BEST NONPARAMETRIC BOUNDS ON DEMAND RESPONSES

Richard Blundell
Martin Browning
Ian Crawford

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP12/05

Best Nonparametric Bounds on Demand Responses

Richard Blundell
University College London and IFS

Martin Browning
Department of Economics, University of Oxford

Ian Crawford
University of Surrey & IFS

January 2007

Abstract

This paper uses revealed preference inequalities to provide the tightest possible (best) nonparametric bounds on consumer responses to price changes using consumer level data over a finite set of relative price changes. These responses are allowed to vary nonparametrically across the income distribution. This is achieved by combining the theory of revealed preference with the semiparametric estimation of consumer expansion paths (Engel curves). We label these expansion path based bounds as E-bounds. Deviations from revealed preference restrictions are measured by preference perturbations which are shown to usefully characterise taste change and to provide a stochastic environment within which violations of revealed preference inequalities can be assessed.

Key Words: Demand responses, relative prices, revealed preference, semiparametric regression, changing tastes.

JEL Classification: D12, C14, C43.

Acknowledgements: An earlier version of this paper was given as the Walras-Bowley lecture by the first author to the North American Summer Meetings of the Econometric Society at UCLA. We are grateful to participants at that meeting, the editor and three annonimous referees, and to seminar participants at Berkeley, Chicago, LSE, Northwestern and UCL/IFS for helpful comments. The research is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Funding from the ESRC, grant number R000239865 and from the Danish National Research Foundation through its grant to CAM is gratefully acknowledged. Material from the FES made available by the ONS through the ESRC Data Archive has been used by permission of the controller of HMSO. Neither the ONS nor the ESRC Data Archive bear responsibility for the analysis or the interpretation of the data reported here. The usual disclaimer applies.

1 Introduction

This paper develops a new approach to measuring demand responses in the study of consumer behaviour. It concerns the commonly occurring empirical setting in which there is only a relatively small number of market prices but a large number of consumers within each of those markets. This research builds on the earlier results in Blundell, Browning and Crawford (2003) where a powerful method for detecting revealed preference violations was advanced which was used to provide tight nonparametric bounds on welfare costs. The contribution here is to use rich within-market consumer level data together with the minimum of restrictions from revealed preference theory to provide the best bounds on consumer responses to new relative prices. These *E-bounds* are shown to be much tighter than those derived from standard revealed preference analysis.

A common situation in applied economics is that we have a set of observations on agents in a fixed environment with particular realised economic variables and we wish to predict their behaviour in the same environment but with new values for the economic variables. For example, we observe demands at particular sets of prices and total expenditures and we wish to predict demands at a new set of prices and total expenditure. With no other structure, the observed behaviour is totally uninformative about the new situation and literally anything that is logically possible is an admissible prediction. One way around this is to use a parametric statistical model and interpolate (or extrapolate). An alternative is adopt a theoretical position on what generates the observed behaviour and to use the theory and the previous observations to make predictions. Usually this leads to bounds on predicted behaviour rather than point predictions. Demand responses are set identified in the sense of Manski (2003). The relevant question then becomes: how plausible is the theory and how tight are the bounds?

In this paper we derive bounds on predicted demand behaviour from observations on expansions paths (Engel curves) for a finite set of prices and the imposition of the basic (Slutsky or revealed preference) integrability conditions from economic theory. The plausibility of the latter derives from them being, effectively, the observable restrictions from assuming transitivity which is the bedrock of consumer theory in economics. Moreover, the theory implies testable restrictions so it is potentially rejectable. We give the tightest possible bounds on demands given observed expansion paths and the basic (nonparametric) theory, if the latter is not rejected by the former. We find that the data and the theory give surprisingly tight bounds if we consider new situations that are within the span of the observed data.

To introduce our methodology, imagine facing a set of individual consumers with a sequence of relative prices and asking them to choose their individual demands, given some overall budget that each can expend. If they behave according to the axioms of revealed preference their vector of demands at each relative price will satisfy certain well known inequalities (see Afriat (1973) and Varian (1982)). If, for any individual, these inequalities are violated then that consumer can be deemed to have failed to behave according to the optimisation rules of revealed preference. This is a very simple and potentially powerful experimental setting for assessing the applicability RP theory. If, as in an experiment, one can choose the budget at which individuals face each price vector then Proposition 1 of Blundell, Browning and Crawford (2003) shows that there is a unique sequence of such budgets, corresponding to the sequence of relative prices, which maximises the chance of finding such a violation. This is the Sequential Maximum Power path. If experimental data are not available then the Blundell, Browning and Crawford (2003) study also shows how to use expansion paths to mimic the experimental choice of this optimal sequence. Thus providing a powerful method of detecting RP violations in observational as well as experimental studies. In this paper we extend the previous analysis in three ways. The first of these is the derivation of the tightest possible bounds on predicted demands for given relative prices and total outlay, for observational data of the type collected in consumer expenditure surveys. To do this we find it convenient to use the Strong Axiom of Revealed Preference (SARP) rather than the more general GARP condition used in Blundell *et al* (2003). Second, we show exactly when having more data (more observed relative price regimes) is informative in the specific sense of tightening predicted bounds. The third innovation concerns how to deal with rejection of the RP conditions. We show that we can find minimal local perturbations to the expansion paths such that the perturbed data do satisfy the RP conditions and how these perturbations may be interpreted in terms of taste changes. We also discuss explicitly how our analysis relates to the emerging literature on set identification (or partial identification, see Manski (2003)).

To construct bounds we extend the analysis introduced in Varian (1983) by considering expansion paths for given relative prices rather than demands at some point. We label these 'expansion path based bounds' as *E-bounds*. The advantages of the *E-bounds* method developed here are that it can describe the complete demand response to a relative price change for any point in the income distribution without recourse to parametric models of consumer behaviour and it gives the tightest possible bounds, given the data and the theory. The measurement of such price responses are at the centre of applied welfare economics, they are a vital ingredient of tax policy reform analysis and is also key to the measurement of market power in modern empirical industrial economics. Robust measurement is therefore a prerequisite of reliable analysis in these fields of applied microeconomics.

In our empirical analysis, the relative price variation occurs over time and we consider consumer behaviour as it is recorded in standard repeated consumer expenditure surveys such as the US Consumers Expenditure Survey and the UK Family Expenditure Survey. The latter is the source for our empirical analysis. We observe samples of consumers, each of a particular household type, at specific points in time. Assuming consumers are price-takers, we can recover expansion paths by estimating Engel curves at each point in time. We present E-bounds for own and cross price responses using these expansion paths.

Since the expansion paths are estimated, albeit by semiparametric techniques, they are subject to sampling variation. Consequently, violations of the revealed preference inequalities may simply reflect sampling variation rather than rejections by the individuals in the population under study. We develop a test statistic for the revealed preference inequalities and a method for drawing inferences on the estimated demand bounds. Examining our consumer expenditure data, we consider whether revealed preference inequality restricted expansion paths can be found that are not rejected by the data. We find that preferences are generally consistent with RP theory over sub-sequences of time periods in our data but that rejections over longer sequences do occur. Where significant rejections occur, there are a plethora of alternatives to the simple model which has stable preferences for the household (the unitary model). Some of these concern the supplementary assumptions we have to make on aggregation across households, aggregation of goods, the choice of an annual time period etc.. Other alternatives are more fundamental. For example, one alternative is that the household does have transitive preferences but these change over time. We present an explicit measure of such taste changes based on estimated perturbations to preferences. These provide a natural metric against which to measure taste change. Another alternative is that since our sample is for many-person households, the unitary assumption is incorrect and it is the individuals in the household who have stable transitive preferences. In this regard Browning and Chiappori (1998) present evidence, based on a parametric model, that couples do reject the usual Slutsky conditions but not those for a non-unitary collective model. An important rationale for our RP approach is that we can be sure that any rejections of the RP conditions for the unitary model are not due to the choice of functional form. Where significant rejections do occur, the RP inequalities approach can be extended to allow for a collective model; see Cherchye *et al* (2006).

The E-bounds on demand responses we construct are found to be informative. The advantage of adding in more relative price variation is carefully explored, both theoretically and empirically. We show that it is the combination of the new prices *and* the quantity choice implied by the new expansion path that determines whether the new observation is informative. We discuss precisely how such information tightens the bounds. Empirically we show the value of allowing for sampling variation and of introducing perturbations. Bounds on demands are improved and we are also able to detect slow changes in tastes. These bounds on demand responses and the changes in tastes are found to differ across the income distribution.

Freeing-up the variation in relative price responses across the income distribution is one of the key contributions of this research. Historically parametric specifications in the analysis of consumer behavior have been based on the Working-Leser or Piglog form of preferences that underlie the popular Almost Ideal and Translog demand models of Deaton and Muellbauer (1980) and Jorgenson, Lau and Stoker (1982). Even though more recent empirical studies have suggested further nonlinear income terms, (see, for example, Hausman, Newey, Ichimura and Powell (1995), Lewbel (1991), Blundell, Pashardes and Weber (1993), Banks, Blundell and Lewbel (1998)), responses to relative prices at different incomes for these parametric forms remain unnecessarily constrained.

The remainder of the paper is as follows: In section 2 we examine bounds on demand responses and develop a method for generating the best bounds. We also consider how additional data impacts the bounds and in particular the circumstance under which new data are informative. In section Section 3 we describe how we apply our approach to the household level data in the UK Family Expenditure Survey. We examine the semiparametric estimation of expansion paths and the method used to detect revealed preference violations and to impose revealed preference restrictions. In section 4 we estimate E-bounds on cross-price and own-price responses and show that these can be quite narrow. In section 5 we consider imposing revealed preference restrictions and introduce the idea of preference perturbations. Although we find we can reject stability of preferences over the whole period from 1975 to 1999, we can find sub-periods over which stable preferences cannot be rejected. This is found to substantially improve the bounds on demand responses. We also estimate bounds on demands at different percentiles of the income distribution and show that these can differ in important ways. Section 6 concludes.

2 Expansion Path Bounds on Demands

2.1 Defining E-bounds.

We shall be concerned with predicting demands given particular budgets. To this end, we assume that every agent responds to a given budget $(\mathbf{p}, x)$, where $\mathbf{p}$ is a J-vector of prices and x is total expenditure, with a unique, positive demand J-vector:

Assumption 1. Uniqueness of demands: for each agent there exists a set of demand functions $\mathbf{q}(\mathbf{p}, x) : \mathbb{R}^{J+1}_{++} \to \mathbb{R}^{J}_{++}$ which satisfy adding-up: $\mathbf{p}'\mathbf{q}(\mathbf{p}, x) = x$ for all prices $\mathbf{p}$ and total outlays x.

For a given price vector $\mathbf{p}_t$ we denote the corresponding J-valued function of x as $\mathbf{q}_t\left(x\right)$ (with $q^j_t\left(x\right)$ for good j) and refer to this vector of Engel curves as an *expansion path* for the given prices. We shall also have need of the following assumption:

Assumption 2. Weak normality: if $x > x'$ then $q^j_t(x) \geq q^j_t(x')$ for all j and all $\mathbf{p}_t$.

This rules out inferior goods. Adding up and weak normality imply that at least one of the inequalities in this assumption is strict and that expansion paths are continuous.

The question we address is: given a budget $\{\mathbf{p}_0, x_0\}$ and a set of observed prices and expansion paths $\{\mathbf{p}_t, \mathbf{q}_t\left(x\right)\}_{t=1,..T}$, what demands are consistent with these observed demands and utility maximisation? Since we are working with a finite set of observed prices, we characterise consistency with utility maximisation in terms of revealed preference axioms. Since we are requiring that demands be single valued (and not correspondences) we work with the Strong Axiom of Revealed Preference (SARP) rather than the more usual Generalised Axiom (GARP).[1] If at prices $\mathbf{p}_t$ the agent chooses $\mathbf{q}_t$ and we have $\mathbf{p}_t'\mathbf{q}_t \geq \mathbf{p}_t'\mathbf{q}_s$ then we say that $\mathbf{q}_t$ is *directly revealed weakly preferred* to $\mathbf{q}_s$: $\mathbf{q}_t R^0 \mathbf{q}_s$ If we have a chain $\mathbf{q}_t R^0 \mathbf{q}_u$, $\mathbf{q}_u R^0 \mathbf{q}_v$, ...$\mathbf{q}_w R^0 \mathbf{q}_s$ then we say that $\mathbf{q}_t$ is *revealed weakly preferred* to $\mathbf{q}_s$: $\mathbf{q}_t R \mathbf{q}_s$. Given this we have:

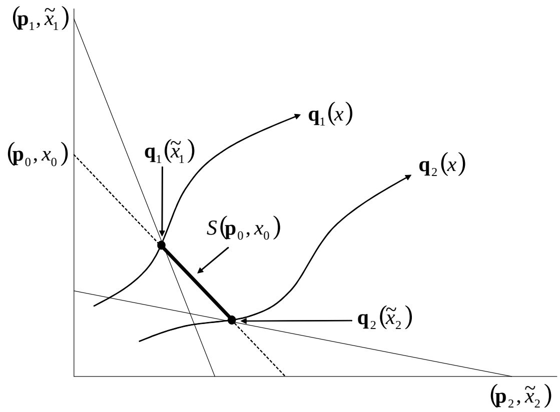
Definition 1 ***SARP:*** $\mathbf{q}_t R \mathbf{q}_s$ *and* $\mathbf{q}_t \neq \mathbf{q}_s$ *implies not* $\mathbf{q}_s R^0 \mathbf{q}_t$ *for all* s, t.

The definition of SARP does not rule that we might have the same demand for two different price vectors.

The basic idea behind our analysis is shown in Figure 1 for a two good, two expansion path example. In this example, the two expansion paths are shown as $\mathbf{q}_1(x)$ and $\mathbf{q}_2(x)$. These intersect the new budget line $\{\mathbf{p}_0, x_0\}$ at $\mathbf{q}_1(\tilde{x}_1)$ and $\mathbf{q}_2(\tilde{x}_2)$ respectively so that $\mathbf{p}_0'\mathbf{q}_1(\tilde{x}_1) = \mathbf{p}_0'\mathbf{q}_2(\tilde{x}_2) = x_0$. We term demand vectors $\mathbf{q}_t(\tilde{x}_t)$ which satisfy $\mathbf{p}_0'\mathbf{q}_t(\tilde{x}_t) = \mathbf{x}_0$ *intersection demands*; the two assumptions on demand above ensure that a unique intersection demand exists for any $\{\mathbf{p}_0, x_0\}$ and $\mathbf{q}_t(x)$. We also show the two budget lines at the intersection demands, labelled $\{\mathbf{p}_1, \widetilde{x}_1\}$ and $\{\mathbf{p}_2, \widetilde{x}_2\}$ respectively. As drawn, the two intersection demands satisfy SARP since neither is revealed weakly preferred to the other. The final step is to display the set of points on the new budget line $\{\mathbf{p}_0, x_0\}$ that are consistent with these intersection points and with SARP. This is shown as the interval labelled $S(\mathbf{p}_0, x_0)$; this set includes the intersection demands and, for two goods, it is closed. We term this set the *support set* for $\{\mathbf{p}_0, x_0\}$. Any point on the new budget that is in the support set $S(\mathbf{p}_0, x_0)$ satisfies SARP for the intersection demands and any point outside fails. For example, a point $\mathbf{q}_0$ within the interior of the support set is weakly revealed preferred to the intersection demands (since $\mathbf{p}_0'\mathbf{q}_0 = x_0 \geq \mathbf{p}_0'\mathbf{q}_t(\tilde{x}_t)$ for $t = 1, 2$), it is distinct from them but the intersection demands are not directly weakly preferred to $\mathbf{q}_0$. Conversely, consider a point $\mathbf{q}_0$ that is not in $S(\mathbf{p}_0, x_0)$. In this case SARP fails immediately since $\mathbf{q}_1(\tilde{x}_1) R^0 \mathbf{q}_0$ (which implies $\mathbf{q}_1(\tilde{x}_1) R \mathbf{q}_0$), $\mathbf{q}_1(\tilde{x}_1) \neq \mathbf{q}_0$ and $\mathbf{q}_0 R^0 \mathbf{q}_1(\tilde{x}_1)$. Finally, the intersection points satisfy SARP and hence are in the support set.

[1] Varian (1982) provides a discussion of the relationship between SARP and GARP; in brief, SARP requires single valued demand curves, whilst GARP allows for set-valued demand correspondences (so that SARP implies GARP).

Figure 1: The Support Set



Given Figure 1 and the definition of intersection demands it is straightforward to define the support set algebraically.[2] Given a budget $\{\mathbf{p}_0, x_0\}$ the set of points that are consistent with observed expansion paths $\{\mathbf{p}_t; \mathbf{q}_t(\tilde{x}_t)\}_{t=1,...,T}$ and utility maximisation is given by the *support set*:

$$S(\mathbf{p}_0, x_0) = \left\{ \mathbf{q}_0 : \begin{array}{c} \mathbf{q}_0 \geq \mathbf{0},\ \mathbf{p}_0'\mathbf{q}_0 = \mathbf{x}_0 \\ \{\mathbf{p}_0, \mathbf{p}_t; \mathbf{q}_0, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,...,T} \text{ satisfy SARP} \end{array} \right\}$$

This differs from the support set definition given in Varian (1982) in two major respects. The Varian definition was based on T observed demand bundles whereas the present definition makes use of T expansion paths. Furthermore this support set is defined using expansion paths evaluated at specific budget levels; the intersection demands. We refer to the intervals defined by expansion paths in this way as *E-bounds* - expansion curve based demand bounds. These bounds on demands for the new budget are *best* in the sense that tighter bounds cannot be found without either observing more expansion paths, imposing some additional theoretical structure over and above utility maximisation (such as quasi-homotheticity or separability) or assuming a functional form for preferences. To show this we define an alternative support set that uses points on the expansion paths that are not necessarily intersection points:

$$S'(\mathbf{p}_0, x_0) = \left\{ \mathbf{q}_0 : \begin{array}{c} \mathbf{q}_0 \geq \mathbf{0},\ \mathbf{p}_0'\mathbf{q}_0 = \mathbf{x}_0 \\ \{\mathbf{p}_0, \mathbf{p}_t; \mathbf{q}_0, \mathbf{q}_t(x_t)\}_{t=1,...,T} \text{ satisfy SARP} \end{array} \right\}$$

[2] In all that follows we assume that the observed prices $\{\mathbf{p}_1, ..., \mathbf{p}_T\}$ are relatively distinct in the sense that $\mathbf{p}_t \neq \lambda \mathbf{p}_s$ for all s, t and any $\lambda > 0$.

The next proposition states that this set is always at least as large as the support set; (the proof is given in the Appendix):

Proposition 1 *If demands are weakly normal then* $S'(\mathbf{p}_0, x_0) \supseteq S(\mathbf{p}_0, x_0)$.

Thus there do not exist alternative bounds (derived from the same data) which are tighter than the E-bounds. The E-bounds therefore make maximal use of the data and the basic nonparametric theory in predicting in a new situation. The properties of the support set are given in the following proposition:

Proposition 2 *(1)* $S(\mathbf{p}_0, x_0)$ *is non-empty if and only if the data set* $\{\mathbf{p}_t,\ \mathbf{q}_t(\tilde{x}_t)_{t=1,\dots T}\}$ *satisfies SARP. (2) If the data set* $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1\dots T}$ *satisfies SARP and* $\mathbf{p}_0 = \mathbf{p}_t$ *for some* t *then* $S(\mathbf{p}_0, x_0)$ *is the singleton* $\{\mathbf{q}_t(\tilde{x}_t)\}$. *(3)* $S(\mathbf{p}_0, x_0)$ *is convex.*

The first statement establishes that there are some predicted demands for $\{\mathbf{p}_0, x_0\}$ if and only if the intersection demands satisfy SARP. The second statement shows that the support set is a single point if the new price vector is one that has been observed. Our decision to consider SARP rather than GARP is largely to give this property; for GARP we would have an interval prediction even for a previously observed price. The convexity is useful when it comes to solving numerically for E-bounds. Note that, contrary to what Figure 1 suggests, with more than two goods the support set is not necessarily closed.

The empirical analysis below requires that we compute E-bounds for given data but the definition of $S(\mathbf{p}_0, x_0)$ is not particularly suited to empirical implementation as it stands. The second set we define gives a set of conditions that allow us to do this in a simple way using linear programming (LP) techniques. If $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\dots,T}$ satisfies SARP we define:

$$S^{LP}(\mathbf{p}_0, x_0) = \left\{ \mathbf{q}_0 : \begin{array}{c} \mathbf{q}_0 \geq \mathbf{0},\ \mathbf{p}_0'\mathbf{q}_0 = x_0, \\ \mathbf{p}_t'\mathbf{q}_0 \geq \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t), t = 1, 2 \dots T \end{array} \right\} \quad (1)$$

The set S^{LP} is closed and convex. We now show that these this set is the same as the support set, except (perhaps) on the boundary of the latter.[3] If we denote the closure of S by $cl(S)$ then we have:

[3] If we had considered GARP rather than SARP then we would have $S = S^{LP}$.

Proposition 3 *(1)* $cl\left(S\left(\mathbf{p}_0, x_0\right)\right) = S^{LP}\left(\mathbf{p}_0, x_0\right)$. *(2)* $S^{LP}\left(\mathbf{p}_0, x_0\right) \backslash S\left(\mathbf{p}_0, x_0\right) = \{\mathbf{q} \in S^{LP}\left(\mathbf{p}_0, x_0\right) : \mathbf{p}_t'\mathbf{q} = \tilde{x}_t$ *and* $\mathbf{q} \neq \mathbf{q}_t\left(\tilde{x}_t\right)$ *for some* $t\}$

As we have seen, for two goods $S\left(\mathbf{p}_0, x_0\right)$ is closed so that it coincides with $S^{LP}\left(\mathbf{p}_0, x_0\right)$ but for more than two goods the set on the right hand side of the second statement is non-empty (so long as $S\left(\mathbf{p}_0, x_0\right)$ is non-empty). $S^{LP}\left(\mathbf{p}_0, x_0\right)$ gives us a feasible algorithm for displaying E-bounds. We first define intersection demands and test for SARP on the intersection demands. If the intersection demands pass SARP, we can then display bounds for each good. For example, to find the supremum predicted value for good j we maximise q_0^j subject to the constraints in (1). This is a standard linear programming problem.

2.2 When is a new observation informative?

We turn now to a consideration of when and how additional observations on expansion paths lead to an improvement in our bounds. We consider the situation in which we have T observed prices $\{\mathbf{p}_1, \mathbf{p}_2, ...\mathbf{p}_T\}$. Take a hypothetical budget $\{\mathbf{p}_0, x_0\}$ and suppose that the corresponding intersection demands satisfy SARP; denote the support set by $S^T\left(\mathbf{p}_0, x_0\right)$. Suppose now that we add one more observed price and expansion path, $\{\mathbf{p}_{T+1}, \mathbf{q}_{T+1}\left(x\right)\}$, find the corresponding intersection demand $\mathbf{q}_{T+1}\left(\tilde{x}_{T+1}\right)$ and compute the new support set $S^{T+1}\left(\mathbf{p}_0, x_0\right)$.

We begin with the following observations. Firstly, the support set cannot increase with the introduction of a new intersection demand; that is $S^{T+1}\left(\mathbf{p}_0, x_0\right) \subseteq S^T\left(\mathbf{p}_0, x_0\right)$ so that additional information weakly shrinks the support set. Secondly, the introduction of a new budget plane and corresponding intersection demand might cause a violation of SARP. If it does then the new support set will be empty (by Proposition 2) and therefore, trivially, we know that the support set will strictly shrink: $S^{T+1}\left(\mathbf{p}_0, x_0\right) = \varnothing \subset S^T\left(\mathbf{p}_0, x_0\right)$. For the rest of this section we will set this possibility aside and assume that the new observation does not cause a violation. Given this we ask when a new observation will be informative and lead to a strict shrinkage of the support set. The first result is trivial but is worth formally recording.

Proposition 4 *If* $\mathbf{p}_{T+1} = \mathbf{p}_0 \neq \mathbf{p}_t$ *for* $t = 1, ...T$, $S^T\left(\mathbf{p}_0, x_0\right)$ *is non-empty and* $\mathbf{q}_t\left(\tilde{x}_t\right) \neq \mathbf{q}_s\left(\tilde{x}_s\right)$ *for some* t *and* s *then* $S^T\left(\mathbf{p}_0, x_0\right) \supset S^{T+1}\left(\mathbf{p}_0, x_0\right)$.

This shows that if the newly observed price just happens to coincide with $\mathbf{p}_0$ then the new support set will be smaller. The proof of this proposition, along with part 2 of proposition 2, establishes that if the intersection points are distinct (which they will almost surely be) then the set of predicted points is a singleton only if the new price $\mathbf{p}_0$ is equal to one of the observed prices. More interesting is the case in which $\mathbf{p}_T \neq \mathbf{p}_0$. To present the characterisation for this, we need one more definition:

Definition 2 *The budget plane* $\{\mathbf{p}_{T+1}, \tilde{x}_{T+1}\}$ *intersects with* $S^T(\mathbf{p}_0, x_0)$ *if there exists some* $\mathbf{q}_0 \in S^T(\mathbf{p}_0, x_0)$ *such that* $\mathbf{p}'_{T+1}\mathbf{q}_0 = \tilde{x}_{T+1}$.

We now present conditions for strict shrinkage of the support set.

Proposition 5 *Given* $S^{T+1}(\mathbf{p}_0, x_0) \neq \varnothing$ *then* $S^{T+1}(\mathbf{p}_0, x_0) \subset S^T(\mathbf{p}_0, x_0)$ *iff the new budget plane* $\{\mathbf{p}_{T+1}, \tilde{x}_{T+1}\}$ *intersects with* $S^T(\mathbf{p}_0, x_0)$.

This says that a new observation is only informative, in the sense that it will strictly shrink the support set if the new budget plane intersects with the initial support set. It is therefore the intersection with the initial support set which is the important feature of any new information rather than the closeness of any new price observation to the $\mathbf{p}_0$ vector of interest. The following three good example serves to illustrate this proposition and to emphasize the point that, if the intersection condition does not hold then a new observation will be uninformative regardless of how close the new price vector is to the hypothetical price vector. Consider the following data for three goods and three periods:

$$\begin{aligned} \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\} &= \begin{bmatrix} 0.64 & 0.19 & 0.90 \\ 0.26 & 0.77 & 0.89 \\ 1 & 1 & 1 \end{bmatrix} \\ \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\} &= \begin{bmatrix} 1.895 & 1.768 & 0.399 \\ 1.571 & 1.141 & 1.901 \\ 1.267 & 1.545 & 1.850 \end{bmatrix} \end{aligned} \tag{2}$$

and take the hypothetical budget given by $[p_0^1, p_0^2, p_0^3] = [0.5, 0.5, 1]$ and $x_0 = 3$.[4] Suppose now that we observe a new price $\mathbf{p}_4$ with an intersection demand:

$$\mathbf{q}_4 = [1, 1, 2]' \tag{3}$$

[4]Note that values for the quantities have been rounded and do not exactly satisfy the intersection demand condition $\mathbf{p}'_0\mathbf{q}_t = x_0$.

We ask: what values of $\mathbf{p}_4$ lead to a strict contraction of the support set? With the values given it is easy to show that any:

$$\mathbf{p}_4 = \mathbf{p}_0 - [\varepsilon, \varepsilon, 0]' \tag{4}$$

does not give a strict contraction for any $\varepsilon > 0$. Thus we can take an new price vector that is arbitrarily close to the hypothetical prices but does not lead to an improvement in the bounds. Conversely, any price vector:

$$\mathbf{p}_4 = \mathbf{p}_0 + [0, \varepsilon, 0]' \tag{5}$$

gives a strict contraction for any $\varepsilon > 0$, even if ε is large. That is, new prices that are far from the hypothetical prices may give a strict contraction of the support set.

As we have shown, adding a new data point may tighten bounds. But it may also lead to a rejection of SARP so that more information is not an unmixed blessing. In the framework presented so far violations of SARP leads to an empty support set so that we are unable to make predictions about the demand curve. In the next section we consider how econometric estimation of expansion paths might provide a stochastic structure in which we can make progress in such a situation.

3 Estimating Bounds on Demand Responses

3.1 Data

In this analysis we take three broad consumption goods: food, other nondurables, and services[5] and examine the E-bounds on demand responses. For this we draw on 25 years of British Family Expenditure Surveys from 1975 to 1999. In many contexts these three consumption goods represent an important grouping as the price responsiveness of food relative to services and to other non-durables is of particular interest. For example, the price responsiveness at different income levels is a key parameter in the indirect tax debate. Although food is largely free of value added tax (VAT) in the UK, the discussions over the harmonisation of indirect tax rates across Europe and the implications of a flat expenditure tax raised uniformly across all consumption items requires a good understanding of food demand responses across the

[5] See the Data Appendix.

income distribution. It is also important in general discussions of cost of living changes across the income distribution. Relative food prices saw some abrupt rises as the tariff structure and food import quotas were changed in Europe early in the period under study. To study further disaggregations of goods with any precision some form of separability has to be assumed.

The Family Expenditure Survey (FES) is a repeated cross-section survey consisting of around 7,000 households in each year. From these data we draw a relatively homogeneous sub-sample of couples with children who own a car. This gives us between 1,421 and 1,906 observations per year and 40,731 observations over the entire period. We use total spending on non-durables to define our total expenditure variable. Table A1 in the Data Appendix provides descriptive statistics for these data. Figure 2 illustrates the trends in mean budget shares over the period. As can be seen, the mean budget share for food exhibits a large fall whereas services are rising steadily over our data period.

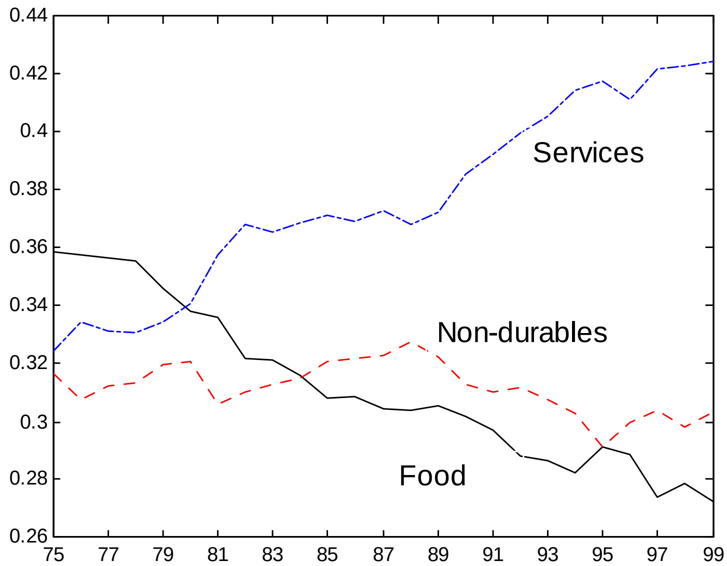


Figure 2: Mean budget shares.

The substantial relative price variation seen in the dated points in Figure 3. This figure shows the scatter plot of the prices of food and services relative to non-durables. The dashed lines in the figure illustrate the convex hull of the relative price data. The relative prices

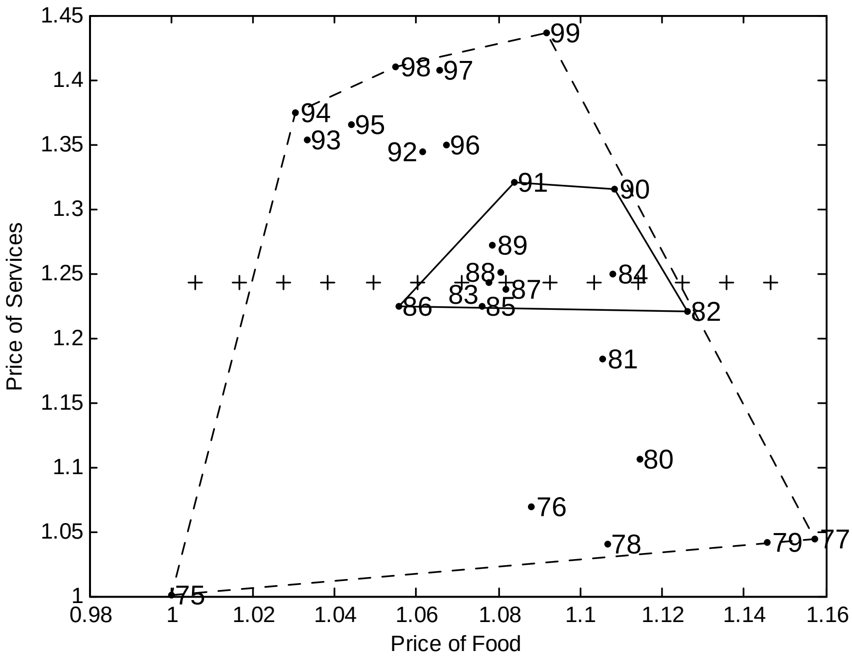


Figure 3: Relative prices, $1975 - 1999$.

show a dramatic change in the mid to late-1970's. The annual price indices for these commodity groups are taken from the annual Retail Prices Index. Nondurables are treated as the numeraire good. We see a steadily rising price for services relative to food and non-durables.

To compute the *E-bounds* on demand responses below we will consider variations in relative prices around a central $\mathbf{p}_0$ vector defined by the mean price vector. We explore a sequence of relative price changes in which the price of food is varied whilst the prices of non-durables and services are held at their mean values. The line of crosses in Figure 3 shows the particular sequence of the $\mathbf{p}_0$ vector we use. Note that this passes through a dense part of the relative price distribution where we might expect (subject to the discussion in section 2) to be able to produce quite informative bounds on demand responses. The path also starts and finishes in areas of very sparse price information outside the convex hull of the prices where, without extrapolation, we would not expect to have much to say about likely demand responses. The solid lines which make up the smaller hull in Figure 3 describe a set of contiguous periods over which revealed preference conditions are not rejected. We return to this case in our analysis of E-bounds bounds on demand responses, we first lay out the estimation of the expansion paths.

3.2 Empirical Expansion Paths

Consumers observed in the same time period and location are assumed to face the same relative prices. In this analysis, relative prices are assumed to vary exogenously across markets but we do allow for the endogeneity of total expenditure. Under this assumption, Engel curves for each location and period correspond to expansion paths for each price regime. Blundell and Duncan (1998) have shown the attraction of nonparametric Engel curves when trying to capture the shape of income effects on consumer behaviour across a wide range of the income distribution. The analysis we present here is applicable to fully nonparametric Engel curves. To make sure we have sufficient support across the income distribution it is often helpful to pool across different household types. As in Blundell, Browning and Crawford (2003) we adopt a shape invariant semiparametric specification for pooling over different demographic groups of households. This semiparametric specification for Engel curves turns out to be a parsimonious, yet accurate, description of behaviour.

Let $\mathbf{d}^i$ represent a $(D \times 1)$ vector of household composition variables relating to household $i = 1, ..., n$. Our specification takes the form

$$w_j^i = g_j \left(\ln x_i - \phi(\mathbf{d}_i' \boldsymbol{\alpha}) \right) + \mathbf{d}_i' \boldsymbol{\gamma}_j + \varepsilon_j^i \tag{6}$$

where w_j^i is the expenditure share for household i on good j.[6]

The x variable is a measure of total outlay by the household on the set of goods under analysis in period t. This is very likely to be jointly determined with the expenditure shares. To account for the endogeneity of $\ln x$ we adopt the control function approach (see Blundell and Powell (2003)).[7] In particular, we specify

$$\ln x_i = \mathbf{z}_i' \pi + v_i \tag{7}$$

where $\mathbf{z}$ are a set of variables which include the demographic variables $\mathbf{d}_i$ and earned income

[6] Throughout this analysis we assume the nonseparable error form (6). As we note below generalisations of this error specification are an important direction for future research.

[7] This is analysed in Blundell, Chen and Kristensen (2003) and compared to a the fully nonparametric instrument variables (NPIV) case. It is found to account quite well for the endogeneity of total expenditure in comparison to a full NPIV approach.

as an excluded instrument. The control function approach assumes that the error term for each consumption good j satisfies:

$$E(\varepsilon_j^i | \ln x_i, \mathbf{d}_i, v_i) = 0. \tag{8}$$

Following Newey, Powell and Vella (1999), semiparametric regression of each share equation using an augmented equation (6) that includes v_i will produce consistent estimates of g_j, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ (see also Blundell and Powell (2003)).

In general, our interest is in consumer behaviour described by the vector of share equations $\mathbf{w} = \mathbf{g}(\ln x, \ln \mathbf{p}, \mathbf{d}, \boldsymbol{\varepsilon})$ where $\boldsymbol{\varepsilon}$ is a $J-1$ vector of unobservable heterogeneity that enters nonseparably in the share equation. An important problem for future research is to estimate the *distribution* of demands across the heterogeneity distribution and not focus on average demands as we do in this paper. In this case, global invertibility is required to identify the complete distribution of demands, see Brown and Matzkin (1998) and Beckert and Blundell (2005). Moreover generalisations of quantile regression are required for estimation, see Matzkin (2006). For the more limited case of local average demands considered in this paper, there is nevertheless a general condition, due to Lewbel (2001), that allows interpretation even in the case of nonseparable unobserved heterogeneity. Assume $F(\varepsilon| \ln x, \ln \mathbf{p}, \mathbf{d}) = F(\boldsymbol{\varepsilon}|\mathbf{d})$ so that preference heterogeneity conditional on demographics is independent of prices and total outlay. The covariance between budget shares and the responsiveness of these to changes in log total outlay, conditional on the observable determinants of demand is defined as

$$H(\ln x, \ln \mathbf{p}, \mathbf{d}) = cov\left(\frac{\partial \mathbf{g}}{\partial \ln x}, \mathbf{g}' \mid \ln x, \ln \mathbf{p}, \mathbf{d}\right)$$

Lewbel (2001) shows that *average* demands of rational consumers satisfy integrability conditions *iff* $H(.)$ is symmetric and positive semidefinite.[8] If H is small relative to the the *Slutsky* matrix for these average demands, then the system will be 'close' to integrable.

3.3 Testing Revealed Preference Restrictions

Because the expansion paths are estimated they will be subject to sampling variation. Consequently violations of SARP may simply be due to estimation error and we now consider the

[8]For example, in the Almost Ideal Demand system (Deaton and Muellbuaer, 1980), heterogeneity in the $a(p)$ parameters would automatically satisfy this condition.

use the stochastic structure of the estimated Engel curves to allow for this. The starting point is the suggestion by Varian (1985) for testing optimising behaviour. We develop this idea by using the precision of the semiparametrically estimated expansion paths at the specific income levels corresponding to the intersection demands. We can then construct a significance test for violations of the revealed preference conditions and provide a measure of the precision for the estimated E-bounds.

Let Σ denote the set of all intersection demands which are SARP-consistent for prices $\{\mathbf{p}_t\}_{t=1,\ldots,T}$ and the new budget $\{\mathbf{p}_0, x_0\}$. If the intersection demands violate SARP then $\{\mathbf{q}_t(\widetilde{x}_t)\}_{t=1,\ldots,T} \notin \Sigma$. In order to estimate the restricted intersection demands, $\mathbf{q}_t^*$, that satisfy the revealed preference conditions, consider the solution to following the minimum distance problem

$$\begin{aligned} \min_{\{\mathbf{q}_t^*\}_{t=1,\ldots T}} \{L\left(\{\mathbf{q}_t^*\}_{t=1,\ldots T}\right) &= \sum_{t=1}^{T} \left(\mathbf{q}_t^* - \mathbf{q}_t(\widetilde{x}_t)\right)' \Omega_t^{-1} \left(\mathbf{q}_t^* - \mathbf{q}_t(\widetilde{x}_t)\right)\} \qquad (9) \\ \text{subject to } \{\mathbf{q}_t^*\}_{t=1,\ldots T} &\in \Sigma, \qquad \mathbf{q}_t^* \geq 0, \qquad \mathbf{p}_0' \mathbf{q}_t^* = x_0 \ \forall \ t. \end{aligned}$$

This weighted distance function defines a L^2 loss function where the weight matrix Ω_t^{-1} is the inverse of the variance-covariance matrix of the estimated unrestricted intersection demands $\mathbf{q}_t(\widetilde{x}_t)$. These estimated demands are defined by the semiparametric conditional mean estimator of the Engel curves for each t, evaluated at $\widetilde{x}_t$ and using n consumer level observations in each period. The solution to (9) defines the nearest set of nonnegative intersection demands which are consistent with SARP. The support set can then be estimated using the restricted intersection demands.

The distance function (9) evaluated at the restricted intersection demands provides a test statistic for the revealed preference null hypothesis. The revealed preference conditions are a set of moment inequalities and if the intersection demands satisfy the revealed preference restrictions then the objective function will be minimised at zero. There will therefore be a range of estimates at which the statistic can be identically zero. This statistic falls in to the class of criterion functions for set identified models investigated in Chernozhukov, Hong and Tamer (2006). To carry out inference we construct a $b_n < n$ sub-sample bootstrap critical value for this statistic (see Andrews and Guggenberger (2006)) and also use the sub-sample bootstrap to provide inference on the estimated E-bounds.

An interpretation of the restricted intersection demands $\mathbf{q}_t^*$ is as local perturbed demands that conform to SARP. That is they measure the minimum perturbation to tastes necessary to ensure preference stability. Consequently, the perturbations, $\mathbf{q}_t^* - \mathbf{q}_t(\widetilde{x}_t)$, themselves are likely to be of interest: random taste behaviour would be reflected in a corresponding random pattern in perturbations; slowly changing tastes would be reflected by a systematic evolution of these perturbations. We investigate the estimated perturbations in our empirical analysis below.

4 Empirical E-Bounds on Demand Responses

To construct E-bounds in our application to the FES data, we first estimate the three-good Engel curve system as described in the previous section. Using the estimated expansion paths we recover the intersection demands for each $\{\mathbf{p}_0, x_0\}$ and check the revealed preference conditions for $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}$. Perhaps unsurprisingly these data contain some violations of SARP. Searching for contiguous periods over which we cannot reject stable preferences we find the periods 1982 through 1991 satisfy SARP. The potential cost of discarding other periods can be seen by looking back to the smaller convex hull in Figure 3 which shows the price data corresponding to the subset of SARP-consistent intersection demands. A comparison of the two convex hulls shows the reduction in the space spanned once SARP-violating intersection demands have been dropped.

In Figure 4 we present the E-bounds on the own demand curve for food at the median income using the reduced set of SARP-consistent observations. As can be seen from a comparison with Figures 3, the bounds on the demand curve are particularly tight when the $\mathbf{p}_0$ vector is in the dense part of the observed price data. Outside the convex hull of the data the E-Bounds widen and we cannot rule out extreme responses (such as households not buying food if the price rises by more than 5%).

In Figures 5 and 6 we present the corresponding E-bounds for cross price responses. These figures show the power of E-Bounds: through the use of revealed preference inequalities and without appealing to parametric models or extrapolation we have been able to construct tight bounds on own and cross price responses. They also show the limitations in the sense that

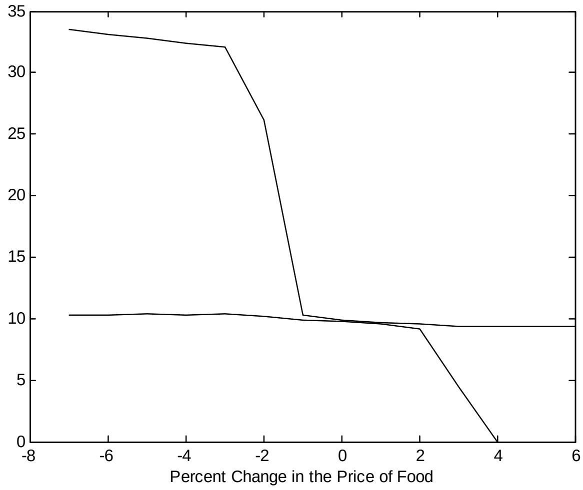


Figure 4: Own price demand bounds for food.

price experiments (by the standards typical of many policy simulation studies) can easily take on values outside the range of observed price variations and produce bounds which are necessarily very wide.

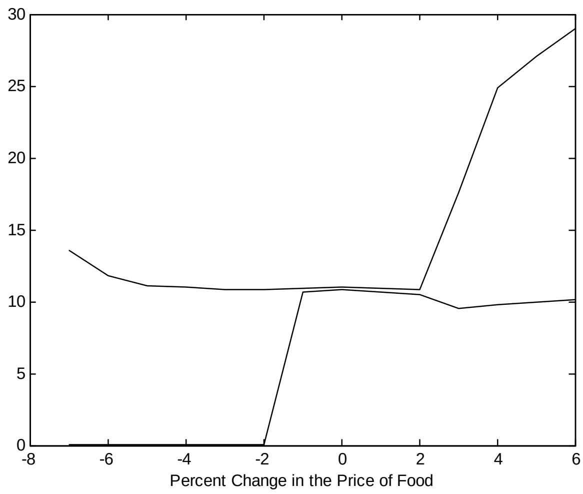


Figure 5: Cross-price demand bounds for non-durables.

To construct E-bounds on demand curves we have exploited movements along the estimated expansion paths and it is reasonable to ask whether this involved comparisons across a wide range of incomes. In fact we find that these comparisons do not require implausibly

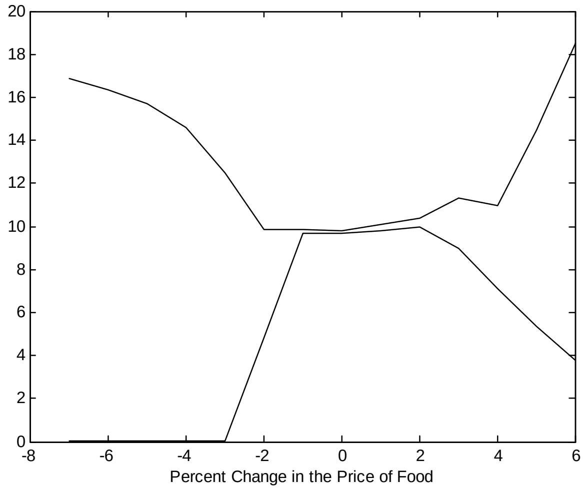


Figure 6: Cross-price demand bounds for services.

wide variations across income levels. For example, to construct the curve in Figure 4, 10 intersection demands are required. The range of income went from the 56th percentile in 1982 to the 40th percentile in 1991. This shows a further attractive feature of the local nature of this analysis: nonparametric Engel are only required over a limited range of the income distribution when constructing a specific demand bound at a particular income percentile.

5 Imposing Revealed Preference Restrictions

5.1 Constrained E-Bounds

In the previous section we searched for a contiguous period of SARP-consistent demands and simply discarded those intersection demands which caused violations. In this section we investigate the improvements which can be made if we impose SARP-consistency across relative prices where violations occur using the criterion function (9). In principle this should further tighten the bounds because (i) it will expand the convex hull of the prices in use thereby potentially increasing the range over which we can tightly bound the demand curves, and (ii) the extra information may include budget planes which intersect with the support sets which underlie Figure 4, 5 and 6. By Proposition 5 this will strictly shrink the bounds.

We begin our examination of SARP-constrained E-bounds by constraining intersection demands at all the relative prices such that they are theory consistent at each $\{\mathbf{p}_0, x_0\}$ budget

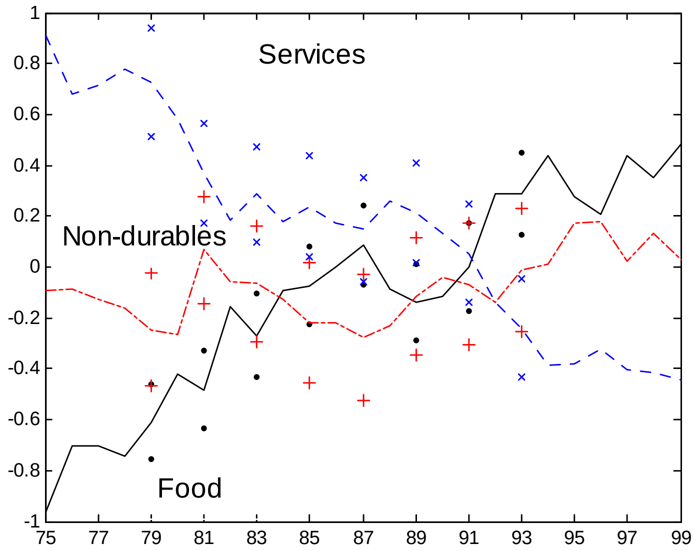


Figure 7: Demand perturbations

along the demand curve. We do this using the weighted minimum distance procedure (9) using the inverse of the pointwise variance covariance matrix of the estimated expansion paths (evaluated at the intersection demands values) as the weight matrix.

Figure 7 illustrates the perturbations $\mathbf{q}_t^* - \mathbf{q}_t(\widetilde{x}_t)$ by good and period in the center of the demand curves at the median income. Since there are three goods and 25 intersection demands (one each of the 25 annual Engel curves) there are 75 perturbations. No structure is imposed on these perturbations other than that the restricted intersection demand are nonnegative and satisfy SARP. The figure also contains 95% pointwise confidence intervals for some periods which suggests an extended period in the centre of this range where we may be able to find a stable representation of preferences.

If demand behaviour were completely random, or if it were rational but contaminated with classical measurement error, then we might expect that the perturbations would reflect this. Slowly changing tastes on the other hand would be reflected by a systematic evolution of these perturbations. In fact, the adjustments needed to make these data theory-consistent seem to follow a reasonably systematic pattern. The perturbation to food demand, for example, is generally increasing over time. It is negative in the early data indicating that the earlier food demands needs to be adjusted downwards and the later observations need to be adjusted upwards. This can be interpreted as the perturbation necessary to adjust for a slow change in preferences away from food towards services. Comparing this to Figure 2 we can see that this adjustment would go some way to slowing the apparent decline in the food share over the period.

The resulting own demand curve E-bounds are illustrated in Figures 8 and 9 along with, for comparison, the E-bounds recovered by dropping SARP rejections (Figure 4: the solid lines). As can be seen, there is an improvement/narrowing of the bounds when all of the observations are used and constrained to be revealed preference consistent, compared to the case in which some data points are just dropped. Nevertheless, the improvement is quite small in the central part of the demand curve (see Figure 9) where the existing bounds were already fairly tight. Note also that there is no reason for the new bounds to lie everywhere inside the old bounds. Whilst the addition of theory-consistent data always weakly tightens the bounds, the data being added here contains violations and has been perturbed as a result. Consequently the restricted intersection demands can lead to the bounds widening at some relative price points. The general pattern of the bounds are similar however, with typically wider bounds the further the new price vector is from the most dense part of the observed price distribution.

As before it is useful to examine the range of incomes (total budgets) over which comparisons have been made to construct these E-bounds for the median income consumer. Again the range is quite limited going from a maximum of the 60 percentile in the mid-1970s to the 40th percentile at the end of the 1990s.

5.2 Price Responses Across the Income Distribution

The demand bounds on price responses presented above have been constructed at the median income (expenditure). But we might expect demand responses to vary with income levels. Figure 10 shows how the demand bounds vary according to the total budget. Three sets of bounds are calculated corresponding to the 25th, 50th and 75th percentiles of the x_0 distribution (the solid lines for the median are identical to the dashed lines in the preceding figure over this range). It is clear from this figure that there is not a single elasticity that

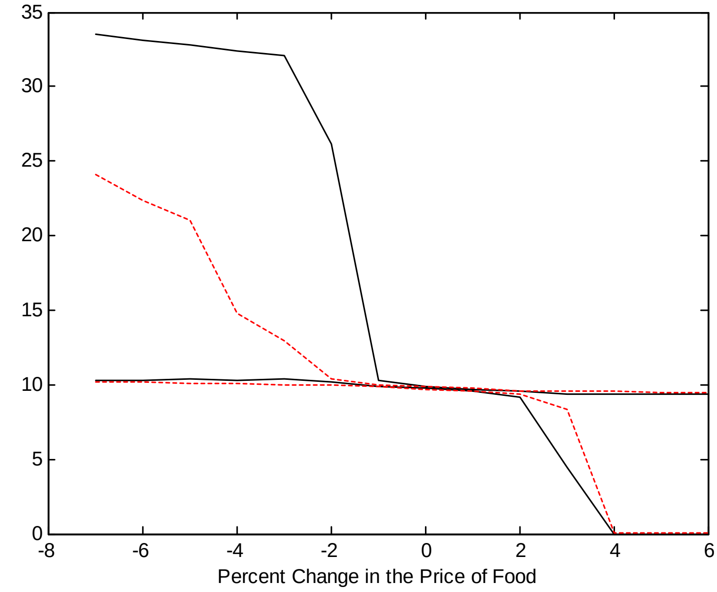


Figure 8: Constrained E-Bounds for Food

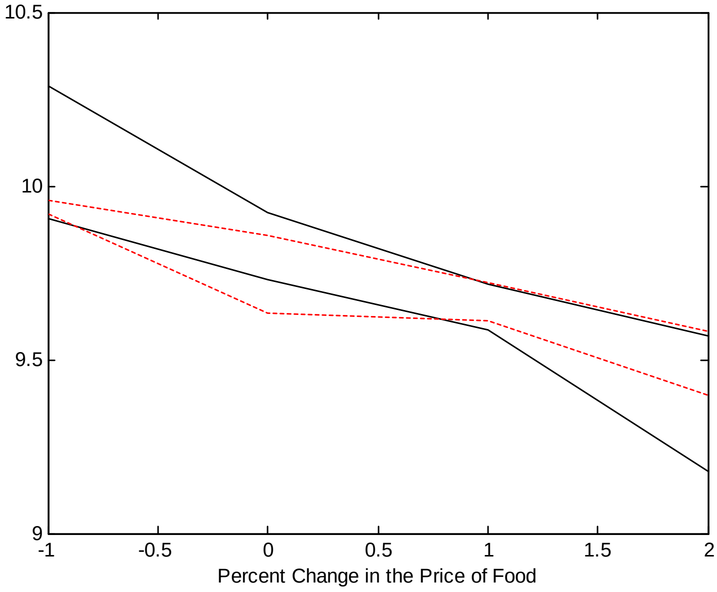


Figure 9: Constrained E-Bounds for Food - Detail

summarises price response behaviour. Price responses appear to be quite variable both along each demand curve and also across income levels. The range of price responsiveness highlights the local nature of our nonparametric analysis. The price responsiveness are local to both income and relative prices. Unlike in the Stone-Geary model, for example, there is no reason why price elasticities should not be increasing or decreasing with income. For some broad aggregates such as food a price elasticity which is increasing with income would seem sensible while for more disaggregated food items - rice and potatoes, for example - the reverse could equally well be true.

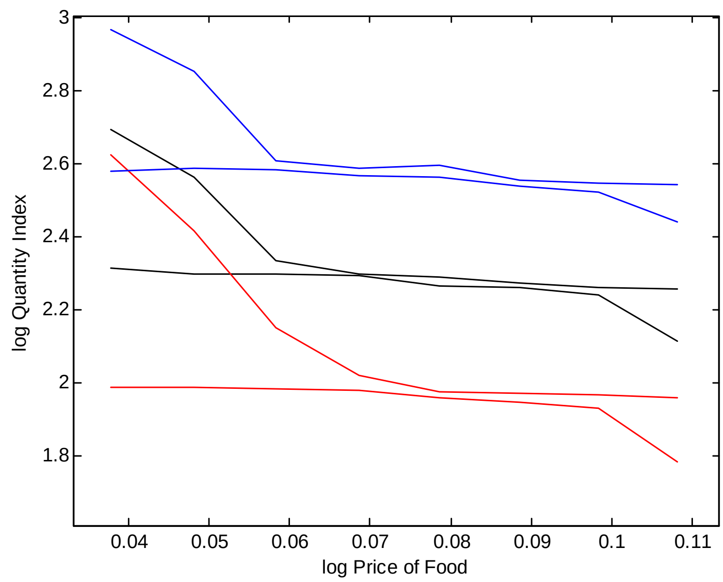


Figure 10: Demand Bounds for Food By Budget Percentile (log-log)

5.3 Revealed Preference Violations and Best RP-Consistent Demands

In the analysis so far we have investigated two approaches to dealing with violations of revealed preference: dropping offending intersection demands and imposing SARP restrictions on all of the data. We have seen that the perturbations required to make the intersections demands SARP-consistent are trended and are consistent with a story of systematic taste change over the period. Using a 20% subsampling critical value for the statistic (9) convincingly rejects

SARP with a p-value very close to zero.[9] Simply imposing SARP across all the periods in this data is clearly invalid.

We therefore return to our SARP-consistent dataset. As the perturbations in Figure 7 suggest it may be possible to add additional intersection demands outside this period without rejecting the SARP restrictions. Using the criteria (9) we found that expanding the set of intersection demands by adding the periods 93-95 did not reject SARP, the 20% subsample p-value was 0.08. The extended convex hull of the relative price space spanned by these periods is shown in Figure 11.

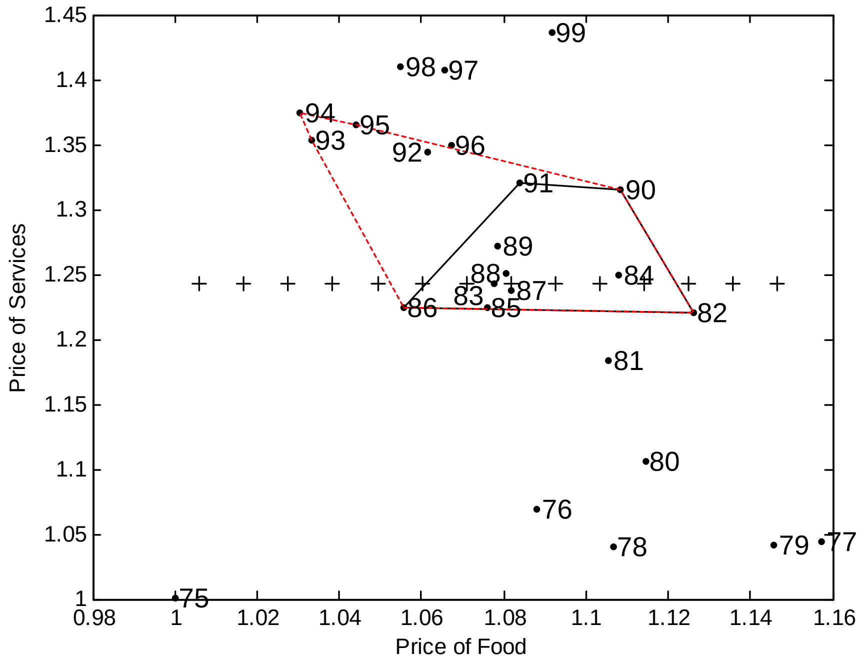


Figure 11: Price scatter plot of the extended period

Using this extended set of intersection demands the resulting E-bounds on the own price demand curve is in Figures 12 and 13. Figure 12 shows the demand curve using the original SARP-consistent subset of the data (solid lines), and the demand curve obtained by imposing SARP on the extended demand subset of the data.

Figure 13 gives a detailed view of the central part of the demand curve. At '0' the E-bounds using the extended period are [9.6742, 9.8694] and the sub-sample 95% confidence

[9]Rejection also occurs using a 25% and a 15% subsample.

region is [9.4987, 9.9516]. As in our discussion of Figure 8, because the extended period uses restricted intersection demands the new E-bounds do not necessarily lie everywhere inside the bounds that simply use the 82-91 period.

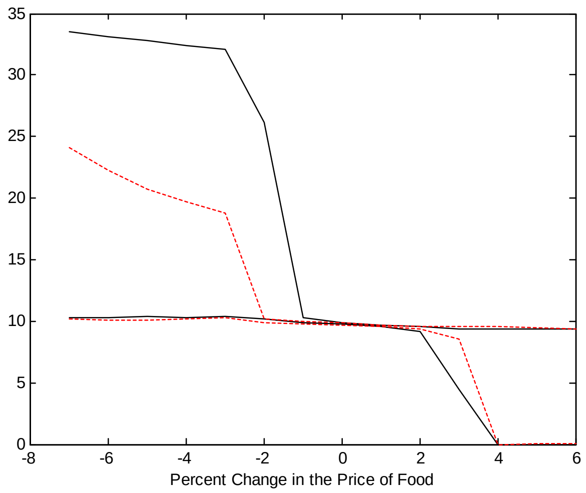


Figure 12: Best RP-Consistent E-Bounds for Food

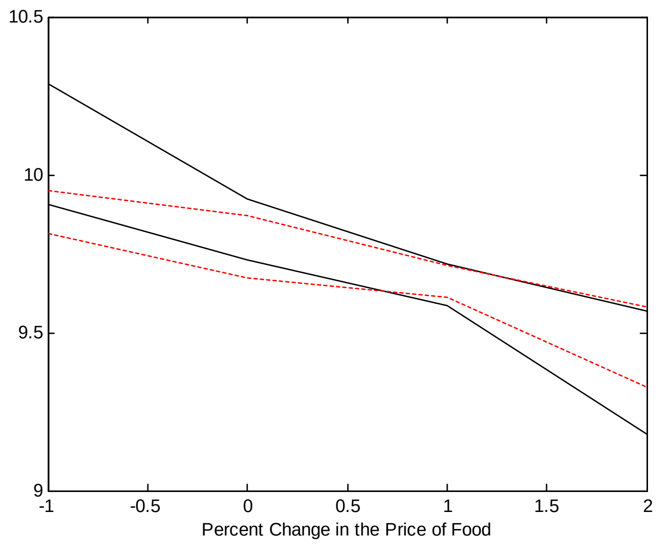


Figure 13: Best RP-Consistent E-Bounds for Food: Detail

6 Summary and Conclusions

The aim of this paper has been to bound demand responses for new sets of relative prices and total expenditure using revealed preference inequalities alone. We have focussed on the situation where we see only a relatively small number of market prices but a large number of consumers in each of these markets. Our approach has been to make use of this rich within-market consumer-level data to estimate income expansion paths conditional on prices. We have then shown how to derive best bounds on predicted demand behaviour from a combination of observations on expansions paths and the imposition of the basic (Slutsky or revealed preference) integrability conditions from economic theory. We find that these *E-bounds* give surprisingly tight bounds, especially where we consider new situations that are within the span of the relative price data in observed markets.

The *E-bounds* approach to measuring consumer behaviour allows price responses to vary nonparametrically across the income distribution by exploiting micro data on consumer expenditures and incomes over a finite set of discrete relative price changes. We have introduced the concept of preference perturbations, local to each income percentile, which characterise the degree of congruence with RP conditions and provide a useful metric for describing taste change.

References

[1] Afriat, S.N. (1973),"On a System of Inequalities in Demand Analysis: An Extension of the Classical Method", *International Economic Review,* 14, 460-472.

[2] Andrews, D.W.K. and P. Guggenberger (2005), 'The Limit of Finite Sample Size and a Problem with Subsampling', mimeo, June, Yale.

[3] Arrow, K. (1958), "The Measurement of Price Changes", Joint Economic Committee, US Congress, Washington, D.C. Reprinted in Collected Works of Kenneth J. Arrow, Vol. 3: Individual Choice under Certainty and Uncertainty, Harvard University Press, Cambridge.

[4] Banks, J. Blundell, R.W. and Lewbel, A. (1996), "Tax Reform and Welfare Measurement: do we need Demand System Estimation?", *Economic Journal*, 106, 1227-1241.

[5] Banks, J. Blundell, R.W. and Lewbel, A. (1997), "Quadratic Engel Curves, Indirect Tax Reform and Welfare Measurement", *Review of Economics and Statistics*, LXXIX, 4, 527-539.

[6] Beckert, W. and R.W. Blundell (2005), "Invertibility of Nonparametric Stochastic Demand Functions", Cemmap Working Paper CWP09/05, July.

[7] Bierens, H. and H. Pott-Butter (1990): "Specification of Household Engel Curves by Nonparametric Regression", *Econometric Reviews* 9: 123-184.

[8] Blundell, R., M. Browning and I. Crawford (2003), "Nonparametric Engel Curves and Revealed Preference" *Econometrica,* Vol.71, No.1. January, 205-240.

[9] Blundell, R., Chen, X. and D. Kristensen, (2003) "Semiparametric Engel Curves with Endogenous Expenditure", CeMMAP Working Paper, IFS, CWP 15/03.

[10] Blundell, R. and Duncan, A. (1998) "Kernel Regression in Empirical Microeconomics" *Journal of Human Resources*, 33, 62-87.

[11] Blundell, R. P. Pashardes and G. Weber, (1993), "What do we Learn about Consumer Demand Patterns from Micro Data?", *American Economic Review*, 83, No.3, 570-597

[12] Blundell, R. and J. Powell, (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models", Chapter 8 in Dewatripont, M., Hansen, L. and S. J. Turnovsky (eds), *Advances in Economics and Econometrics*, Cambridge University Press, ESM 36, pp 312-357.

[13] Brown, D.J. & R.L. Matzkin (1998) "Estimation of Simultaneous Equations Models, with an Application to Consumer Demand" Cowles Foundation Discussion Paper.

[14] Cherchye L, B. De Rock and F. Vermeulen (2006), "The collective model of household consumption: a nonparametric characterization", CentER Discussion Paper, 2004-76, Tilburg, CentER.

[15] Chernozhukov, V. H. Hong and E. Tamer (2006), 'Estimation and Confidence Regions for Parameter Sets', mimeo, MIT, May.

[16] Deaton A. S. and J. Muellbauer (1980),"An Almost Ideal Demand System", *American Economic Review*, 70, 312-36.

[17] Härdle W. and Jerison, M. (1988), "The Evolution of Engel Curves over time", Discussion paper no. A-178. SFB 303, University of Bonn.

[18] Härdle W. and Marron J. (1990), "Semiparametric Comparison of Regression Curves," *Annals of Statistics*, 18, 63-89.

[19] Jorgenson, D.W., L.J.Lau, and T.M.Stoker (1982),"The Transcendental Logarithmic Model of Aggregate Consumer Behavior", in *Advances in Econometrics*, Vol1, ed. by R. Basmann and G. Rhodes, Greenwich, Connecticut: JAI Press.

[20] Leser, C. E. V. (1963), "Forms of Engel Functions", *Econometrica*, 31, 694-703.

[21] Lewbel, A. (1991), "The rank of demand systems: theory and nonparametric estimation", *Econometrica*, 59, 711-730.

[22] Lewbel, A (1996), " Demand systems with and without errors: reconciling econometric, random utility and GARP models", *American Economic Review.*

[23] Matzkin (2006), "Heterogeneous Choice", forthcoming in R. Blundell, W. Newey and T. Persson (eds), Advance in Econometrics: Proceedings of the 9th World Congress; Econometric Society Monograph Series, CUP, Cambridge.

[24] Manski, C. (2003), "Partial Identification of Probability Distributions", Springer-Verlag.

[25] Muellbauer, J. (1976), "Community Preferences and the Representative Consumer", *Econometrica*, 44, 525-543.

[26] Newey, W.K. and J. Powell (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565-1578.

[27] Newey, W. Powell, J. and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-604.

[28] Varian, H.(1982), "The Nonparametric Approach to Demand Analysis ", *Econometrica*, 50, 945-974.

[29] Varian, H.(1983), "Nonparametric Tests of Consumer Behaviour", *Review of Economic Studies*, 99-110.

[30] Varian, H.(1986), "Nonparametric Analysis of Optimizing Behavior with Measurement Error", *Journal of Econometrics*30, 445-459.

[31] Working, H., (1943), "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, 38, 43-56.

Appendix: Proofs of Propositions

Proof of Proposition 1.
Let $S'(\mathbf{p}_0, x_0)$ denote the support set

$$S'(\mathbf{p}_0, x_0) = \left\{ \mathbf{q}_0 : \begin{array}{c} \mathbf{p}_0'\mathbf{q}_0 = x_0,\ \mathbf{q}_0 \geq \mathbf{0} \text{ and} \\ \{\mathbf{p}_0, \mathbf{p}_t; \mathbf{q}_0, \mathbf{q}_t(x)\}_{t=1,\ldots,T} \text{ satisfies SARP} \\ \text{and } x_t \neq \tilde{x}_t \text{ for some } t \end{array} \right\}$$

where the $\mathbf{q}_t(x)$ data are demands on expansion paths at arbitrary budget levels. Suppose that there exists some demand vector $\mathbf{q}_0 \geq \mathbf{0}$ and $\mathbf{p}_0'\mathbf{q}_0 = x_0$ such that $\mathbf{q}_0 \in S(\mathbf{p}_0, x_0)$ but $\mathbf{q}_0 \notin S'(\mathbf{p}_0, x_0)$. Then by definition of $S'(\mathbf{p}_0, x_0)$ it must be the case that $\{\mathbf{p}_0, \mathbf{p}_t; \mathbf{q}_0, \mathbf{q}_t(x)\}_{t=1,\ldots,T}$ contains a violation of SARP. That is there is some element of $\{\mathbf{q}_t(x)\}_{t=1,\ldots,T}$ (call it $\mathbf{q}_t(x)$) such that either $\mathbf{q}_t(x) R \mathbf{q}_0$ and $\mathbf{q}_0 R^0 \mathbf{q}_t(x)$ or $\mathbf{q}_0 R \mathbf{q}_t(x)$ and $\mathbf{q}_t(x) R^0 \mathbf{q}_0$. Consider the first case where $\mathbf{q}_0 R^0 \mathbf{q}_t(x)$. If demands are weakly normal then the corresponding intersection demand $\mathbf{q}_t(\tilde{x}_t)$ used to define $S(\mathbf{p}_0, x_0)$ must be such that $\mathbf{q}_t(\tilde{x}_t) R^0 \mathbf{q}_t(x)$. But $\mathbf{q}_t(x) R \mathbf{q}_0$ and hence $\mathbf{q}_t(x) R \mathbf{q}_t(\tilde{x}_t)$ and there is a contradiction of SARP. Now consider the second case where $\mathbf{q}_t(x) R^0 \mathbf{q}_0$. Since $\mathbf{q}_0 \in S(\mathbf{p}_0, x_0)$ we know that by definition $\mathbf{p}_t'\mathbf{q}_0 \geq \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t)$ and hence $\mathbf{q}_t(x) R^0 \mathbf{q}_t(\tilde{x}_t)$. Therefore we have another contradiction of SARP. Hence $\mathbf{q}_0 \notin S'(\mathbf{p}_0, x_0) \Rightarrow \mathbf{q}_0 \notin S(\mathbf{p}_0, x_0)$.$\blacksquare$

Proof of Proposition 2.
(1) $S(\mathbf{p}_0, x_0)$ *is non-empty if and only if the data set* $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\ldots T}$ *satisfies SARP.*
If $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\ldots T}$ fail SARP than so does $\{\mathbf{p}_0, \mathbf{p}_t; \mathbf{q}_0, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\ldots,T}$ for any $\{\mathbf{p}_0; \mathbf{q}_0\}$ so that the support set is empty. Conversely, if $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\ldots T}$ pass SARP then these points satisfy the conditions for inclusion in $S(\mathbf{p}_0, x_0)$ which is thus non-empty.
(2) $S(\mathbf{p}_0, x_0)$ *is the singleton* $\mathbf{q}_t(\tilde{x}_t)$ *if* $\mathbf{p}_0 = \mathbf{p}_t$ *and the data set* $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,\ldots T}$ *satisfies SARP.*
Let $\mathbf{p}_0 = \mathbf{p}_t$ and suppose there is a $\mathbf{q}_0 \in S(\mathbf{p}_0, x_0)$ with $\mathbf{q}_0 \neq \mathbf{q}_t(\tilde{x}_t)$. We have $\mathbf{p}_0'\mathbf{q}_0 = x_0$. By construction $\mathbf{q}_t(\tilde{x}_t) R^0 \mathbf{q}_0$ which implies $\mathbf{q}_t(\tilde{x}_t) R \mathbf{q}_0$. Since $\mathbf{q}_0$ satisfies SARP and $\mathbf{q}_0 \neq \mathbf{q}_t(\tilde{x}_t)$ we have *not* $(\mathbf{q}_0 R^0 \mathbf{q}_t(\tilde{x}_t))$ which is equivalent to $\mathbf{p}_0'\mathbf{q}_0 < \mathbf{p}_0'\mathbf{q}_t(\tilde{x}_t) = \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t)$. Since both sides of this strict inequality are equal to x_0 this gives a contradiction.
(3) $S(\mathbf{p}_0, x_0)$ *is convex.*
Let the support set contain $\check{\mathbf{q}}_0$ and $\tilde{\mathbf{q}}_0$. The convex combination $\lambda\check{\mathbf{q}}_0 + (1-\lambda)\tilde{\mathbf{q}}_0$ for $\lambda \in [0,1]$ satisfies the non-negativity constraint and $\mathbf{p}_0'(\lambda\check{\mathbf{q}}_0 + (1-\lambda)\tilde{\mathbf{q}}_0) = \lambda\mathbf{x}_0 + (1-\lambda)\mathbf{x}_0 = \mathbf{x}_0$. Finally, we have $\mathbf{p}_t'\check{\mathbf{q}}_0 \geq \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t)$ and $\mathbf{p}_t'\tilde{\mathbf{q}}_0 \geq \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t)$ so that $\mathbf{p}_t'(\lambda\check{\mathbf{q}}_0 + (1-\lambda)\tilde{\mathbf{q}}_0) \geq \mathbf{p}_t'\mathbf{q}_t(\tilde{x}_t)$.$\blacksquare$

Proof of Proposition 3.
If $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,2\ldots T}$ fails SARP then both sets are empty and the proposition holds trivially. In the following we shall assume that $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,2\ldots T}$ passes SARP. We shall first show $S^{LP} \supseteq S$, then part 2 of the proposition and then $cl(S) \supseteq S^{LP}$.
$S^{LP}(\mathbf{p}_0, x_0) \supseteq S(\mathbf{p}_0, x_0)$.
Take any $\mathbf{q}_0 \in S(\mathbf{p}_0, x_0)$. We have $\mathbf{q}_0 \geq \mathbf{0}$ and $\mathbf{p}_0'\mathbf{q}_0 = x_0$ and $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,2\ldots T}$ satisfies SARP. Thus we only need to check the last condition in S^{LP}. Since $\mathbf{p}_0'\mathbf{q}_0 = x_0 = \mathbf{p}_0'\mathbf{q}_t$ we have $\mathbf{q}_0 R^0 \mathbf{q}_t$ which implies $\mathbf{q}_0 R \mathbf{q}_t$. The definition of SARP then gives $\mathbf{p}_t'\mathbf{q}_t < \mathbf{p}_t'\mathbf{q}_0$ which is the condition in the definition of $S^{LP}(\mathbf{p}_0, x_0)$.

For part 2 of the proposition we have:

$$S^{LP} \setminus S = \left\{ \begin{array}{c} \mathbf{q}_0 : \mathbf{q}_0 \geq 0, \mathbf{p}_0' \mathbf{q}_0 = x_0, \\ \mathbf{p}_t' \mathbf{q}_0 \geq \mathbf{p}_t' \mathbf{q}_t(\tilde{x}_t), t = 1, 2, ...T \\ \{\mathbf{p}_0, \mathbf{p}_t, \mathbf{q}_0, \mathbf{q}_t(\tilde{x}_t)\}_{t=1,...T} \text{ fails SARP} \end{array} \right\}$$

If $\mathbf{q}_0 = \mathbf{q}_t(\tilde{x}_t)$ $\mathbf{q}_0 \in S$ so that we only need to consider $\mathbf{q}_0 \neq \mathbf{q}_t(\tilde{x}_t)$ for all t. This and the failure of SARP implies either:
(A) $\mathbf{q}_t(\tilde{x}_t) R \mathbf{q}_0$ and $\mathbf{p}_0' \mathbf{q}_0 \geq \mathbf{p}_t' \mathbf{q}_t(\tilde{x}_t)$ for some t. The first statement requires that there is some s such that $\mathbf{q}_s(\tilde{x}_s) R^0 \mathbf{q}_0$ which implies $\mathbf{p}_s' \mathbf{q}_s(\tilde{x}_s) \geq \mathbf{p}_s' \mathbf{q}_0$. Combining this with the condition $\mathbf{p}_s' \mathbf{q}_0 \geq \mathbf{p}_s' \mathbf{q}_s(\tilde{x}_s)$ gives $\mathbf{p}_s' \mathbf{q}_0 = \mathbf{p}_s' \mathbf{q}_s(\tilde{x}_s)$ as in the statement in the proposition.
or:
(B) $\mathbf{q}_0 R \mathbf{q}_t(\tilde{x}_t)$ and $\mathbf{p}_t' \mathbf{q}_t(\tilde{x}_t) \geq \mathbf{p}_0' \mathbf{q}_0$. In this case the latter statement and $\mathbf{p}_t' \mathbf{q}_0 \geq \mathbf{p}_t' \mathbf{q}_t(\tilde{x}_t)$ give the statement in the proposition.
$cl(S) \supseteq S^{LP}$.
We have just shown that it is only boundary of S^{LP} that are not in S. Thus the closure of S contains S^{LP}.■

Proof of Proposition 4.
Since $\mathbf{p}_{T+1} = \mathbf{p}_0$ we have that S^{T+1} is a singleton (by part 2 of proposition 2). Since S^T is convex and there are two distinct intersection points in S^T, there are a continuum of points in S^T. Hence S^Tstrictly includes S^{T+1}.■

Proof of Proposition 5.
1) We first show that intersection of the budget plane $\{\mathbf{p}_{T+1}, x_{T+1}\}$ with $S^T(\mathbf{p}_0, x_0)$ implies that $S^{T+1}(\mathbf{p}_0, x_0) \subset S^T(\mathbf{p}_0, x_0)$. The definition of intersection between the new budget plane $\{\mathbf{p}_{T+1}, x_{T+1}\}$ and $S^T(\mathbf{p}_0, x_0)$ implies that $\mathbf{q}_{T+1}(\tilde{x}_{T+1}) R^0 \mathbf{q}_0$. Since $\mathbf{q}_0 \in S^T(\mathbf{p}_0, x_0)$ the definition of an intersection demand implies $\mathbf{q}_0 R^0 \mathbf{q}_{T+1}(\tilde{x}_{T+1})$. This gives a violations of SARP in the dataset $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=0,...,T+1}$. Therefore $\mathbf{q}_0 \notin S^{T+1}(\mathbf{p}_0, x_0)$ and hence $S^{T+1}(\mathbf{p}_0, x_0) \subset S^T(\mathbf{p}_0, x_0)$.

2) We now show that $S^{T+1}(\mathbf{p}_0, x_0) \subset S^T(\mathbf{p}_0, x_0)$ implies intersection of the budget plane $\{\mathbf{p}_{T+1}, x_{T+1}\}$ with $S^T(\mathbf{p}_0, x_0)$. Suppose $S^{T+1}(\mathbf{p}_0, x_0) \subset S^T(\mathbf{p}_0, x_0)$. This implies that there exists at least one $\mathbf{q}_0 \in S^T(\mathbf{p}_0, x_0)$ such that $\mathbf{q}_0 \notin S^{T+1}(\mathbf{p}_0, x_0)$. In the following $\overline{R^0}$ denotes "not R^0". Since $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=0,...,T}$ satisfies SARP, and since $\mathbf{q}_0 R^0 \{\mathbf{q}_t(\tilde{x}_t)\}_{t=1,...,T}$ by the definition of intersection demands, this implies that $\{\mathbf{q}_t(\tilde{x}_t)\}_{t=1,...,T} \overline{R^0} \mathbf{q}_0$. Since $\mathbf{q}_0 \notin S^{T+1}(\mathbf{p}_0, x_0)$ the dataset $\{\mathbf{p}_t, \mathbf{q}_t(\tilde{x}_t)\}_{t=0,...,T+1}$ violates SARP. Given $\{\mathbf{q}_t(\tilde{x}_t)\}_{t=1,...,T} \overline{R^0} \mathbf{q}_0$ and the assumption that $S^{T+1}(\mathbf{p}_0, x_0) \neq \varnothing$ this violation must result from $\mathbf{q}_{T+1}(\tilde{x}_{T+1}) R^0 \mathbf{q}_0 \Rightarrow x_{T+1} \geq \mathbf{p}_{T+1}' \mathbf{q}_0$. Hence $\mathbf{q}_0$ must lie in the intersection of the convex set $S^T(\mathbf{p}_0, x_0)$ and the closed half-space $\mathbf{p}_{T+1}' \mathbf{q}_0 \leq x_{T+1}$. If there exists some $\mathbf{q}_0 \in S^T(\mathbf{p}_0, x_0)$ such that $\mathbf{p}_{T+1}' \mathbf{q}_0 < x_{T+1}$ then there must also exist some $\mathbf{q}_0 \in S^T(\mathbf{p}_0, x_0)$ such that $\mathbf{p}_{T+1}' \mathbf{q}_0 = x_{T+1}$ and therefore the new budget plane $\{\mathbf{p}_{T+1}, x_{T+1}\}$ intersects with $S^T(\mathbf{p}_0, x_0)$.■

7 Appendix: Data Descriptives

Commodity Groups

"Food": {bread,cereals,biscuits & cakes, beef, lamb, pork, bacon, poultry, other meats & fish, butter, oil & fats, cheese, eggs, fresh milk, milk products, tea, coffee, soft drinks, sugar & preserves, sweets & chocolate, potatoes, other vegetables, fruit, other foods, canteen meals, other restaurant meals and snacks}.

"Non-durables": {beer, wine & spirits, cigarettes, other tobacco, household consumables, petcare, mens outer clothes, women's outer clothes, children's outer clothes, other clothes, footwear, chemist's goods, audio visual goods, records and toys,book & newspapers, gardening goods}

"Services": {domestic fuels, postage & telephone, domestic services, fees & subscriptions, personal services, maintenance of motor vehicles, petrol and oil, vehicle tax and insurance, travel fares, tv licences, entertainment}.

Table A1. Descriptive Statistics, 1975 to 1999

	Budget Shares			Total Exp.	Prices		Children	n
	F	ND	S		F	S		
1975	0.3587	0.3166	0.3247	33.7838	1.0000	1.0000	1.9893	1873
1976	0.3577	0.3076	0.3347	32.5127	1.0881	1.0687	1.9702	1642
1977	0.3564	0.3124	0.3312	32.3477	1.1574	1.0447	1.9429	1770
1978	0.3556	0.3136	0.3308	32.5452	1.1067	1.0398	.1.8828	1681
1979	0.3458	0.3196	0.3346	36.4990	1.1457	1.0414	1.8893	1689
1980	0.3384	0.3208	0.3408	36.6857	1.1145	1.1061	1.8619	1781
1981	0.3363	0.3061	0.3576	35.7316	1.1056	1.1836	1.8751	1906
1982	0.3218	0.3101	0.3681	35.8705	1.1262	1.2199	1.8539	1876
1983	0.3214	0.3129	0.3657	35.6571	1.0775	1.2429	1.8571	1743
1984	0.3162	0.3151	0.3688	37.5016	1.1081	1.2492	1.8438	1671
1985	0.3081	0.3207	0.3712	37.8100	1.0759	1.2242	1.8323	1622
1986	0.3088	0.3221	0.3692	38.4100	1.0556	1.2239	1.8645	1587
1987	0.3043	0.3228	0.3730	39.0197	1.0819	1.2372	1.8713	1632
1988	0.3042	0.3278	0.3680	41.5325	1.0807	1.2512	1.8744	1648
1989	0.3054	0.3222	0.3724	41.5346	1.0786	1.2713	1.8662	1652
1990	0.3017	0.3129	0.3854	44.2983	1.1084	1.3150	1.8966	1538
1991	0.2972	0.3103	0.3925	42.6966	1.0839	1.3207	1.8351	1510
1992	0.2882	0.3121	0.3997	41.5212	1.0616	1.3445	1.9068	1578
1993	0.2866	0.3077	0.4057	41.3798	1.0332	1.3533	1.8895	1511
1994	0.2825	0.3029	0.4146	40.9660	1.0305	1.3748	1.8838	1489
1995	0.2912	0.2912	0.4176	39.6002	1.0439	1.3645	1.8622	1502
1996	0.2889	0.2999	0.4112	41.8850	1.0671	1.3491	1.8638	1476
1997	0.2741	0.3041	0.4218	45.2517	1.0655	1.4071	1.8410	1421
1998	0.2788	0.2981	0.4230	44.0626	1.0551	1.4102	1.9099	1432
1999	0.2722	0.3032	0.4245	47.1033	1.0918	1.4367	1.8774	1501

Notes: F=Food, ND=Non-durables, S=Services