function is consistent and the counterfactual hedonic prices can be predicted using the hedonic price function and ignoring the error term. Second, it could reflect unmeasured

product characteristics that are observed by buyers and sellers and hence that affect market prices. In this case, if the unmeasured product characteristics are mean independent of the observed characteristics we can still estimate the hedonic price function consistently but we must use care in predicting counterfactual prices. Counterfactual prices rely an assumption about what value of unmeasured characteristics is assumed for the counterfactual goods. Transforming a good from organic to non-organic holding everything else constant requires holding the unobservable constant as well. Alternative counterfactuals can be generated under different assumptions about the level of unobserved characteristics that is forecast for the counterfactual good. Finally, the error term in the regression could reflect pricing errors. In this case, the analysis is similar to that in the unobserved characteristics case. Alternative counterfactuals can be generated under different assumptions about the level of the “pricing error” for the counterfactual product. In our analysis, we include the error term because we believe that in our data measurement error is relatively minimal, while the other two considerations may be more important.

4 Data and empirical results

The data come from the TNS Homescan panel and cover the period 8 November 2003 to 7 November 2004. Households record purchases of all items 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 all food items purchased for home consumption by 17,162 households. The sample contains data on more than sixteen million purchases. The characteristics data includes information on a large number of product characteristics judged to be important by market researchers as well as store identity and location. We discuss these in more detail in section 4.2. Demographic information and information on a range of consumer attitudes (including attitudes towards health, quality, the environment and organic) is updated annually by a telephone survey. Further details on how the data are collected are given in the Data Appendix.

4.1 Organic market shares

We consider all food products that are purchased by households in the UK and brought into the home. Individual food products (defined by a unique bar code) are grouped into categories such as “fresh lamb”, “tea”, “olives”.⁶ In total there are 146 such groups. We restrict our analysis to those food categories where more than one food product is available as organic and where we observe more than ten purchases of these goods over the year - this includes 98 food products.

Total expenditure on all 146 food categories in our sample of households is £22.5m, while spending on the food categories we focus on is £21m. The 98 food categories we consider are shown in Table 1, along with total expenditure, the share on organic, the number of purchases, and the share of purchases that are organic. The first column of Table 1 shows annual expenditure grossed up to the national level using household sampling weights. The bold rows summarises these data under 19 broad categories. On average 1.4% of expenditure is on products that have the organic characteristic. This ranges from fish, which is infrequently organic, to fruit and vegetables, for which 2.7% of expenditure is on organic products. There is a large amount of additional interesting information in Table 1 which we leave the reader to consider on their own.

Table 4 breaks these market shares down to the household level. Around 18% of households never buy any organic products, and over one-third buy only a very small amount (less than one-quarter of one percent of their total expenditure). However, 26% of households spend more than 1% of their budget on organic products, and there are a small number of households (1.67%) that spend over 10% 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 part of the population.

⁶We use the categories used by the market researchers.

4.2 Organic price premia

Our interest in this paper is to estimate bounds on household willingness to pay for organic and to investigate how these bounds vary with demographic and attitudinal variables. To accomplish the first task we compare the price of organic products with the price of non-organic products in the same category, conditioning on a large range of other characteristics. This price difference provides a bound on willingness to pay for the discrete choice described in section 2.1. As an example, Table 2 displays results for the bacon rashers food category. The first column displays results from a univariate regression showing that on average organic bacon rashers are 85% more expensive than non-organic. If the organic characteristic cannot be unbundled from all other characteristics, then this measure provides a bound on willingness to pay for organic (and all other characteristics bundled together with organic). Note that the adjusted R^2 in this regression is 0.01. The organic characteristic explains only 1% of the variation in the log price of bacon rashers. The average organic/non-organic price difference is measured precisely with a standard error of 0.114. In this regression each observation is a transaction (an occasion when a household bought a product from a shop). There are 152,158 such transactions in our data for purchases of bacon rashers. The standard errors are clustered at the barcode level as prices of a specific product may be correlated across stores and time periods.

The single variable regression is very poor at predicting the price of bacon rashers because it omits many other characteristics that are important determinants of the price. Moreover, as we show in section 4.3.1, it appears that non-organic alternatives to organic bacon, identical in every dimension save organic, are available for nearly all bacon products. In such circumstances, a multivariate regression of bacon prices on characteristics is required to provide a bound on willingness to pay for organic. The second column of Table 2 reports results controlling for a large number of other important characteristics of bacon rashers - the package size, whether the product is branded, own brand or private label, whether the bacon is back or streaky, where it originates from, whether it is packaged or loose, whether it was

purchased on sale, what type of store it was purchased in, which month it was bought it, and where in the country it was purchased. Controlling for these other characteristics reduces the coefficient on organic to 0.48 and increases the adjusted R^2 to 0.98. After accounting for other characteristics, organic bacon rashers are 48% more expensive than non-organic. The multivariate hedonic regression explains 98% of the variation in bacon rasher prices. As a group, the characteristics included are extremely good predictors of the price of bacon rashers. Using the estimated markup from column (2) of Table 2, equation (3) and data on quantities we can calculate the lower bound on the willingness to pay for organic bacon rashers.

We repeat this analysis for each of the 98 food categories in the data by running 98 separate regressions of the log of the unit price on a large number of characteristics of the products. Each regression includes a set of characteristics that is common to all categories (organic, branded, fascia effects, package size, time and region effects) as well as a set of category specific characteristics. The characteristics included as regressors in each regression are summarised in Table 3. There is interesting variation in the number and type of characteristics that differentiate products in different food categories, for example, lamb appears fairly homogenous, with four origins, two package types and two varieties, whereas yoghurt (with 259 flavours) and soups (with 325 flavours) are very differentiated. Note that all of these regressors are discrete valued variables. The organic coefficients and their standard error along with the adjusted R^2 's from each regression are also shown in Table 3. Each row in the table shows results from a separate regression.

The first notable feature of Table 3 is that the adjusted R^2 are very high. Across the 98 food groups the adjusted R^2 is above 90% in 71 cases and above 80% in all but 8. The observed characteristics in this dataset do a very good job of explaining the price variation in these markets.

Another notable feature of the results is that the estimated organic price premium is almost always positive and varies significantly across product categories. Of the 98 coefficients,

72 are significantly different from zero at the 5% level. The mean of the price premia is 0.33 and the median is 0.28. The coefficients range from -0.09 (four coefficients are negative, but none of these is significantly different from zero) to 1.44.

How much difference does it make to control for other observable characteristics? If we had simply looked at the mean differences in price between organic and non-organic goods we would substantially over-estimate the mean premium at 0.56 and the median at 0.54. Figure 1 shows the distribution of the estimated coefficients across product categories when we control only for organic (the lower density further to the right) and controlling for all product characteristics (the higher density further to the left). As can be seen in the figure, controlling for all characteristics shifts the distribution to the left and reduces the cross-category variation.

Despite the reduction in variance due to including more controls there still is significant cross category variation in the price premium for organic products. An open question is how much of this variation is explained by variations in costs of production and how much by variation in market structure and pricing and variations in willingness to pay. On the cost side, we expect organic price variation due to variation in costs of organic production across categories. On the demand side, we might expect variation in the organic price premium due to variations in the relationship between organic and things that households actually value such as quality, health and the environment. For example, it might be that organic is a stronger indicator of quality for products with high price premia than for products with low price premia. As a result, demand for these products might be less price elastic resulting in higher price premia.

As discussed in section 2.3, the organic price premia are bounds on willingness to pay for individual organic items under the assumption that the organic characteristic can be bundled (and unbundled) from other characteristics. Under the same assumption, these item specific bounds can be combined with data on quantities purchased to estimate household specific bounds on willingness to pay for baskets of organic products and to investigate how these

vary across households. We study these bounds and their variation in sections 4.4 and 4.6.1. First, however we investigate empirically whether the bundling assumption is reasonable.

4.3 Bundling

Our analysis assumes that for every organic product, a non-organic alternative identical in every dimension other than organic can be purchased on the market, and that it is technologically feasible to convert every non-organic product into an organic product with all other characteristics held constant. These assumptions allow us to construct predictions of prices for goods that are not observed and to bound willingness to pay. We consider two empirical questions:

1. Does the data support the assumption that for every organic product, we can find a non-organic product identical in every other (measured) dimension?
2. Is it technologically feasible to produce organic versions of all existing non-organic products? In particular, for every non-organic product, can we find an organic product identical in every (measured) dimension other than organic?

4.3.1 Non-organic alternatives to existing organic products

To look at the first question we consider all organic product in our data and calculate the minimum distance in characteristics space to each non-organic product, ignoring differences in time and size (we assume that organic is not bundled together with time and size characteristics).

Formally, let B_g^o and B_g^n be the sets of organic and non-organic products in category g . For every pair $(i, j) \in B_g^o \times B_g^n$, we define the distance between them as

$$d_1(i, j) = \sum_{k \in C_g^1} \mathbf{1}(z_i(k) \neq z_j(k))$$

where C_g^1 is the set of characteristics of products in category g excluding organic, time and

size. For each organic product $o \in B_g^o$, we compute

$$d_o = \min_{\{n \in B_g^n\}} \{d_1(o, n)\}.$$

Table 7 details the results. In the first two columns we include fascia in the set of characteristics, while in the second two columns we also exclude fascia from the characteristics set. There are 3383 organic products in our data, a number of which are sold across several fascia, giving us 7323 product-fascia combinations. From the table we can see that 81% of the organic products in the sample have at least one identical non-organic alternative that is sold within the same fascia. Most of the other organic products have a non-organic alternative that differs in only one dimension, with around 0.6% having a closest alternative that differ in two dimensions (with one differing on three dimensions). When we consider all products in the market we see that 87% of organic products have a non-organic alternative that is the same on all characteristics, and only 14 products differ in more than one characteristic. Clearly, the data support the assumption that the organic characteristic can be unbundled from other characteristics for almost all products. The largest number of organic products for which there is not a perfect non-organic substitute are fruit and vegetables, where for example organic varieties of a specific type of fruit might come from one country while non-organic varieties come from a different location, so the products will differ in the origin characteristic.⁷

4.3.2 Feasibility of converting non-organic products to organic

In a similar spirit, we consider whether the products in our sample provides evidence about the availability of organic alternatives to existing non-organic products. For every pair $(i, j) \in B_g^o \times B_g^n$, we define the distance

$$d_2(i, j) = \sum_{k \in C_g^2} \mathbf{1}(z_i(k) \neq z_j(k))$$

⁷In cases where no non-organic option is available, we can estimate an alternative price expenditure premia by comparing each organic item to a non-organic item that is closest in characteristics space and price. This method is similar in spirit to the replacement methods that statistical authorities use when sampling prices for price indexes when they cannot find a product whose price has previously been sampled.

where C_g^2 is the set of characteristics excluding organic, time, and size, and were we again consider the products sold within the same fascia and in the entire market. We consider the latter to tell us about the technological feasibility of producing organic alternatives, while the former also reflects the choice of the store over whether to carry that range.

The results are shown in Table 8. There are 87,370 non-organic products in our sample, a number of which are sold across multiple fascia, giving us 213,885 product-fascia combinations. For 41% of these products we observe at least one organic product sold in the market that is the same in all other characteristics, whereas only 29% of these are sold within the same fascia. There are a large number of non-organic products where there is no exact organic equivalent, but in most cases there are products that are close, i.e. that differ only by a few characteristics. The single product that differs on 7 characteristics is an own-label soft cheese that is sold in KwikSave, a store that sells relatively few organic products. In the case where we look across the whole market, the two products that differ on six characteristics are both tinned fruits - one is pineapple chunks and the other mandarin orange slices.

4.4 Bounds on willingness to pay

We now turn to our estimate of the lower bounds on individual households' willingness to pay for organic foods, as given by equation (3). Tables 5 and 6 and Figure 2 show the distribution of this variable. Table 5 shows the distribution of the lower bound measured in 2004 pounds sterling. Just under 20% of households either bought no organic products, or bought only a small amount whose price premium was below zero, revealing that their willingness pay for organic may very well have been zero or negative. The remaining 80% of households were willing to pay at least some positive amount for organic goods. Around 30% were willing to pay less than a pound a year, while over 50% were willing to pay more than £1 a year, around 8% were willing to pay more than £10 and 98 households were willing to pay more than £100 a year. As fractions of annual expenditure, these numbers are quite small. That is because expenditure on organic goods represents only 1.4% of all expenditure, and the expenditure

premium on organic is by construction only a fraction of the organic share. To see how important the expenditure premium is relative to the expenditure share, Table 6 expresses these numbers as a share of households' annual expenditure on organic foods. From these numbers, we can see that the expenditure premia represent relatively large fractions of total organic expenditure; 73% of households have a willingness to pay that is more than 10% of the organic food expenditure. Figure 2 shows the density of the distribution of the lower bound on willingness to pay (excluding the zeros and the top percentile). Most households are willing to spend very little on organic foods, but a small fraction are willing to spend a relatively large amount.

We also compute estimates of upper bounds on household's willingness to pay for organic based on equation (4). Figure 3 show the distribution of this. These upper bounds are very large and represent an extreme case - what is the most that a household would be willing to pay to convert all the goods purchased to organic.

4.5 Aggregate willingness to pay

What do these individual bounds on willingness to pay suggest about the aggregate willingness to pay? We report all figures grossed up to UK population levels using household demographic weights, based on the Annual Census. The aggregate lower bound is £121 million. The total expenditure that we observe in our sample is £37 billion, of which £535 million is on organic products. The lower bound thus represents 22% of annual expenditure on organic products. If all organic products were removed from the market, the loss to the economy would be at least £121 million. This number assumes that the organic characteristic can be unbundled for all organic products. If we focus only on organic products that have an identical non-organic alternative already in the market (total expenditure on organic products in this group is £478 million) we find that the aggregate estimated lower bound is £109 million. The error introduced by ignoring bundling issues in this case is not very large, about 10% of the total.

Our estimate of the aggregate upper bound in our sample, again grossed up, is £11,757 million or roughly 32% of grocery expenditure on non-organic products. This is an upper bound on the gain to the economy that would arise by requiring all products to switch to organic production. This number is extreme. However, it represents an upper bound for a very extreme counterfactual - changing the economy from one in which 1.4% of food expenditure is on organic products to one in which 100% is organic.

We consider two more restrictive counterfactuals for an upper bound based on "feasible" organic baskets. The first one is where we consider the counterfactual of converting only those non-organic products where an organic alternative is already available in the market (from any supplier). Total grossed up expenditure on these products was £17,727 million, and the estimated upper bound on willingness to pay for organic in this group was £5,174 million or 29% of all grocery expenditure. Our second feasible basket is to consider only non-organic products that are also currently available in the organic form in the same supermarket fascia (i.e. from the same supplier) that the household purchased the non-organic product from. Total grossed up expenditure on these products is £11,820 million, and the estimated upper bound on willingness to pay for organic on these products is £3,207 million or 27% of total expenditure.

It is worth emphasising that, in general, our lower bound and our upper bound are not strictly comparable, because they are literally bounds on willingness to pay for organic apples and oranges. However, there are some special cases where they are comparable. Suppose indirect utility is separable and takes the form

$$v(x_h, g_1(z_1, p_1), \dots, g_B(z_B, p_B)).$$

where (p_{bs}, z_{bs}) is the price and characteristics vector for good (b, s) . Then, for each good, the tradeoff between characteristics and price is independent of all other goods. Moreover, if the functions g_{bs} for all (b, s) are identical, then the tradeoff is the same for all (b, s) . Under these two conditions, household x_h will have a unique willingness to pay for organic - a willingness to pay that is the same for all goods. For all goods this consumer buys, the

consumer will buy organic if the price premium is less than the willingness to pay. Under these conditions, every organic purchase decision is independent and identical; there is a single threshold. Unfortunately, these assumptions are immediately rejected by the data. There is not a single threshold that explains organic purchase decisions. Nevertheless, if one assumed that the separability plus unique willingness to pay assumptions were approximately true, then we could argue that range from 22% to 32% represents an estimate of the lower and upper bounds on willingness to pay for organic.

4.6 Analysis

In this section we analyse two examples of how these bounds can be informative. We first consider how the lower bound in willingness to pay varies with household characteristics, beliefs and attitudes. Secondly, we evaluate the implications of our bounds for potential revenue that a store could earn from introducing a new organic product line.

4.6.1 Variation in willingness to pay

Why are households willing to pay for organic food? We combine our estimates with demographic information and survey response data on attitudes towards health, the environment and product quality as described in section 2.4 to shed light on this question. We consider the variation in our estimated lower bound on willingness to pay for organic foods for households that report different preferences and beliefs, controlling for a number of demographic characteristics. We exploit qualitative survey data that are collected by TNS and consider three main factors that have been highlighted in the literature as being reasons why people value organic, and on which we have data - benefits to the environment, health benefits, and better quality food.

In the survey households are asked to indicate the degree to which they agree with each of the following statements:

1. Organic products are healthier

2. I try to buy a healthy range of foods these days
3. Organic foods are friendlier to the environment
4. I try to buy environmentally friendly products
5. Organic foods are better quality
6. I don't mind paying for quality

For each statement, respondents are asked to choose one response from the list

{Agree strongly, Agree, Neither agree nor disagree, Disagree, Strongly disagree}.

We treat these responses as qualitative measures of household preferences and beliefs and investigate the statistical relationships between the responses and the lower bounds to household willingness to pay. For each of the three factors (environment, health, quality), we have one response that provides a qualitative measure of beliefs and a second that provides a qualitative measure of preferences.

We first report cross-tabulations of responses to these survey questions, household organic expenditure shares and the lower bound on their willingness to pay for organic. In order to reduce the dimensionality of our tables, we report results that pool the five possible survey responses into two groups, (Agree strongly, Agree) and (Neither agree nor disagree, Disagree, Strongly disagree). Tables 9-11 show these cross-tabulations.

Table 9 shows that 2196 households say both that they try to buy healthy foods and that they think organic products are healthier. For these households organic products make up 3.4% of total expenditure, and these household have a mean lower bound on their willingness to pay that represents 0.8% of their total food expenditure. Agreement with both statements is correlated with high expenditure shares. In contrast, 7503 households do not particularly try to buy healthy foods and do not think organic foods are healthier. Disagreement with both statements is negatively correlated with organic expenditure shares; organic

products make up 0.9% of total expenditure in this group and their estimated lower bound on willingness to pay for organic is 0.3% of total expenditure on food. Tables 10 and 11 display similar figures for the questions related to the environment and quality. A total of 1957 households both feel that organic products are good for the environment and try to buy environmentally friendly products. These households spend 2.7% of their budget on organic foods and are willing to pay at least 0.6% of total expenditure. A smaller number of households, 1345, believe organic products are higher quality and do not mind paying for quality; these households spend a larger share (4.3%) on organic products and have a higher estimated lower bound on willingness to pay of 1% of total food expenditure.

We use these survey responses to try and disentangle what motivates households to purchase organic foods by looking at how the lower bounds on the willingness to pay vary with households' responses to these questions. We regress the households' lower bounds on willingness to pay on the attitudinal responses discussed above and a range of other household characteristics including: family structure, total annual expenditure on food and non-food items (as a proxy for income), the household's social class, where A (upper middle class - higher managerial, administrative or professional), B (middle class - intermediate managerial, administrative or professional) C1 (lower middle class - supervisory or clerical, junior managerial, administrative or professional) or C2 (skilled working class - skilled manual workers) (the omitted category is D (working class - semi and unskilled manual workers) and E (those at lowest level of subsistence - state pensioners or widows (no other earner), casual or lowest grade workers)), and whether anyone in the household is a vegetarian. The means of these variables are shown in Table 13 (most are discrete variables).

In Table 12 we investigate how the lower bound on willingness to pay varies with household preferences and attitudes. In column (1) we include dummies that are equal to one if the household is in the upper left-hand quadrant of Tables 9, 10 or 11 respectively, as well as an indicator of whether the household's responses to these questions were missing. Households that care a lot about organic and health have on average a lower bound on willingness

to pay that is £6.00 higher than households that do not (i.e. are in any of the other three quadrants of Table 9). Households that care a lot about organic and the environment have a lower bound that is £1.61 higher and those that care about quality have a lower bound that is £9.83 higher. Responses to each of the attitude questions are positively correlated with both organic market shares and lower bounds on willingness to pay for organic. Moreover, the average lower bound is highest amongst the group for whom quality is important, next highest amongst the "health is important" group, and next highest amongst the "environment is important" group.

In column (2) of Table 11, we also include indicators of households that are in the upper right and lower left quadrants, with little effect. In column (3) we include the other demographic controls, again with little effect. In columns (4)-(6) we repeat this specification for each of the main social classes separately, and see broadly similar patterns. The estimates in Table 12 allow us to calculate the contribution to willingness to pay lower bounds of each of the three concerns - health, environment and quality. In all cases we see that valuing quality is the characteristic that is associated with households that have the highest lower bound on their willingness to pay, followed by health with environment contributing the least. If we want to know the aggregate lower bound on the valuation of these three concerns then we need to consider not only the mean lower bound on the valuation for those who have the preferences and attitudes that the particular issue is important, but also the number of households that fall into that group. Combining those two pieces of information we get estimates that suggest that the total lower bound on willingness to pay for health is around £17m, for the environment around £5m, and for quality around £20m. These results are interesting and may be surprising to some people. Quality and health seem to be much more important factors in determining the amount (or at least the lower bound) households are willing to pay for organic products. This has implications for the regulation of organic labelling, and for the way that firms may want to advertise organic products.

4.6.2 Introduction of a new organic product line

The upper bound on willingness to pay can also be informative. Suppose a firm were considering whether to convert some set of products to organic. What would be the potential impact? If we are willing to make specific assumptions about preferences, the form of utility and costs then we can obtain precise estimates about the impact on profits. But what if we are concerned about the validity of these assumptions? Our results allow us to estimate an upper bound on the potential revenue implications of such a move without imposing strong assumptions. This bound is the firm specific component of our previously computed upper bounds. To illustrate, we compute upper bounds on revenue gains from converting two specific food categories to organic - eggs and fruit. In a similar way, we could calculate results for any other food categories.

In considering such a strategy, basically three factors drive differences in projected revenue across stores - the baseline expenditure on the category in each store, the current proportion that is organic, and the price premium on organic. We can easily calculate the first two from our data (and they are figures that a store would readily know). To get the third we need to have the hedonic regressions. We re-estimate the hedonic regressions allowing the organic price premium to vary across the major stores. We include a separate organic price premium for Asda, Safeway, Sainsbury, Tesco, Marks & Spencer, Waitrose, the large discounters (Aldi, KwikSave, Lidl, Netto and Iceland), other chains and Small stores.

We find significant variation in the price premia charged by different stores. For eggs, the premia range from 29% to over 49%. Similarly, the organic share of the eggs category varies substantially from less than 1% to over 20% . There is also a large variance in the total expenditure on eggs. Putting these together, for each store we estimate an upper bound on the revenue increase from converting all eggs to organic. These estimates range from an increase in revenue of around 30% for Large Discounters to a maximum increase of just over 60% for Sainsbury's and Waitrose.

Looking at fruit we see an even wider variance in the price premia on organic across fascia

- ranging from zero to 46%. In contrast, the organic share of the fruit category displays less variation, ranging from less than 1% to over 7%. The projected maximum revenue gain from converting all fruit to organic ranges from zero to over 60% increase.

When considering product line changes, a supermarket could compare these maximum revenue projections to expected costs and begin to make first-order judgements as to which changes might be profitable. These could be used to rule out unprofitable changes and allow the supermarkets to focus more detailed analysis on categories that are potentially profitable. While these estimates do not provide point estimates on revenues or profits, they require very few assumptions about consumer preferences or behaviour and so are quite robust. Further work and either more data or more assumptions would then be needed to estimate more precise consumer substitution responses and to gauge rivals responses.

5 Summary and conclusions

Rich data on consumer spending behaviour are now widely available in a number of countries. These data offer great potential to learn about consumers' willingness to pay for many different characteristics. In this paper we have used rich data on households' purchases of food in the UK to estimate lower and upper bounds on willingness to pay for baskets of organic foods as an example. We have made very few assumptions in doing this. Our results suggest that there is enormous heterogeneity in willingness to pay for organic products and that in aggregate willingness to pay for organic foods in England in 2004 was at least £114m and that willingness to pay for a change to a 100% organic economy was less than £10.9bn.

It is interesting to compare our results to Blow, Browning, and Crawford (2008). In Denmark, the organic share of the milk market is much higher than in Britain (25.3% vs. 2.1%) while the price premium is comparable (20.4% on average in Denmark not controlling for other characteristics vs. 16.1% in Britain). Their approach illustrates how additional assumptions (separability, no time varying demand shocks) and panel data can be used to obtain point estimates of willingness to pay, at least for a fraction of the population.

The willingness to pay estimates they obtain have a median of about 18.2% of the average milk price. These estimates, while similar in magnitude to our estimates, are difficult to directly compare because they are reported only for a fraction of the population and rely on aggregation of purchases to the monthly level. It would be useful in future work to investigate the panel dimension of our data and obtain more precise estimates of structural demand parameters. This will require at the very least an explicit model of store choice and of the dynamics of household preference shocks. It will also require further assumptions to make the problem empirically tractable.

We have also shown that the reasons that households are willing to pay for organic differ, with quality being the most important, health concerns coming second, and environmental concerns lagging far behind. We have also shown how to usefully calculate an upper bound on the revenue implications of introducing a new product range. These are both applications that have direct practical relevance.

A Data appendix

The data come from the TNS Homescan panel (www.tnsorfres.com). We use data for the period 8 November 2003 to 7 November 2004. We observe expenditure for 17,162 households [give distribution of period of time each HH in sample] and more than sixteen million purchases. Households are able to leave the survey voluntarily at any time and are refreshed to keep the sample representative of the British population. Households are issued with electronic hand held scanner in their homes. They scan the barcodes of all grocery purchases (foodstuffs, alcohol, bathroom products, medicines, pet food etc.) that come into the house. Information on purchases is downloaded once a week by TNS. Prices are collected from till receipts, which households are required to send to TNS. In addition, information on loose weight items such as vegetables and fruit is collected by households scanning barcodes in a book and keying in the weight data. Purchases from all store types –supermarkets, corner stores, online, local speciality shops etc. –are in principle covered by the survey. For larger

stores, the exact store of purchase is recorded; for smaller stores the store type is known. The data includes information on the characteristics of the product –price, brand, pack size, whether the item was bought on promotion and crucially for this work information on a large number of characteristics of the product. Demographic information about the household is collected by means of an annually-updated telephone survey. Participants are rewarded with points redeemable for a range of products and services (though limited to items which shouldn't directly affect grocery consumption patterns).

References

- [1] Bajari, Patrick, and Benkard, C. Lanier (2005), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113(6): 1239-1276.
- [2] Akerberg, Daniel (2001) "Empirically Distinguishing Informative and Prestige Effects of Advertising" *RAND Journal of Economics*, Vol. 32, No. 2
- [3] Blow, Laura, Martin Browning and Ian Crawford (2006) "Revealed Preference Analysis of Characteristics Models", *Oxford Department of Economics Discussion Paper*.
- [4] Chay, K.Y., and Greenstone, M. (2005), "Does Air Quality Matter? Evidence from the Housing Market", *Journal of Political Economy*, 113(2): 376-424.
- [5] Chernozhukov, V., Hong, H., and Tamer, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5): 1243-1284.
- [6] Dube, Jean-Pierre (2005), "Product Differentiation and Mergers in the Carbonated Soft Drink Industry," *Journal of Economics and Management Strategy*, 14(4): 879-904.
- [7] Epple, Dennis (1987), "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 95(1): 59-80.
- [8] Epple, Dennis, and Sieg, Holger (1999), "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107(4): 645-681.
- [9] Erickson, Timothy, and Pakes, Ariel (2007), "An experimental component index for the CPI: from annual computer data to monthly data on other goods, *unpublished manuscript*.
- [10] Freeman, A.M. (1995), "Hedonic Pricing Methods," in *The Handbook of Environmental Economics*, ed. Daniel Bromley, Blackwell Publishers Ltd. Oxford, pages 672-686.

- [11] Griliches, Zvi (1961), "Hedonic price indexes for automobiles: an econometric analysis of quality change," in *The Price Statistics of the Federal Government, NBER Staff Report No. 3*, General Series No. 73, New York: 173-196.
- [12] Hausman, Jerry (2003), "Sources of Bias and Solutions to Bias in the Consumer Price Index," *Journal of Economic Perspectives*, 17(1): 23-44.
- [13] Heckman, James, and Scheinkman, Jose (1987), "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies*, 54(2): 243-255.
- [14] Kanemoto, Y. (1988), "Hedonic prices and the benefits of public projects," *Econometrica*, 56: 981-989.
- [15] Leeth, J.D., Ruser, J. (2003), "Compensating Wage Differentials for Fatal and Nonfatal Injury Risk by Gender and Race," *Journal of Risk and Uncertainty*, 27(3): 257-277.
- [16] Manski, C.F. (2003), *Partial Identification of Probability Distributions*, Springer.
- [17] Nevo, Aviv (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry" *Econometrica*, Vol 69, No. 2 (March 2001), 307-342
- [18] Pakes, Ariel (2003), "A Reconsideration of Hedonic Price Indexes with an Application to PC's," *The American Economic Review*, 93(5) 1578-1596.
- [19] Pesendorfer, Martin (2002) "Retail Sales: A Study of Pricing Behavior in Supermarkets" *Journal of Business*, Vol. 75, No. 1, January 2002
- [20] Pollak, Robert (1989), *The theory of the cost of living index*, Oxford University Press: New York.
- [21] Scotchmer, Susanne (1985), "Hedonic prices and cost/benefit analysis," *Journal of Economic Theory*, 37: 55-75.

- [22] Sattinger M. (1993), "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2): 831–880.
- [23] Scotchmer, Susanne (1986), "The short run and long run benefits of environmental improvement," *Journal of Public Economics*, 30: 61-81.
- [24] Sieg, H., Smith, VK, Banzhaf, S., Walsh, R. (2004), "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods," *International Economic Review*, 45(4): 1047-69.
- [25] Smith, Howard (2004), "Supermarket Choice and Supermarket Competition in Market Equilibrium," *Review of Economic Studies*, 71(1): 235–263.
- [26] Smith, V. Kerry, and Huang, Ju-Chin (1995), "Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models", *Journal of Political Economy*, 103(2): 209-27.
- [27] Triplett, Jack (2005), "Handbook on hedonic indexes and quality adjustments in price indexes: special application to information technology products," *OECD Science, Technology and Industry Working Papers*, 2004/9, OECD Publishing.
- [28] Varian, Hal (1982), "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50(4): 945–73.

Table 1: Expenditure by food category

	grossed up annual expenditure in £m	total observed expenditure in £	share of total expenditure on organic	total number of purchases
Alcohol	3987	2204941	0.5%	355197
Chilled Foods	3641	2047994	0.4%	1247368
Chilled Dips	58	33072	0.9%	31824
Chilled/Prepared Frt and Veg	327	179748	1.4%	150530
Cooked Meats	1203	688607	0.2%	492143
Defined Chilled Meat and Veg Prds	793	438503	0.7%	222179
Fresh P/P Meat and Veg Pastry Prds	499	285258	0.0%	191331
Fresh Pasta	64	34083	0.4%	22538
Fresh Processed Poultry	697	388724	0.1%	136823
Fish	455	243649	0.1%	89991
Fruit and Vegetables	4871	2692585	2.7%	3015369
Fruit	2206	1211694	2.3%	1078484
Vegetable	2665	1480892	3.0%	1936885
Meat	2629	1507259	0.9%	585853
Bacon Rashers	586	335438	0.4%	152158
Fresh Beef/Veal	985	560710	1.0%	182469
Fresh Lamb	268	151177	1.8%	35215
Fresh Pork	428	247365	0.7%	82655
Fresh Sausages	294	169143	0.6%	107576
Frozen Sausages	68	43427	0.4%	25780
Bakery Products	2533	1448038	0.7%	1601182
Christmas Puddings	25	14733	1.2%	5480
Bread	890	505176	1.8%	647145
Cakes and Pastries	899	518417	0.0%	452598
Morning Goods	719	409712	0.1%	495959
Biscuits	1345	784361	1.4%	843088
Canned Goods	828	484427	0.9%	805739
Baked Beans	170	101200	0.7%	157628
Canned Fruit	113	63600	0.9%	94628
Canned Milk Puddings	24	14036	1.0%	23231
Canned Pasta	88	55129	0.0%	92251
Other Canned Veg	39	22343	0.3%	49883
Prepared Peas And Beans	66	37834	1.9%	107133
Soups In Tins and Cartons	214	122862	1.2%	161741
Sweetcorn (Canned)	32	19536	0.2%	32379
Tinned Tomatoes, Passata	82	47889	1.7%	86865
Confectionery	1469	843773	2.0%	650719

Table 1 continued

Dairy Products	5294	3010922	2.2%	2445957
Butter	247	136996	1.6%	108535
Chilled Desserts (excl Formage Frais)				
Frais)	328	189676	0.5%	184359
Hens Eggs	361	202374	5.7%	192100
Margarine	364	208415	0.3%	184672
Desserts, Long Life	26	15484	2.2%	18517
Total Cheese (excl Formage Frais)	1387	792086	1.0%	504362
Cream, inc. Flavoured	141	78399	1.3%	98333
Total Milk	1623	930530	2.1%	730080
Yoghurt	647	362812	5.6%	380404
Yoghurt Drinks and Juices, wet	139	76948	1.6%	37110
Yoghurt Drinks and Juices, dry	31	17204	0.8%	7485
Frozen Foods	2579	1529941	0.5%	946448
Defined Processed Poultry Prods	198	124053	0.2%	60278
Frozen Confectionery	191	112158	0.1%	69393
Frozen Meat Products	740	437358	0.4%	245778
Frozen Pizzas	288	173711	0.1%	93273
Frozen Vegetables	436	263116	1.2%	216498
Other Frozen Foods excl. Pizza	256	150879	0.5%	100995
Ice Cream	470	268667	0.8%	160233
Hot Beverages	952	536615	2.2%	267346
Food Drinks	102	56811	1.4%	42024
Fruit And Herbal Teas	26	14301	3.8%	13197
Instant Coffee	423	238267	1.4%	92384
Coffee Beans	71	36566	13.0%	15868
Tea	330	190671	1.4%	103873
Breakfast Foods	1277	738977	1.6%	534046
Breakfast Cereals	949	552569	1.4%	339142
Chocolate Spread	11	7098	1.4%	7506
Honey	40	22177	4.1%	11585
Pastes and Spreads	145	83376	0.9%	92353
Peanut Butter	21	12616	1.4%	11855
Preserves	111	61141	3.0%	71605
Pickles and Sauces	965	551509	1.3%	541825
Mustard	15	8696	0.5%	11592
Olives	20	11000	0.6%	8850
Pickles	48	27424	0.6%	31151
Salad Accompaniments, wet	85	49771	0.4%	43521
Salad Accompaniments, dry	47	26444	1.8%	27749
Sauces and Ketchup (excl Worcester)	119	70963	1.5%	71261
Sauces For Cooking and Condiment, wet	546	308462	1.5%	294987
Sauces For Cooking and Condiment, dry	17	9706	2.3%	10564
Sour and Speciality Pickles	68	39044	0.3%	42150

Table 1 continued

Take Home Savouries	948	561387	1.0%	474400
Crisps	429	254599	0.2%	212194
Nuts	136	74801	1.2%	58590
Savoury Snacks and Reconstitutes	383	231987	1.9%	203616
Carbohydrate and Savoury Snacks	369	211997	1.1%	267109
Rice, Pasta, Cous Cous	211	122183	0.7%	134178
Dried Soup	65	36459	0.1%	50041
Dry Meat Substitutes	2	1347	9.0%	1227
Dry Pasta	91	52007	2.6%	81663
Savoury Home Cooking	600	342471	1.4%	345889
Defined Mixes and Breadcrumbs	40	23693	0.2%	25694
Ethnic Ingredients	80	45040	0.1%	33484
Flour	52	28677	4.9%	46345
Meat Extracts	166	96836	1.2%	96730
Packet Stuffing	21	12779	0.3%	19931
Pepper Herbs and Spices	70	37986	0.6%	37987
Salt	16	9523	0.6%	14954
Vinegar	22	12437	3.6%	18667
White Fats And Oils	133	75500	1.7%	52097
Soft Drinks	1655	961825	2.2%	730395
Bottled Non-Lemonade	244	143069	6.1%	131515
Canned Colas	171	97233	0.0%	34986
Canned Non-Lemonade	91	52923	1.0%	30739
Fruit Juices	590	326195	2.8%	266801
One Shot Drinks	241	147864	1.5%	110848
Fruit Squash	318	194541	0.4%	155506
Sweet Home Cooking	450	253372	1.1%	269682
Home Baking	163	88155	2.1%	74980
Lemon And Lime Juices	7	4336	1.2%	7045
Mincemeat (Sweet)	6	3631	0.4%	2981
Powdered Desserts and Custard	82	47803	0.2%	63729
Sugar	183	104139	0.5%	114878
Syrup and Treacle	9	5308	3.1%	6069
Total	36847	20,956,043	1.4%	16017603

Note: Data include 17,162 households over the period 8 November 2003 to 7 November 2004. A purchase is a household-store-day transaction (if a household buys two of the exact same product in one day at the same store this is one transaction, if they buy the same product at a different store or a different product at the same store that counts as a separate transaction).

Table B.1 shows the breakdown by category.

Table 2: Hedonic regressions for bacon rashers

Dep var: ln(price per kg)	(1)	(2)
Organic	0.0847 [0.114]	0.483 [0.076]
size quintile 2		-0.177 [0.033]
size quintile 3		-0.194 [0.035]
size quintile 4		-0.233 [0.048]
size quintile 5		-0.393 [0.036]
Budget Private Label		-0.431 [0.054]
Standard Private Label		0.094 [0.031]
Streaky		-0.298 [0.037]
origin Ireland		0.151 [0.104]
origin Danish		-0.061 [0.033]
origin Dutch		-0.168 [0.031]
origin Other foreign		-0.152 [0.042]
packaging: Prepacked		0.056 [0.039]
Ticket price reduction		-0.163 [0.022]
Multi-purchase deal		-0.263 [0.034]
Extra free deal		-0.031 [0.056]
Asda		1.989 [0.047]
Large Discount Store		1.977 [0.063]
MandS		2.577 [0.059]
Waitrose		2.415 [0.051]
Other Chain Store		2.165 [0.050]
Safeway		2.027 [0.043]
Sainsbury		2.246 [0.061]
Tesco		2.084 [0.042]
Small store		2.019 [0.043]
Time and region effects	no	yes
R-squared	0.01	0.98

Note: Regression include 152,158 observations. Standard errors in [] are clustered at the barcode level and allow for general correlation..

Table 3: Organic coefficient, R² and characteristics for all food categories

Food category	organic coefficient	standard error	Adjusted R2	characteristics							
				brand	size	fat	flavour	origin	package	variety	obs
1 Alcohol											
Alcohol	0.138	0.048	0.95	3	5					43	355197
2 Chilled Foods											
Chilled Dips	0.356	0.047	0.98	3	5		89		3		31824
Chilled/Prepared Frt and Veg	0.352	0.098	0.95	3	5				2	8	150530
Cooked Meats	0.580	0.092	0.97	3	5				2	15	492143
Defined Chilled Meat and Veg Prods	0.220	0.084	0.93	3	5				5		222179
Fresh Meat and Veg Pastry Prds	0.341	0.064	0.95	3	5						191331
Fresh Pasta	0.088	0.071	0.99	3	5					29	22538
Fresh Processed Poultry	0.396	0.086	0.97	3	5		4		2		136823
3 Fish											
Wet/smoked fish	0.097	0.057	0.98	3	5				2	40	89991
4 Fruit and Vegetables											
Fruit	0.230	0.043	0.91	3	5			25	2	95	1078484
Vegetable	0.404	0.025	0.83	4	5			16	2	135	1936885
5 Meat											
Bacon Rashers	0.483	0.076	0.98	3	5			6	2	4	152158
Fresh Beef/Veal	0.247	0.123	0.91	3	5			5	2	5	182469
Fresh Lamb	0.142	0.084	0.97	3	5			4	2	2	35215
Fresh Pork	0.376	0.073	0.96	3	5			3	2	2	82655
Fresh Sausages	0.541	0.063	0.95	3	5		132		2	14	107576
Frozen Sausages	0.715	0.082	0.88	3	5		23			11	25780
8 Bakery Products											
Christmas Puddings	0.492	0.201	0.97	3	5				3	5	5480
Bread	0.193	0.019	0.76	4	5				4	22	647145
Cakes and Pastries	0.537	0.067	0.91	3	5				2	84	452598
Morning Goods	0.098	0.049	0.90	4	5				2	34	495959

Table 3 continued

9 Biscuits											
Biscuits	0.047	0.058	0.92	3	5				10	843088	
10 Canned Goods											
Baked Beans	0.415	0.095	0.96	3	5	2			37	157628	
Canned Fruit	0.505	0.151	0.67	3	5				13	94628	
Canned Milk Puddings	0.201	0.073	0.82	3	5				2	23231	
Canned Pasta	0.622	0.105	0.71	3	5		40		8	92251	
Other Canned Veg	0.283	0.172	0.90	3	5				30	49883	
Prepared Peas And Beans	0.509	0.072	0.88	3	5				12	107133	
Soups In Tins and Cartons	0.280	0.064	0.85	3	5		325	10	2	161741	
Sweetcorn (Canned)	0.165	0.084	0.87	3	5				2	32379	
Tinned Tomatoes, Passata	0.332	0.095	0.82	3	5				5	86865	
11 Confectionery											
Chocolate and sugar	0.221	0.061	0.96	3	5				31	650719	
12 Dairy Products											
Butter	0.197	0.058	0.99	3	5			9		108535	
Chilled Desserts	0.413	0.165	0.94	3	5		187		19	184359	
Hens Eggs	0.418	0.028	0.98	4	5				4	192100	
Margarine	0.316	0.114	0.70	3	5				6	184672	
Desserts, Long Life	-0.022	0.019	0.93	2	5				5	18517	
Total Cheese (excl Formage Frais)	0.240	0.033	0.98	3	5	4	154	12	2	39	504362
Cream, inc. Flavoured	-0.088	0.094	0.97	3	5	2	28		5	14	98333
Total Milk	0.161	0.021	0.98	3	5				4	12	730080
Yoghurt	0.167	0.036	0.91	3	5	4	259			4	380404
Yoghurt Drinks and Juices, wet	0.480	0.040	0.99	2	5		20				37110
Yoghurt Drinks and Juices, dry	-0.095	0.157	0.94	3	5		55				7485
13 Frozen Foods											
Defined Processed Poultry Prods	0.232	0.107	0.96	3	5		4			4	60278
Frozen Confectionery	0.087	0.107	0.95	3	5	2				13	69393
Frozen Meat Products	0.066	0.049	0.95	3	5					10	245778
Frozen Pizzas	0.044	0.095	0.97	3	5		139			12	93273
Frozen Vegetables	0.127	0.094	0.76	3	5					18	216498
Other Frozen Foods excl. Pizza	0.375	0.183	0.96	3	5					10	100995

Table 3 continued

Ice Cream	-0.100	0.235	0.91	3	5			15	160233	
14 Hot Beverages										
Food Drinks	0.630	0.063	0.98	3	5			12	2	42024
Fruit And Herbal Teas	0.245	0.078	1.00	3	5	145			2	13197
Instant Coffee	0.302	0.094	0.99	3	5				10	92384
Coffee Beans	0.028	0.098	0.98	3	5			4	97	15868
Tea	0.281	0.078	0.99	3	5				98	103873
16 Breakfast Foods										
Breakfast Cereals	0.225	0.056	0.97	3	5	177			3	339142
Chocolate Spread	0.392	0.109	0.97	3	5	19		3		7506
Honey	0.243	0.040	0.99	3	5				38	11585
Pastes and Spreads	0.165	0.055	0.98	3	5	318		3	8	92353
Peanut Butter	0.575	0.054	0.97	3	5				9	11855
Preserves	0.228	0.082	0.85	3	5	121			14	71605
18 Pickles and Sauces										
Mustard	0.411	0.088	0.98	3	5	18		5	15	11592
Olives	0.088	0.096	0.95	3	5			7	7	8850
Pickles	0.450	0.108	0.96	3	5				141	31151
Salad Accompaniments, wet	0.734	0.068	0.97	3	5	2	68	4	70	43521
Salad Accompaniments, dry	0.443	0.310	0.90	3	5	2	140	7	142	27749
Sauces and Ketchup (excl Worcester)	0.166	0.081	0.91	3	5				3	71261
Sauces For Cooking and Condiment, wet	0.644	0.151	0.99	3	5			13	13	294987
Sauces For Cooking and Condiment, dry	0.114	0.065	0.96	2	5			4	94	10564
Sour and Speciality Pickles	0.491	0.191	0.90	3	5			6	37	42150
19 Take Home Savouries										
Crisps	0.195	0.098	0.99	3	5	212			13	212194
Nuts	0.324	0.113	0.88	3	5			13		58590
Savoury Snacks and Reconstitutes	0.227	0.109	0.98	3	5	183				203616

Table 3 continued

20 Carbohydrate and Savoury Snacks										
Rice, Pasta, Cous Cous	0.040	0.080	0.94	3	5				235	134178
Dried Soup	0.599	0.194	0.99	3	5	2	144	5	2	50041
Dry Meat Substitutes	0.266	0.082	0.99	2	5				19	1227
Dry Pasta	0.354	0.105	0.81	3	5				83	81663
21 Savoury Home Cooking										
Defined Mixes and Breadcrumbs	0.939	0.212	0.96	3	5		74	9	129	25694
Ethnic Ingredients	0.183	0.080	0.95	3	5					33484
Flour	0.100	0.096	0.97	3	5				25	46345
Meat Extracts	0.206	0.158	0.97	3	5					96730
Packet Stuffing	0.618	0.026	0.99	3	5		35		35	19931
Pepper Herbs and Spices	0.123	0.071	0.97	3	5					37987
Salt	0.766	0.439	0.88	3	5			9	6	14954
Vinegar	0.073	0.179	0.87	3	5		21	2	27	18667
White Fats And Oils	0.434	0.119	0.86	3	5				3	52097
22 Soft Drinks										
Bottled Non-Lemonade	0.737	0.169	0.93	3	5	2	172			131515
Canned Colas	0.155	0.094	0.68	3	5	2	5			34986
Canned Non-Lemonade	0.119	0.156	0.78	3	5	2	81			30739
Fruit Juices	0.332	0.051	0.88	3	5	2	167	2	7	266801
One Shot Drinks	0.389	0.045	0.68	3	5	2	233	10	2	110848
Fruit Squash	0.820	0.180	0.82	3	5	5	110		4	155506
23 Sweet Home Cooking										
Home Baking	0.162	0.060	0.93	3	5					74980
Lemon And Lime Juices	0.476	0.059	0.97	2	5				2	7045
Mincemeat (Sweet)	0.397	0.117	0.87	3	5				2	2981
Powdered Desserts and Custard	0.837	0.172	0.94	3	5				14	63729
Sugar	0.327	0.155	0.90	3	5				10	114878
Syrup and Treacle	1.445	0.108	0.88	2	5				2	6069

Notes: Each row represents a separate hedonic regression. An observation is a transaction. The coefficient and standard error are those on a dummy for whether the specific product (bar code) is organic. Standard errors are clustered at the product level. The adjusted R^2 is from the overall regression. The numbers indicate the number of characteristics of each type appear in the hedonic regression. For example, for Bacon Rashers there are 3 brand characteristics (Branded, Budget or Standard Private Label), 4 origin characteristics (Britain, Ireland, Northern Europe and Other) and 2 variety characteristics (smoked or unsmoked). In all regressions there are 5 size categories, 8 store indicators (see Table 2), time and region effects. The final column shows the number of observations.

Table 4: share of household expenditure on foods with organic characteristic

share of household expenditure on foods with organic characteristic	Number of households	% of sample	Cumulative % of sample
zero	3,132	18.25	18.25
less than 0.25%	3,407	19.85	38.10
btwn 0.25%-0.5%	2,949	17.18	55.28
btwn 0.5%-1%	3,198	18.63	73.92
btwn 1%-5%	3,739	21.79	95.71
btwn 5%-10%	451	2.63	98.33
over 10%	286	1.67	100.00
Total	17,162	100.00	

Table 5: estimated household lower bound on willingness to pay for organic over a year

household expenditure on foods with organic characteristic	Number of households	% of sample	Cumulative % of sample
zero	3,225	18.79	18.79
less than £1	5,122	29.85	48.64
btwn £1-£5	5,915	34.47	83.10
btwn £5-£10	1,457	8.49	91.59
btwn £10-£50	1,216	7.09	98.68
btwn £50-£100	129	0.75	99.43
over £100	98	0.57	100.00
Total	17,162	100.00	

Table 6: estimated household lower bound on willingness to pay for organic over a year as a share of expenditure on organic products

lower bound on willingness to pay for organic as a share of household expenditure on organic foods	Number of households	% of sample	Cumulative % of sample
zero	3,225	18.79	18.79
less than 10%	1,353	7.88	26.68
btwn 10-15%	1,707	9.95	36.62
btwn 15-20%	3,856	22.47	59.09
btwn 20-25%	3,396	19.79	78.88
btwn 25-50%	3,405	19.84	98.72
over 50%	220	1.28	100.00
Total	17,162	100.00	

Table 7: bundling of organic products

Number of characteristics that differ to nearest non-organic product	with fascia		whole market	
	Number of organic products	% of products	Number of organic products	% of products
0	5908	80.68	2957	87.41
1	1370	18.71	412	12.18
2	44	0.60	14	0.41
3	1	0.01		
Total	7323	100.00	3383	100

Table 8: bundling of non-organic products

Number of characteristics that differ in nearest organic product	with fascia		whole market	
	Number of non-organic products	% of products	Number of non-organic products	% of products
0	61318	28.67	35806	40.98
1	84984	39.73	34165	39.10
2	32731	15.30	9832	11.25
3	17834	8.34	5135	5.88
4	8322	3.89	1993	2.28
5	8076	3.78	437	0.50
6	619	0.29	2	0.00
7	1	0.00		
Total	213885	100.00	87370	100.00

Table 9: Health, share of expenditure on organic, lower bound on wtp as share of total expenditure and number of households

	Organic Products Are Healthier	
	Agree Strongly/Agree	Agree Nor Disagree/Disagree /Strongly Disagree
I Try To Buy a Healthy Range Of Foods These Days	Agree Strongly/Agree	1.2%
	Agree Nor Disagree/Disagree /Strongly Disagree	0.3%
	(2196)	(518)
	Agree Strongly/Agree	0.9%
	Agree Nor Disagree/Disagree /Strongly Disagree	0.3%
	(4166)	(7503)

Note: In each cell the % indicates that share of total household expenditure that is on organic products, and the number in () indicates the number of households that selected the indicated response.

Table 10: Environment, share of expenditure on organic, lower bound on wtp as share of total expenditure and number of households

	Organic Products Are Friendlier To The Environment	
	Agree Strongly/Agree	Agree Nor Disagree/Disagree /Strongly Disagree
I Try To Buy Environmentally Friendly Products	Agree Strongly/Agree	2.0%
	Agree Nor Disagree/Disagree /Strongly Disagree	0.5%
	(1957)	(1731)
	Agree Strongly/Agree	0.9%
	Agree Nor Disagree/Disagree /Strongly Disagree	0.2%
	(2089)	(8200)

Note: In each cell the % indicates that share of total household expenditure that is on organic products, and the number in () indicates the number of households that selected the indicated response.

Table 11: Quality, share of expenditure on organic and number of households

	Organic Foods Are Better Quality	
	Agree Strongly/Agree	Agree Nor Disagree/Disagree /Strongly Disagree
I Don't Mind Paying For Quality	4.3%	1.8%
	1.0%	0.4%
	(1345)	(420)
	0.9%	0.9%
	0.2%	0.2%
	(4739)	(7558)

Note: In each cell the % indicates that share of total household expenditure that is on organic products, and the number in () indicates the number of households that selected the indicated response.

Table 12: Determinants of lower and upper bound on willingness to pay for organic

Dep var: lower bound on willingness to pay for organic in £	(1) All households	(2) All households	(3) All households	(4) Class AB	(5) Class C	(6) Class DE
Health important	6.008 [0.544]	5.528 [0.561]	4.883 [0.520]	7.445 [1.977]	4.637 [0.663]	4.255 [0.863]
Health (upper right quadrant of Table 8)		-0.987 [0.613]	-1.293 [0.628]	-5.959 [3.392]	-1.327 [0.879]	-0.026 [0.675]
Health (bottom left quadrant of Table 8)		-0.706 [0.156]	-1.238 [0.201]	-3.034 [1.332]	-1.453 [0.231]	-0.216 [0.183]
Environment important	1.606 [0.694]	2.089 [0.605]	1.539 [0.555]	2.052 [2.452]	1.428 [0.688]	1.678 [0.815]
Environment (upper right quadrant of Table 9)		0.668 [0.490]	0.287 [0.515]	3.749 [2.475]	0.164 [0.664]	-0.337 [0.528]
Environment (bottom left quadrant of Table 9)		-0.132 [0.171]	-0.51 [0.188]	-2.089 [1.097]	-0.312 [0.241]	-0.395 [0.233]
Quality important	9.831 [1.216]	9.566 [1.249]	8.983 [1.178]	9.749 [4.165]	10.249 [1.533]	5.374 [1.469]
Quality (upper right quadrant of Table 10)		-0.619 [0.834]	-0.38 [0.796]	3.061 [4.587]	-0.699 [0.812]	0.202 [1.437]
Quality (bottom left quadrant of Table 10)		-0.281 [0.179]	-0.997 [0.225]	-3.007 [1.383]	-0.743 [0.232]	-0.811 [0.341]
Single young			3.202 [0.780]	6.344 [5.863]	3.162 [0.949]	2.318 [0.636]
Single with kids			1.836 [0.586]	9.757 [6.599]	1.151 [0.718]	1.537 [0.528]
Single pensioner			2.918 [0.736]	5.566 [6.207]	1.988 [0.982]	2.526 [0.560]
Couple no kids			2.442 [0.564]	2.785 [3.574]	2.646 [0.723]	1.742 [0.741]
Couple with kids			0.866	-0.576	1.203	0.195

			[0.460]	[2.816]	[0.659]	[0.420]
Couple pensioner			1.621	0.652	2.332	1.228
			[0.518]	[3.187]	[0.928]	[0.425]
Other no kids			1.574	0.003	1.148	2.515
			[0.584]	[3.438]	[0.714]	[0.887]
At least one vegetarian in the household			4.993	6.815	6.335	0.783
			[1.740]	[5.826]	[2.382]	[1.116]
Annual expenditure on alcohol, food, toiletries and cleaning products			0.003	0.008	0.003	0.002
			[0.001]	[0.004]	[0.000]	[0.000]
Household Class A or B			3.255			
			[0.822]			
Household class C1 or C2			0.474			
			[0.266]			
Response to attitudinal question missing	-1.155	-1.111	-0.967	-3.585	-1.039	-0.642
	[0.284]	[0.286]	[0.309]	[1.705]	[0.460]	[0.412]
Constant	2.623	2.943	-5.326	-12.056	-4.217	-3.013
	[0.116]	[0.157]	[1.253]	[9.883]	[1.218]	[0.686]
Observations	14383	14383	14383	1430	8309	4644
R-squared	0.06	0.06	0.09	0.1	0.09	0.12

Notes: Standard errors in [] are robust. See Table 12 for means of variables.

Table 13: mean of demographic variables

Variable	mean (s.d.)
Family type = Single young	0.082 (0.275)
Family type =Single parent	0.065 (0.247)
Family type = Single pensioner	0.075 (0.264)
Family type = Couple no children	0.136 (0.343)
Family type = Couple with children	0.395 (0.489)
Family type = Pensioner couple	0.115 (0.319)
Family type = Others no children	0.083 (0.276)
Family type = Others with children	0.0483 (0.137)
Annual expenditure on alcohol, food, toiletries and cleaning products	2002.95 (1052.78)
Household Class A or B	0.099 (0.299)
Household class C1 or C2	0.578 (0.494)
At least one vegetarian in the household	0.0220 (0.147)
Demographics or attitudes missing	0.035 (0.183)

Notes: Social class is A (upper middle class - higher managerial, administrative or professional), B (middle class - intermediate managerial, administrative or professional) C1 (lower middle class - supervisory or clerical, junior managerial, administrative or professional) or C2 (skilled working class - skilled manual workers) (the omitted category is D (working class - semi and unskilled manual workers) and E (those at lowest level of subsistence - state pensioners or widows (no other earner), casual or lowest grade workers).

Figure 1: Distribution of estimated price premia with and without controls for other observed characteristics

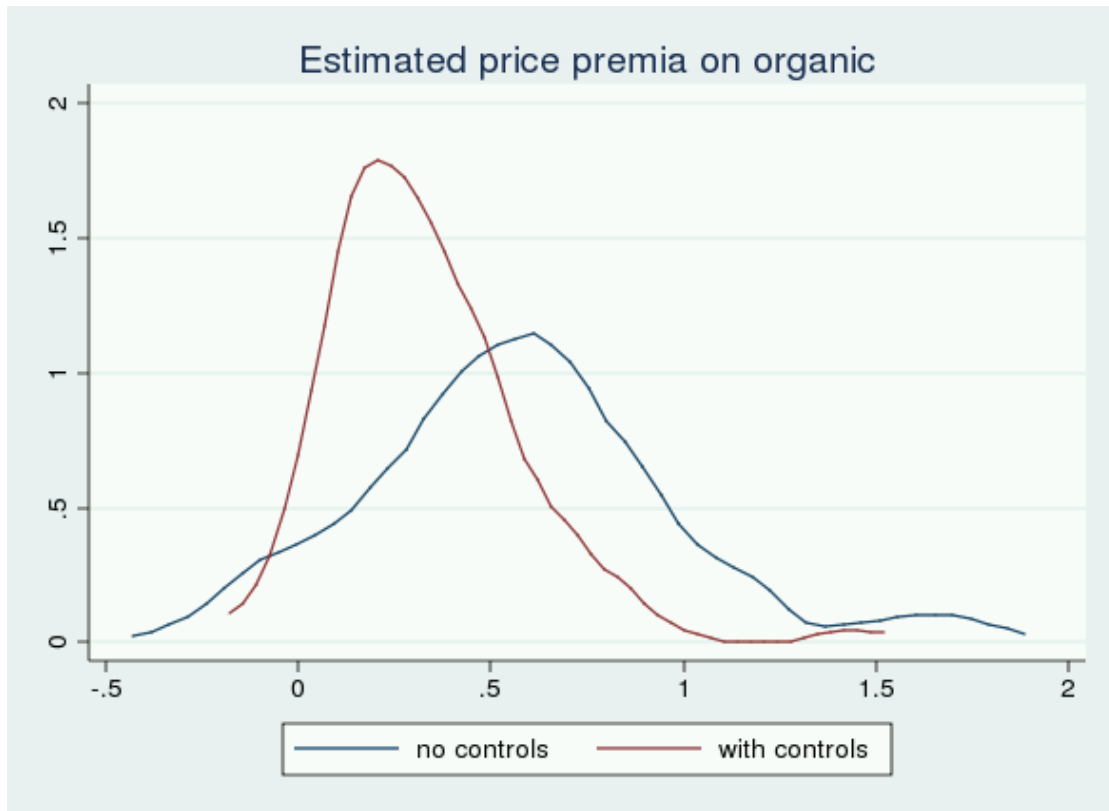


Figure 2: density of lower bound on households' willingness to pay for organic food

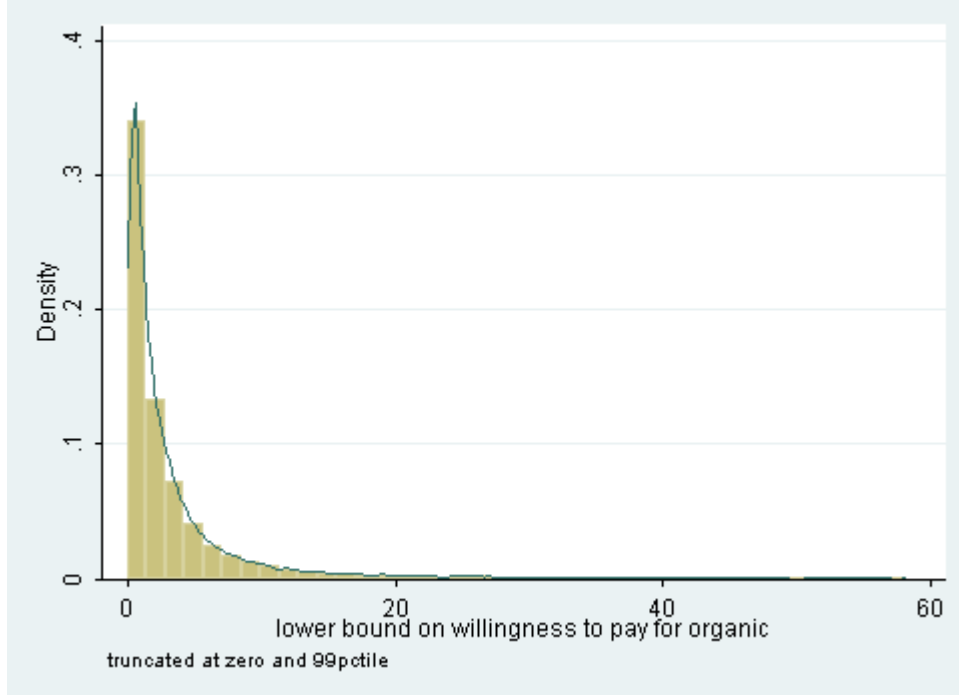


Figure 3: density of upper bound on households' willingness to pay for organic food

