Valuing A New Good

Laura Blow[*] and Ian Crawford[†]

January 3, 2001

Abstract

This paper presents a revealed preference method for calculating a lower bound on the virtual or reservation price of a new good and suggests a way to improve these bounds by using budget expansion paths. This allows the calculation of cost-of-living and price indices when the number of goods available changes between periods. We apply this technique to the UK National Lottery and illustrate the effects of its inclusion in measures of inflation.

Key Words: Cost-of-living indices, New goods, GARP.

JEL Classification: C43, D11.

Acknowledgements: We very are grateful to two anonymous referees for their helpful comments on an earlier draft. We are also grateful to Orazio Attanasio, Richard Blundell, Martin Browning, J. Peter Neary, Ian Preston, David Ulph, Ian Walker and seminar participants at University College Dublin, University College London, the University of Toulouse and the Bank of England for their comments. This study was jointly funded by the Leverhulme Trust (Grant Ref: F/386/H) and as part of the research program of the ESRC Centre for the Microeconomics of Fiscal Policy at IFS. All errors are the sole responsibility of the authors.

[*]Institute for Fiscal Studies

[†]Institute for Fiscal Studies and University College London.

Address for correspondence: Institute for Fiscal Studies, 7 Ridgmount St., London, WC1E 7AE. laura_blow@ifs.org.uk or ian_crawford@ifs.org.uk

1. Introduction

There has been much recent interest in the extent to which official price indices may mis-measure the true rate of inflation. This has been particularly so for the Consumer Price Index in the US, where a study of the possible sources of bias was commissioned, with a report of the findings, the Boskin Report, published in 1996[1]. One of the major sources of bias identified in the Boskin Report was the bias associated with the arrival of new goods.

New goods bias refers to the failure to incorporate properly into a cost-of-living index the effect which the arrival of a new good has on economic welfare. The arrival of a new good is potentially welfare-improving because it expands the set of choices available to the consumer. This means that some reference level of utility may now be available at a lower cost than previously. It is well known that the way to deal with new goods in a cost-of-living index which spans a period before and after the introduction of a new good is to impute a price for the new good in the period before it exists. This price should be the price which would just have driven the consumer's demand for the good to zero in that period, i.e. the 'virtual' price[2] or the 'reservation' price.

The most common approach to calculating the virtual price of a new good is the parametric estimation of, and extrapolation from, demand curves[3]. This requires the imposition of a particular functional form for preferences, upon which the results of the extrapolation will be heavily dependent. This paper presents an alternative revealed preference method for calculating the virtual price of a good. This method is consistent with the maximisation of a well-behaved utility function which is stable over time, with no further restrictions on the exact form of preferences necessary.

[1] Boskin, M. J. *et al* (1996).
[2] The term is due to Rothbarth (1941).
[3] For example, Hausman (1997a).

The plan of the paper is as follows. In section 2 we more formally state the problem that new goods pose for the calculation of an individual consumer's cost-of-living index, and we review a simple framework for the valuation of new goods at the individual level. The section ends with a discussion of how these individual cost-of-living indices with individual-level virtual prices can be combined into an aggregate cost-of-living measure. In section 3 we describe a revealed preference method for calculating a lower bound on the virtual price using observable choice outcomes generated by an individual consumer. We also describe a way of improving this bound which requires knowledge of the consumer's budget expansion paths. Section 4 describes an empirical application to a time series of cross section household level data on the UK National Lottery, which was introduced in November 1994. In section 4.1 we discuss a framework for implementing these ideas on microdata using Engel curves conditional on total expenditure and a list of demographic variables. Because price data at the individual level are not available for a comprehensive list of goods in the UK, we have to assume that households observed at the same point in time face a common vector of prices. Under this assumption, within-period Engel curves correspond to budget expansion paths. In keeping with the nonparametric focus of the revealed preference ideas we aim to estimate these Engel curves nonparametrically. However, reliable multivariate nonparametric regression typically requires a very large number of observations which we do not have, and we therefore opt for semi-parametric extended partially linear specification[4] in which the effects of changes in the total budget are estimated nonparametrically, while household characteristics variables are parametrised. We also discuss how we deal with the endogeneity of the total available budget and the issue of selection on zero demands. Section 4.2 discusses the problems caused by violations of GARP and we discuss how statistical tests of revealed preference restrictions can be constructed from the estimated Engel curves. In section 4.3 we describe the results. We calculate the virtual price of

[4] Härdle and Marron (1990), Blundell, Duncan and Pendakur (1998).

the UK National Lottery one year prior to its introduction and examine the effect of including this new good in some measures of annual non-housing inflation rates over the period. Section 5 concludes.

2. New goods and index numbers

In what follows we are interested in the case in which a single new good appears. Our aim is to calculate a cost-of-living index which compares the cost to an individual consumer of reaching some reference level of welfare in the period before the introduction of the new good, with the cost of reaching the same level of welfare in a period after its introduction. We are faced with two immediate issues. Firstly, can we be sure that under these circumstances there exists a cost function, consistent with a stable set of preferences, which will allow such a comparison to be made? Secondly, what are the relevant price vectors? In particular, how should we price the new good in the period before it first exists?[5]

A new good is usually thought of as a special case of a rationed good: non-existence is treated like a ration level of zero. Hicks (1940) and Rothbath (1941) and more recently Neary and Roberts (1980) discuss the question of how to deal with rationed goods in economic problems, and in particular how to price goods when the consumer is free to purchase goods in some markets, but forced to purchase certain levels of other goods in other markets. They show how the properties of demands under these circumstances can be expressed in terms of unrationed demands by allowing free choice over all goods but replacing the observed market prices with a vector of 'virtual' prices or 'support' prices. These support prices are such that this unrationed choice would generate exactly the same demand vector as the one generated by the observed prices under the rationing constraint. Neary and Roberts (1980) show that convexity, continuity and strict monotonicity of the consumer's preferences are sufficient to ensure that there always exists a set of

[5] Only in the case where the reference utility level is set equal to what the consumer's actual utility level was in the period before the new good was introduced is this not a problem. This is because (assuming cost minimising behaviour) the minimum cost of reaching this reference level of utility is the consumers observed total expenditure for that period.

strictly positive support prices consistent with any set of demands. They also show that the virtual or support prices for the unrationed goods are identical to their actual prices[6]. The term 'virtual' is therefore usually reserved for the support prices of the rationed goods only.

To place the new goods problem in a simple rationing context we suppose that there are $T+1$ periods, $t = 0, ..., T$, and $K+1$ goods, $k = 0, ..., K$. The 0th good is subject to a ration level of $\overline{q}$ in period $t = 0$ but is freely available from period 1 onwards. All other goods are freely available in every period. We denote by $\mathbf{q}_t^K$ and $\mathbf{p}_t^K$ the $(K \times 1)$ vectors of quantities and prices of the $k = 1, .., K$ goods in period t. Consider the consumer's problem

$$\begin{array}{lll} \max_{\mathbf{q}_t} u(\mathbf{q}_t) & \text{subject to} & \mathbf{p}_t'\mathbf{q}_t \leq x_t \\ & \text{and} & q_0^0 = \overline{q} \end{array}$$

where x_t denotes the available budget in period t. The first order conditions are

$$\left[\begin{array}{c} u'\left(q_0^0\right) \\ u'\left(\mathbf{q}_0^K\right) \end{array} \right] = \lambda_0 \left[\begin{array}{c} p_0^0 + \frac{\mu_0}{\lambda_0} \\ \mathbf{p}_0^K \end{array} \right]$$

for period 0 and

$$u'(\mathbf{q}_t) = \lambda_t \mathbf{p}_t$$

for $t \neq 0$. The scalar λ_0 is the marginal utility of income and μ_0 is the shadow value of the rationing constraint. This is positive or negative according to whether the consumer would like to purchase more or less than the constrained level of the rationed good. The vectors $\boldsymbol{\pi}_0 = \left[p_0^0 + \frac{\mu_0}{\lambda_0}, \mathbf{p}_0^K\right]'$ for period 0 and $\boldsymbol{\pi}_t = \mathbf{p}_t$ for $t \neq 0$ are the support price vectors. The vector of period 0 support prices is made up of the virtual price of the rationed good $\left(p_0^0 + \frac{\mu_0}{\lambda_0}\right)$ and the list of observed prices for the other goods. The support price vectors for all the other periods are simply the observed prices. The support prices are such that the outcome of the rationed model is identical to the outcome of the unrationed choice generated by

$$\max_{\mathbf{q}_t} u(\mathbf{q}_t) \text{ subject to } \boldsymbol{\pi}_t'\mathbf{q}_t \leq x_t$$

[6] Neary and Roberts (1980), p.27-9.

The cost function associated with the unrationed choice is defined as

$$c\left(\boldsymbol{\pi}_t, u\right) = \min_{\mathbf{q}_t}\left[\boldsymbol{\pi}_t'\mathbf{q}_t : u\left(\mathbf{q}_t\right) \geq u\right] \tag{2.1}$$

with the associated indirect utility function $v\left(\boldsymbol{\pi}_t, x\right)$, and the cost function when the consumer is forced to set $q_0^0 = \overline{q}$ is defined as

$$\overline{c}\left(\mathbf{p}_t, u, \overline{q}\right) = \min_{\mathbf{q}_t}\left[\mathbf{p}_t'\mathbf{q}_t : u\left(\mathbf{q}_t\right) \geq u;\ q_0^0 = \overline{q}\right] \tag{2.2}$$

with the associated indirect utility function $\overline{v}\left(\mathbf{p}_t, x, \overline{q}\right)$. We note that while (2.1) is defined for all u contained in the image of the consumption set, (2.2) is defined only if demands $\mathbf{q}_0^K$ can be found such that $u\left(0, \mathbf{q}_0^K\right) \geq u$. Neary and Roberts (1980) show that relationship between (2.1) and (2.2) is given by

$$\overline{c}\left(\mathbf{p}_0, u, \overline{q}\right) = c\left(\boldsymbol{\pi}_0, u\right) + \left(p_0^0 - \left[p_0^0 + \frac{\mu_0}{\lambda_0}\right]\right)\overline{q} \tag{2.3}$$

in period 0 and

$$\overline{c}\left(\mathbf{p}_t, u, \overline{q}\right) = c\left(\boldsymbol{\pi}_t, u\right)$$

for the unrationed periods $t \neq 0$. Differentiating (2.3) with respect to the ration level gives $\left(p_0^0 - \left[p_0^0 + \frac{\mu_0}{\lambda_0}\right]\right)$ as an exact measure of the benefit to the consumer of a change in the constraint $\overline{q}$. In the case of a new good the ration level is $\overline{q} = 0$ and so (2.3) simplifies to

$$\overline{c}\left(\mathbf{p}_0, u, \overline{q}\right) = c\left(\boldsymbol{\pi}_0, u\right) \tag{2.3$'$}$$

The cost-of-living index linking the base period 0 (before the new good exists) with period t (after its introduction) can then be defined in terms of the cost function associated with the unrationed problem with support prices as arguments.

$$\frac{c\left(\boldsymbol{\pi}_t, u\right)}{c\left(\boldsymbol{\pi}_0, u\right)} \tag{2.4}$$

Thus the price of the new good in period 0 is the price which would just have driven demand for the good to zero, i.e. the virtual price. This approach captures the introduction of a new good by imagining that its price has reached its period t value from a level in period 0 which was just above the maximum value of the good to the consumer and no higher.

So far we have considered a single consumer. Suppose that we have a population of consumers with identical preferences but different incomes. It is well known that homotheticity of the consumers' preferences is sufficient for there to exist a unique cost-of-living index[7]. For the virtual price of the rationed good to be independent of income requires $\frac{\partial[\mu_0/\lambda_0]}{\partial x} = 0$ where $\mu_0 = \frac{\partial \overline{v}(\mathbf{p}_0, x, \overline{q})}{\partial \overline{q}}$ and $\lambda_0 = \frac{\partial \overline{v}(\mathbf{p}_0, x, \overline{q})}{\partial x}$. Since $\frac{\partial[\mu_0/\lambda_0]}{\partial x} = 0$ implies $\frac{\partial \mu_0}{\partial x} = \frac{\partial \lambda_0}{\partial x}\frac{\mu_0}{\lambda_0}$, it is therefore sufficient for either λ_0 or μ_0 to be independent of x and so homotheticity is also sufficient for there to be a unique virtual price for the new good. However, even from the very earliest studies of household spending patterns there has been strong empirical evidence against homotheticity[8]. With a population consisting of many heterogeneous individuals, we would expect them each to have a different virtual price for the new good not least because of income variation, but also due to differences in tastes. Households which value it highly will have relatively high virtual prices compared to those who do not. It is possible that for some households the new good is something that they would never want to buy at any price. For these households the virtual price will be zero.

In this paper we assume that consumers have common, probably nonhomothetic preferences, and that differences in tastes are due to differences in their characteristics. In order to calculate a group cost-of-living index based on individual specific virtual prices and individual specific cost-of-living indices we require some scheme for aggregating these data into a group cost-of-living index. In accordance with most of the literature[9], and the current practice in the calculation of the UK Retail Prices Index[10] and many other country's consumer price indices[11], we use a weighted arithmetic mean of the individual cost-of-living indices in our applied work. These weights are the individual's share out of total expenditure (known as plutocratic weights). However we note that there are a

[7] Deaton and Muellbauer (1980).
[8] See Engel (1895) for a very early example and Banks *et al* (1996) for more recent evidence.
[9] Prais (1959), Pollak (1981).
[10] Baxter, (ed) (1998).
[11] Ruiz-Castillo *et al* (1999).

number of other schemes which have been suggested including the unweighted arithmetic mean[12] (known as democratic weights) and also geometric versions of these two schemes[13].

Returning to the problem of the single consumer, the remaining issue is one of missing data. All of the support prices are observable, except for the element $\pi_0^0 = \left[p_0^0 + \frac{\mu_0}{\lambda_0}\right]$ which is unknown. In order to construct the individual's cost-of-living index, which can then be combined with those of others to form a group cost-of-living index, we need a way of calculating the individual's virtual price.

The most usual approach to calculating the virtual price of a new good has been the parametric estimation of demand curves. A particular functional form for demand is assumed which is consistent with maximisation of a particular form for the utility function which is assumed to be common to all consumers. A system of demand equations is then estimated using data from periods in which all goods are available, and this is used to predict $\pi_0^0 = \left[p_0^0 + \frac{\mu_0}{\lambda_0}\right]$ i.e. the lowest price which would result in zero demand for good 0 in period 0 either for a representative consumer when aggregate data is used, or for each individual in a micro-level dataset. A recent example of this sort of technique is Hausman (1997a) who estimates the welfare gains associated with the launch of new varieties of breakfast cereals.

One possible problem associated with this approach is that the estimate of the virtual price will be heavily dependent on the maintained hypothesis concerning functional form as parametric methods are reliant upon (possibly suspect) out-of-sample predictions of the demand curve to solve for π_0^0. This is because it is usually necessary to extrapolate the demand curve across regions over which relative price variations have never been observed in the data (i.e. to a very high relative price for the new good). A second problem is that parametric models usually require a good deal of observed relative price variation in order to capture price effects accurately and this may not always be available.

[12] Prais (1959) and Muellbauer (1974).

[13] Diewert (1984).

3. A revealed preference approach

In this paper we propose using a revealed preference technique. The first attraction of revealed preference conditions is that they apply to any well behaved utility function and, beyond this, they require no additional restrictions on the precise form of preferences underlying consumer demands. This property is set out in Afriat's Theorem[14] which shows that, if consumers' observed choices, given the prices they face, satisfy the Generalised Axiom of Revealed Preference (GARP) (defined below), then these choices could have been generated by the maximisation of any well behaved utility function. The second attraction of the revealed preference approach which we are proposing is that it is computationally very simple. Finally, as we show, it can make effective use of very few post-introduction price observations.

Following Varian (1982) we set out the following definitions of revealed preference conditions;

Definition 1. $\mathbf{q}_t$ is directly revealed preferred to $\mathbf{q}$, written $\mathbf{q}_t R^0 \mathbf{q}$, if $\boldsymbol{\pi}_t' \mathbf{q}_t \geq \boldsymbol{\pi}_t' \mathbf{q}$.
Definition 2. $\mathbf{q}_t$ is strictly directly revealed preferred to $\mathbf{q}$, written $\mathbf{q}_t P^0 \mathbf{q}$, if $\boldsymbol{\pi}_t' \mathbf{q}_t > \boldsymbol{\pi}_t' \mathbf{q}$.
Definition 3. $\mathbf{q}_t$ is revealed preferred to $\mathbf{q}$, written $\mathbf{q}_t R \mathbf{q}$, if $\boldsymbol{\pi}_t' \mathbf{q}_t \geq \boldsymbol{\pi}_t' \mathbf{q}_s$, $\boldsymbol{\pi}_s' \mathbf{q}_s \geq \boldsymbol{\pi}_s' \mathbf{q}_r$,...,$\boldsymbol{\pi}_m' \mathbf{q}_m \geq \boldsymbol{\pi}_m' \mathbf{q}$, for some sequence of observations $(\mathbf{q}_t,\ \mathbf{q}_s, ..., \mathbf{q}_m)$. In this case, we say that the relation R is the transitive closure of the relation R^0.
Definition 4. $\mathbf{q}_t$ is strictly revealed preferred to $\mathbf{q}$, written $\mathbf{q}_t P \mathbf{q}$, if there exist observations $\mathbf{q}_s$ and $\mathbf{q}_m$ such that $\mathbf{q}_t R \mathbf{q}_s,\ \mathbf{q}_s P^0 \mathbf{q}_m,\ \mathbf{q}_m R \mathbf{q}$.
Definition 5. Data can be said to satisfy GARP if $\mathbf{q}_t R \mathbf{q}_s \Rightarrow \boldsymbol{\pi}_s' \mathbf{q}_s \leq \boldsymbol{\pi}_s' \mathbf{q}_t$. Equivalently, the data satisfy GARP if $\mathbf{q}_t R \mathbf{q}_s$ implies not $\mathbf{q}_s P^0 \mathbf{q}_t$.

Our aim is to use the restrictions imposed by revealed preference theory to place a lower bound on the virtual price of the new good in period 0 in the following way. We have data on prices and demands in period 0, $(\boldsymbol{\pi}_0, \mathbf{q}_0)$, with the missing price π_0^0, and data on prices and demands after the introduction of the new good,$(\boldsymbol{\pi}_t, \mathbf{q}_t)$, $t \neq 0$, with no missing variables. If a post-introduction demand bundle, $\mathbf{q}_s$ say, is revealed preferred to $\mathbf{q}_0$, then (if these choices have been generated by the maximisation of a stable, well-behaved utility function) $\mathbf{q}_0$

[14] Afriat (1965) and (1973). Varian (1982) provides a proof.

cannot be strictly preferred to $\mathbf{q}_s$, and this gives us a restriction on the value that the price of good 0 in period 0 can take.

Since, in placing a bound on π_0^0, we exploit the assumption that the data were generated by stable preferences, our first step should be to test this hypothesis for the data from the post-introduction period $(\boldsymbol{\pi}_t, \mathbf{q}_t)$, $t \neq 0$. By Afriat's Theorem, we can do this by testing whether the data $(\boldsymbol{\pi}_t, \mathbf{q}_t)$, $t \neq 0$ satisfy GARP. If the subset of data $(\boldsymbol{\pi}_t, \mathbf{q}_t)$ for $t \neq 0$ satisfy GARP, then we can go on to use it to place restrictions on the set of possible values that π_0^0 can take as described above. If this subset of data was not internally consistent with GARP, then there exists no value of π_0^0 which can rationalise the data.

Assuming for the moment that $(\boldsymbol{\pi}_t, \mathbf{q}_t)$ for $t \neq 0$ satisfy GARP, the bound we choose for π_0^0 will be smallest value of π_0^0 such that the entire data set, i.e. now including $(\boldsymbol{\pi}_0, \mathbf{q}_0)$ with $q_0^0 = 0$, satisfies GARP. This will give the smallest value for π_0^0 which makes the choice of $q_0^0 = 0$ consistent with the unrationed maximisation of a stable utility function, i.e. precisely the virtual price of the new good in period 0 that we wish to calculate. If the subset of data, $(\boldsymbol{\pi}_t, \mathbf{q}_t)$ for $t \neq 0$, did not satisfy GARP, then, of course, there could not exist a π_0^0 which would rationalise $(\boldsymbol{\pi}, \mathbf{q})$. We first set out the general idea in more detail and then discuss a way of improving the bound by means of expansion paths.

3.1. Bounding the virtual price

We observe the support prices $\boldsymbol{\pi}_t$ (equal to the actual prices $\mathbf{p}_t$) for all goods from period 1 onwards, and for all goods in the 0th period except good 0 ($\pi_0^k = p_0^k$ for $k \neq 0$). If the data from periods $t \neq 0$ satisfy GARP, then we can calculate the lower limit on π_0^0 in the following way. We require the entire dataset $(\boldsymbol{\pi}, \mathbf{q}), t = 0, ..., T$ to be consistent with non-violation of GARP. Denote the set of consumption bundles which are revealed preferred to $\mathbf{q}_0$ by $RP(\mathbf{q}_0)$. With $K + 1 > 1$ and $T + 1 > 1$, and $(\boldsymbol{\pi}_t, \mathbf{q}_t)$ for $t \neq 0$ satisfying GARP, then for non-violation of GARP for the entire data set $(\boldsymbol{\pi}, \mathbf{q})$, we cannot have $\mathbf{q}_0 P^0 \mathbf{q}_s$ for $\mathbf{q}_s \in RP(\mathbf{q}_0)$. For each $\mathbf{q}_s \in RP(\mathbf{q}_0)$, this requirement implies:

$$\begin{aligned} & \boldsymbol{\pi}_0'\mathbf{q}_s \geq \boldsymbol{\pi}_0'\mathbf{q}_0 \\ \Rightarrow \quad & \pi_0^0\left(q_s^0 - q_0^0\right) \geq \boldsymbol{\pi}_0^{k\prime}\left(\mathbf{q}_0^K - \mathbf{q}_s^K\right) \\ \Rightarrow \quad & \pi_0^0 \geq \boldsymbol{\pi}_0^{k\prime}\left(\mathbf{q}_0^K - \mathbf{q}_s^K\right)\left(q_s^0\right)^{-1} \qquad \text{if } q_s^0 > 0 \end{aligned}$$

Note that if the consumer chooses not to buy any of the new good after the introduction either then $q_s^0 = 0$ for all s implies $\left(q_s^0 - q_0^0\right) = 0$. Non-violation of GARP requires $\boldsymbol{\pi}_0^{K\prime}\mathbf{q}_s^K \geq \boldsymbol{\pi}_0^{K\prime}\mathbf{q}_0^K$ which does not depend on the new good and so is testable. If GARP is not violated for the other goods then there is implies $\pi_0^0 \times 0 \geq \boldsymbol{\pi}_0^{K\prime}\left(\mathbf{q}_0^K - \mathbf{q}_s^K\right)$ where the right hand side is negative. Assuming away the possibility of negative prices then for such a consumer the lower bound is zero. Of course if GARP is violated $\left(\boldsymbol{\pi}_0^{K\prime}\mathbf{q}_s^K < \boldsymbol{\pi}_0^{K\prime}\mathbf{q}_0^K\right)$ then the right hand side is positive and no value of π_0^0 can be found which is consistent with utility maximisation.

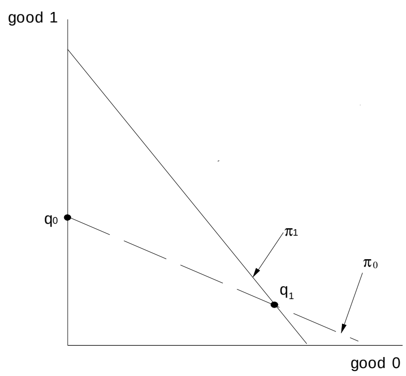
As each $\mathbf{q}_s \in RP\left(\mathbf{q}_0\right)$ gives a lower bound on π_0^0 — call this set $\pi_0^0(\mathbf{q}_s)$. The highest value in this set encompasses all the other lower bounds and is the lower limit on π_0^0 given the data. This is proved below. Denote $\max\left\{\pi_0^0(\mathbf{q}_s)\right\}$ by $\overline{\pi}_0^0$.

Proposition 3.1. *Any $\pi_0^0 < \overline{\pi}_0^0$ violates GARP for $(\boldsymbol{\pi}, \mathbf{q})$.*

Proof.
(1) Denote $\overline{\boldsymbol{\pi}}_0 = \left(\overline{\pi}_0^0, \pi_0^1, ..., \pi_0^K\right)$
(2) $\overline{\pi}_0^0$ is such that $\overline{\boldsymbol{\pi}}_0'\overline{\mathbf{q}}_s = \overline{\boldsymbol{\pi}}_0'\mathbf{q}_0 = x_0$ where $\overline{\mathbf{q}}_s \in RP\left(\mathbf{q}_0\right)$
(3) Suppose $\underline{\pi}_0^0 < \overline{\pi}_0^0$, where $\underline{\boldsymbol{\pi}}_0 = \left(\underline{\pi}_0^0, \pi_0^1, ..., \pi_0^K\right)$
(4) Then from (2) and (3) $\underline{\boldsymbol{\pi}}_0'\overline{\mathbf{q}}_s < \overline{\boldsymbol{\pi}}_0'\overline{\mathbf{q}}_s = \overline{\boldsymbol{\pi}}_0'\mathbf{q}_0 = \underline{\boldsymbol{\pi}}_0'\mathbf{q}_0$ (since $q_0^0 = 0$) $\Rightarrow \mathbf{q}_0 P^0 \overline{\mathbf{q}}_s$ which is a violation of GARP.
■

A two good, two period case is illustrated in figure 3.1. The budget line in period 1 is given by $\boldsymbol{\pi}_1$, the period 1 bundle by $\mathbf{q}_1$ and the corner solution in period 0 by $\mathbf{q}_0$. Clearly, $\mathbf{q}_1 P^0 \mathbf{q}_0$. As a result, any period 0 price ($\boldsymbol{\pi}_0$) shallower than the line connecting $\mathbf{q}_0$ and $\mathbf{q}_1$ would violate GARP for the data set $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1; \mathbf{q}_0, \mathbf{q}_1)$. So π_0^0/π_0^1 must be greater than or equal to the gradient of the $\mathbf{q}_0$ to $\mathbf{q}_1$ line.

Figure 3.1: A two-period, two-good example



One problem with using the bundles observed in actual data is that[15], because movements of the budget line between periods are generally large and relative price changes are typically small, budget lines seldom cross. As a result, data may lack power either to reject, or to usefully invoke GARP. This means that, when applying revealed preference restrictions to observed bundles, it is possible that the lower bounds we can recover are not particularly enlightening. For example, if the bundle $\mathbf{q}_1$ contains more of *both* commodities than $\mathbf{q}_0$, then since $\mathbf{q}_1$ lies in the interior of the $RP(\mathbf{q}_0)$ set by monotonicity of preferences, the data contain no additional information on the shape of the indifference curve through $\mathbf{q}_0$ and *any* non-negative value for π_0^0 can rationalise the data (giving $\overline{\pi}_0^0 = 0$ as the lower bound).

The second problem with this approach is that, unlike parametric models, we cannot use data for periods when $\mathbf{q}_t \notin RP(\mathbf{q}_0)$. This is because these periods do not provide any revealed preference restrictions at all on π_0^0.

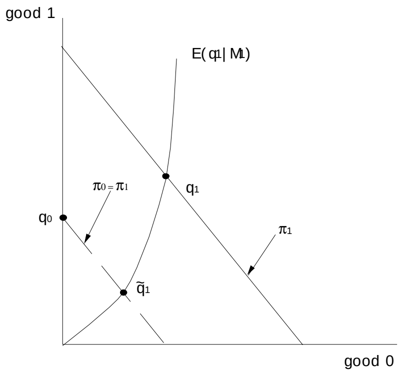
3.2. Improving the bounds

Both the problems just mentioned occur either because the budget constraint for period t lies a long way out from $\mathbf{q}_0$ or because $\mathbf{q}_0$ lies outside the period t

[15] As pointed out by, amongst others, Varian (1982) and Blundell *et al* (2000).

budget constraint. It is intuitively obvious from this that, in order to use the data from periods $1, ..., T$ to provide a tighter bound on π_0^0, we would like to move the budget planes closer to $\mathbf{q}_0$. We would like to find the smallest value of period t expenditure, $\mathbf{p}_t'\mathbf{q}_t$, which still yields a $\mathbf{q}_t$ that is revealed preferred to $\mathbf{q}_0$. That is, we would like to pass the period t budget constraint through the period 0 consumption bundle. The use of the consumer's expansion path is illustrated for the two good, two period case in figure 3.2. The curve $E(\mathbf{q}_1|x)$ is the consumer's expansion path through the bundle chosen in period 1 ($\mathbf{q}_1$). This shows how demands change with the consumer's total budget holding prices constant at $\boldsymbol{\pi}_1$.

Figure 3.2: A two-period, two-good example with an expansion path



Revealed preference restrictions applied to $\mathbf{q}_0$ and $\mathbf{q}_1$ would simply revealed the bound $\overline{\pi}_0^0 = 0$. However, the dashed line shows the budget constraint which makes $\mathbf{q}_0$ just affordable at period 1's prices, and the bundle which would be chosen under these circumstances, $\widetilde{\mathbf{q}}_1$, is given by the intersection of the consumer's expansion path with this budget constraint. As $\widetilde{\mathbf{q}}_1 R \mathbf{q}_0$, the line through $\widetilde{\mathbf{q}}_1$ and $\mathbf{q}_0$ gives the lowest value for π_0^0 consistent with GARP for the data set $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1; \mathbf{q}_0, \widetilde{\mathbf{q}}_1)$. As is evident from the illustration, in the two good case, the lower bound obtained for π_0^0 is simply that the price of good 0 relative to good 1 must be greater than or equal to the period 1 relative price. Therefore, with only two goods, the lower bound for π_0^0 is the highest relative price at which we have since observed it being bought — which is not particularly insightful. However, for more than two goods, the lower bound on the period 0 relative price for good 0 is equal to the highest subsequent observed relative price for good 0 *only* if there is no relative price movement in the other $k = 1, .., K$ goods between period 0 and period t.

By shifting the budget constraint inwards in this manner, we improve the upper bound on the indifference curve passing through $\mathbf{q}_0$. In addition, we can now use information from all periods in which the full price vector is observed rather than just the subset of these periods which are revealed preferred to $\mathbf{q}_0$. That is, we can move budget lines *out* as well as in. We apply this procedure to *all* periods in which the full vector of prices is observed thereby defining a K-dimensional convex set representing the boundary of the $RP(\mathbf{q}_0)$ set (of which all $\widetilde{\mathbf{q}}_t$ are members). As we know that $\widetilde{\mathbf{q}}_t R^0 \mathbf{q}_0$ (since by construction, $\boldsymbol{\pi}'_t \widetilde{\mathbf{q}}_t = \boldsymbol{\pi}'_t \mathbf{q}_0$, and so $\widetilde{\mathbf{q}}_t$ was chosen when $\mathbf{q}_0$ was affordable), we can use the set $\widetilde{\mathbf{q}}_t \in RP(\mathbf{q}_0)$ where $t = 1, ..., T$ bundles to compute an improved lower bound on π_0^0 by the same argument as before. That is, for non-violation of GARP, $\widetilde{\mathbf{q}}_t R \mathbf{q}_0$ implies not $\mathbf{q}_0 P^0 \widetilde{\mathbf{q}}_t$, and so

$$
\begin{aligned}
& \boldsymbol{\pi}'_0 \widetilde{\mathbf{q}}_t \geq \boldsymbol{\pi}'_0 \mathbf{q}_0 \\
\Rightarrow \quad & \pi_0^0 \left(\widetilde{q}_t^0 - q_0^0\right) \geq \boldsymbol{\pi}_0^{k\prime} \left(\mathbf{q}_0^K - \widetilde{\mathbf{q}}_t^K\right) \\
\Rightarrow \quad & \pi_0^0 \geq \boldsymbol{\pi}_0^{k\prime} \left(\mathbf{q}_0^K - \widetilde{\mathbf{q}}_t^K\right) \left(\widetilde{q}_t^0\right)^{-1} \qquad \text{if } \widetilde{q}_t^0 > 0.
\end{aligned}
$$

As discussed in section 3.1 there is no restriction if $\widetilde{q}_t^0 = 0$ for all $t \neq 0$. Thus, each $\widetilde{\mathbf{q}}_t \in RP(\mathbf{q}_0)$ gives a lower bound on π_0^0 — call this set $\pi_0^0(\widetilde{\mathbf{q}}_t)$. As with the $\pi_0^0(\mathbf{q}_s)$ set, this will contain a highest value $\left(\max\left\{\pi_0^0\left(\widetilde{q}_t\right)\right\} \equiv \overline{\pi}_0^0\right)$, and it is this value which should be taken as the lower limit on π_0^0.

Proposition 3.2. *Any* $\pi_0^0 < \overline{\pi}_0^0$ *violates GARP for* $(\boldsymbol{\pi}, \widetilde{\mathbf{q}})$.

Proof.
The proof is analogous with that for Proposition 1.
■

The lower bound obtained by this method of using expansion paths is always an improvement over the bound obtained from raw data (unless $\{\widetilde{\mathbf{q}}_t\} = \{\mathbf{q}_s\}$).

Proposition 3.3. $\max\{\pi_0^0(\widetilde{\mathbf{q}}_t)\} \geq \max\{\pi_0^0(\mathbf{q}_s)\}$.

Proof.
(1) $\boldsymbol{\pi}'_s\mathbf{q}_s \geq \boldsymbol{\pi}'_s\mathbf{q}_0 = \boldsymbol{\pi}'_s\widetilde{\mathbf{q}}_s \Rightarrow \mathbf{q}_s R\widetilde{\mathbf{q}}_s \;\forall\; \mathbf{q}_s \in RP(\mathbf{q}_0)$
(2) The bound $\pi_0^0(\widetilde{\mathbf{q}}_s)$ comes from setting $\boldsymbol{\pi}'_0\widetilde{\mathbf{q}}_s = \boldsymbol{\pi}'_0\mathbf{q}_0 = x_0$
(3) Denote $\widetilde{\boldsymbol{\pi}}_0 = \left(\pi_0^0(\widetilde{q}_s), \pi_0^1, ..., \pi_0^K\right)$
(4) The bound $\pi_0^0(q_s)$ comes from setting $\boldsymbol{\pi}'_0\mathbf{q}_s = \boldsymbol{\pi}'_0\mathbf{q}_0 = x_0$
(5) Suppose that $\pi_0^0(\widetilde{q}_s) < \pi_0^0(q_s)$
(6) Since $\widetilde{\pi}_0^k = \pi_0^k \;\forall\; k \neq 0$ steps (2), (4) and (5) imply that $\widetilde{\boldsymbol{\pi}}'_0\mathbf{q}_s < x_0 = \widetilde{\boldsymbol{\pi}}'_0\widetilde{\mathbf{q}}_s \Rightarrow \widetilde{\mathbf{q}}_s P^0\mathbf{q}_s$, but this is a violation of GARP, from (1)
(7) $\pi_0^0(\widetilde{q}_s) \geq \pi_0^0(q_s) \;\forall\; s \;\Rightarrow\; \max\{\pi_0^0(\widetilde{q}_s)\} \geq \max\{\pi_0^0(q_s)\}$
(8) Since $\{\widetilde{\mathbf{q}}_s\} \subset \{\widetilde{\mathbf{q}}_t\}$, $\max\{\pi_0^0(\widetilde{q}_t)\} \geq \max\{\pi_0^0(\widetilde{q}_s)\} \Rightarrow \max\{\pi_0^0(\widetilde{q}_t)\} \geq \max\{\pi_0^0(q_s)\}$
■

The improved $RP(\mathbf{q}_0)$ set that comes from using expansion paths to calculate $\widetilde{\mathbf{q}}_t$ such that $\boldsymbol{\pi}'_t\,\widetilde{\mathbf{q}}_t = \boldsymbol{\pi}'_t\,\mathbf{q}_0 \;\forall\; t \neq 0$ may not give the tightest upper bound on the indifference curve through q_0 that we can obtain. This can be seen by considering the following. Amongst the $RP(\mathbf{q}_0)$ set, we may be able to find one or more members $\widetilde{\mathbf{q}}_i$ for which there exists some $\widetilde{\mathbf{q}}_j \in RP(\mathbf{q}_0)$, $j \neq i$, such that $\widetilde{\mathbf{q}}_i P^0 \widetilde{\mathbf{q}}_j$, i.e. $\boldsymbol{\pi}'_i\widetilde{\mathbf{q}}_i > \boldsymbol{\pi}'_i\widetilde{\mathbf{q}}_j$. In this case, we can use expansion paths to find a $\widehat{\mathbf{q}}_i$ for each $\widetilde{\mathbf{q}}_i$ such that $\boldsymbol{\pi}'_i\widehat{\mathbf{q}}_i = \boldsymbol{\pi}'_i\widetilde{\mathbf{q}}_j$, i.e. $\widehat{\mathbf{q}}_i R^0 \widetilde{\mathbf{q}}_j$. Since $\widetilde{\mathbf{q}}_j R^0 \mathbf{q}_0$ and $\widehat{\mathbf{q}}_i R^0 \widetilde{\mathbf{q}}_j$ this implies that $\widehat{\mathbf{q}}_i R \mathbf{q}_0$. In addition $\boldsymbol{\pi}'_i\widetilde{\mathbf{q}}_i > \boldsymbol{\pi}'_i\widetilde{\mathbf{q}}_j = \boldsymbol{\pi}'_i\widehat{\mathbf{q}}_i$ tells us that $\widetilde{\mathbf{q}}_i P^0 \widehat{\mathbf{q}}_i$. Hence $\widetilde{\mathbf{q}}_i P^0 \widehat{\mathbf{q}}_i R \mathbf{q}_0$, which implies that $\widehat{\mathbf{q}}_i$ tightens the boundary on the indifference curve passing through $\mathbf{q}_0$ as compared to $\widetilde{\mathbf{q}}_i$. It may be possible to iterate this procedure several times as each improvement may introduce new $\mathbf{q}_i P^0 \mathbf{q}_j$ relationships, where $\mathbf{q}_i$ and $\mathbf{q}_j$ are members of the current best $RP(\mathbf{q}_0)$ set. It might seem that this would allow us to further improve the bound on π_0^0. However, this proves not to be the case as the following proposition shows.

Proposition 3.4. *None of these further boundary improvements on the original improved $RP(\mathbf{q}_0)$ set will enable us to tighten the lower bound on π_0^0.*

Proof.
(1) Take $\widetilde{\mathbf{q}}_i$, $\widehat{\mathbf{q}}_i \in RP(\mathbf{q}_0)$ where $\widetilde{\mathbf{q}}_i P^0 \widehat{\mathbf{q}}_i$

(2) Then $\exists$ a $\mathbf{q}_j \in RP(\mathbf{q}_0)$ s.t. $\boldsymbol{\pi}_i'\widetilde{\mathbf{q}}_i > \boldsymbol{\pi}_i'\mathbf{q}_j = \boldsymbol{\pi}_i'\widehat{\mathbf{q}}_i$, i.e. $\widetilde{\mathbf{q}}_i P^0 \widehat{\mathbf{q}}_i R^0 \mathbf{q}_j R \mathbf{q}_0$
(3) Denote the bound on π_0^0 from setting $\boldsymbol{\pi}_0'\mathbf{q}_j = \boldsymbol{\pi}_0'\mathbf{q}_0 = x_0$ by $\pi_0^0(\mathbf{q}_j)$
(4) Let $\boldsymbol{\pi}_{0j}$ be the price vector for period 0 when π_0^0 is set to $\pi_0^0(\mathbf{q}_j)$
(5) Denote the bound on π_0^0 from setting $\boldsymbol{\pi}_0'\widehat{\mathbf{q}}_i = \boldsymbol{\pi}_0'\mathbf{q}_0 = x_0$ by $\pi_0^0(\widehat{\mathbf{q}}_i)$
(6) Let $\boldsymbol{\pi}_{0i}$ be the price vector for period 0 when π_0^0 is set to $\pi_0^0(\mathbf{q}_i)$
(7) Suppose that $\pi_0^0(\mathbf{q}_j) < \pi_0^0(\widehat{\mathbf{q}}_i)$
(8) Since $\pi_{0i}^k = \pi_{0j}^k \; \forall \; k \neq 0$ steps (3), (5) and (7) imply that $\boldsymbol{\pi}_{0j}'\widehat{\mathbf{q}}_i < \boldsymbol{\pi}_{0i}'\widehat{\mathbf{q}}_i = x_0 = \boldsymbol{\pi}_{0j}'\mathbf{q}_0 \Rightarrow \mathbf{q}_0 P^0 \widehat{\mathbf{q}}_i$, which is a violation of GARP
(9) Hence $\pi_0^0(\mathbf{q}_j) \geq \pi_0^0(\widehat{\mathbf{q}}_i)$, so improving the boundary point from $\widetilde{\mathbf{q}}_i$ to $\widehat{\mathbf{q}}_i$ cannot give a higher lower bound on π_0^0 than can already be obtained from $\mathbf{q}_j$
■

In this section we have described how revealed preference restrictions can be used to bound the virtual price of a new good. This is shown to be the lowest price consistent with the assumption that the data have been generated by a well-behaved utility function. We have also shown how knowledge of the consumer's budget expansion paths can improve the bound and how the levels at which budget planes should be set to give the tightest bound.

4. An empirical application

The problem of new goods is a much-studied one empirically – see for example the papers collected in Bresnahan and Gordon (1997) and references therein. To be able to solve the empirical problem successfully ideally requires data with the following characteristics. Firstly, the data should reflect the introduction of the new good in a timely manner. Many new goods are not separately identified in datasets on consumer expenditure until some time after their launch; usually not until they have proved themselves sufficiently important. Take for example a classic and frequently examined new good: the personal computer. Purchases of computers by households were not recorded in the US Consumer Expenditure Survey until the first quarter of 1982, and they did not appear in the UK Family Expenditure Survey until January 1985. Commercial data sources are usually better but even these suffer lags. An example is Hausman (1997b) where his data on US cellular phones begin in 1985, two years after the cellular phone became a commercial reality. Secondly, in order that preferences might be correctly identified, a period of post-introduction stability is desirable. Much post-introduction quality change or much learning about the good by consumers, for example, complicates the task of estimating stable preferences. Many hi-tech goods are probably subject to *both* rapid learning by consumers and rapid quality improvement quickly after their initial appearance.

Because it satisfies most of these requirements, the particular example of a new good which we have chosen to examine is the UK National Lottery. Spending on the Lottery appeared as a separately identified expenditure category in our data source, the UK Family Expenditure survey (FES), immediately upon commencement in November 1994. This is comparatively rare since spending data on most new goods are usually allocated to residual categories of miscellaneous expenditures. The National Lottery, however, was recognised as interesting enough at the time of its launch (November 1994) for it to warrant separate recording immediately. This makes the effects of its introduction much cleaner in the data. Secondly, unlike many new goods, particularly technological goods, in the time period covered by our dataset the Lottery has not been subject to much change in quality since its introduction[16] — its characteristics have remained largely unaltered (but see the qualification regarding variations in expected value below).

Finally, and probably most importantly, studies of new goods should be interesting. We think the National Lottery is interesting partly because it is not currently included in the Retail Prices Index, and partly because the average budget share for the lottery is significant at around 1%. This budget share is bigger than other categories of consumer expenditure which are more often the subject of new good studies: audio-visual equipment and breakfast cereals are 0.7% and 0.5% respectively. This means that allowing for the welfare effects of its introduction on a cost-of-living index is potentially empirically more important for the lottery than for, say, a new breakfast cereal.

The FES is an annual random sample of around 7,000 households. The Na-

[16] There are now two weekly draws which may have affected the demand for the initial Saturday only draws, however our data ends before these were introduced.

tional Lottery was launched in the middle of November 1994 so, as we do not have a full month's observations, we drop November 1994 from the sample. Note also that, rather unfortunately, in April 1995 the FES stopped distinguishing purchases of National Lottery tickets from other similar products (in particular scratchcards sold by the same organisation that runs the Lottery). This means that we cannot use data past March 1995. This gives us only four months during which the full set of goods, including the Lottery, were available (December 94 to March 95).

There is no household or regional variation in our price data — nor is any reliable price data published at such a level in the UK. The monthly prices for the goods in our data are given by sub-indices of the UK Retail Prices Index (RPI). The construction of the price data for the National Lottery requires some discussion. The expected value of a lottery ticket depends upon the size of the jackpot, the number of tickets sold, the probability that a ticket matches the balls drawn (6 out of 49 draw without replacement); and the size of the jackpot depends on the amount of accumulated undistributed prize money "rolled-over" from the previous week, the proportion of sales revenue used as prize money and the number of tickets sold (see Farrell and Walker (1999) for a description of the Lottery design). Taking all of these factors into account, the expected value of a UK National Lottery ticket is usually around £0.45. Assuming individuals are risk averse or risk neutral, we would not expect people to take part in the National Lottery since it is an unfair bet (£1 for a ticket with expected value of less than £1). But it seems reasonable to assume that participating in the lottery generates some entertainment value that individuals are prepared to pay for. If we assume, following Farrell and Walker (1999), that individuals are locally risk neutral, then the price of the Lottery is the difference between the price of a ticket and its expected value. The assumption of local risk neutrality is plausible for the Lottery since the expected value is so small compared to most incomes. In the four month time period we are looking at, there were thirteen non-roll-over draws, with sufficient sales to make every expected value close to £0.45 (they vary between £0.442 and £0.447). We have four roll-over weeks, with expected values ranging between £0.474 and £0.591. We take the monthly price of the Lottery to be the (sales-weighted) average of the weekly prices over the month[17]. We treat each draw as being in the month in which the Saturday of the draw falls, although of course, not every single purchase of a ticket will occur in that month, particularly for draws at the very beginning of a month.

We take December 1993 as our 0th period, and calculate the reservation price of the National Lottery just under one year before its introduction. We allocate the RPI definition of total non-housing household spending to 23 commodity groups including spending on the National Lottery and we use the published item price indices and weights for the RPI to compute price indices (using RPI construction methodology) for the 22 RPI groups. Details of the components of the groups and summary statistics are provided in the appendix.

4.1. Estimation issues

We assume that households have common, probably nonhomothetic preferences, and that differences in tastes are due to differences in their characteristics. The approach that we propose, therefore, has to be adopted at the level of each individual household to recover household-specific virtual prices. In order to apply the approach we need to observe household demands in the base period, and also to be able to either observe or estimate the budget expansion paths (demands conditional on the household's budget , given prices and characteristics) for each of the post-introduction periods. In sympathy with the nonparametric focus of the revealed preference ideas we would wish to estimate these paths nonparametrically. Our aim is to be able to answer the counterfactual: how will a household's expenditure share patterns change for some *ceteris paribus* change in total expenditure?

The first issue to note is that our dataset is not a panel. Rather it is a time

[17] We are very grateful to Lisa Farrell and Ian Walker for providing us with their data.

series of cross sections and we must use data on different households in different periods to estimate budget expansion paths from which we then predict demands for base period households given their observed characteristics and total budget in each of the post-introduction price regimes. The second issue to note is that in the post-introduction period, on average, one third of the sample does not buy the new good during the two week diary period. We would not expect all of the base period sample to have positive demands after the introduction of the new good and, as discussed in sections 2 and 3, there is no GARP restriction on the virtual price for these households. We therefore need to take account of the possibility of zeros. The final important factor is that we have 2818 observations in all (between 577 and 540 in each period) and this limits the flexibility we have in modelling demands nonparametrically.

In this section we discuss the estimation issues. We begin with a brief outline of the general method which is the estimation period-by-period budget share systems conditional on the log budget by means of kernel regression. We then discuss how to allow for household characteristics bearing in mind the constraints imposed by the data and the constraints which different approaches place on preferences. We also discuss allowing for zero demand for the new good and for the endogeneity of the total budget in such a system.

4.1.1. Nonparametric estimation of the budget share system

The general framework is as follows. Let $\{(\ln x^i, w_j^i)\}_{i=1}^N$ represents a sequence of N household observations on the log of total expenditure $\ln x^i$ and on the jth budget share w_j^i, for each household i facing the same relative prices. For each commodity j, budget shares and total outlay are related by the stochastic Engel curve

$$w_j^i = g_j(\ln x^i) + \varepsilon_j^i \tag{4.1}$$

where we assume that, for each household i, the unobservable term ε_j^i satisfies

$$E(\varepsilon_j^i | \ln x) = 0 \,\text{and}\, Var(\varepsilon_j^i | \ln x) = \sigma_j^2(\ln x) \;\forall \text{ goods } j = 1, ..n \tag{4.2}$$

so that the nonparametric regression of budget shares on log total expenditure estimates $g_j(\ln x)$. We use the following unrestricted Nadaraya-Watson kernel regression estimator for the j'th budget share

$$\widehat{g}_j(\ln x) = \frac{\widehat{r}_j^h(\ln x)}{\widehat{f}^h(\ln x)} \equiv \widehat{w}_j(\ln x) \tag{4.3}$$

in which

$$\widehat{r}_j^h(\ln x) = \frac{1}{N}\sum_{l=1}^{N} K_h\left(\ln x - \ln x_l\right) w_{lj}, \tag{4.4}$$

and

$$\widehat{f}^h(\ln x) = \frac{1}{N}\sum_{l=1}^{N} K_h\left(\ln x - \ln x_l\right), \tag{4.5}$$

where h is the bandwidth and $K_h(\cdot) = h^{-1}K(\cdot/h)$ for some symmetric kernel weight function $K(.)$ which integrates to one. We assume the bandwidth h satisfies $h \to 0$ and $Nh \to \infty$ as $N \to \infty$. Under standard conditions the estimator (4.3) is consistent and asymptotically normal[18]. Provided the same bandwidth is used to estimate each $g_j(\ln x)$, adding-up across the share equations will be automatically satisfied for each $\ln x$ and there is no efficiency gain from combining equations. To compute the demand by household i for commodity j some given total expenditure level from the Engel curve, we utilise our common price regime assumption (dropping the bandwidth)

$$\widehat{q}_j^i = E\left(q_j | \ln x^i, \pi_j\right) = \widehat{g}_j(\ln x^i)\left(\frac{x^i}{\pi_j}\right).$$

4.1.2. Demographic composition and semiparametric estimation

Household expenditures typically display a high degree of variation with respect to demographic composition and we wish to take account of this. Let $\mathbf{z}^i$ represent a $(D \times 1)$ vector of household composition variables relating to household i. A general specification of the within-period Engel curve which took account of this would be

$$w_j^i = G_j(\ln x^i, \mathbf{z}^i) + \varepsilon_j^i \tag{4.6}$$

[18] See Härdle (1990) and Härdle and Linton (1994).

with

$$E(\varepsilon_j^i|\mathbf{z}^i, \ln x^i) = 0 \, \text{and} \; Var(\varepsilon_j^i|\mathbf{z}^i, \ln x^i) = \sigma_j^2(\ \mathbf{z}^i, \ln x^i). \tag{4.7}$$

There are a number of approaches to estimating (4.6). We might wish to estimate a multivariate nonparametric regression. However, the estimation of multivariate densities requires a huge amount of data[19] and the *curse of dimensionality* rules this out here (recall that we have no more than 577 observations in any period)[20]. One simple alternative is to stratify by each distinct outcome of $\mathbf{z}^i$ and estimate separate Engel curves for different groups (we are already doing a version of this by estimating separate within-period/price regime Engel curves). However, the success of this clearly depends on the number of possible outcomes of $\mathbf{z}$ and the number of observations in our dataset. In our case, many of the variables we wish to take account of are continuous (age, years of education etc.) and splitting the sample on grouped versions of these variables would leave cell sizes which are too small for within-group nonparametric regression to be successful.

Placing a simple additive structure on the model we could estimate.

$$w_j^i = \sum_{d=1}^{D} g_{jd}^z \left(z_d^i \right) + g_j^x \left(\ln x^i \right) + \varepsilon_j^i \tag{4.8}$$

This greatly reduces the amount of data required because univariate smoothers can be used to estimate the g_{jd}^z functions and g_j^x thereby avoiding the curse of dimensionality[21]. A particularly simple version of this model is the popular Robinson (1988) partially linear specification

$$w_{ij} = g_j(\ln x^i) + \mathbf{z}^i \boldsymbol{\gamma}_j + \varepsilon_{ij} \tag{4.9}$$

in which w_{ij} is the within-period budget share for the jth commodity in the ith household $\boldsymbol{\gamma}_j$ represents a finite parameter vector of household composition effects for commodity j and $g_j(\ln x^i)$ is some unknown function as in (4.1). The

[19] Silverman (1986) , chapter 4.

[20] An important implication of this, for estimators based on local averaging procedures, is that in high dimensions "local" neighbourhoods are, almost surely, empty and neighbourhood which are not empty are almost surely not "local", see Simonoff (1996).

[21] Hastie and Tibshirani (1990), Linton and Nielsen (1995).

benefit of this partially linear additive approach is that it allows us to condition on demographics and keep the mean response conditional on the total budget flexible (recall that our procedure which recovers the virtual price for a household involves predicting how its demands vary as the total household budget changes, holding the household's characteristics fixed). Assuming that the additive structure is correct (if it isn't then the estimator $\widehat{w}_j^i$ need not even be consistent) then this may be quite attractive. However, this model has been shown to be consistent with utility maximisation only if either the effects of demographics on budget shares are restricted, or if preferences are Piglog and hence budget shares are linear in $\ln x$ for all goods (Blundell, Browning and Crawford (2000), proposition 6).

One generalisation which has been suggested is the extended partially linear model

$$w_{ij} = g_j\left(\ln x_i - \phi(\mathbf{z}_i'\boldsymbol{\alpha})\right) + \mathbf{z}_i'\boldsymbol{\gamma}_j + \varepsilon_{ij} \tag{4.10}$$

in which $\phi(\mathbf{z}_i'\boldsymbol{\alpha})$ is some known function of a finite set of parameters $\boldsymbol{\alpha}$ and can be interpreted as the log of a general equivalence scale for household i (see Blundell, Duncan and Pendakur (1998) and Pendakur (1998)). This extended partially linear model is the shape invariance specification considered in the work on pooling nonparametric regression curves in Härdle and Marron (1990), Pinske and Robinson (1995) and Pendakur (1998). Blundell, Duncan and Pendakur (1998) estimate (4.10) with ϕ set to be the unit function by means of a grid search algorithm over $\boldsymbol{\alpha}$. In their application they estimate Engel curves for a sub-sample of couples with either one or two children and the only demographic variation is the number of children. They are therefore searching over one parameter and z is a dummy. In our application we allow for many household characteristics, some discrete, some continuous and we were unable to apply their grid search approach successfully to a multi-dimensional problem. As an alternative we implement the following. We first estimate within-period quadratic almost ideal demand

(QuAIDS) models

$$w_{ij} = \mathbf{z}_i' \boldsymbol{\alpha}_j + \beta_j \left(\ln x_i - \ln a\left(\mathbf{z}_i\right) \right) + \lambda_j \left[\ln x_i / a\left(\mathbf{z}_i\right) \right]^2 + \varepsilon_{ij} \qquad (4.11)$$

where[22]

$$\ln a\left(\mathbf{z}_i\right) = \alpha_0 + \mathbf{z}_i' \boldsymbol{\alpha}_j$$

to get starting values for $\phi(\mathbf{z}_i'\boldsymbol{\alpha})$. We then conduct a grid search to estimate the term ϕ given the single index $\mathbf{z}_i'\boldsymbol{\alpha}$ as well as estimating g_j and $\boldsymbol{\gamma}_j$.[23] The benefit of the extended partially linear model is that it provides a preference-consistent method for estimating Engel nonparametric curves, but does not imposed the strong, and probably unreasonable, restrictions on preferences implicit in the partially linear model (4.9). As a check on sensitivity we have carried out, but do not present, the empirical analysis using the parametric model (4.11), within-groups nonparametric regression stratified on a rather rough partition of $\mathbf{z}$ (essentially a within-groups version of (4.1), the partially linear model (4.9) as well as the extended partially linear model (4.10). Both the median and mean values for the reservation price were hardly different from those reported below under any of the approaches. Compared to the results reported here, the standard errors were rather wider in the stratified bivariate Engel curve model, and less wide in the fully parametric QuAIDS model.

There remains the issue of unobserved heterogeneity in the cross-section data. In particular we are interested in the effects which unobserved heterogeneity will have on the expected welfare effects of price changes. The model in (4.10) is supposed to give the expected budget shares conditional on the budget and demographics, given the current price vector. We re-introduce the dependence on prices and let $\mathbf{u}$ be the vector of heterogeneity terms with $E[\mathbf{u}|\ln \mathbf{x}, \mathbf{z}, \ln \boldsymbol{\pi}] = 0)$. Blundell, Browning and Crawford (2000) show that a necessary condition for

[22] See Banks, Blundell and Lewbel (1997) fo a description of the QuAIDS model, and Blundell and Robin (1998) for a discussion of estimation methods.

[23] See Blundell, Duncan and Pendakur (1998) and Pinkse and Robinson (1995) for further details.

the expected budget shares recovered by a cross-section analysis of the general type discussed above to be equal to average budget shares is that $w_j = F_j(\ln x, \mathbf{z}, \ln \boldsymbol{\pi}) + \kappa_j(\ln x, \mathbf{z}, \ln \boldsymbol{\pi})' \mathbf{u}$. Given this combination of functional form restrictions and distributional assumptions, our nonparametric analysis recovers $F_j(\ln x, \mathbf{z}, \ln \boldsymbol{\pi})$. This allows for different tastes across agents. In particular, the first-order income responses for agents can vary in any way as can the price responses. Thus a good may be a luxury for one household and a necessity for another. Letting $c = c(\boldsymbol{\pi}, u, z)$ be the cost function, we can show that the effect of a nonmarginal price change $\Delta \ln \pi_j$ on expected welfare can be given as

$$E\left[\frac{\Delta \ln c}{\Delta \ln \pi_j}\right] = w_j + \frac{1}{2}\left(\frac{\partial F_j}{\partial \ln \pi_j} + \frac{\partial F_j}{\partial \ln x} F_j\right) \Delta \ln \pi_j + \frac{1}{2} \frac{\partial \kappa_j}{\partial \ln x} V_u \kappa_j \Delta \ln \pi_j \quad (4.12)$$

where $E[\mathbf{uu}' | \boldsymbol{\pi}, x, z] = V_u$. The third term indicates the bias and from this we can see that this heterogeneity structure gives an exact first order welfare effect and also gives a correct second order effect if the either V_u is zero, or if the heterogeneity term κ_j is independent of the total budget. These conditions are sufficient, weaker ones would allow these terms to cancel, or for them to be small.

4.1.3. Zero demands

As discussed in section 3 and at the beginning of this section, not all households buy the new good after it's introduction and GARP gives us no restrictions on the virtual price for these sorts of households. Any positive price for the new good will support observed behaviour giving a lower bound of zero. As we will only observe demands for households whose reservation prices in greater than the actual price, we have a standard selection bias problem. We need to take account of this in our applied work. In our data 673 households out of 2241 in the post introduction period do not buy the new good. We would expect a roughly similar proportion of our pre-introduction sample of 577 households not to buy tickets after the new good becomes available (typically we might think that this is because the price is too high). To account for this we adopt a two step strategy which is a semiparametric analogue of Heckman (1979). We first estimate a simple parametric linear index probability model in which we define a binary indicator $\left(\delta^i\right)$ to be one if the household has a positive expenditure on the Lottery, zero otherwise. We assume

$$\Pr\left[\delta^i = 1\right] = \Pr\left[\boldsymbol{\psi}'\mathbf{h}^i + e^i \geq 0\right] \tag{4.13}$$

where $\mathbf{h}^i = \left[\ln x^i, \mathbf{z}^{i\prime}, m^i\right]'$ is a vector made up of the log total budget and the household level characteristics included in the Engel curve (4.10) plus the additional variable m^i which embodies our identification restriction. In this case we use years of education.

Under normality the parameters $\boldsymbol{\psi}/\sigma_e$ (where σ_e is the standard deviation of the error term e) can be estimated consistently by the standard probit maximum likelihood estimator. The two step procedure amounts to the substitution of the sample selection correction term $\left(l^i = \phi\left(\boldsymbol{\psi}'\mathbf{h}^i/\sigma_e\right)/\Phi\left(\boldsymbol{\psi}'\mathbf{h}^i/\sigma_e\right)\right)$ computed from the probit, into equation (4.10) as an additional regressor by means of the Robinson (1988) method described above. For discussion of this estimator and an example of this approach see Blundell and Windmeijer (2000).

4.1.4. Endogeneity of the total budget

To adjust for endogeneity we adapt the control function or augmented regression technique (see Holly and Sargan (1982), for example) to our semiparametric Engel curve framework. To avoid cluttered notation we drop the demographic variables in the following discussion. Suppose $\ln x$ is endogenous:

$$E(\varepsilon_j^i | \ln x^i) \neq 0 \text{ or } E(w_j^i | \ln x^i) \neq g_j(\ln x^i). \tag{4.14}$$

In this case the nonparametric estimator will not be consistent for the function of interest. In the application below we take the log of disposable income as the excluded instrumental variable for log total expenditure, and assume that this instrumental variable ζ^i is such that

$$\ln x^i = \eta'\zeta_i + v_i \text{ with } E(v_i|\zeta_i) = 0. \tag{4.15}$$

We make the following key assumption

$$E(w_j^i|\ln x^i, \zeta_i) = E(w_j^i|\ln x^i, v_i) \tag{4.16}$$

$$= g_j(\ln x^i) + v_i\rho_j \ \forall \ j. \tag{4.17}$$

This implies the augmented regression model along the lines of (4.10)

$$w_j^i = g_j(\ln x^i) + v_i\rho_j + \varepsilon_j^i \ \forall \ j \tag{4.18}$$

with

$$E(\varepsilon_j^i|\ln x^i) = 0 \ \forall \ j. \tag{4.19}$$

Note that the unobservable error component v in (4.18) is unknown. In estimation v is replaced with the first stage reduced form residuals

$$\widehat{v}_i = \ln x^i - \widehat{\eta}'\zeta_i \tag{4.20}$$

where $\widehat{\eta}$ is the least squares estimator of η. This is a semi-parametric version of the idea proposed in Newey, Powell and Vella (1999).

4.2. Violations of GARP and inference

To estimate the reservation price π_0^0 we are invoking revealed preference conditions to fill in the missing price observation in the household level dataset $(\boldsymbol{\pi}_t, \widehat{\mathbf{q}}_t^i)$. We are exploiting the maintained assumption that the data were generated by a stable, well-behaved utility function. This assumption is the key to identifying the bound we are interested in. It is, of course, untestable because of the missing price (Varian (1982)). However, the idea that the post introduction period is consistent with the existence of stable preferences is testable because all of the data is available for these periods. And if the household-level dataset $(\boldsymbol{\pi}_t, \widehat{\mathbf{q}}_t^i)$ where $t \neq 0$, did not itself satisfy GARP then the validity of the whole exercise is in question and indeed, no virtual price exists which can rationalise the data $\{\boldsymbol{\pi}_t, \boldsymbol{\pi}_0; \widehat{\mathbf{q}}_t^i, \widehat{\mathbf{q}}_0^i\}$ where $t \neq 0$.

Constructing the test requires that we check that the data contains no cases where $\widehat{\mathbf{q}}_s^i P^0 \widehat{\mathbf{q}}_t^i$ and $\widehat{\mathbf{q}}_t^i R \widehat{\mathbf{q}}_s^i$ where $s, t \neq 0$. GARP tests, in experimental contexts (see Sippel (1997)) typically have a yes/no type of character. In a non-experimental setting subject to sampling variation, as here, we need a stochastic structure which will allow us to assess whether rejections of GARP are statistically significant. We use the idea proposed by Blundell, Browning and Crawford (2000) who use nonparametric predictions of demands at the household level to test for violations of GARP. They use the fact that the nonparametric Engel curve has a pointwise asymptotic standard error so we can evaluate the distribution of each $\widehat{g}_j(\ln x)$ (dropping the demographics etc for ease of notaiton) at any point. Briefly, for bandwidth choice h and sample size N the variance can be well approximated at the point $\ln x$ for large samples by

$$var(g_h(\ln x)) \simeq \frac{\sigma_j^2(\ln x) c_K}{Nh\widehat{f}_h(\ln x)}$$

where c_K is a known constant and $g_h(\ln x)$ is an (estimate) of the density of $\ln x$ and

$$\sigma_j^2(\ln x) = N^{-1} \sum_i \left(\frac{K_h(\ln x - \ln X_i)}{N^{-1} \sum_i K_h(\ln x - \ln X_i)} \right) (w_j^i - \widehat{g}_j(\ln x))^2.$$

Since we can easily compute the pointwise covariance matrix of $\mathbf{g}(\ln x)$ at any comparison point we choose, we can test the significance of GARP violations by formulating GARP conditions in terms of weighted sums of kernel regressions. Note further that we can also use the pointwise covariance matrix to calculate asymptotic standard errors for the reservation price.

Consider the GARP comparison between consumption bundles $\widehat{\mathbf{q}}_s^i$ and $\widehat{\mathbf{q}}_t^i$. This can be written as a comparison of price-weighted sums of kernel regressions. For example, writing the predictions from the Engel curves for household i in period t as $\widehat{\mathbf{w}}(\ln x_t^i)$, the GARP comparison $\boldsymbol{\pi}'_s \widehat{\mathbf{q}}_s^i > \boldsymbol{\pi}'_s \widehat{\mathbf{q}}_t^i$ which implies $\widehat{\mathbf{q}}_s^i P^0 \widehat{\mathbf{q}}_t^i$ can be written more conveniently (for the purpose of constructing a test) in terms of budget shares as

$$\frac{x_s^i}{x_t^i} > \left[\boldsymbol{\pi}_s \frac{1}{\boldsymbol{\pi}_t} \right]' \widehat{\mathbf{w}}(\ln x_t^i)$$

where the total expenditure levels x_s^i/x_t^i can be chosen by the investigator and where $\boldsymbol{\pi}_s(1/\boldsymbol{\pi}_t)$ are known weights. Thus each part of the GARP condition can be tested using a one-sided test against the null $\frac{x_s^i}{x_t^i} = \left[\boldsymbol{\pi}_s \frac{1}{\boldsymbol{\pi}_t}\right]' \widehat{\mathbf{w}}(\ln x_t^i)$. If we reject this null in favour of $\frac{x_s^i}{x_t^i} > \left[\boldsymbol{\pi}_s \frac{1}{\boldsymbol{\pi}_t}\right]' \widehat{\mathbf{w}}(\ln x_t^i)$ and we cannot reject the null $\frac{x_t^i}{x_s^i} = \left[\boldsymbol{\pi}_t \frac{1}{\boldsymbol{\pi}_s}\right]' \widehat{\mathbf{w}}(\ln x_s^i)$ in favour of $\frac{x_t^i}{x_s^i} < \left[\boldsymbol{\pi}_t \frac{1}{\boldsymbol{\pi}_s}\right]' \widehat{\mathbf{w}}(\ln x_s^i)$ then we conclude that $\widehat{\mathbf{q}}_s^i P^0 \widehat{\mathbf{q}}_t^i$ and $\widehat{\mathbf{q}}_t^i R^0 \widehat{\mathbf{q}}_s^i$ and that we have a violation of GARP for some size of test (this is a similar procedure to the approaches used in the literature on tests of distributional dominance, see for example, Beach and Davidson (1983) and Bishop, Formby and Smith (1991)). To check transitivity, we follow Varian (1982) and use these tests to fill in a $(T \times T)$ matrix where a one in the tth row and the sth column indicates that $\boldsymbol{\pi}_t' \widehat{\mathbf{q}}_t^i \geq \boldsymbol{\pi}_t' \widehat{\mathbf{q}}_s^i$ with a zero otherwise. Varian (1982) shows that transitivity can be checked inexpensively using this matrix by means of Warshall's algorithm, and we apply this approach here.

If rejections of GARP for the $t \neq 0$ periods are insignificant for some acceptable size of test, we can proceed with the ideas outlined in above. If there is a significant rejection then we cannot and drop that household. While it is obvious that, given violations in the $t \neq 0$ periods, there cannot exist values for the reservation price π_0^0 which can rationalise the whole dataset *exactly*, the reservation price we calculate will not introduce any *further* violations on top of the statistically insignificant ones that already exist, and so will be consistent with the idea that the data do not statistically reject GARP. To see this imagine that the dataset $(\boldsymbol{\pi}_t, \widehat{\mathbf{q}}_t^i)$ where $t \neq 0$, contained instances of GARP violations, but none which were statistically significant. We then compute the virtual price π_0^0 as described in section 3.2. Can the dataset $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_t; \widehat{\mathbf{q}}_0^i, \widehat{\mathbf{q}}_t^i)$ contain any significant violations of GARP? By propositions 3.1 and 3.2 we know that even if two bundles $\widehat{\mathbf{q}}_t^i$ and $\widehat{\mathbf{q}}_s^i$ violate GARP when compared to each other, the virtual price π_0^0 is chosen such that $\widehat{\mathbf{q}}_0^i$ cannot violate GARP in a direct comparison with either $\widehat{\mathbf{q}}_t^i$ or $\widehat{\mathbf{q}}_s^i$, or any other of the demand bundles (even without allowing for sampling variation in the comparison). It is also the case that $\widehat{\mathbf{q}}_0^i$ cannot violate

GARP when indirectly (transitively) compared to any other bundle either. To see this suppose that we have $\widehat{\mathbf{q}}_s^i R\ \widehat{\mathbf{q}}_t^i$ and $\widehat{\mathbf{q}}_t^i P^0\ \widehat{\mathbf{q}}_s^i$ which is a violation of GARP in our dataset. Suppose that π_0^0 is derived by setting $\boldsymbol{\pi}_0'\widehat{\mathbf{q}}_0^i = \boldsymbol{\pi}_0'\widehat{\mathbf{q}}_s^i$ which implies $\widehat{\mathbf{q}}_0^i R^0 \widehat{\mathbf{q}}_s^i$. We already have by construction $\widehat{\mathbf{q}}_s^i R^0 \widehat{\mathbf{q}}_0^i$. We therefore have $\widehat{\mathbf{q}}_0^i R^0 \widehat{\mathbf{q}}_s^i R\ \widehat{\mathbf{q}}_t^i$ which implies $\widehat{\mathbf{q}}_0^i R\ \widehat{\mathbf{q}}_t^i$, and we also have $\widehat{\mathbf{q}}_t^i P^0 \widehat{\mathbf{q}}_s^i R^0 \widehat{\mathbf{q}}_0^i$ which implies $\widehat{\mathbf{q}}_t^i P\ \widehat{\mathbf{q}}_0^i$. However the strict, but indirectly revealed, preference for $\widehat{\mathbf{q}}_t^i$ over $\widehat{\mathbf{q}}_0^i$ is not a violation of GARP. GARP is only violated if $\widehat{\mathbf{q}}_t^i$ is directly revealed strictly preferred to $\widehat{\mathbf{q}}_0^i$.

4.3. Results

To recap the estimation procedure. In order to apply the ideas outline in section 3 we need to estimate household demands, conditional on household characteristics, at levels of the total budget chosen to give the tightest revealed preference bounds, for each post-introduction set of prices. We estimate Engel curves using the extended partially linear specification, conditional on household characteristics and the log total budget, for the sub-sample of households with positive expenditure on the Lottery, separately within each of the four post-introduction periods, taking account of the endogeneity of the budget and the sample selection. Using the probit model we predict which of our base-period households will buy the new good after its introduction. For households predicted not to consume the new good after its introduction, in the absence of GARP restrictions, we set the lower bound on their virtual price at zero. For the rest of the households we use the semiparametric Engel curves to predict their demands, holding their other characteristics constant, given the set of prices in each period with their total budget in each period set such that they could just afford their base demand bundle (so that all bundles are directly revealed preferred to the base bundle). We then test GARP for each household using the post introduction data. If these data reject GARP at the 95% confidence level then we conclude that no virtual price exists which would support their predicted demands and these households are dropped. For the remaining households (if there are any) we then compute their individual virtual price for the new good as described in section 3.2.

To conserve space we have placed tables, in pdf downloadable format, giving results from the probit (4.13), the reduced form equations (4.15), and the extended partially linear model (4.10) (by period) on the worldwide web at

http://www.ifs.org.uk/staff/ian_c/newgoods.shtml.

Table 4.1 lists the variables used in (4.13), (4.15) and (4.10) and tables A.1 and A.2 in the appendix gives descriptive statistics of the budget shares and explanatory variables by period.

Table 4.1: Variable definitions.

	Variable
$\ln x^i$	Log total household budget.
$\mathbf{z}^i$	[mean age of adults, mean age of adults squared, mean age of children, number of adults, number of children, head of household employee (dummy), head of household retired (dummy), owner-occupier (dummy)]
m^i	Years of education (head of household)
ς^i	Log household income.

The base period is December 1993, one year before our first post-introduction period. We have also investigated the use of other base periods, specifically December 1992 (two years beforehand), and October 1994 (one month before the introduction). For both of these alternatives the mean of the virtual prices recovered was not statistically different from those presented below, nor were the effects of including versus excluding the new good in inflation measures qualitatively different. We also investigated different groupings of goods. In particular we looked at whether the results were sensitive to grouping goods into fewer categories of expenditure. Again we found no significant effect on the mean virtual price, at least amongst households for which a virtual price could be found. However, we did find that the number of households whose demands rejected GARP increased as we grouped commodities together. Testing GARP, then grouping goods and re-testing provides a tests of (weak) separability iff the price and quantity indices of the new groups satisfy Afriat inequalities (Varian (1983) provides an algorithm). One way to investigate this further would therefore be to attempt to compute Afriat numbers for the new groups (if such numbers exist then this means that there exists a sub-utility function which can rationalise demands and prices within the group). We did investigate a couple of groups in this way (grouping fuels together and foods together) but rejected the existence of separable sub-utility functions for these goods.

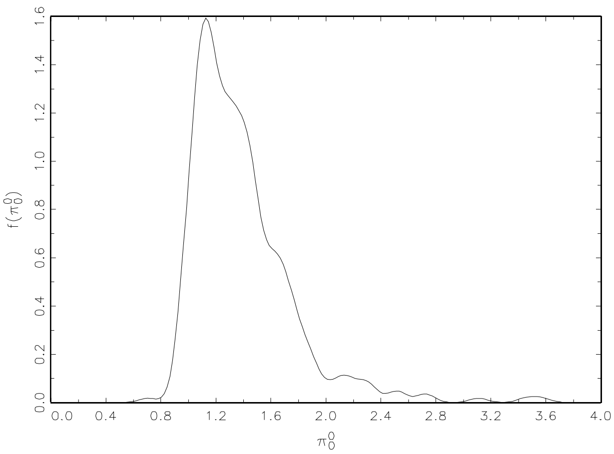
For our base period sample of households (577 in all), we predict that 183 households will have zero demands after the introduction of the Lottery. For these households we set $\pi_0^0 = 0$. For the remaining 394 households we set their budgets in each of the post introduction periods such that they could just afford their base-period bundle and predict their demands given their characteristics. we then use these data to test GARP for each household. There were 49 statistically significant violations of GARP at the 95% confidence level among these households (12% of the sample). These households were then dropped – there existing no virtual price which could rationalise their demands. For the remaining 345 households we calculate their individual virtual price for the Lottery each with an individual standard error. Table 4.2 shows the basic descriptive statistics for the distribution of virtual prices (normalised so that the price of the Lottery in March 1995 is one). The first column is for all households (including those with zero demands, excluding those which reject GARP), and the second concentrates on those expected to have a positive demand. Recall that the means and standard errors are plutocratically weight averages.

Taking all households the average virtual price is 1.334 (i.e. roughly a third higher than the price in March 1995 and also – because there was little change in the price of the lottery – roughly one third higher than the price on introduction) with a standard error of 0.462. The bottom 35% of this distribution all have a

Table 4.2: Virtual price, descriptive statistics, 03/95=1.

π_0^0 (03/95=1)	All households ($n = 528$)	Non-zero demands ($n = 345$)
Mean *(Std Error)*	1.334 *(0.462)*	1.660 *(0.576)*
5th percentile	0.000	0.994
50th percentile	1.123	1.305
95th percentile	1.886	2.167

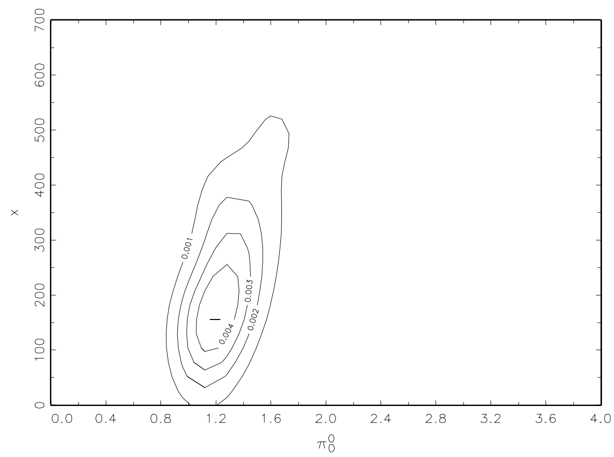
Figure 4.1: The density of the virtual price distribution, non zero values only.



virtual price of zero because they are not expected to buy the new good, however, taking this into account the median for the sample is 1.123 and the 95th percentile is 1.886. Dropping the zeros results in the figures given in the right-hand column. the mean virtual price for households expected to buy the new good is 1.66 (two thirds higher than the introduction price, with a median of 1.305 and 90% of the distribution taking values between 0.994 and 2.167.

Figure 4.1 shows an estimate of the probability density for π_0^0 for the non-zero part of the distribution. This shows a relatively long right-hand-side tail, which is partly due to households with bigger total budgets. Some evidence for this can been seen in figure 4.2 which shows a contour map of the bivariate distribution of virtual prices and (base period) total expenditure, also for the consuming households. This indicates that households with bigger total budgets tend, on

Figure 4.2: The density of the bivariate distribution of the virtual prices and the total budget, non-zero values only.



average, to be predicted to have higher virtual prices.

We next use these virtual prices to measure inflation in the year to December 1994. We present the three indices, the Paasche, the Laspeyres and the Törnqvist calculated inclusive and exclusive of the virtual price of the new good[24]. Note that these are calculated at the household level using household-specific weights from the Engel curves and household-specific virtual prices. Table 4.3 reports the (plutocratically weighted) mean rate of inflation in the year to December 1994 for each of the three measures[25] (each household is weighted by their share out of total expenditure). The Laspeyres, as it is base-weighted and hence gives the fall from

[24] Bounds on true cost-of-living indices can be derived nonparametrically (see for example Varian (1983)). Blundell, Browing and Crawford (2000) show how to derive tightest bounds using revealed preference restrictions and nonparametric expansion paths. In the present case, the upper bound is available and this corresponds to the Laspeyres index (we are grateful to a referee for pointing this out). A lower bound cannot be derived by their method because an upper bound on the virtual price is not available. Blundell, Browning and Crawford (2000) set out the data requirements for the two-sided bounds. They also note that inflation measures based on the Törnqvist index perform the best out of a range of price indices formulae studied (in the sense that it stays between their nonparametric bounds).

[25] The official non-housing inflation rate in the year to December 1994 was 2.3%. Our measure differs because the RPI for that period was based on average weights from the period July 1992 to June 1993 (i.e. the RPI is not a true Laspeyres index) and the RPI uses weight data from a number of sources other than the FES (see Baxter (1997))

the reservation price to the observed end-period price zero weight, is unaffected by the inclusion of the lottery. This is one of the major criticisms of a cost-of-living index interpretation of the Laspeyres-type indices like the UK's RPI. The Paasche, which uses end-period weights, shows a 0.44 percentage point effect. The Törnqvist, which is based on a preferred model of household behaviour[26] which allows for non-homotheticity of preferences and commodity substitution shows an upward bias of 0.156 percentage points caused by excluding the new good.

Table 4.3: Inflation in the year to 12/94, descriptive statistics.

	Mean rate of inflation *(Std Err)*	
Year to 12/94	Including	Excluding
Laspeyres	1.997 *(0.063)*	1.997 *(0.063)*
Törnqvist	1.826 *(0.047)*	1.982 *(0.045)*
Paasche	1.523 *(0.082)*	1.967 *(0.066)*

Figure 4.3 shows the probability density functions for the three prices indices: the Paasche (solid line, to the left), the Laspeyres (sold line, to the right) and the Törnqvist index (dashed line, centre). Figure 4.4 concentrates on the Törnqvist index and shows evidence of non-homotheticity of preferences by illustrating the contours of the bivariate density of the Törnqvist index and total expenditure. This indicates that lower inflation rates were associated with households with higher total expenditures. This is partly to do with the general pattern of relative price changes over the period and the changing pattern of budget shares as the total budget changes, but it is also to do with the cross-sectional variation in the virtual price. That households with higher total expenditure tend to have a higher virtual price was shown in figure 4.2 hence the price fall for the new good over the period is greater for these households, and the inflation rate correspondingly lower. This is further reinforced by the fact that the Engel curve for the Lottery is

[26] See Deiwert (1976).

Figure 4.3: The densities of the distributions of the Paasche, Laspeyres and Tornqvist price indices; all households

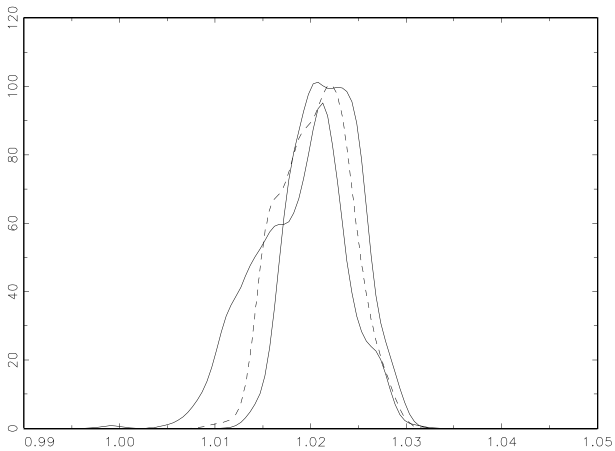
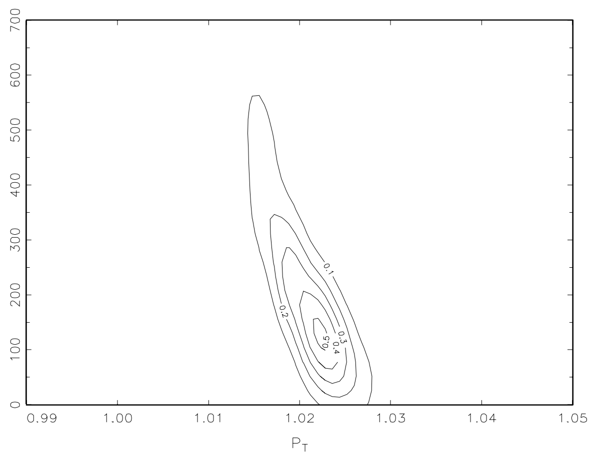


Figure 4.4: The bivariate density of the total budget and the Tornqvist price index; all households.



upward sloping so the weight attached to this price fall is greatest for households with larger total budgets.

5. Conclusions

This paper presents a revealed preference method of calculating the lower bound on the reservation price of a new good for a period prior to the one in which it first exists. This bound is chosen such that the data are consistent with the Generalised Axiom of Revealed Preference and, therefore, it is also consistent with the maximisation of a well-behaved utility function. As a result this bound encompasses all parametric solutions which arise from the estimation of integrable demand systems from the same data. We also present a method for improving the bounds recoverable by predicting household demands conditional on household characteristics, at particular levels of total expenditure given the set of prices in each of the post-introduction period. We argue that this approach has three principal merits compared to parametric estimation. First, it does not require a maintained assumption regarding the form of the utility function. Second, it is computationally simple. Thirdly it can make efficient use of very few post-introduction price observations. We illustrate our technique with UK Family Expenditure Survey data on the National Lottery and compute its reservation price, one year before its introduction. We describe the distribution of the virtual price and provide evidence that the welfare increases associated with the arrival of the Lottery were higher for better-off households. We also show how measures of inflation over this period are affected by the inclusion of the new good and describe how the distributional effects of inflation were more strongly pro-rich when the new good is allowed for.

References

[1] Afriat, S.N. (1965), "The construction of utility functions from expenditure data", *International Economic Review*, **8**, 67-77.

[2] Afriat, S.N. (1973), "On a system of inequalities in demand analysis: an extension of the classical method", *International Economic Review*, **14**, 460-472.

[3] Banks, J. Blundell, R.W. and Lewbel, A. (1996), 'Quadratic Engel Curves and Consumer Demand", *Review of Economics and Statistics*, **4**, 769-788.

[4] Baxter, M. (ed) (1998) *The Retail Prices Index Technical Manual*, London: HMSO.

[5] Beach, C. and R. Davidson, (1983), "Distribution-free statistical inference with Lorenz curves and income shares", *Review of Economic Studies*, **50**, 723-35.

[6] Bishop,J., Formby,J. and W. Smith, (1991), "Lorenz dominace and welfare: Changes in the US income distribution, 1967-1986" *Review of Economics and Statistics*, **73**, 134-39.

[7] Blundell, R., Browning, M. and I. Crawford (2000), "Revealed preference and non-parametric Engel curves", *Institute for Fiscal Studies*, mimeo.

[8] Blundell, R. Duncan, A. and Pendakur, K.(1998), "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, **13**, 435-461.

[9] Blundell, R. and F. Windmeijer (2000), "Identifying demand for health resources using waiting time information", forthcoming, *Health Economics*, August.

[10] Boskin *et al* (1996), "Toward a more accurate measure of the cost of living", *Final Report to the Senate Finance Commission from the Advisory Commission to Study the Consumer Price Index.*

[11] Bresnahan, T. and R. J. Gordon (eds.) (1997), *The Economics of New Goods*, Chicago, University of Chicago Press.

[12] Deaton A. S. and J. Muellbauer (1980), *Economics and Consumer Behaviour*, Cambridge, Cambridge University Press.

[13] Diewert, W. E. (1976), "Exact and superlative index numbers", *Journal of Econometrics*, **4**, 312-36.

[14] Diewert, W. E. (1984), "Group cost of living indexes: Approximations and axiomatics", *Methods of Operations Research*, **48**, 23-45.

[15] Efromovich, S., (1999), *Nonparametric Curve Estimation, Methods, Theory and Applications*, New York: Springer Verlag, Springer Series in Statistics.

[16] Engel, E. (1895), "Die lebenskosten Belgischer arbeiter-familien früher und jetzt", *International Statistical Institute Bulletin*, **9**, 171-80.

[17] Farrell, L and I. Walker, (1999), "The welfare effects of lotto: evidence from the UK", forthcoming *Journal of Public Economics.*

[18] Famulari, M. (1996), "A household-based, nonparametric test of demand theory", *The Review of Economics and Statistics*, **77**, 372-382.

[19] Gorman, W.M. (1956), "A possible procedure for analysing quality differentials in the egg market", *London School of Economics*, mimeo.

[20] Härdle, W. (1990), *Applied Non-Parametric Regression*, Cambridge: Cambridge University Press.

[21] Härdle W. and O.B. Linton, (1994), "Applied nonparametric methods", *Handbook of Econometrics*, vol IV, D.F. McFadden and R.F. Engle (eds), Amsterdam: North Holland.

[22] Härdle W. and J. Marron, (1990), "Semiparametric comparison of regression curves," *Annals of Statistics*, **18**, 63-89.

[23] Hausman, J. (1997a), "Valuation of new goods under perfect and imperfect competition", in Bresnahan, T. & Gordon, R.J., eds., *op cit.*

[24] Hausman, J. (1997b), “Celluar telephones, new goods and the CPI”, *NBER* Working Paper Series, 5982.

[25] Heckman, J.(1979), ”Sample selection bias as a specification error”, *Econometrica*, **47**, 153-161.

[26] Hicks, J. (1940), “The valuation of the social income”, *Economica*, **7**, 105-124.

[27] Lancaster, K.J. (1966), “A new approach to consumer theory”, *Journal of Political Economy*, **74**, 132-157.

[28] Linton, O.B. and J.P.Nielsen., (1995) “A kernel method for estimating structured nonparametric regression based on marginal integration”, *Biometrika*, **82**, 93-100.

[29] Muellbauer, J., (1974), “The political economy of price indices”, *Birkbeck College*, Discussion Paper, 22.

[30] Neary, J.P. and K.W.S. Roberts, (1980), “The theory of household behaviour under rationing”, *European Economic Review*, **13**, 25-42.

[31] Newey, W. Powell, J. and Vella, F. (1999), “Nonparametric estimation of triangular simultaneous equations models”, *Econometrica*, **67**, 565-604.

[32] Pendakur, K. (1998), “Semiparametric etimates and tests of base-independent equivalence scales”, *Journal Of Econometrics*, **88**, 1-40.

[33] Pinske, C. and Robinson, P. (1995), “Pooling nonparametric estimates of regression functions with a similar shape”, *Advances in Econometrics and Quantative Economics*, eds. G. Maddala, P. Phillips and T.N. Srinivisan, 172-195.

[34] Pollak, R.A., (1981), “The social cost of living index”, *Journal of Public Economics*, **15**, 135-150.

[35] Prais, S., (1959), "Whose cost of living?", *Review of Economic Studies*, **26**, 126-134.

[36] Robinson, P.M. (1988), "Root n-consistent semiparametric regression", *Econometrica*, **56**, 931-954.

[37] Rothbarth, E. (1941), "The measurement of changes in real income under conditions of rationing", *Review of Economic Studies*, **8**, 100-107.

[38] Ruiz-Castillo, J,. Ley, E., and M. Izquierado, (1999), "The plutocratic bias in the CPI: Evidence from Spain", FEDEA, Working Paper, Madrid.

[39] Sen, A.K. (1985), *Commodities and Capabilities*, Amsterdam: North-Holland.

[40] Simonoff, J.S. (1996), *Smoothing Methods in Statistics*, New York: Springer Verlag, Springer Series in Statistics.

[41] Sippel, R. (1997), "An Experiment on the Pure Theory of Consumer's Behaviour", *Economic Journal*, **107**, 1431-1444.

[42] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[43] Varian, H. (1982), "The non-parametric approach to demand analysis", *Econometrica*, **50**, 945-973.

[44] Varian, H. (1983), "Non-parametric tests of consumer behaviour", *Review of Economic Studies*, **L**, 99-110.

Appendices

A. Summary statistics

Table A.1: Descriptive statistics, budget shares, by period.

	Month/Year				
Commodity Group	**12/93**	**12/94**	**1/95**	**2/95**	**3/95**
National Lottery	0.0000	0.0083	0.0113	0.0108	0.0118
Wheat	0.0326	0.0325	0.0324	0.0347	0.0351
Meat	0.0531	0.0542	0.0519	0.0543	0.0579
Dairy	0.0323	0.0340	0.0379	0.0373	0.0399
Fruit & Veg	0.0340	0.0353	0.0417	0.0426	0.0430
Other Foods	0.0544	0.0465	0.0463	0.0481	0.0477
Food Out	0.0484	0.0501	0.0586	0.0595	0.0559
Beer	0.0321	0.0354	0.0339	0.0318	0.0327
Wines & Spirits	0.0322	0.0311	0.0167	0.0190	0.0191
Tobacco	0.0313	0.0318	0.0342	0.0378	0.0325
Electricity	0.0379	0.0391	0.0460	0.0474	0.0458
Gas	0.0264	0.0259	0.0343	0.0353	0.0397
Other Fuels	0.0098	0.0041	0.0086	0.0078	0.0070
H'hold Goods	0.1069	0.0966	0.0989	0.0921	0.0924
H'hold Services	0.0611	0.0621	0.0724	0.0650	0.0716
Men's Clothes	0.0193	0.0196	0.0095	0.0065	0.0074
Women's Clothes	0.0254	0.0263	0.0186	0.0174	0.0164
Other Clothes and Shoes	0.0420	0.0411	0.0314	0.0266	0.0294
Personal Goods and Services	0.0576	0.0579	0.0447	0.0493	0.0501
Motoring	0.1121	0.1201	0.1304	0.1390	0.1290
Fares and Travel	0.0243	0.0244	0.0324	0.0269	0.0258
Leisure Goods	0.0749	0.0716	0.0481	0.0543	0.0556
Leisure Services	0.0520	0.0520	0.0595	0.0563	0.0541
n	577	577	564	560	540

Table A.2: Descriptive statistics, explanatory variables, by period.

	Month/Year				
Commodity Group	**12/93**	**12/94**	**1/95**	**2/95**	**3/95**
ln(Income)	5.3766	5.4495	5.4603	5.4920	5.4177
ln(Total Spending)	5.2406	5.2997	5.0738	5.1130	5.1147
Mean age, adults	48.2915	47.1672	45.7382	46.5186	49.2125
Mean age, children	2.6727	2.6527	2.5859	2.3250	2.7287
No. of adults	1.8059	1.8943	1.7996	1.8732	1.8037
No. of children	0.6742	0.5633	0.6294	0.5875	0.6556
Head employed==1	0.4454	0.4818	0.5035	0.4536	0.4481
Head retired==1	0.0849	0.0728	0.0745	0.0857	0.0704
Owner-occupier==1	0.6620	0.6759	0.6188	0.6429	0.7148
Years education>16	0.4454	0.4818	0.5035	0.4536	0.4666
n	577	577	564	560	540