Consumer Theory

Joshua Lanier
St. Peter's College
Department of Economics
University of Oxford
1 March, 2018

Abstract

This thesis consists of three self-contained chapters covering topics in consumer theory. The first chapter presents an estimator for a Marshallian demand function which not only obeys all of the standard assumptions of consumer theory but will also converge to any true demand function which also obeys these standard assumptions. The second chapter explores Giffen behavior in the context of financial assets. The chapter finds that an agent with Maxmin preferences almost always displays Giffen behavior in some financial environments. Giffen behavior is also characterized for other classes of preferences. The last chapter, coauthored with one of my supervisors: John K.-H. Quah, develops a revealed preference test for weakly separable preferences in the spirit of Sydney Afriat. Unlike previous tests, ours does not impose concavity and applies to nonlinear budgets.

Acknowledgments

I would like to thank my two supervisors John K.-H. Quah and Ian Crawford. John has provided outstanding guidance throughout my time at Oxford and this thesis would not have been possible without him. Ian eagerly stepped in as my supervisor after John left Oxford. He helped tremendously in developing the content of the chapter on demand estimation. I am very grateful for all the times he pushed me to share my ideas with other members of the department. My work benefited greatly from these experiences.

I would also like to thank my supporting wife Xiwen. Her beautiful spirit has kept me uplifted throughout this process.

Chapter 1

Theory Consistent Nonparametric Demand Estimation

Theory Consistent Non-Parametric Demand Estimation

Abstract

We develop a non-parametric method for estimating a Marshallian demand function. The estimator satisfies two desirable properties that no other estimator has been shown to satisfy. First, the estimated demand function will be generated by a well-behaved utility function in the sense that there exists a utility function whose maximization generates the demand. Second, the estimator converges in probability to the true demand function provided that the true demand function is generated by a well-behaved utility function. The technique simultaneously estimates the demand function and the preferences of the consumer. This makes it easy to compare the welfare of the consumer under different budget constraints.

1 Introduction

We present a new non-parametric technique for estimating a Marshallian demand function. The estimator satisfies the following two desirable properties:

Theory Consistency: There is a well-behaved utility function[1] whose maximization generates the estimated demand function.

Statistical Consistency: The estimator converges in probability to the true demand function provided the true demand is generated by a well-behaved utility function.

No other estimator has been shown to satisfy both of these two properties. The technique simultaneously estimates the consumer's demand and preferences.

Much of consumer theory operates on the premise that a consumer's behavior is the outcome of the maximization of some utility function. Most notions of consumer welfare (compensating variation, equivalent variation, etc.) are motivated by or require the existence of a utility function which explains a consumer's behavior. Theory consistency states that an estimator of the demand function of a consumer ought to also be generated

[1] This will be precisely defined later.

by some utility function. When an estimator of demand fails this property then one of two undesirable approaches must be taken; either one must give up on speaking about the preferences of the consumer or one must get an estimate of the preferences of the consumer which will *not* generate the estimated demand (that is, the estimated demand is not generated by the estimated preferences).

While consumer theory assumes the existence of a utility function for a consumer, the theory usually places very few restrictions on this function. It is common to assume that a utility function is increasing, quasi-concave, and continuous but any other requirements are departures from the standard theory. The property of statistical consistency states that if the true demand function satisfies the usual assumptions of consumer theory then its estimator should converge to the true demand as the sample grows.

Estimators which fail statistical consistency unjustifiably restrict what the estimator can report. That is, there are preferences and demand functions which the consumer may possess which are perfectly consistent with economic theory but will never be recovered by the estimation technique which fails statistical consistency. This misspecification could have serious implications when estimating demand responses and the welfare consequences of price changes. Parametric models cannot satisfy this property. Thus, the Translog specification of Christensen, Jorgenson, and Lau (1975), the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980), and the Quadratic Almost Ideal Demand System (QUAIDS) of Banks, Blundell, and Lewbell (1997) are not statistically consistent in our sense. Models that do not satisfy statistical consistency also include semi-parametric specifications such as the EASI demand system in Lewbel and Pendakur (2009) and the semi-parametric model of Pendakur and Sperlich (2010).

Estimators which are not flexible enough will fail statistical consistency while estimators which are too flexible will fail theory consistency. There are in fact very few known demand specifications which are theory consistent over the entire domain of the function. Past demand system specifications which can be made theory consistent over their entire domain restrict behavior in undesirable ways (see Deaton and Muellbauer (1980) and Barnett and Serletis (2008)). Instead of seeking global theory consistency many of the parametric models are able to impose theory consistency in some neighborhood of a point in price space or on some grid of points in price space and even then most approaches only allow one to impose some of the restrictions of theory consistency. This remark applies to the previously mentioned models (see again Barnett and Serletis (2008)) but also to existing non-parametric approaches; specifically we have in mind the Fourier Flexible Form of Gallant (1981), the Asymptotically Ideal Model (AIM) of Barnett and Jonas (1983), the specifications of Blundell, Horowitz, and Parey (2012) and Blundell, Horowitz, and Parey (2017) and the specification of Haag, Hoderlein, and

Pendakur (2009).

As eluded to, the property of statistical consistency requires the parameter space to be large enough (great enough flexibility of the estimator) while theory consistency requires the parameter space to not be too big (a restriction on the flexibility of the estimator). Typically parametric estimators fail statistical consistency while non-parametric estimators fail theory consistency. We are able to find a non-parametric estimator which can satisfy both properties.

To discuss further we introduce some notation. Let $L \in \mathbb{N}$ denote the number of goods (as usual we assume $L \geq 2$) available for consumption. Let $c \in \mathbb{R}^L_+$ denote a consumption bundle where c_ℓ denotes the amount of good ℓ consumed. We let $p \in \mathbb{R}^L_{++}$ denote a price vector for the L goods and $w \in \mathbb{R}_{++}$ represent an amount of expenditure. A demand function x is a function which reports the bundle demanded by a consumer for price vectors and expenditure levels. That is $x(p, w) \in \mathbb{R}^L_+$ is the bundle a consumer chooses when prices are p and expenditure is w. It is commonly assumed that a consumer's demand function arises as the result of the maximization of some utility function $U : \mathbb{R}^L_+ \to \mathbb{R}$ which represents this consumer's preferences over the various bundles. That is,

$$x(p, w) = \underset{c \in \{\tilde{c} \in \mathbb{R}^L_+ :\ p \cdot \tilde{c} \leq w\}}{\operatorname{argmax}} U(c) \tag{1}$$

Assume we have T observations $(y^1, p^1), \ldots, (y^T, p^T)$ where $y^t \in \mathbb{R}^L_+$ represents the observed purchases in period t for a consumer and $p^t \in \mathbb{R}^L_{++}$ represents the observed prices which the consumer paid in period t. We assume that the data is generated according to

$$y^t = x(p^t, p^t \cdot y^t) + \varepsilon^t$$

where $x : \mathbb{R}^L_{++} \times \mathbb{R}_{++} \to \mathbb{R}^L_+$ is a demand function generated by a utility function $U : \mathbb{R}^L_+ \to \mathbb{R}$ according to equation (1) and ε^t is an error term. We are interested in estimating the demand function x. To do this we consider estimating the indirect utility function for the consumer.

An indirect utility function v is a function which reports the utility a consumer associates to each budget set. That is, $v(p, w)$ is the amount of utility a consumer receives when the consumer spends w and prices are p. If the agent has utility function U and demand function x then $v(p, w) = U(x(p, w))$. If we know a consumer's indirect utility v it is straightforward to derive the corresponding demand function using Roy's identity, which states

$$x(p, w) = -\frac{\nabla_p v(p, w)}{\nabla_w v(p, w)} \tag{2}$$

Our approach exploits (2). We estimate an indirect utility function v for the consumer which allows us to also estimate the demand for the agent with Roy's identity. Our method is not the first to use Roy's identity to estimate demand, but it is the first to impose globally the shape restrictions implied by economic theory onto the indirect utility function while maintaining enough flexibility to consistently estimate any demand function. Other methods which estimate the indirect utility function (or the cost function) of the consumer are the previously mentioned Translog, AIDS, QUAIDS, the EASI system, the Fourier Flexible Form, and the AIM system.

At the heart of our method is the idea of building complicated indirect utility functions from simple ones. To illustrate, suppose we have a family of indirect utility functions $f(\cdot,\cdot,\beta)$, parameterized by β, which are fairly simple.[2] We shall specify an "aggregator" g which allows us to smooth N simple indirect utility functions together so that the function v defined by

$$v(p,w) = g\Big(f(p,w,\beta_1),\dots,f(p,w,\beta_N)\Big) \tag{3}$$

is a more complicated indirect utility function. Let $x(p,w,\beta_n)$ be the demand function corresponding to the indirect utility function $f(p,w;\beta_n)$. Then using (2) we may show that the demand corresponding to v defined in (3) can be represented as

$$x(p,w) = \sum_{n=1}^{N} \alpha_n x(p,w,\beta_n) \tag{4}$$

where $\alpha_1,\dots,\alpha_N$ are weights which are functions of p and w. So, we see that the demand x is a weighted average of the demand functions from each of the simple indirect utility functions. Our estimator of the demand function will take the form specified in (4). One of the challenges is to choose $f(\cdot,\cdot,\beta)$ and g in such a way that v defined in (3) is an indirect utility function consistent with economic theory. This will guarantee that the estimator is theory consistent. To discuss making the estimator statistically consistent we need to interpret our estimator as a type of sieve estimator.

The method of sieves, introduced in Grenander (1981), is based on the idea of allowing the number of parameters of the estimator to increase with the sample size. For example, the estimator may consist of 3 parameters when the sample size is 10, 4 parameters when the sample size is 20, and 50 parameters when the sample size is 1,000. For a fixed sample size the estimator is essentially parametric. The method achieves its non-parametric status as the number of parameters tends to infinity as the sample size

[2] We shall be more precise in the body of the paper

tends to infinity. Applying this method is how we achieve the statistical consistency of our estimator.

Referring back to equation (3) one might observe that there is no reason why the number of parameters $\beta_1, \ldots, \beta_N$ cannot be allowed to increase with the sample size provided the domain of g is appropriately defined. This is indeed the approach we take. Importantly, we must ensure that the space of functions of the form (4) become dense in the space of all demand functions as the number of parameters increases.

With our method it is a simple matter to calculate measures of consumer welfare which depend on the cost function as we obtain an indirect utility function through our estimation method. Here we are referring to the well-known equivalent variation, compensating variation, and deadweight loss.

The remainder of the paper proceeds as follows. In Section 2 we review the duality between direct and indirect utility. We comment on the usefulness of estimating a demand function by first estimating an indirect utility function. Section 3 provides a description of the assumptions being made on the data. In Section 4 we provide a detailed description of the new estimator which relies in part on the duality discusses in Section 2. This is followed by Section 5 where we show how the estimator can be understood as a type of sieve estimator. Having set the stage Section 6 presents the main result, which is a theorem showing that the new estimator is both statistically consistent and theory consistent. We follow this with a brief numeric example of the method and end with a brief conclusion section.

2 Indirect Utility and Roy's Identity

As in the introduction, L will denote the number of goods which a consumer may purchase. We say that a function $f : \mathbb{R}^L_+ \to \mathbb{R}$ is *increasing* if $f(c') > f(c)$ for $c' \gg c$. We say that f is *quasi-concave* if $f(c'') \geq \min(f(c), f(c'))$ for $c'' = \alpha c + (1-\alpha)c'$ and $\alpha \in [0, 1]$. f is *quasi-convex* if $-f$ is quasi-concave. A vector $c \in \mathbb{R}^L_+$ will represent a bundle of goods where element ℓ of c (written as c_ℓ) represents the quantity of good ℓ. We assume that each good has a price which we collect in a price vector $p \in \mathbb{R}^L_{++}$. We say that $U : \mathbb{R}^L_+ \to \mathbb{R}$ is a *utility function* if it is (i) continuous, (ii) increasing, and (iii) quasi-concave. We say that a function $x : \mathbb{R}^L_{++} \to \mathbb{R}^L_+$ is *generated by a utility function* $U : \mathbb{R}^L_+ \to \mathbb{R}$ if

$$x(p) = \operatorname*{argmax}_{c \in \{\tilde{c} \in \mathbb{R}^L_+ :\ p \cdot \tilde{c} \leq 1\}} U(c), \qquad \text{for all } p \in \mathbb{R}^L_{++} \tag{5}$$

That is, x is the demand function which arises from the maximization of the utility function U. Notice that we have normalized the total value of expenditures to 1. We shall stick to this normalization throughout the paper. Of course, it is a normalization done for notational convenience and has no consequence on the results. We say that U is a *well-behaved utility function* if x defined in (5) is single-valued.

A direct utility function U returns the amount of utility associated with different consumption bundles $c \in \mathbb{R}^L_+$ while an indirect utility function $v : \mathbb{R}^L_{++} \to \mathbb{R}$ returns the amount of utility associated with different budget sets. Under certain regularity conditions there is a duality between direct and indirect utility functions in the following sense. For a utility function U let U^* denote the indirect utility function formed by associating a budget with the highest utility bundle contained in it. That is,

$$U^*(p) = \sup_{c \in \left\{\bar{c} \in \mathbb{R}^L_+ :\ p \cdot \bar{c} \leq 1\right\}} U(c)$$

For an indirect utility function v let v^* denote the direct utility formed by associating a bundle of goods with the lowest utility budget which contains the bundle. That is,

$$v^*(c) = \inf_{p \in \left\{\bar{p} \in \mathbb{R}^L_{++} :\ \bar{p} \cdot c \leq 1\right\}} v(p) \tag{6}$$

Then, under certain regularity conditions we have $U = U^{**}$ and $v = v^{**}$. That is, the direct and indirect utility functions provide the exact same information about the preferences of the consumer. For more on this duality see Diewert (1974) and Blackorby, Primont, and Russell (1978).

We say that $v : \mathbb{R}^L_{++} \to \mathbb{R}$ is a *well-behaved indirect utility function* if v^* defined by (6) is a well-behaved utility function. It is well-known that if v is a well-behaved indirect utility function then it is (i) decreasing, (ii) quasi-convex, (iii) continuously differentiable. However, these properties are not sufficient to guarantee that v is well-behaved. In general, a function v which is increasing, quasi-convex, and continuously differentiable will yield a function v^* which is increasing, has single-valued demand, and is *upper semi-continuous.* Thus, v^* can fail to be continuous even when v is continuous. We shall return to this later.

One of the principle advantages of working with indirect utility functions (instead of direct utility functions) is the ease with which the demand function in (5) can be derived. Indeed, if v is an indirect utility function satisfying certain regularity conditions then a result known as Roy's identity states that

$$x(p) = \frac{\nabla_p v(p)}{p \cdot \nabla_p v(p)} \tag{7}$$

where x is the demand function generated by the utility function v^* (in the sense of (5)). We proceed by estimating an indirect utility function v which we use to estimate a demand function by exploiting (7). Before presenting the estimator we explain the assumptions placed on the data.

3 The Data

Suppose we observe a consumer making purchasing decisions over L goods. We collect T observations $(y^1, p^1), \ldots, (y^T, p^T)$ where $y^t \in \mathbb{R}^L_+$ represents the bundle of goods the consumer purchased in period t and $p^t \in \mathbb{R}^L_{++}$ represents the unit prices which the consumer paid in period t. We maintain the following assumption.

Assumption 1. *The data satisfies*

$$y^t = x_0(p^t) + \varepsilon^t \tag{8}$$

where $\varepsilon^t \in \mathbb{R}^L$*,* $p^t \cdot \varepsilon^t = 0$*, and* $x_0(p^t) = \mathbf{E}[y^t|p^t]$*. Observations* (p^t, ε^t) *are independently and identically distributed across* t*. The demand function* x_0 *is generated by a well-behaved utility function* $U_0 : \mathbb{R}^L_+ \to \mathbb{R}$ *in the sense of equation* (5)*.*

The setup of equation (8) is adopted in most of the parametric models mentioned in the introduction (see Barnett and Serletis (2008) for a good summary). It is worth noting that recently there has been interest in relaxing the assumption that the error term ε^t enters the model additively. This is an especially worthwhile endeavor when ε^t is thought to represent preference heterogeneity. This is often the case when the data is cross-sectional. In this case the additive error specification has been shown to be undesirably restrictive (see Brown and Walker (1989) and Lewbel (2001)). When rejecting the additive error assumption one must develop alternative assumptions which allow for the identification of the function(s) of interest. Such alternative identification strategies are pursued in Hoderlein and Stoye (2014), Blundell, Kristensen, and Matzkin (2014), and Blundell et al. (2017). In the present paper we do not pursue relaxing the additive error assumption as this would take us too far afield from our primary goals. We view the relaxation of the additive error assumption as a fruitful direction of future research. For this reason it is best to think that the dataset we are working with is a panel with a medium-sized number of time periods. This allows us to estimate demand for each consumer and avoids issues of preference heterogeneity.

4 The Estimator

The goal of this paper is to develop an estimator $\hat{x} : \mathbb{R}^L_{++} \to \mathbb{R}^L_+$ which estimates a true demand function x_0 under Assumption 1 where the estimator $\hat{x}$ satisfies the following two properties.

Theory Consistency: There is a well-behaved utility function $\hat{U} : \mathbb{R}^L_+ \to \mathbb{R}$ which generates the estimated demand function $\hat{x}$ in the sense of (5).

Statistical Consistency: The estimator converges in probability to the true demand function. That is $||x_0 - \hat{x}|| \xrightarrow{p} 0$ under some suitably chosen norm $|| \cdot ||$.

As we have noted in the introduction both of these properties are in some sense indispensable properties for an estimator of a Marshallian demand function to satisfy and yet no estimator in the literature satisfies both of these properties.

The estimation technique we develop is based on the idea of combining many simple indirect utility functions into a more complicated form. The goal will be to allow the more complex indirect utility to become sufficiently flexible so as to allow the corresponding demand function (obtained via Roy's identity: equation (7)) to approximate any demand function which is generated from the maximization of a well-behaved utility function (in the sense of (5)). We now present what we have in mind by simple indirect utility functions.

Definition 1. Let $\Theta = \mathbb{R} \times \left(\mathbb{R}^{2L}_+ \backslash \{0\}\right)$. Let $f : \mathbb{R}^L_{++} \times \Theta \to \mathbb{R}$ be defined by

$$f(p, u, a, b) = u - \sum_{\ell=1} a_\ell p_\ell - \sum_{\ell=1}^{L} b_\ell \sqrt{p_\ell} \tag{9}$$

A function $f(\cdot, u, a, b) : \mathbb{R}^L_{++} \to \mathbb{R}$ for some $(u, a, b) \in \Theta$ will be referred to as a *basic indirect utility function.*[3]

It will guide intuition to think of the basic indirect utility functions as the simple building blocks by which our class of indirect utility functions shall be defined. It should be remarked that the specific form we have given f is somewhat arbitrary. What we really require is for f to contain the set of all indirect utility functions which are negative affine transformations of normalized prices. So, we could just as well work with the simple indirect utility functions $f(\cdot, u, a, 0)$. We made f slightly more complicated than it needs to be because this form seems to perform better in our numerical demonstration.

[3]Note that for a function $F : X \times Y \to Z$ with generic arguments x and y we write $F(\cdot, y)$ to represent the function from X to Z defined by $F(x, y)$. We shall maintain this notation throughout. Note that under this convention $F(x, y)$ is a value of Z while F, $F(\cdot, y)$, and $F(x, \cdot)$ are functions

We define a function g which will allow several of the basic indirect utility functions to be combined into a new indirect utility function. First, for $\alpha \in \mathbb{R}$ let $(\alpha)_+ = \max(\alpha, 0)$. Next, fix $M \geq 2$ and let g be defined implicitly by

$$\sum_{i=1}^{N} \Big(\nu_i - g(\nu)\Big)_+^M = 1, \qquad \text{for } \nu \in \mathbb{R}^N \tag{10}$$

The next proposition demonstrates that g is a function.

Proposition 1. *The correspondence g defined by* (10) *is a function.*

Proof. Let $\nu \in \mathbb{R}^N$ and define $F : \mathbb{R} \to \mathbb{R}$ by

$$F(a) = \sum_{i=1}^{N} \Big(\nu_i - a\Big)_+^M - 1$$

If $F(a) = 0$ then $a \in g(\nu)$ so we must show that there is only one such a. Let $\bar{\nu}$ be the largest value in the vector ν. So, $F(\bar{\nu}) = -1$. Note F is strictly decreasing on $(-\infty, \bar{\nu}]$ and $\lim_{a \to -\infty} F(a) = \infty$. Now, the fact that F is continuous establishes the result. $\square$

The function g allows for the creation of new indirect utility functions from the basic indirect utility functions. Define the function $h : \mathbb{R}^L_{++} \times \Theta^N \to \mathbb{R}^N$ by

$$h(p, \theta) = (f(p, \theta_1), \ldots, f(p, \theta_N)) \tag{11}$$

where $\theta = (\theta_1, \ldots, \theta_N)$ and $\theta_n \in \Theta$. So, $h(\cdot, \theta)$ is an N vector of basic indirect utility functions. Define a function $v : \mathbb{R}^L_{++} \times \Theta^N \to \mathbb{R}$ by

$$v(p, \theta) = g(h(p, \theta)) \tag{12}$$

We show that the function v in (12) is an indirect utility function.

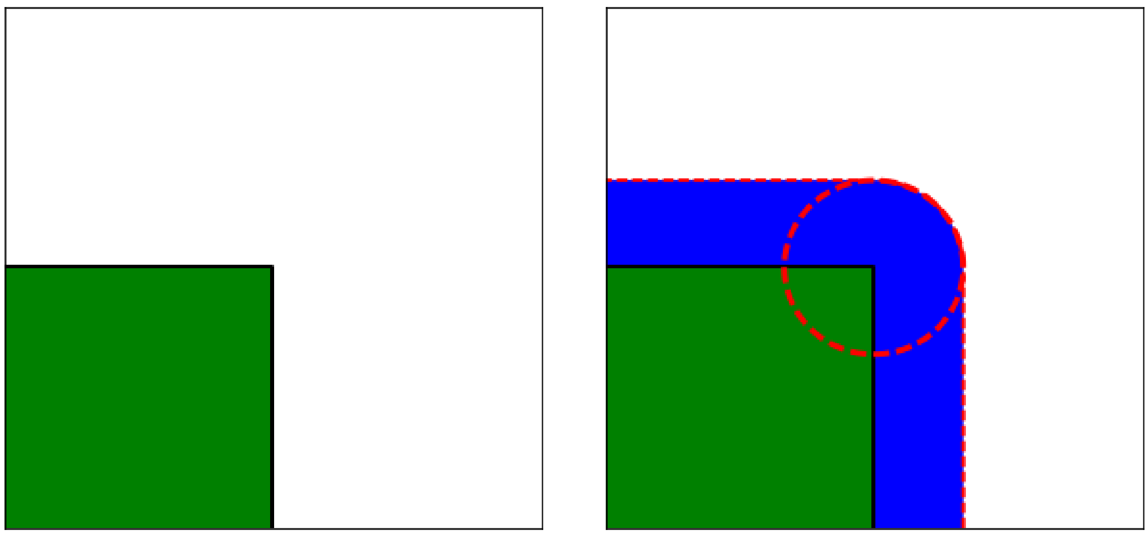
Proposition 2. *Let $\theta \in \Theta^N$. The function $v(\cdot, \theta) : \mathbb{R}^L_{++} \to \mathbb{R}$ defined by* (12) *is an $M - 1$ times continuously differentiable well-behaved indirect utility function.*

The proof appears in the appendix along with the proof of all lemmas.

We shall attempt to help the reader gain an understanding of the function g. To do this, first picture a lower contour set of the $\max : \mathbb{R}^2 \to \mathbb{R}$ function (the function which returns the largest value of the elements of the input). We graph one such lower contour set in Figure 1 below. Place a ball of radius 1 over each point in the lower contour set to create a new lower contour set. These newly created lower contour sets are the lower contour sets of g. We also graph this newly created lower contour set in Figure 1. In this figure the green represents the lower contour set of the max function while the blue represents the new lower contour set. We also draw a ball centered at the corner of the max function to help give an idea of how the new lower contour set is formed.

Figure 1: Lower Contour Sets

(a) Max Function (b) New Function



It is nice that g allows for the creation of new indirect utility functions from the basic indirect utility functions but of course it is only one of many possible aggregation techniques. For example, merely taking convex combinations of basic indirect utility functions would also create new indirect utility functions. The advantage of g is that it allows for the creation of a large enough set of indirect utility functions so as to span the entire space of interest. We shall return to this point momentarily.

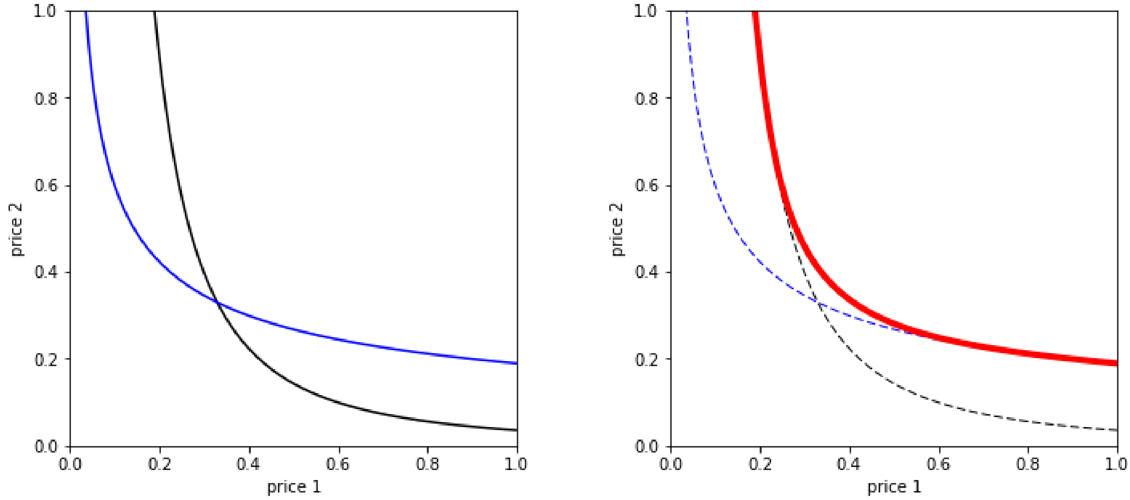
One way of understanding how g combines basic indirect utility functions is through understanding what the indifference curves look like. That is, suppose we have two basic indirect utility functions f_1, f_2. How do the indifference curves for f_1 and f_2 compare to the indifference curves of $g(f_1, f_2)$? We demonstrate the answer in Figure 2 and 3. Figure 2 shows the indifference curves in price space. The first graph depicts two indifference curves each corresponding to a different basic function. The red line in the second graph shows the corresponding indifference curve for $g(f_1, f_2)$.

Figure 3 shows the same indifference curves in consumption space. Again the first graph depicts the two indifference curves of the two different basic functions. The red line in the second graph shows the corresponding indifference curve for $g(f_1, f_2)$.

Let $x : \mathbb{R}^L_{++} \times \Theta^N \to \mathbb{R}^L_+$ be the function defined so that $x(\cdot, \theta)$ is the demand function

Figure 2: Price Indifference Curves

(a) Original Curves (b) Smoothed Curves



corresponding to indirect utility function $v(\cdot, \theta)$. We shall use Roy's identity to get an expression for $x(p, \theta)$. Define $G : \mathbb{R}^N \to \mathbb{R}^N$ by

$$G_i(\nu) = \big(\nu_i - g(\nu)\big)_+^{M-1} \tag{13}$$

Next, define $H : \mathbb{R}^L_{++} \times \Theta^N \to \mathbb{R}^{L \times N}$ by

$$H(p, \theta) = -\partial_p h(p, \theta) \tag{14}$$

Using Roy's identity we see

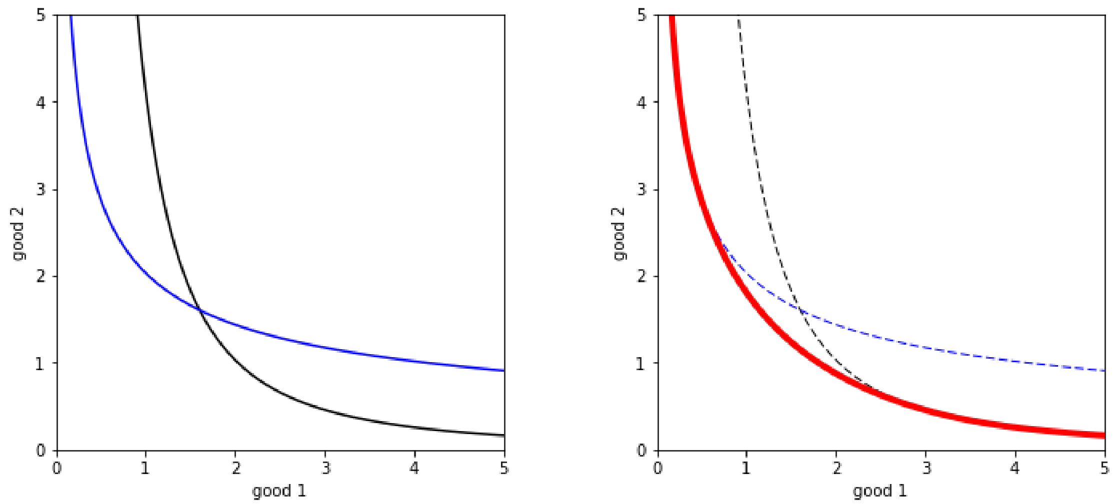
$$x(p, \theta) = \frac{H(p, \theta)' G(h(p, \theta))}{p' H(p)' G(h(p, \theta))} \tag{15}$$

We may also interpret $x(p, \theta)$ as a convex combination of the demand functions generated by the basic indirect utility functions. To see this notice that for $(u, a, b) \in \Theta$ the function $x(\cdot, u, a, b)$ is the demand function generated by the indirect utility function $f(\cdot, u, a, b)$. Let $\Delta^{N-1} = \{\sigma \in \mathbb{R}^N_+ : \sum_{n=1}^N \sigma_n = 1\}$. Define $\beta : \mathbb{R}^L_{++} \times \Theta^N \to \Delta^{N-1}$ by

$$\beta_n(p, \theta) = \frac{[\nabla_p f(p, \theta_n)] \cdot p \; G_n(h(p, \theta))}{\sum_{k=1}^N p' H(p, \theta)' G(h(P, \theta))}$$

Figure 3: Consumption Indifference Curves

(a) Original Curves (b) Smoothed Curves



Now, we rearrange the expression of x in (15).

$$x(p,\theta) = \frac{H(p,\theta)'G(h(p,\theta))}{p'H(p)'G(h(p,\theta))} = \frac{\sum_{n=1}^{N} G_n(h(p,\theta))\nabla_p f(p,\theta_n)}{\sum_{n=1}^{N} G_n(h(p,\theta))[\nabla_p f(p,\theta_n)]\cdot p}$$
$$= \frac{\sum_{n=1}^{N} G_n(h(p,\theta))[\nabla_p f(p,\theta_n)]\cdot p \frac{\nabla_p f(p,\theta_n)}{[\nabla_p f(p,\theta_n)]\cdot p}}{\sum_{n=1}^{N} G_n(h(p,\theta))[\nabla_p f(p,\theta_n)]\cdot p} = \sum_{n=1}^{N} \beta_n(p,\theta)x(p,\theta_n) \quad (16)$$

So, we see that $x(p,\theta)$ is a convex combination of the demands generated by the basic indirect utility functions $f(\cdot,\theta_1),\dots,f(\cdot,\theta_N)$.

5 The Sieve Spaces

The demand function in (15) is indexed by a parameter $\theta \in \Theta^N$. We shall allow N to grow with the sample size. This enables us to understand the estimator as a type of sieve estimator. For more details on the method of sieve estimation see Chen (2007).

For $k \in [1,\infty]$ let $||\cdot||_k$ denote the usual k-norm. So, if $z \in \mathbb{R}^N$ then $||z||_k = (\sum_{n=1}^{N} |z|^k)^{1/k}$ for $k \in [1,\infty)$ and $||z||_k = \max(|z_1|,\dots,|z_N|)$ for $k=\infty$. Let $\mathscr{F}$ be the set of demand functions which are generated by well-behaved utility functions in the sense of (5) and define a function $\mathscr{L}_T : \mathscr{F} \to \mathbb{R}_+$ by

$$\mathscr{L}_T(x) = \frac{1}{T}\sum_{t=1}^{T} \left\| y^t - x(p^t) \right\|_2^2 \quad (17)$$

where the T in $\mathscr{L}_T$ signifies that the function depends on the sample. We assume there is some true demand function $x_0 \in \mathscr{F}$ which is related to the observations via (8). One might be tempted to estimate x_0 by trying to select the element of $\mathscr{F}$ which minimizes the sum of squared residuals (that is, choose $\hat{x} \in \mathscr{F}$ to minimize $\mathscr{L}_T$). This approach suffers from two problems. First, it may be computationally demanding (or impossible) to select the desired element of $\mathscr{F}$. Second, even if it were possible to find the element of $\mathscr{F}$ which minimizes $\mathscr{L}_T$ this estimator may fail to converge to the true demand function. This problem arises from the non-compactness of $\mathscr{F}$ (see Chen (2007)).

To overcome these problems one may construct sets $\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3, \ldots$ which are subsets of $\mathscr{F}$. If the sets $\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3, \ldots$ are less complicated than $\mathscr{F}$ then it is possible to overcome both of the previously identified problems of choosing an element of $\mathscr{F}$ to minimize $\mathscr{L}_T$. That is, the estimator defined by choosing the $\hat{x}$ from $\mathscr{F}_T$ to minimize $\mathscr{L}_T$ can be computationally feasible and tend to the true demand function in probability even when the estimator defined by selecting the element of $\mathscr{F}$ to minimize $\mathscr{L}_T$ is not computationally feasible or not statistically consistent. We refer to the sets $\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3, \ldots$ as *sieve spaces*.

We let our sieve spaces consist of demand functions of the form (15) where we allow the number of basic indirect utility functions used to increase with the sample size. We shall also take steps to bound the derivatives of the estimator and relax these bounds as the sample size increases. Let

$$\Theta(\eta) = \Big\{(u, a, b) \in \Theta : |u| \leq \eta, \text{ and } \tfrac{1}{\eta} \leq \max\Big(a_1, \ldots, a_L, b_1, \ldots, b_L\Big) \leq \eta\Big\}$$

Note that if $\eta \leq \eta'$ then $\Theta(\eta) \subseteq \Theta(\eta')$. If $\theta \in \Theta(\eta)$ then we know upper and lower bounds on the partial price derivatives of $f(p, \theta)$. We shall follow that convention that if A is a set and F is a function then $F(A) = \{b : b = F(a),\ a \in A\}$. Under this convention $\mathscr{F}(\eta, N) = x(\cdot, \Theta(\eta)^N)$ is the set of all demand functions $x(\cdot, \theta)$ where $\theta \in \Theta(\eta)^N$. Let $(N_T)_{T \in \mathbb{N}}$ be an increasing sequence of integers and let $(\eta_T)_{T \in \mathbb{N}}$ be an increasing sequence of real numbers which are all greater than 1. Our sieve spaces are $\mathscr{F}(\eta_T, N_T) = x(\cdot, \Theta(\eta_T)^T)$. Define $\hat{\theta}$ and $\hat{x}$ by

$$\hat{\theta} = \underset{\theta \in \Theta(\eta_T)^{N_T}}{\operatorname{argmin}} \ \mathscr{L}_T\Big(x(\cdot, \theta)\Big) \tag{18}$$

$$\hat{x} = x(\cdot, \hat{\theta}) \tag{19}$$

In the case that the argmin in (18) is not unique then just pick any element from the argmin set to be $\hat{\theta}$. Clearly, $\hat{x} \in \mathscr{F}(\eta_T, N_T)$ and $\mathscr{L}_T(\hat{x}) \leq \mathscr{L}_T(x)$ for all $x \in \mathscr{F}(\eta_T, N_T)$.

6 The Main Theorem

We maintain the following assumption for our main result.

Assumption 2. *The normalized price space has compact support $P \subseteq \mathbb{R}^L_{++}$.*

We need to introduce the metric under which convergence will take place. If μ is a measure on a set Z and $f : Z \to \mathbb{R}^K$ is a μ-measurable function on Z then for $k \in [1, \infty)$ define $||f||_{k,\mu}$ by

$$||f||_{k,\mu} = \left(\int \|f(z)\|_k^k \, d\mu(z) \right)^{1/k}$$

and define $||f||_{\infty,\mu}$ to be the essential supremum of $||f(z)||_\infty$. We now state our main result.

Theorem 1. *The estimator $\hat{x}$ defined in* (19) *is generated by a well-behaved utility function $\hat{U}$ in the sense of* (5)*. Let μ be the distribution of normalized price p^t. Suppose Assumptions 1, 2 hold, that $N_T \to \infty$, $\eta_T \to \infty$, and*

$$\frac{N_T \ln(\eta_T N_T)}{T} \to 0 \tag{20}$$

then

$$||x_0 - \hat{x}||_{2,\mu} \xrightarrow{p} 0 \tag{21}$$

This theorem shows that our estimator $\hat{x}$ satisfies the desired properties of statistical and theory consistency. Its proof relies on two lemmas

Lemma 1. *Suppose Assumptions 1 and 2 hold. Then there is a sequence of demand functions $(x^T)_{T \in \mathbb{N}}$ such that $x^T \in \mathscr{F}(\eta_T, N_T)$ and $||x_0 - x^T||_{\infty,\mu} \to 0$.*

Lemma 1 shows that any true demand function $x_0 \in \mathscr{F}$ can be approximated by a sequence of demand functions from $\mathscr{F}(\eta_T, N_T)$. The lemma says nothing about the estimator $\hat{x}$ though. It states that there is a good approximating sequence (x^T) it does not say that $\hat{x}$ will be this sequence.

Let $\sigma_\varepsilon = \mathbf{E}[\varepsilon^t \cdot \varepsilon^t]$ and define $\mathscr{L} : \mathscr{F} \to \mathbb{R}_+$ by

$$\mathscr{L}(x) = \mathbf{E}\Big[\mathscr{L}_T(x)\Big] \tag{22}$$

The following lemma shows that $\mathscr{L}_T$ converges uniformly to $\mathscr{L}$.

Lemma 2. *Suppose Assumptions 1 and 2 hold and that* (20) *holds. Then*

$$\sup_{x \in \mathscr{F}(\eta_T, N_T)} \Big| \mathscr{L}(x) - \mathscr{L}_T(x) \Big| \xrightarrow{p} 0$$

Lemma 2 shows that as T becomes large the function $\mathscr{L}_T$ becomes very close to $\mathscr{L}$. Thus, it might be thought that the minimizer of $\mathscr{L}_T$ should be close to the minimizer of $\mathscr{L}$. Further, it is straightforward to see that if Assumption 1 holds then

$$\mathscr{L}(x) = ||x_0 - x||_{2,\mu}^2 + \sigma_\varepsilon^2$$

Thus, if the minimizer of $\mathscr{L}_T$ becomes close to the minimizer of $\mathscr{L}$ then Lemma 1 (which shows that x_0 can be approximated well by functions in $\mathscr{F}(\eta_T, N_T)$) grants the desired convergence property.

Proof of Theorem 1. We first comment on the measurability of $\hat{x}$. White and Wooldridge (1991) show that if (i) $\mathscr{L}_T(x)$ is a measurable function of the data for any $x \in \mathscr{F}$ and (ii) the function $\mathscr{L}_T$ is continuous for any data then $\hat{x}$ is measurable. Clearly, $\mathscr{L}_T$ possesses these properties so $\hat{x}$ is measurable.

Next, the fact that $\hat{x}$ is generated by a well-behaved utility function follows from the fact that $\hat{x}$ is generated from a well-behaved indirect utility function as established in Proposition 2.

Define $\mathscr{L}$ by equation (22). Clearly, equation (21) holds if and only if $\mathscr{L}(\hat{x}) \xrightarrow{p} \sigma_\varepsilon^2$. So, it suffices to prove the latter. Let $(x^T)_{T\in\mathbb{N}}$ denote the sequence in Lemma 1. Now,

$$\begin{aligned}\sigma_\varepsilon^2 \le \mathscr{L}(\hat{x}) &\le \left|\mathscr{L}(\hat{x}) - \mathscr{L}_T(\hat{x})\right| + \mathscr{L}_T(\hat{x}) \le \left|\mathscr{L}(\hat{x}) - \mathscr{L}_T(\hat{x})\right| + \mathscr{L}_T(x^T) \\ &\le \left|\mathscr{L}(\hat{x}) - \mathscr{L}_T(\hat{x})\right| + \left|\mathscr{L}_T(x^T) - \mathscr{L}(x^T)\right| + \mathscr{L}(x^T) \xrightarrow{p} \sigma_\varepsilon^2\end{aligned}$$

where the first two terms in the last expression go to 0 by Lemma 2 and the final expression goes to σ_ε^2 by Lemma 1. □

It is perhaps worth noting the lack of restrictions placed on the true demand function x_0. The only requirement that we have placed on this function is that it is generated by a well-behaved utility function. Under this condition x_0 is continuous but it is not even necessarily differentiable. An interesting fact is that even though the indirect utility function (as defined in (12)) we use to estimate x_0 will be convex we are able to estimate consistently demand functions which are not generated by convex indirect utility functions. This follows from the fact that not all well-behaved utility functions will have corresponding indirect utility functions which are convex (as we have noted well-behaved utility maximization only guarantees that the indirect utility function is quasi-convex).

7 A Numerical Demonstration

In this section we provide a simple numerical demonstration of the estimation technique. Because it is easier to present the demonstration in 2 dimensions we fix the number of goods at 2. We generate 50 observations for a consumer with indirect utility function

$$v(p) = -\ln(p_1 + 1) - \ln(p_2 + 1) \tag{23}$$

While this is indeed a simple specification it has two properties which make the estimation problem non-trivial. First, this specification is not contained within our set of basic indirect utility functions so that we will need to combine multiple basic functions together to get a good estimate. The second thing to notice is that the Engle curves corresponding to the indirect utility function in (23) are not linear nor is the indirect utility function homothetic. This means that the estimated demand function $\hat{x}$ will not do well under the many parametric specifications which assume these properties.

As mentioned we generate a dataset of 50 observations. To do this, we normalize expenditure to 1 and draw 50 random prices from a uniform distribution on $[0.1, 2] \times [0.1, 2]$. We then generate 50 demand bundles $y^1, \ldots, y^{50}$ using the demand function corresponding to the indirect utility function in (23). We then try to estimate the true demand function x_0 by selecting 4 different basic functions in order to minimize the sum of squared errors. This entails choosing $4 \times 5 = 20$ parameters. To do this we use the least_squares function in SciPy.

We present the results graphically. The first two figures display the quantities demanded of the first good as prices change.

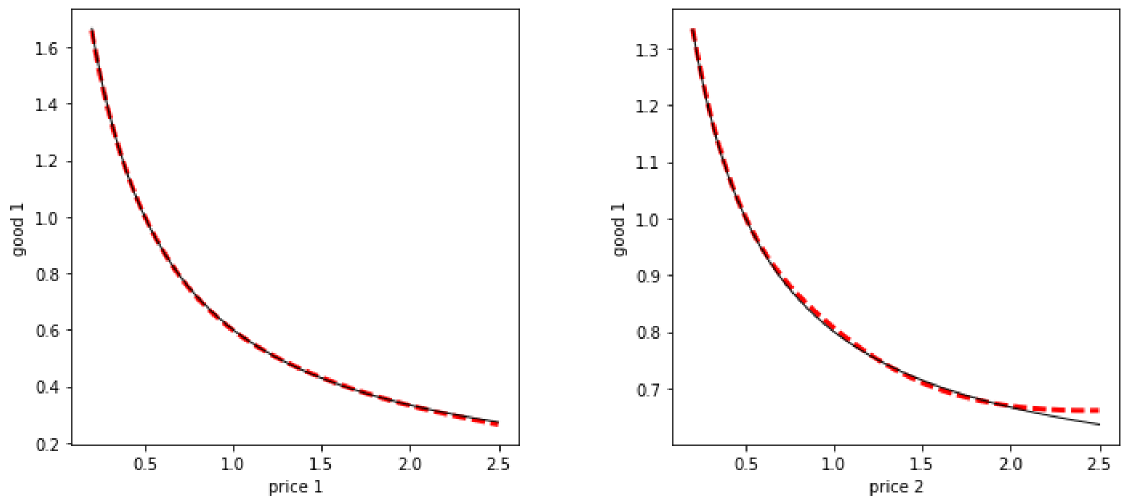
The first figure displays how the demand for good 1 changes as the price of good 1 changes and the price of good 2 is fixed at 0.5. The black line is the real demand (the demand generated by (23)) while the dotted red line is the estimated demand. The second figure displays how the demand for good 1 changes as the price of good 2 changes and the price of good 1 is fixed at 0.5. Again, the black line represents the real demand and the dotted red line represents the estimated demand. There seems to be a very good fit for the demand function with the exception of the bottom right side of the other price figure. This area is outside of the convex hull of the data and so it is not surprising that the fit is worse here.

Next, we graph some indifference curves. Figure 5 presents several indifference curves in consumption space. The black lines represent indifference curves corresponding to the indirect utility function in (23) while the red dashed lines represent the estimated indifference curves. Again there appears to be a very good fit.

Figure 4: Demand Graphs

(a) Own Price

(b) Other Price



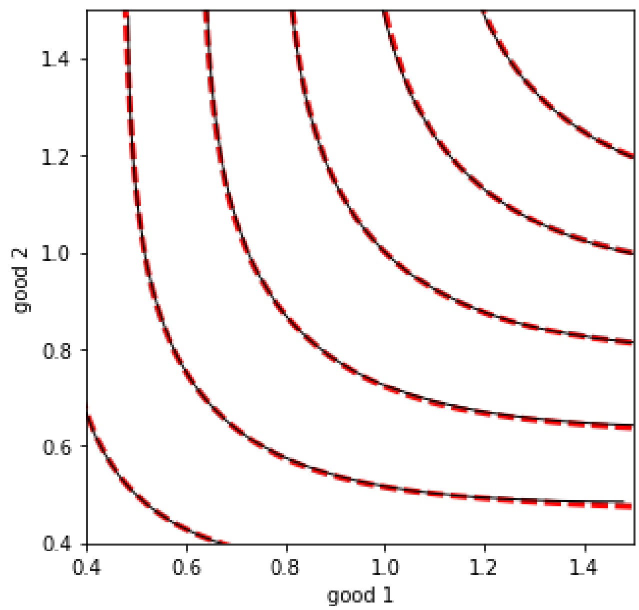
While we believe that there is reason to be encouraged by the results displayed we temper our optimism by remarking that the form in (23) might be easy to estimate compared to cases with more complicated wealth expansion paths. We also observe that the two good case that we consider here may be considerably easier than the three-plus good case. Future work will include estimating demand functions with more than two goods and cases with complicated wealth expansion paths.

8 Conclusion

Theorem 1 establishes that our estimator of demand satisfies two properties which no other estimator has been shown to possess. Our estimator will converge in probability to any true demand function which is generated by well-behaved utility maximization while also being generated by a well-behaved utility function itself. Our numerical demonstration shows that the estimator is able to provide a very good approximation to not only the true demand function but also the true indifference curves. This shows that the estimator may prove very useful for not only estimating demand but also estimating the welfare consequences of price changes.

Throughout the course of this paper we have already suggested several areas for future research. Here we reiterate these. First, there may be times when the additive error specification in (8) is not appropriate. This is most likely the case when the error term includes unobservable preference heterogeneity. In this case an alternative strategy for identifying the demand functions of interest must be employed. It will be

Figure 5: Real and Estimated Indifference Curves



very interesting to make our proposed technique robust to this setup.

The numerical demonstration we present suggests that the method can estimate demand and indifference curves well. There is a lot of work that needs to be done to prove the robustness of this claim. To do this we hope to show that the estimator performs well when the true demand function has complicated wealth expansion paths and when the number of goods is greater than two.

We believe that the two properties of the estimator established in Theorem 1 will allow our estimator to provide new insights into the welfare of consumers when applied to the right kind of data. In particular, the method would seem well-suited to panel data with a lot of price variation. Panel seems the best type of data to use because, as we have mentioned, cross sectional data may not satisfy the assumptions we have placed on the observed demand, i.e. equation (8). When there is insufficient price variation in the dataset we are in essence viewing an Engle curve of the consumer. In this case, the hypothesis of utility maximization imposes very few restrictions on observable behavior. We believe the estimator will be best used in contexts where the consequences of utility maximization are most heavily felt.

Appendix

A Proving Proposition 2

To prove Proposition 2 we establish two lemmas. The first lemma establishes certain properties of the g function defined in (10). These properties are sufficient to ensure that g is able to transform several basic indirect utility functions into a new indirect utility function.

Lemma 3. *The function g defined in* (10) *is increasing, $M-1$ times continuously differentiable, and convex.*

Proof. Verifying that g is increasing is trivial. Showing that g is $M-1$ times continuously differentiable is straightforward with the implicit function theorem. So, we focus on showing that g is convex. Let $||\cdot||_M$ denote the M norm, let $\iota \in \mathbb{R}^N$ be the vector of all 1s, and for $a \in \mathbb{R}^N$ let $(a)_+ = (\max(a_1, 0), \ldots, \max(a_N, 0))$. Let ν, $\nu' \in \mathbb{R}^N$, and let $\nu'' = (1-\alpha)\nu + \alpha\nu'$ for some $\alpha \in (0, 1)$. We have

$$
\begin{aligned}
1 &= (1-\alpha)\Big\|(\nu - g(\nu)\iota)_+\Big\|_M + \alpha\Big\|(\nu' - g(\nu')\iota)_+\Big\|_M \\
&= \Big\|(1-\alpha)(\nu - g(\nu)\iota)_+\Big\|_M + \Big\|\alpha(\nu' - g(\nu')\iota)_+\Big\|_M \\
&\geq \Big\|(1-\alpha)(\nu - g(\nu)\iota)_+ + \alpha(\nu' - g(\nu')\iota)_+\Big\|_M \\
&\geq \Big\|\Big(\nu'' - \big((1-\alpha)g(\nu) + \alpha g(\nu')\big)\iota\Big)_+\Big\|_M
\end{aligned}
$$

Which demonstrates that $(1-\alpha)g(\nu) + \alpha g(\nu') \geq g(\nu'')$ and so g is convex. □

The next step is to present a lemma which provides sufficient conditions for a function $v : \mathbb{R}^L_{++} \to \mathbb{R}$ to be a well-behaved indirect utility function. While it is well-known that if v is decreasing, quasi-convex, and continuously differentiable then v^* is increasing, quasi-concave, has single-valued demand, and upper semi-continuous; it is perhaps less well-known that these conditions are not sufficient ensure that v^* is continuous. This is the reason why the next lemma is required. If one were happy to redefine the notion of well-behaved utility to mean increasing, upper semi-continuous, with single-valued demand, then the following lemma is not required.

To state the lemma we require some additional notation. Let $\bar{\mathbb{R}} = [-\infty, \infty]$ be the extended real line with its usual topology. Let $\bar{\mathbb{R}}^L = \bar{\mathbb{R}} \times \ldots \times \bar{\mathbb{R}}$ be the L times Cartesian product of $\bar{\mathbb{R}}$. Let f be a real-valued function whose domain X is a subset of $\bar{\mathbb{R}}^L$. The *closure of f on* $\bar{\mathbb{R}}^L$ is the function $\operatorname{cl} f : \bar{X} \to \bar{\mathbb{R}}$ (here $\bar{X}$ denotes the closure of X in

$\bar{\mathbb{R}}^L$) defined by

$$\operatorname{cl} f(p) = \liminf_{\tilde{p} \to p} f(\tilde{p})$$

This definition of a closure of a function agrees with the definition in Rockafellar (1970) when the domain of the function is restricted to $\mathbb{R}^L$.

Lemma 4. *Suppose that $v : \mathbb{R}^L_{++} \to \mathbb{R}$ is continuously differentiable, quasi-convex, decreasing, and $\operatorname{cl} v$ is continuous. Then v is a well-behaved indirect utility function.*

Proof. Define v^* by (6). Berge's Maximum Theorem can be applied to $\operatorname{cl} v$ to see that v^* is continuous on $\mathbb{R}^L_+$. The fact that v^* is increasing follows from v decreasing. As v is continuously differentiable we see that the demand correspondence is single-valued (by Roy's identity). Let c and c' be elements of $\mathbb{R}^L_+$ and let $c'' = \alpha c + (1-\alpha)c'$ for $\alpha \in [0, 1]$. From the definition of v^* there is a sequence $(p_n)_{n\in\mathbb{N}}$ which satisfies $p_n \cdot c'' = 1$ and $\lim v(p_n) = v^*(c'')$. For each p_n it must be the case that either $p_n \cdot c \leq 1$ or $p_n \cdot c' \leq 1$ which means $v(p_n) \geq \min(v^*(c), v^*(c'))$. Take the limit as n goes to ∞ to see that $v^*(c'') \geq \min(v^*(c), v^*(c'))$. This shows that v^* is quasi-convex. □

We are ready to present the proof of the proposition.

Proof of Proposition 2. It is straightforward to show that $v(\cdot, \theta)$ is $M - 1$ times continuously differentiable and decreasing. We will show that $v(\cdot, \theta)$ is quasi-convex and that $\operatorname{cl} v(\cdot, \theta)$ is continuous. From here Lemma 4 shows that $v(\cdot, \theta)$ is well-behaved.

We shall show that $v(\cdot, \theta)$ is quasi-convex by showing that it is convex. Let $\alpha \in (0, 1)$, p, $p' \in \mathbb{R}^L_{++}$. Let $p'' = (1-\alpha)p + \alpha p'$. Let $\nu = h(p, \theta)$ and $\nu' = h(p', \theta)$. Then

$$\begin{aligned}
v(p'', \theta) &= g\Big(h(p'', \theta)\Big) \\
&\leq g\Big((1-\alpha)\nu + \alpha\nu'\Big) && \text{by } h_1(\cdot, \theta), \dots, h_N(\cdot, \theta) \text{ convex and } g \text{ increasing.} \\
&\leq (1-\alpha)g(\nu) + \alpha g(\nu') && \text{by } g \text{ convex.} \\
&= (1-\alpha)v(p, \theta) + \alpha v(p', \theta)
\end{aligned}$$

which shows that $v(\cdot, \theta)$ is convex.

Finally, we show that $\operatorname{cl} v(\cdot, \theta)$ is continuous. It is straightforward to verify that $\operatorname{cl} v(p, \theta) = \operatorname{cl} g(\operatorname{cl} h(p, \theta))$. Further, it is easy to see that $\operatorname{cl} g$ and $\operatorname{cl} h(\cdot, \theta)$ are continuous. Thus, $\operatorname{cl} v(\cdot, \theta)$ is continuous and the result follows. □

B Proving Lemma 1

This section is devoted to proving Lemma 1. Throughout this section we maintain Assumption 1 and define $v_0 : \mathbb{R}^L_{++} \to \mathbb{R}$ by $v_0(p) = U_0(x_0(p))$. Let $f : \mathbb{R}^L_{++} \times \Theta \to \mathbb{R}$ be defined by (9). For $(u, a, b) \in \Theta$ we shall write $f(u, a)$ to mean $f(u, a, 0)$. In this section we let $||\cdot||$ denote the norm $||\cdot||_\infty$ on $\mathbb{R}^L$. To prove Lemma 1 we need three new lemmas. The first lemma gives us a condition under which two demanded bundles are the same.

Lemma 5. *Suppose for $p, p' \in \mathbb{R}^L_{++}$ we have $v_0(p) \geq v_0(p')$ and $x_0(p) \cdot (p' - p) \leq 0$. Then $x_0(p) = x_0(p')$*

Proof. Notice that $x_0(p) \cdot (p' - p) = p' \cdot (x_0(p) - x_0(p')) \leq 0$. Because $x_0(p)$ is weakly preferred to $x_0(p')$, $x_0(p')$ is chosen when $x_0(p)$ is available, and x_0 is single-valued we see that $x_0(p) = x_0(p')$. □

The next lemma gives a condition which ensures that two demanded bundles are close.

Lemma 6. *Let $\varepsilon > 0$. There exists a $\delta > 0$ such that for all $p, p', p'' \in P$ satisfying (i) $||p - p'|| \leq \delta$, (ii) $v_0(p'') \geq v_0(p')$, (iii) $x_0(p'') \cdot (p - p'') \leq \delta$ we have $||x_0(p) - \frac{x_0(p'')}{p \cdot x_0(p'')}|| \leq \varepsilon$.*

Proof. Suppose not. Then there is a sequence $(\delta_n, p_n, p'_n, p''_n)_{n \in \mathbb{N}}$ in $\mathbb{R}_{++} \times P^3$ such that $\delta_n \to 0$ and (i) $||p_n - p'_n|| \leq \delta_n$, (ii) $v_0(p''_n) \geq v_0(p'_n)$, and (iii) $x_0(p''_n) \cdot (p_n - p''_n) \leq \delta_n$ but $||x_0(p_n) - \frac{x_0(p''_n)}{p_n \cdot x_0(p''_n)}|| > \varepsilon$. As P is compact there is a convergent subsequence $(\delta_{n_k}, p_{n_k}, p'_{n_k}, p''_{n_k})_{k \in \mathbb{N}}$ where $p_{n_k} \to p$, $p'_{n_k} \to p'$, and $p''_{n_k} \to p''$ for some $p, p', p'' \in P$. This implies (i) $||p - p'|| = 0$, (ii) $v_0(p'') \geq v_0(p')$, (iii) $x_0(p'') \cdot (p - p'') \leq 0$ and (iv) $||x_0(p) - \frac{x_0(p'')}{p \cdot x_0(p'')}|| \geq \varepsilon$. Item (i) shows that $p = p'$. Item (ii) then states that $v_0(p'') \geq v_0(p)$. Using this with item (iii) and Lemma 5 we see that $x_0(p) = x_0(p'')$. But this contradicts item (iv). We have achieved a contradiction and so the lemma is proved. □

The next lemma is important for constructing the basic indirect utility functions which will be used to prove Lemma 1.

Lemma 7. *Let $p_1, \ldots, p_N \in \mathbb{R}^L_{++}$, $c_1 = x_0(p_1), \ldots, c_N = x_0(p_N)$, and let $\delta > 0$. There exists $u_1, \ldots, u_N \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R}_{++}$ so that for all $m > n$ and $p \in P$*

$$c_m \cdot (p - p_m) \leq 0 \implies f(p, u_m, \lambda_m c_m) \geq f(p, u_n, \lambda_n c_n) + 1 \tag{24}$$

$$c_m \cdot (p - p_m) \geq \delta \implies f(p, u_n, \lambda_n c_n) \geq f(p, u_m, \lambda_m c_m) + 1 \tag{25}$$

Proof. We shall find $\bar{u}_1, \ldots, \bar{u}_N \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R}_{++}$ so that for all $m > n$, and $p \in P$

$$c_m \cdot (p - p_m) \leq 0 \implies \bar{u}_m - \lambda_m c_m \cdot (p - p_m) \geq \bar{u}_n - \lambda_n c_n \cdot (p - p_n) + 1 \tag{26}$$

$$c_m \cdot (p - p_m) \geq \delta \implies \bar{u}_n - \lambda_n c_n \cdot (p - p_n) \geq \bar{u}_m - \lambda_m c_m \cdot (p - p_m) + 1 \tag{27}$$

We get the desired result by putting $u_n = \bar{u}_n - \lambda_n$ as

$$\bar{u}_n - \lambda_n c_n \cdot (p - p_n) = u_n - \lambda_n - \lambda_n c_n \cdot p + \lambda_n = u_n - \lambda_n c_n = f(p, u_n, \lambda_n c_n)$$

which after plugging into (26) and (27) we see (24) and (25).

Let $\bar{u}_1 = 0$ and $\lambda_1 = 1$. Let $\bar{u}_2$ be large enough so that $\bar{u}_2 \geq -c_1 \cdot (p - p_1) + 1$ for all $p \in P$. Such a number will exist because P is compact. Next, let $\lambda_2 \in \mathbb{R}_{++}$ be large enough so that $-x_0(p_1) \cdot (p - p_1) \geq \lambda u_2 - \lambda_2 c_2 \cdot (p - p_2) + 1$ for $p \in P$ such that $c_2 \cdot (p - p_2) \geq \delta$. Equations (26) and (27) holds for $n = 1$ and $m = 2$. We shall now use an induction argument to finish the proof.

Suppose that we have found $\bar{u}_1, \ldots, \bar{u}_{\bar{N}-1}$ and $\lambda_1, \ldots, \lambda_{\bar{N}-1}$ such that (26), (27) hold. Let $\bar{u}_{\bar{N}}$ be large enough so that $\bar{u}_{\bar{N}} \geq \bar{u}_n - \lambda_n c_n \cdot (p - p_n) + 1$ for all $p \in P$. Such a number exists because P is compact. Assign $\lambda_{\bar{N}} > 0$ large enough so that $\bar{u}_n - \lambda_n c_n \cdot (p - p_n) \geq \bar{u}_{\bar{N}} - c_{\bar{N}} \cdot (p - p_{\bar{N}})$ for all $p \in P$ such that $c_{\bar{N}} \cdot (p - p_{\bar{N}}) \geq \delta$. Equations (26) and (27) hold for $n < \bar{N}$ and $m = \bar{N}$. Thus, we have shown that we can find the desired numbers. □

Readers familiar with Afriat (1967) and Varian (1982) will notice similarities between our equations (26), (27) and the so-called Afriat numbers. The spirit is indeed similar but it should be noted that equations (26) and (27) are required to hold for all $p \in P$ whereas the Afriat numbers are defined with respect to a finite grid of points. We now prove Lemma 1.

Proof of Lemma 1. Let $\varepsilon > 0$. Let $\delta > 0$ be chosen so that for all p, p', $p'' \in P$

$$||p - p'|| \leq \delta, \quad v_0(p'') \geq v_0(p'), \quad x_0(p'') \cdot (p - p'') \leq \delta \implies \left\| x_0(p) - \frac{x_0(p'')}{p \cdot x_0(p'')} \right\| \leq \varepsilon \tag{28}$$

Such a δ is guaranteed to exist by Lemma 6. Because P is compact we may find an $N \in \mathbb{N}$ and sequence of prices $p_1, \ldots, p_N \in \mathbb{R}^L_{++}$ such that $v_0(p_N) \geq \ldots \geq v_0(p_1)$ and for all $p \in P$ there is an n so that $||p - p_n|| \leq \delta$ and $p_n \geq p$. For all n let $c_n = x_0(p_n)$.

Lemma 7 shows that we may find numbers $u_1, \ldots, u_N \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R}_{++}$ so that (24) and (25) hold. There will be a $\bar{T} \in \mathbb{N}$ so that $N \leq N_{\bar{T}}$ and $(u_n, \lambda_n c_n, 0) \in$

$\Theta(\eta_{\bar{T}})$. Let $T \geq \bar{T}$. Now, $\tilde{\theta} = (u_1, \lambda_1 c_1, 0, \ldots u_N, \lambda_N c_N, 0) \in \Theta(\eta_T)^N$ and thus $\theta^T = (\tilde{\theta}, u_N, \lambda_N c_N, 0, \ldots, u_N, \lambda_N, 0) \in \Theta(\eta_N)^{N_T}$ where $u_N, \lambda_N c_N, 0, \ldots, u_N, \lambda_N, 0$ is just $N_T - N$ copies of $(u_N, \lambda_N c_N, 0)$. Let $x^T : \mathbb{R}^L_{++} \to \mathbb{R}^L_+$ be defined by $x^T(p) = x_0(p, \theta^T)$.

Let $p \in P$. We will show that $||x_0(p) - x^T(p)|| \leq \varepsilon$. From how we selected $p_1, \ldots, p_{N_T}$ we know there is a k so that $||p - p_k|| \leq \delta$ and $p_k \geq p$. From (16) we know that $x^T(p)$ will be a weighted average of $c_n/c_n \cdot p$ for $n \leq N_T$. Further, from the definition of g we see that the weight assigned to $c_n/c_n \cdot p$ is 0 if $f(p, u_k, \lambda_k c_k) \geq f(p, u_n, \lambda_n c_n) + 1$. Thus, $x^T(p)$ is a weighted average of $c_n/c_n \cdot p$ such that it is not the case that

$$f(p, u_k, \lambda_k c_k) \geq f(p, u_n, \lambda_n c_n) + 1 \tag{29}$$

We shall show that for any p_n not satisfying (29) we have $||x_0 - c_n/c_n \cdot p|| \leq \varepsilon$ which will establish that $||x_0(p) - \hat{x}^T(p)|| \leq \varepsilon$. Now, suppose that p_n does not satisfy (29). Consider three cases: (1) $c_k = c_n$, (2) $k > n$, $c_k \neq c_n$, (3) $n > k$, $x(p_k) \neq x(p_n)$. Consider case (1). The conclusion follows by setting $p = p$, $p' = p_k$, and $p'' = p_n$ in equation (28).

Suppose case (2) holds. As $p_k \geq p$ we have $c_k \cdot (p - p_k) \leq 0$ and so equation (24) and the fact that $k > n$ shows that $f(p, u_k, \lambda_k c_k) \geq f(p, u_n, \lambda_n c_n) + 1$ which contradicts (29). Thus, case (2) never occurs. Suppose case (3) holds. Equation (25) and the fact that (29) does not hold shows that $c_n \cdot (p - p_n) < \delta$. Setting $p = p$, $p' = p_k$, and $p'' = p_n$ in (28) demonstrates that $||x_0 - c_n/c_n \cdot p|| \leq \varepsilon$.

So, $||x_0(p) - \hat{x}^T(p)|| \leq \varepsilon$. But, $p \in P$ was chosen arbitrarily and so $||x_0 - x^T||_{\mu,\infty} \leq \varepsilon$ for all $T \geq \bar{T}$. The result follows by sending ε to 0. $\square$

C Proving Lemma 2

Throughout this section was maintain Assumptions 1 and 2. There is an $\alpha \geq 0$ such that

$$\alpha = \max_{p \in P} \left(\max(p_1, \ldots, p_L, \tfrac{1}{p_1}, \ldots, \tfrac{1}{p_L}) \right)$$

We leave this α fixed throughout this section. In this section f is defined by (9), g is defined by (10), G is defined by (13), h is defined by (11), and H is defined by (14).

We break the section into three parts. First, we establish that x defined in (15) is Lipschitz on $P \times \Theta^N(\eta)$. Second we prove a lemma and a corollary relating to the loss function defined in (17). The third subsection uses these result and a result in Pollard (1984) to prove Lemma 2.

C.1 The Estimator is Lipschitz

This subsection is devoted to proving x is Lipschitz. In what follows we shall refer to a generic element of $P \times \Theta(\eta)^N$ as ξ.

Lemma 8. *For all $\tilde{\xi}$, $\xi \in P \times \Theta(\eta)^N$*

$$\left|x_\ell(\tilde{\xi}) - x_\ell(\xi)\right| \leq 96\alpha^6\eta^5L^2N^2M||\tilde{\xi} - \xi||_\infty \tag{30}$$

This will be shown through a series of lemmas which explore the properties of g, G, h, and H. The following lemma collects facts about these functions.

Lemma 9. *Let $\eta > 1$ and $N \in \mathbb{N}$. Then,*

$$||g(\tilde{\nu}) - g(\nu)|| \leq ||\tilde{\nu} - \nu||_\infty, \qquad \text{for all } \tilde{\nu}, \nu \in \mathbb{R}^N \tag{31}$$

$$||G(\tilde{\nu}) - G(\nu)||_\infty \leq 2(M-1)||\tilde{\nu} - \nu||_\infty, \qquad \text{for all } \tilde{\nu}, \nu \in \mathbb{R}^N \tag{32}$$

$$N \geq ||G(\nu)||_1 \geq 1, \qquad \text{for all } \nu \in \mathbb{R}^N \tag{33}$$

$$||h(\tilde{\xi}) - h(\xi)||_\infty \leq 5\alpha\eta L||\tilde{\xi} - \xi||_\infty, \qquad \text{for all } \xi,\ \tilde{\xi} \in P \times \Theta(\eta)^N \tag{34}$$

$$||G(h(\tilde{\xi})) - G(h(\xi))||_\infty \leq 10M\alpha\eta L||\tilde{\xi} - \xi||_\infty, \qquad \text{for all } \xi,\ \tilde{\xi} \in P \times \Theta(\eta)^N \tag{35}$$

$$2\eta\alpha \geq |H_{k,\ell}(\xi)| \geq 2(\eta\alpha)^{-1}, \qquad \text{for all } \xi \in P \times \Theta(\eta)^N \tag{36}$$

$$|H_{k,\ell}(\tilde{\xi}) - H_{k,\ell}(\xi)| \leq 3\alpha^2\eta||\tilde{\xi} - \xi||_\infty, \qquad \text{for all } \xi,\ \tilde{\xi} \in P \times \Theta(\eta)^N \tag{37}$$

Proof. We shall address the above equations (mostly) in order. Let $\bar{\nu}, \nu \in \mathbb{R}^N$. Let $\delta = ||\bar{\nu} - \nu||_\infty$. Notice that if for all $i \leq N$ we have $\bar{\nu}_i - \nu_i = \delta$ then $g(\bar{\nu}) - g(\nu) = \delta$. It is clear that if there exists an i where $\bar{\nu}_i - \nu_i < \delta$ then $g(\bar{\nu}) - g(\nu) < \delta$. Equation (31) follows.

Let $f_i = (\nu_i - g(\nu))_+$ and $\bar{f}_i = (\bar{\nu}_i - g(\bar{\nu}))_+$. Equation (31) shows $|\bar{f}_i - f_i| \leq 2\delta$. Now using the fact that $|f_i(\nu)| \leq 1$ (this follows from the definition of g)

$$G_i(\bar{\nu}) - G_i(\nu) = \left|\bar{f}_i^{M-1} - f_i^{M-1}\right| = \left|\bar{f}_i - f_i\right| \left|\sum_{k=0}^{M-2} \bar{f}_i^k f_i^{M-2-k}\right| \leq 2\delta(M-1)$$

which establishes (32).

For $x \in \mathbb{R}^N$ let $||x||_M = \sum_{i=1}^N |x|^M$. It is well known that

$$N^{1/(M(M-1))}||x||_M \geq ||x||_{M-1} \geq ||x||_M$$

Putting $x = \left(\left(\nu_1 - g(\nu)\right)_+, \ldots, \left(\nu_N - g(\nu)\right)_+\right)$ we may use the above fact as well as the fact that $||x||_M = 1$ to see that (33) holds.

Equation (34) may be established by differentiating h and applying the mean value theorem. This same approach establishes (37). Equation (35) follows immediately from (32) and (37). Finally (36) follows from the definition and plugging in the larger values η and α for θ and p. $\square$

Lemma 10. *Let $\eta > 1$, $N \in \mathbb{N}$, $\xi \in P \times \Theta(\eta)^N$. Then*

$$\left| p' H(p,\theta)' G\big(h(p,\theta)\big) \right| \geq (\alpha\eta)^{-2} \tag{38}$$

Proof. Abbreviating $G = G\big(h(p;\theta)\big)$ and $H = H(p;\theta)$ we have

$$\left| p'H'G \right| = \sum_{i=1}^{N} \sum_{l=1}^{L} H_{i,l} G_i p_l \geq \alpha^{-1} \sum_{i=1}^{N} G_i \sum_{l=1}^{L} H_{i,l} \geq 2\alpha^{-2}\eta^{-1} \sum_{i=1}^{N} G_i \geq \alpha^{-2}\eta^{-2}$$

where the second to last inequality used (36) and the last inequality used (32). $\square$

We are now ready to prove Lemma 8.

Proof of Lemma 8. Let ι_N and ι_L be the vectors of all ones in $\mathbb{R}^N$ and $\mathbb{R}^L$ respectively. Let $\delta = ||\tilde{\xi} - \xi||_\infty$. Make the abbreviations $\tilde{H} = H(\tilde{\xi})$, $H = H(\xi)$, $\tilde{G} = G(h(\tilde{\xi}))$, $G = G(h(\xi))$. Now,

$$\begin{aligned}
&\left| x_\ell(\tilde{\xi}) - x_\ell(\xi) \right| = \left| \frac{\tilde{H}'_\ell \tilde{G}}{\tilde{p}'\tilde{H}'\tilde{G}} - \frac{H'G}{p'H'G} \right| \leq \left(\frac{1}{\tilde{p}'\tilde{H}\tilde{G}p'HG} \right) \left| p'H'G\tilde{H}'_\ell\tilde{G} - \tilde{p}'\tilde{H}'\tilde{G}H'_\ell G \right| \\
&\leq \alpha^2\eta^2 \left| p'H'G\tilde{H}'_\ell\tilde{G} - \tilde{p}'\tilde{H}'\tilde{G}H'_\ell G \right| \qquad \text{by (38)} \\
&= \alpha^2\eta^2 \Big| (\tilde{p}-p)'H'G\tilde{H}'_k\tilde{G} + \tilde{p}'(H-\tilde{H})'G\tilde{H}'_k\tilde{G} + \tilde{p}'\tilde{H}'(G-\tilde{G})\tilde{H}'_k\tilde{G} + \tilde{p}'\tilde{H}'\tilde{G}(H_k - \tilde{H}_k)'\tilde{G} \\
&+ \tilde{p}'\tilde{H}'\tilde{G}\tilde{H}'_k(\tilde{G}-G) \Big| \qquad \text{to see this step just distribute the expressions and cancel terms.} \\
&\leq \alpha^2\eta^2 \Big| 4\alpha^3\eta^2\delta\iota'_L\iota_L\iota'_N G\iota'_N\tilde{G} + 6\alpha^4\eta^2\delta\iota'_L\iota_L\iota'_N G\iota'_N\tilde{G} + 40M\alpha^4\eta^3 L\delta\iota'_L\iota_L\iota'_N\iota_N\iota'_N\tilde{G} \\
&+ 6\alpha^4\eta^2\delta\iota'_L\iota_L\iota'_N\tilde{G}\iota'_N\tilde{G} + 40M\alpha^4\eta^3 L\delta\iota'_L\iota_L\iota'_N\tilde{G} \Big| \qquad \text{uses (32), (36), (37)} \\
&\leq \alpha^5\eta^4\delta \left| 4LN^2 + 6\alpha LN^2 + 40M\alpha\eta L^2N^2 + 6\alpha LN^2 + 40M\alpha\eta L^2 N \right| \qquad \text{uses (33)} \\
&\leq \alpha^6\eta^5 L^2N^2 M\delta \Big(4+6+40+6+40 \Big) \leq 96\alpha^6\eta^5 L^2N^2M\delta
\end{aligned}$$

Which is what we needed to show. $\square$

C.2 The Loss Function

This subsection will establish some important properties of $\mathscr{L}_T$ and $||\cdot||_2$. We first establish two properties of the $||\cdot||_2$ norm.

Lemma 11. *Let $p \in P$ and suppose z, z', and $z'' \in \mathbb{R}^L_+$ satisfy $p \cdot z = p \cdot z' = p \cdot z'' = 1$. Then*

$$||z - z'||_2 \leq 2\alpha \tag{39}$$

$$\Big| \big\| z' - z \big\|_2^2 - \big\| z'' - z \big\|_2^2 \Big| \leq 4\alpha \big\| z' - z'' \big\|_\infty \tag{40}$$

Proof. Note that

$$||z||_1 = |\iota \cdot z| \leq \alpha |p \cdot z| = \alpha \tag{41}$$

Now, using Holder's inequality and (41) we see

$$||z - z'||_2 \leq \Big(||z - z'||_1 ||z - z'||_\infty \Big)^{1/2} \leq \Big((||z||_1 + ||z'||_1)(||z||_\infty + ||z'||_\infty) \Big)^{1/2} \leq 2\alpha$$

So, we have established (39). Next,

$$\begin{aligned} \Big| \big\| z' - z \big\|_2^2 - \big\| z'' - z \big\|_2^2 \Big| &= \Big| (z' - z) \cdot (z' - z) - (z'' - z) \cdot (z'' - z) \Big| \\ &= \Big| z' \cdot z' - z'' \cdot z'' + 2z \cdot (z'' - z') \Big| \\ &\leq \Big| (z' - z'') \cdot (z' + z'') \Big| + 2||z||_1 ||z'' - z'||_\infty \Big| \\ &\leq 2\alpha ||z' - z''||_\infty + 2\alpha ||z'' - z'||_\infty = 4\alpha ||z' - z''|| \end{aligned}$$

where the last step used (41). □

This leads to a useful corollary. Define $\mathscr{Z}$ by $\mathscr{Z} = \{(y, p) \in \mathbb{R}^L_+ \times P,\ p \cdot y = 1\}$.

Corollary 1. *Let $\eta > 1$ and $N \in \mathbb{N}$. Define $m : \Theta(\eta)^N \times \mathscr{Z} \to \mathbb{R}$ by*

$$m(\theta, y, p) = \big\| y - x(p, \theta) \big\|_2^2 \tag{42}$$

For any $\tilde{\theta}$, $\theta \in \Theta(\eta)^N$ and $(p, y) \in \mathscr{Z}$

$$\big| m(\tilde{\theta}, y, p) - m(\theta, y, p) \big| \leq 384 \alpha^7 \eta^5 M L^2 N^2 \big\| \tilde{\theta} - \theta \big\|_\infty$$

Proof. This follows immediately from (30) and (40). □

C.3 Uniform Convergence of the Loss Function

Let $(X, ||\cdot||)$ be a normed space. Let $\mathbf{N}(\delta, X, ||\cdot||)$ be the smallest number of open balls of radius δ needed to cover X (if no finite number of balls satisfies this condition then let $\mathbf{N}(\delta, X, ||\cdot||)$ be ∞). We refer to $\mathbf{N}(\delta, X, ||\cdot||)$ as the *covering number* of X. Let $\mathbf{H}(\delta, X, ||\cdot||) = \ln(\mathbf{N}(\delta, X, ||\cdot||))$. We refer to this number as the *metric entropy* of X.

Let $\mathscr{M}(\eta, N)$ be the set of all functions $m(\theta,\cdot,\cdot) : \mathscr{X} \to \mathbb{R}$ defined by (42) for some $\theta \in \Theta(\eta)^N$.

Lemma 12. *Let $\eta > 1$, $N \in \mathbb{N}$, and $\delta \in (0, 1)$. Then*

$$\mathbf{N}(\delta, \mathscr{M}, ||\cdot||_{\infty,\mu}) \leq \left(800\alpha^7\eta^6 ML^2N^2\delta^{-1}\right)^{(2L+1)N}$$

Proof. For $z \in \mathbb{R}$ let $\lceil z\rceil$ denote the smallest integer that is greater than or equal to z. For $\bar{\delta} > 0$ we may cover an interval $[1/\eta_T, \eta_T]$ with $\lceil \eta/\bar{\delta}\rceil$ balls of size $\bar{\delta}$. This implies that the set $\Theta(\eta)^N$ can be covered with $(\lceil\eta/\bar{\delta}\rceil)^{N(2L+1)}$ balls of size $de\bar{l}ta$. This means that we may cover $\Theta(\eta)^N$ with $(\lceil 400\alpha^7\eta^6ML^2N^2\bar{\delta}^{-1}\rceil)^{N(2L+1)}$ balls of size $400\alpha^7\eta^5ML^2N^2$. From the definition of $\mathscr{M}$ the map $F : \Theta(\eta)^N \to \mathscr{M}$ defined by $F(\theta) = m(\theta,\cdot,\cdot)$ is onto and Corollary 1 shows that f maps a ball of size $400\alpha^7\eta^5ML^2N^2$ in $(\Theta(\eta)^T, ||\cdot||_\infty)$ into a ball of size δ in $(\mathscr{M}, ||\cdot||_{\infty,\mu})$. This demonstrates that $(\mathscr{M}, ||\cdot||_{\infty,\mu})$ can be covered by $(\lceil 400\alpha^7\eta^6ML^2N^2\rceil)^{N(2L+1)}$ balls of size δ. Now,

$$\mathbf{N}(\delta, \mathscr{M}, ||\cdot||_{\infty,\mu}) \leq \left(\lceil 400\alpha^7\eta^6ML^2N^2\delta^{-1}\rceil\right)^{(2L+1)N} \leq \left(800\alpha^7\eta^6ML^2N^2\delta^{-1}\right)^{(2L+1)N}$$

which completes the proof. □

Let $Z_1, Z_2, Z_3, \ldots$ be a sequence of iid observations with state space $\mathbf{A}$. For measurable set $A \subseteq \mathbf{A}$ let $\mu_T(A) = \sum_{t=1}^T \mathbf{1}_A Z_t$ where $\mathbf{1}$ is the indicator function. We refer to μ_T as the *empirical measure*.

Definition 2. Let $\mathscr{F}$ be a set of real-valued functions with domain $A \subseteq \mathbb{R}^{\tilde{N}}_{++}$ bounded in absolute value by $\kappa \geq 0$. Suppose $\mathscr{F}$ is indexed by a compact set $K \subseteq \mathbb{R}^N$. For $\theta \in K$ let $f(\cdot,\theta)$ return the function in $\mathscr{F}$ corresponding to the index θ. We say that $\mathscr{F}$ is a *suitable* class of functions if the map $f : \mathbb{R}^L_{++} \times K \to \mathbb{R}$ is measurable in the product measure.

Lemma 13. *Let $Z_1, Z_2, Z_3, \ldots$ be a sequence of iid random vectors in $A \subseteq \mathbb{R}^{\tilde{N}}_{++}$ with distribution μ. Let $\mathscr{F}$ be a suitable class of functions. For any $\delta > 0$*

$$\mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{1,\mu_T}) \leq \mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{\infty,\mu})$$

with probability 1.

Proof. First note that $\mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{1,\mu_T}) \leq \mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{\infty,\mu_T})$. Next, observe that $\mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{\infty,\mu_T}) \leq \mathbf{N}(\delta, \mathscr{F}, ||\cdot||_{\infty,\mu})$ except maybe on a set of probability 0. □

Write $\mathbf{E}_T$ to refer to the expectation taken with respect to the empirical measure. We will use the following result from Pollard (1984).

Lemma 14 (Pollard). *Let $Z_1, Z_2, Z_3, \ldots$ be a sequence of iid random vectors in $\mathbb{R}_{++}^{\tilde{N}}$ with distribution μ. Let $\mathscr{F}$ be a suitable class of functions bounded above by $\kappa > 0$. Then for $\delta > 0$ and $T \geq 8\kappa\delta^{-2}$*

$$\Pr\left(\sup_{f\in\mathscr{F}}\left|\mathbf{E}[f] - \mathbf{E}_T[f]\right| \geq \delta\right) \leq 8\exp\left(-\frac{T\delta^2}{256\kappa^2}\right) + 4\Pr\left(\mathbf{H}\left(\tfrac{\delta}{8}, \mathscr{F}, ||\cdot||_{1,\mu_T}\right) > \frac{T\delta^2}{256\kappa^2}\right)$$

This result is shown in the proof of Theorem 24 in Pollard (1984) where we have used a more simple version of what he refers to as a permissible class of functions.

Corollary 2. *Let $\delta \in (0,1)$, $\eta > 1$, and $N \in \mathbb{N}$. Let $T \geq 16\alpha^2\delta^{-2}$ and*

$$T \geq 256(2L+1)N\delta^{-2}\ln\left(6400\alpha^7\eta^6 ML^2N^2\delta^{-1}\right)$$

Then

$$\Pr\left(\sup_{\theta\in\Theta(N)^\eta}\left|\mathscr{L}(\theta) - \mathscr{L}_T(\theta)\right| \geq \delta\right) \leq 8\exp\left(-\frac{T\delta^2}{1024\alpha^2}\right) \tag{43}$$

Proof. First, notice that $\mathscr{L}(\theta) = \mathbf{E}[m(\theta)]$ and $\mathscr{L}_T(\theta) = \mathbf{E}_T[m(\theta)]$. Note Lemma 11 shows that the functions $\mathscr{M}(\eta, N)$ are bounded above by 2α. Applying Lemma 14 we have

$$\begin{aligned}
&\Pr\left(\sup_{\theta\in\Theta(N)^\eta}\left|\mathscr{L}(\theta) - \mathscr{L}_T(\theta)\right| \geq \delta\right)\\
&\leq 8\exp\left(-\frac{T\delta^2}{1024\alpha^2}\right) + 4\Pr\left(\mathbf{H}\left(\tfrac{\delta}{8}, \mathscr{M}(\eta,N), ||\cdot||_{1,\mu_T}\right) \geq \frac{T\delta^2}{1024\alpha^2}\right)\\
&\leq 8\exp\left(-\frac{T\delta^2}{1024\alpha^2}\right) + 4\Pr\left(\mathbf{H}\left(\tfrac{\delta}{8}, \mathscr{M}(\eta,N), ||\cdot||_{\infty,\mu}\right) \geq \frac{T\delta^2}{1024\alpha^2}\right) = 8\exp\left(-\frac{T\delta^2}{1024\alpha^2}\right)
\end{aligned}$$

where the second to last inequality came from Lemma 13 and the final equality follows from Lemma 12. □

Using Corollary 2 it is straightforward to prove Lemma 2.

Proof of Lemma 2. Just take $T \to \infty$ in equation (43). □

References

Sydney Afriat. The construction of a utility function from expenditure data. *International Economic Review*, 8, 1967.

James. Banks, Richard Blundell, and Arthur Lewbell. Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79:527–539, 1997.

William A. Barnett and Andrew B. Jonas. The Muntz-Szatz demand system: An application of a globally well behaved series expansion. *Economics Letters*, 11:337–342, 1983.

William A. Barnett and Apostolos Serletis. Consumer preferences and demand systems. *Journal of Econometrics*, 147:210–224, 2008.

Charlies Blackorby, Daniel Primont, and R. Robert Russell. *Duality, Separability, and Function Structure: Theory and Economic Applications*. 1978.

Richard Blundell, Joel L. Horowitz, and Matthias Parey. Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, 3:29–51, 2012.

Richard Blundell, Dennis Kristensen, and Rosa Matzkin. Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics*, 179:112–127, 2014.

Richard Blundell, Joel L. Horowitz, and Matthias Parey. Nonparametric estimation of a nonseparable demand function under the Slutsky inequality restriction. *Review of Economic Studies*, 99(2):291–304, 2017.

Bryan W. Brown and Mary Beth Walker. The random utility hypothesis and inference in demand systems. *Econometrica*, 57(4):815–829, 1989.

Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. In Handbook of Econometrics, editor, *Heckman, James and Leamer, Edward*, volume 6B, pages 5549–5632. 2007.

Laurits R. Christensen, Dale W. Jorgenson, and Lawrence J. Lau. Transcendental logarithmic utility functions. *American Economic Review*, 65:367–383, 1975.

Angus S. Deaton and John N. Muellbauer. An almost ideal demand system. *American Economic Review*, 70:312–326, 1980.

W. Erwin Diewert. Applications of duality theory. In M. Intriligator and D. Kendrick, editors, *Frontiers in Quantitative Economics*, volume 2. 1974.

Ronald A. Gallant. On the bias in flexible functional forms and an essentially unbiased form. *Journal of Econometrics*, 15:211–245, 1981.

Ulf Grenander. *Abstract Inference*. 1981.

Berthold R. Haag, Stefan Hoderlein, and Krishna Pendakur. Testing and imposing Slutsky symmetry in nonparametric demand systems. *Journal of Econometrics*, 153: 33–50, 2009.

Stefan Hoderlein and Jorg Stoye. Revealed preferences in a heterogeneous population. *The Review of Economics and Statistics*, 96(2):197–213, 2014.

Arthur Lewbel. Demand systems with and without errors. *The American Economic Review*, 91(3):611–618, 2001.

Arthur Lewbel and Krishna Pendakur. Tricks with Hicks. *American Economic Review*, 99(3):827–863, 2009.

Krishna Pendakur and Stefan Sperlich. Semiparametric estimation of consumer demand systems in real expenditure. *Journal of Applied Econometrics*, 25:420–457, 2010.

David Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

Hal R. Varian. The non-parametric approach to demand analysis. *Econometrica*, 50(4): 945–973, July 1982.

Halbert White and Jeffrey Wooldridge. Some results on sieve estimation with dependent observations. In William A. Barnett, James Powell, and George Tauchen, editors, *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*. 1991.

Chapter 2

Giffen Behavior Under Risk and Ambiguity

Giffen Behavior Under Risk and Ambiguity

Abstract

We provide necessary and sufficient conditions (in terms of the coefficient of relative risk aversion) for the demand for financial assets to be ordinary (i.e. non-Giffen) when preferences take the Expected Utility, Choquet Expected Utility, and the Maxmin Expected Utility (and α-Maxmin) forms. We show that when an agent has Maxmin Expected Utility preferences they always display Giffen behavior in some financial environment except in the special case of homothetic preferences. Surprisingly, in the closely related Choquet Expected Utility model this conclusion no longer holds and the conditions to preclude Giffen behavior coincide with the Expected Utility case. We also provide some sufficient conditions for assets to be ordinary when preferences display smooth ambiguity aversion.

1 Introduction

We investigate conditions on preferences for risk and ambiguity which allow for or preclude Giffen behavior. Giffen behavior, the increase in quantity demanded by an individual in response to an increase in price, is not ruled out by the usual assumptions (increasing, convexity, continuity, etc.) made on preferences. Yet, Giffen behavior often seems an affront to intuition and is rare empirically (see Jensen and Miller (2008) for discussion and some evidence of Giffen behavior.) There are three reasons why finding conditions on preferences which preclude Giffen behavior is important. First, from a theoretical position it allows for the precise statement of assumptions required to preclude this often unrealistic behavior. Second, suppose the preferences of a consumer has been estimated through empirical means. It is helpful to interpret these preferences through some of their behavioral consequences. If the estimate is taken seriously then it can provide evidence of the existence of Giffen goods. On the other hand, if Giffen behavior is deemed implausible in a given context then the conditions which preclude them may be imposed on the preference estimates. Third, as we explain in this article, the agent's demand displays the property of monotonicity if and only if there are no Giffen assets. Monotonicity is an important property in the general equilibrium model which helps guarantee uniqueness and stability of equilibrium prices. Thus, our results provide conditions which help ensure uniqueness of equilibrium prices.

In this paper we focus on models of risk and ambiguity where an agent may purchase assets which deliver consumption in the next period. This conditional consumption is referred to as contingent consumption. An asset X is a random variable which delivers a specific amount of consumption (possibly negative) in each state. We find conditions on a consumer's preferences which preclude the possibility that the consumer exhibits Giffen behavior when buying assets. The basic ingredients of the analysis are (i) an object Λ representing the agent's beliefs (for example a probability or something more general) which we refer to as an *uncertainty*, (ii) a utility function U mapping beliefs (Λ) and contingent consumption bundles into the real line, (iii) an ordered collection of assets A called an asset matrix. We refer to the pair (A, Λ) as a financial environment.

First, we address the classic case of expected utility preferences. Here, the agent is endowed with some strictly increasing, strictly concave, and twice continuously differentiable function $u : \mathbb{R}_{++} \to \mathbb{R}$ (referred to as a Bernoulli utility function) which allows the agent to assign a utility value $u(a)$ to the consumption level a. The agent seeks to maximize the expected value of the utility of their future consumption. That is, the agent seeks to maximize $\mathbf{E}[u(c)]$, where c is a random variable representing the payoffs of the agent's asset holdings. The object which represents the agent's beliefs in this context is a probability π.

We find necessary and sufficient conditions on u which preclude the existence of financial environments (A, π) (where π is a probability) under which Giffen behavior can occur. These conditions are given in terms of the curvature of u as measured by the coefficient of relative risk aversion. Recall that the coefficient of relative risk aversion is the function $\sigma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ defined by

$$\sigma(a) = -\frac{u''(a)}{u'(a)}a \tag{1}$$

We show that an agent with expected utility preferences will never display Giffen behavior for any asset structure if and only if

$$\sqrt{\sigma(d)} - \sqrt{\sigma(a)} \leq 2, \qquad \text{for all } a,\ d \in \mathbb{R}_{++} \tag{2}$$

This result is a generalization of a result in Quah (2003). We will say more on the connection between this result and Quah (2003) in Section 3.

We next consider some of the generalizations of the expected utility model. Specifically, we investigate Giffen behavior in the presence of ambiguity (that is, where there is uncertainty about the true probabilities of events). There is a plethora of models which allow for the presence of ambiguity. We characterize Giffen behavior in several of the most popular models. Particularly, we consider the Maxmin Expected Utility (MEU) (and also the α-MEU), Choquet Expected Utility (CEU), and the Smooth Ambiguity Aversion (SAA) models.

In the MEU model (axiomatized in Gilboa and Schmeidler (1989)) an agent's beliefs (the uncertainty in our parlance) is a set of probabilities $\mathcal{P}$. Each of these probabilities represents a characterization of the world which the agent considers plausible. As the agent is averse to ambiguity he derives a utility value from contingent consumption c equal to the minimum of all expected utilities derived from each of the probabilities in $\mathcal{P}$. That is, the agent derives utility

$$\min_{\pi \in \mathcal{P}} \mathbf{E}_\pi[u(c)] = \min_{\pi \in \mathcal{P}} \int u(x) d\pi$$

We show that an agent will not display Giffen behavior in any financial environment if and only if u displays constant relative risk aversion. Constant relative risk aversion is a very restrictive assumption. Indeed, it implies preferences are homothetic which means that a doubling of an agent's wealth results in a doubling of the agent's purchased contingent consumption bundle. Whether this assumption is reasonable is an empirical matter. See Chiappori and Paiella (2011) and references therein for discussion.

We also consider a popular generalization of MEU preferences; the α-MEU preferences (discussed in Marinacci (2002)). This class of preferences operates in the same financial environments as MEU. An agent with α-MEU preferences derives utility

$$\alpha \min_{\pi \in \mathcal{P}} \mathbf{E}_\pi[u(c)] + (1-\alpha) \max_{\pi \in \mathcal{P}} \mathbf{E}_\pi[u(c)]$$

from a contingent consumption bundle c. The $\alpha \in [0,1]$ represents the agent's attitude towards ambiguity. We show that provided $\alpha \in (0.5, 1]$ (the agent is sufficiently ambiguity averse) the agent will not display Giffen behavior if and only if u displays constant relative risk aversion. That is, our results do not change when generalizing MEU to α-MEU.

CEU preferences (axiomatized in Schmeidler (1989)) use convex non-additive probabilities to evaluate the utility of asset holdings, where expected values are calculated using the Choquet integral. We will elaborate on the precise meaning of these expressions later. An uncertainty in this framework is a convex non-additive probability P. The agent derives utility of $\mathbf{CE}[u(X)]$ where $\mathbf{CE}$ is the Choquet integral with respect to P. We show that in this model there are no additional restrictions needed on u to preclude Giffen behavior beyond what is required in the EU framework. That is, an agent with CEU preferences will never display Giffen behavior in any financial environment if and only if (2) holds.

The SAA model (presented in Klibanoff, Marinacci, and Mukerji (2005)) is similar to the MEU model in that the agent's knowledge Λ is expressed as a set $\mathcal{P}$ of probabilities. What distinguishes SAA is that there is a probability η on $\mathcal{P}$ where $\eta(\pi)$ may be interpreted as the confidence with which the agent believes that π is the true distribution. The agent is endowed with a function $v : \mathbb{R}_{++} \to \mathbb{R}$ which acts as a Bernoulli utility function over the random variable of certainty equivalents of each possible probability in $\mathcal{P}$. It is assumed that v is more concave than u. To a contingent consumption bundle c the agent assigns a utility value of

$$\mathbf{E}_\eta\left[v\Big(u^{-1}\big(\mathbf{E}_\pi\left[u(c)\right]\big)\Big)\right] = \int_{\mathcal{P}} v\left(u^{-1}\left(\int_\Omega u(c)d\pi\right)\right)d\eta$$

We define a net coefficient of relative risk aversion which is analogous to σ defined in (1). Let $\gamma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ be defined by

$$\gamma(a) = -\left(\frac{v''(a)}{v'(a)} - \frac{u''(a)}{u'(a)}\right)a$$

We refer to γ as the *net coefficient of relative risk aversion*. We show that if u displays increasing relative risk aversion then a sufficient condition to preclude Giffen behavior is for

$$\gamma(d) + \sigma(a) \leq 4, \qquad \text{for all } a, d \in \mathbb{R}_{++} \tag{3}$$

Now we provide a few thoughts on these results. Perhaps the most striking finding is how great a chasm exists between the requirements for an agent with CEU preferences to not display Giffen behavior and those requirements for an agent with MEU preferences. The requirement for an agent with CEU preferences is no more demanding than for a standard EU maximizer. However, when the agent has MEU preferences it is only the extremely special case of u displaying constant relative risk aversion that can prevent Giffen behavior. This looks especially strange in light of the well-known link between CEU and MEU preferences which can be expressed in terms of the core of the convex non-additive probability P. Let $\Delta(S)$ be the set of probabilities on sample space S. The core of P is

$$\text{core}(P) = \{\pi \in \Delta(S) : \pi(A) \geq P(A) \text{ for all } A \in 2^S\}$$

where 2^S is the power set of S. It is well known that

$$\mathbf{CE}[u(X)] = \min_{\pi \in \mathrm{core}(P)} \mathbf{E}_\pi[u(X)]$$

In words, CEU preferences are a special case of MEU preferences where the set of probabilities $\mathcal{P}$ used happens to be the core of some convex non-additive probability P. The results of this paper give some indication as to how restrictive it is to require $\mathcal{P}$ to be consistent with CEU preferences. Indeed, it means the difference between the mild requirement that u satisfies (2) and the severe restriction that u displays constant relative risk aversion when requiring that the agent does not display Giffen behavior. A reason why such a gap exists between these two models relates to where the models allow for kinks in their utility functions. Any kink in the utility function for CEU preferences occurs at locations in consumption space where the payoffs derived in several states are equal (for example when an agent holds only a safe asset). This property does not hold in the case of MEU preferences. It is the fact that kinks in the utility function can occur in many locations in MEU preferences which are not possible in CEU preferences that makes the difference.

The SAA model has the useful property that it allows one to separate beliefs (expressed by the probability η) from aversion to ambiguity (expressed by the curvature of v). This allows for many conditions of the form (3), which depend on both risk aversion as expressed by σ and ambiguity aversion as expressed by γ. This nuance is not achievable in the MEU and SEU models which may somewhat explain the more stark results we attain in these cases.

The remainder of the paper is organized as follows. Section 2 introduces the basic concepts used in our exposition. Section 3 is devoted to the treatment of risk in the absence of ambiguity (or ambiguity aversion). We define expected utility preferences and prove that indeed (2) is necessary and sufficient to preclude Giffen behavior in the absence of ambiguity. Section 4 analyzes the models of ambiguity aversion. Section 5 provides some additional results of interest. In particular, we discuss which assets may become Giffen.

2 Definitions and Concepts

Let $\mathbb{N}$ denote the natural numbers. Let $S \in \mathbb{N}$. We often abuse notation and write S to denote both the number in $\mathbb{N}$ and the set $\{1, \ldots, S\}$. For $S \in \mathbb{N}$ let $\Delta(S)$ be the set of probabilities which are defined on the set S. That is $\pi \in \Delta(S)$ means $\pi = (\pi_1, \ldots, \pi_S)$ and $\sum_{s=1}^{S} \pi_s = 1$. Let $\Delta_+(S)$ be the subset of $\Delta(S)$ where each $\pi \in \Delta_+(S)$ satisfies

$\pi_s > 0$ for all $s \in S$.

We model an agent purchasing assets which entitle this agent to consumption in the future depending on which state of the world occurs. A contingent consumption bundle $c \in \mathbb{R}^S_{++}$ specifies the amount of consumption that the agent will receive contingent upon the state of the world. So, c_s represents the amount of consumption the agent will receive should state s occur.

We shall call an object Λ which completely describes the agent's beliefs about the likelihood of the different states of the world occurring an *uncertainty*. The exact structure of the uncertainty will depend upon which model is being considered. In the simplest case of expected utility an uncertainty is just a probability. In the maxmin expected utility model an uncertainty is a set of probabilities, etc.

An agent is characterized by a utility function U which maps contingent consumption and uncertainty to the real line. We say that U is *well-behaved* if for any uncertainty Λ the map $U(\cdot, \Lambda) : \mathbb{R}^S_{++} \to \mathbb{R}$ is continuous, increasing, and strictly concave.

Suppose contingent consumption has a price $p \in \mathbb{R}^S_{++}$ and the agent is endowed with some money to spend $w \in \mathbb{R}_{++}$. We define the *contingent consumption demand function* $x(\cdot, \cdot; \Lambda) : \mathbb{R}^S_{++} \times \mathbb{R} \to \mathbb{R}^S_{++}$ by

$$x(p, w, \Lambda) = \underset{c \in \{c' \in \mathbb{R}^S_{++} :\ p \cdot c' \leq w\}}{\operatorname{argmax}} U(c, \Lambda) \tag{4}$$

The function x shows what an agent will purchase when every contingent consumption bundle has a price. However, it may not always be the case that every contingent consumption bundle is for sale. In fact, we assume that the agent receives contingent consumption through the purchase of assets and so there may be some contingent consumption bundles which may not be purchased.

An *asset* $X \in \mathbb{R}^S$ is an object which specifies how much the agent is owed or owes in each state. So, $X_s \in \mathbb{R}$ specifies the amount of consumption that the asset delivers in state s. An *asset matrix* A is a matrix whose columns are assets ($A = [X_1, \ldots, X_J]$) where there exists a $\theta \in \mathbb{R}^J$ such that $A\theta \gg 0$ ($b \gg 0$ for $b \in \mathbb{R}^S$ means $b_s > 0$ for all $s \in S$). We shall refer to such a $\theta \in \mathbb{R}^J$ as an asset holding or asset position. Let $\Theta_A = \{\theta \in \mathbb{R}^J : A\theta \gg 0\}$.

A *financial environment* is a pair (A, Λ) where A is an asset matrix and Λ is an uncertainty. We assume that each asset in a financial environment has a price. For an asset matrix $A = [X_1, \ldots, X_J]$ we let $q \in \mathbb{R}^J$ denote a price vector where q_j represents the price of asset j. An arbitrage is an asset holding $\theta \in \mathbb{R}^J$ which satisfies ($q \cdot \theta < 0$ and $A\theta \geq 0$) or ($q \cdot \theta \geq 0$ and $A\theta > 0$). For an asset matrix A let $\mathcal{Q}_A \subseteq \mathbb{R}^J$ be the set of prices for which there is no arbitrage. Throughout, we will only consider prices for which there is no arbitrage. The agent is endowed with a certain amount of money $w \in \mathbb{R}_{++}$ which she uses to purchase assets.

The agent's preferences over contingent consumption U induce preferences over asset holdings. To these induced preferences corresponds a demand function for assets. Let $\mathcal{A} = (A, \Lambda)$ be a financial environment. Define the asset demand function for $\mathcal{A}$ as

$$y(q, w, \mathcal{A}) = \underset{\theta \in \{\theta' \in \Theta_A :\ q \cdot \theta' \leq w\}}{\operatorname{argmax}} U(A\theta, \Lambda) \tag{5}$$

We say that the agent displays Giffen behavior (does not displays ordinary behavior) if there is some financial environment $\mathcal{A}$, price pair $\tilde{q}$, $q \in \mathbb{R}^J$, wealth level $w \in \mathbb{R}_{++}$, and k such that $\tilde{q}_j = q_j$ for $j \neq k$ and

$$(\tilde{q}_k - q_k)(y_k(\tilde{q}, w; \mathcal{A}) - y_k(q, w; \mathcal{A})) > 0 \tag{6}$$

We also say that a demand function that satisfies (6) displays Giffen behavior. We develop conditions on preferences which preclude Giffen behavior. The first class of preferences we consider is expected utility.

3 Risk

This section is called risk as it considers an agent who responds to uncertainty as if she is aware of the probability with which each state of the world occurs. This is in distinction to the next section where the agent is not certain about the probability of each state of the world occurring and is averse to this lack of knowledge. In the previous section we have stated that we shall call any object Λ which characterizes the agent's beliefs about the likelihood of each state of the world occurring an uncertainty. In this section an uncertainty is a probability $\pi \in \Delta_+(S)$.

3.1 Expected Utility

An *expected utility (EU) uncertainty is a probability vector* $\pi \in \Delta_+(S)$. We say that an agent has EU preferences if there is some twice continuously differentiable, strictly concave, and strictly increasing function $u : \mathbb{R}_{++} \to \mathbb{R}$ such that contingent consumption bundle $c \in \mathbb{R}^S_{++}$ and probability $\pi \in \Delta_+(S)$ is awarded utility

$$U(c, \pi) = \mathbf{E}_\pi[u(c)] = \sum_{s=1}^{S} \pi_s u(c_s)$$

We refer to u as a *Bernoulli* utility function.

A pair $\mathcal{A} = (A, \pi)$ where A is an $S \times J$ asset matrix and $\pi \in \Delta_+(S)$ is called an *EU financial environment*. An EU financial environment induces an asset demand function y defined by (5). Our goal in this section is to find conditions on the Bernoulli utility function u which preclude the possibility of Giffen behavior over all EU financial environments. The curvature of u, as expressed in the coefficient of relative risk aversion, will play a major role in the characterization. The *coefficient of relative risk aversion* is a function $\sigma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ defined by

$$\sigma(a) = -\frac{u''(a)}{u'(a)}a$$

We now state the main theorem of the section.

Theorem 1. *Suppose an agent has EU preferences with Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$. The agent will not display Giffen behavior on any EU financial environment if and only if*

$$\sqrt{\sigma(d)} - \sqrt{\sigma(a)} \leq 2, \qquad \text{for all } a,\ d \in \mathbb{R}_{++} \tag{7}$$

The theorem states that if (7) fails then there is an EU financial environment $\mathcal{A} = (A, \pi)$ where the asset demand $y(\cdot, \cdot; \mathcal{A})$ displays Giffen behavior. In a later proposition we shall show that when (7) fails there is an asset structure A with 2 risky assets and a risk free asset where the risk free asset is Giffen. Indeed the later proposition will demonstrate that once (7) fails then almost any asset will be a Giffen asset in some EU financial environment.

The proof of Theorem 1 uses many concepts that will be used in the proof of all theorems in this paper. Our first task is to connect Giffen behavior to the law of demand (aka monotonic demand). Once we have done this then a powerful result due to Milleron (1974) and Mityushin and Polterovich (1978) can be applied to characterize utility functions which display Giffen behavior.

3.2 Monotonic Demand

Monotonicity is a property that a demand function may posses which is closely related to the property of ordinary demand. A demand function $x : \mathbb{R}^S_{++} \times \mathbb{R}_{++} \to \mathbb{R}^S_+$ is *monotonic* if for all price vectors $\tilde{p}$, $p \in \mathbb{R}^S_{++}$ and wealth $w \in \mathbb{R}_{++}$

$$\big(\tilde{p} - p\big) \cdot \big(x(\tilde{p}, w) - x(p, w)\big) \leq 0 \tag{8}$$

Monotonic demand clearly precludes the existence of Giffen goods. When considering the case of asset demand this implication goes both ways in a certain sense.

Proposition 1. *Let $U : \mathbb{R}^S_{++} \to \mathbb{R}$ be an increasing, continuous, and strictly concave utility function. Define $x : \mathbb{R}^S_{++} \times \mathbb{R}_{++} \to \mathbb{R}^S_{++}$ by*

$$x(p, w) = \underset{c \in \{c' \in \mathbb{R}^S_{++} :\ p \cdot c' \leq w\}}{\operatorname{argmax}} U(c)$$

For each $J \in \mathbb{N}$ and each $S \times J$ asset matrix A define a function $y(\cdot, \cdot, A) \to \Theta_A$ by

$$y(q, w, A) = \underset{\theta \in \{\theta' \in \Theta^J_A :\ q \cdot \theta' \leq w\}}{\operatorname{argmax}} U(A\theta)$$

Then, y does not display Giffen behavior for all A if and only if x is monotonic.

Proposition 1 shows that when deriving conditions under which asset demand (equation (5)) is ordinary we may focus on conditions which ensure that contingent consumption demand (as defined in equation (4)) is monotonic. Fortunately, there is a powerful result known as the Miller, Mityushin, and Polterovich (MMP) Theorem which shows exactly when a twice continuously differentiable, strictly concave utility function $U : \mathbb{R}^S_{++} \to \mathbb{R}$ generates a demand function which satisfies monotonicity.

Theorem (MMP)**.** *Suppose $C \subseteq \mathbb{R}^S_{++}$ is convex with non-empty interior. Let $U : C \to \mathbb{R}$ be a well-behaved utility function which generates demand function x. Demand function x satisfies monotonicity if and only if*

$$\frac{\mu \cdot c}{\mu \cdot \mathbf{U}^{-1} \mu} - \frac{c \cdot \mathbf{U} c}{\mu \cdot c} \leq 4, \qquad \textit{for all } c \in C \tag{9}$$

where $\mu = \nabla U(c)$ and $\mathbf{U} = \partial^2 U(c)$. Further, if C has an empty interior then (9) is still a sufficient condition for monotonic demand.

We bring the MMP Theorem into the context of asset demands. Let $u : \mathbb{R}_{++} \to \mathbb{R}$ be a Bernoulli utility function. For a $\pi \in \Delta_+(S)$ and a contingent consumption bundle $c \in \mathbb{R}^S_{++}$ define the *MMP-coefficient* by

$$\mathrm{MMP}(c, \pi) = \frac{\mathbf{E}_\pi[u'(c)c]}{\mathbf{E}_\pi[u'(c)^2/u''(c)]} - \frac{\mathbf{E}_\pi[u''(c)c^2]}{\mathbf{E}_\pi[u'(c)c]}$$

$\mathrm{MMP}(c, \pi)$ is the expression on the left hand side of (9) where we substitute $U(c) = \mathbf{E}[u(c)]$. The following corollary connects monotonicity to ordinary demand when preferences are EU.

Corollary 1. *An agent with EU preferences displays ordinary behavior for all EU financial environments if and only if for all $S \in \mathbb{N}$, contingent consumption bundles $c \in \mathbb{R}^S_{++}$, and $\pi \in \Delta_+(S)$ we have MMP(c) ≤ 4.*

Proof. This is a straightforward consequence of applying the MMP Theorem to the demand for contingent consumption context and then applying Proposition 1. $\square$

The next subsection uses Corollary 1 to establish Theorem 1.

3.3 Proving Theorem 1

Let $b : \mathbb{R}^S_{++} \times \Delta_+(S) \to \Delta(S)$ be defined by

$$b_r(c; \pi) = \frac{\pi_r u'(c_r) c_r}{\mathbf{E}[u(c)c]} \tag{10}$$

It is straightforward to show that $b(c, \pi)$ is the vector of budget shares of contingent consumption when the agent has EU preferences and demands contingent consumption c. For this reason we refer to b defined in (10) as the *budget share measure.*

Let $X \in \mathbb{R}^S_{++}$, $\pi \in \Delta_+(S)$, and let $\psi(X, \pi)$ be defined by

$$\psi(X, \pi) = \mathbf{E}_\pi[X] - \mathbf{H}_\pi[X]$$

where $\mathbf{H}_\pi[X]$ is the harmonic mean. That is,

$$\mathbf{H}_\pi[X] = \Big(\mathbf{E}_\pi[1/X]\Big)^{-1}$$

The following lemma makes MMP(c, π) easier to work with.

Lemma 1. *Let u be a Bernoulli utility function. Let $S \in \mathbb{N}$, $c \in \mathbb{R}^S_{++}$, $\pi \in \Delta_+(S)$, let b be the budget share measure in* (10)*, and let $\sigma = (\sigma(c_1), \dots, \sigma(c_S))$. Then*

$$MMP(c, \pi) = \psi(\sigma, b)$$

The proofs of all lemmas will appear in the appendix. We are now ready to prove Theorem 1.

Proof of Theorem 1. Let $X \in \mathbb{R}^S_{++}$. Let X^M and X^m be two numbers such that $X^M \geq X_s \geq X^m \geq 0$ for all s. It is shown in Shisha and Mond (1967) that for any $b \in \Delta_+(S)$

$$\mathbf{E}_b[X] - \mathbf{H}_b[X] \leq \Big(\sqrt{X^M} - \sqrt{X^m}\Big)^2 \tag{11}$$

Let σ^M and σ^m be defined by

$$\sigma^M = \sup_{a \in \mathbb{R}_{++}} \sigma(a) \qquad \text{and} \qquad \sigma^m = \inf_{a \in \mathbb{R}_{++}} \sigma(a)$$

Suppose that (7) holds and let $c \in \mathbb{R}^S_{++}$ and $\pi \in \Delta_+(S)$. Let b be the corresponding budget share measure as defined in (10) and let $\sigma = (\sigma(c_1), \ldots, \sigma(c_S))$. Applying Lemma 1 and (11) we see

$$\mathrm{MMP}(c, \pi) = \psi(\sigma, b) \leq (\sqrt{\sigma^M} - \sqrt{\sigma^m})^2 \leq 4$$

Applying Corollary 1 we see that the agent has ordinary demand.

Next, suppose there exists d, $a \in \mathbb{R}_{++}$ such that $\sqrt{\sigma(d)} - \sqrt{\sigma(a)} > 2$. Let $S = 2$ and define a $\pi \in \Delta_+(2)$ by

$$\pi_1 = \frac{\frac{\sqrt{\sigma(a)}}{u'(a)a}}{\frac{\sqrt{\sigma(a)}}{u'(a)a} + \frac{\sqrt{\sigma(d)}}{u'(d)d}} \qquad \text{and} \qquad \pi_2 = \frac{\frac{\sqrt{\sigma(d)}}{u'(d)d}}{\frac{\sqrt{\sigma(d)}}{u'(a)a} + \frac{\sqrt{\sigma(d)}}{u'(d)d}}$$

Let c be the contingent consumption bundle on (S, π) defined by $c(1) = a$ and $c(2) = d$. The corresponding budget share measure $b \in \Delta(S)$ is

$$b_1 = \frac{\sqrt{\sigma(a)}}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}} \qquad \text{and} \qquad b_2 \frac{\sqrt{\sigma(d)}}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}}$$

We have

$$\begin{aligned}
\psi(\sigma, b) &= b_1\sigma(a) + b_2\sigma(d) - \Big(b_1/\sigma(a) + b_2/\sigma(b)\Big)^{-1} \\
&= \frac{\sigma(a)^{3/2} + \sigma(d)^{3/2}}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}} - \left(\frac{\sqrt{\sigma(a)}}{\sigma(a)(\sqrt{\sigma(a)} + \sqrt{\sigma(d)})} + \frac{\sqrt{\sigma(d)}}{\sigma(d)(\sqrt{\sigma(a)} + \sqrt{\sigma(d)})}\right)^{-1} \\
&= \frac{\sigma(a)^{3/2} + \sigma(d)^{3/2}}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}} - (\sqrt{\sigma(a)} + \sqrt{\sigma(d)}) \left(\frac{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}}{\sqrt{\sigma(a)\sigma(b)}}\right)^{-1} \\
&= \frac{\sigma(a)^{3/2} + \sigma(d)^{3/2} - \sigma(a)\sqrt{\sigma(d)} - \sigma(d)\sqrt{\sigma(a)}}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}} \\
&= \frac{\sigma(d)(\sqrt{\sigma(d)} - \sqrt{\sigma(a)}) + \sigma(a)(\sqrt{\sigma(d)} - \sqrt{\sigma(a)})}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}} \\
&= \frac{\sigma(d) + \sigma(a)}{\sqrt{\sigma(a)} + \sqrt{\sigma(d)}}\Big(\sqrt{\sigma(d)} - \sqrt{\sigma(a)}\Big) = \Big(\sqrt{\sigma(d)} - \sqrt{\sigma(a)}\Big)^2 > 4
\end{aligned}$$

Using Lemma 1 we see $\mathrm{MMP}(c, \pi) = \psi(\sigma, b) > 4$. Applying Corollary 1 we see that the agent displays Giffen behavior. □

Theorem 1 generalizes a result in Quah (2003) on which we now comment. Mas-Colell (1991) and others noted that a straightforward application of the MMP Theorem to the expected utility case shows that demand is monotonic if the coefficient of relative risk aversion is everywhere less than or equal to 4. Quah showed that the key to monotonic demand in the EU context lay in the *variation* in the coefficient of relative risk aversion. He proved that demand is monotonic if

$$\sigma(d) - \sigma(a) \leq 4, \qquad \text{for any } a, d > 0.$$

Quah (2003) also established the connection between monotonic contingent consumption and monotonic asset demand in a way similar to Proposition 1. Our result is sharper than Quah's. Indeed if $\sigma(d) \geq \sigma(a)$ then

$$\begin{aligned} 4 \geq \sigma(d) - \sigma(a) &\implies 4 \geq \left(\sqrt{\sigma(d)} - \sqrt{\sigma(a)}\right)\left(\sqrt{\sigma(d)} + \sqrt{\sigma(a)}\right) \geq \left(\sqrt{\sigma(d)} - \sqrt{\sigma(a)}\right)^2 \\ &\implies 2 \geq \sqrt{\sigma(d)} - \sqrt{\sigma(a)} \end{aligned}$$

Perhaps more importantly Theorem 1 provides the first necessary and sufficient condition on the coefficient of relative risk aversion for demand to be monotonic.

4 Ambiguity

Thus far we have considered the case where the agent behaves as if each bundle of contingent consumption has attached to it a probability which the agent uses to evaluate the expected utility of the bundle. The famous Ellsberg urns thought experiment (see Ellsberg (1961)) show that this is not always a good assumption. The experiment demonstrates that people do not treat objectively given probabilities and subjectively derived probabilities in the same way. That is, people tend to prefer situations in which they know the true probabilities of events. We now consider the potential for Giffen behavior in several models which allow for aversion to ambiguity in the true distribution of a contingent consumption bundle.

4.1 Maxmin Expected Utility

In the EU model the notion of uncertainty is captured in a probability $\pi \in \Delta_+(S)$. In this section an uncertainty will be a set of probabilities. A Maxmin Expected Utility

(MEU) uncertainty $\mathcal{P}$ is a convex and compact set of probabilities. The interpretation is that the agent believes that any probability in $\mathcal{P}$ might describe the world.

We say that an agent has MEU preferences if there is some Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$ such that a contingent consumption bundle $c \in \mathbb{R}^S_{++}$ and MEU uncertainty $\mathcal{P}$ is awarded utility

$$U(c, \mathcal{P}) = \min_{\pi \in \mathcal{P}} \ \mathbf{E}_\pi[u(c)]$$

The interpretation is that because the agent does not like the ambiguity of not knowing the true probability, the agent evaluates a contingent consumption bundle c according to the most pessimistic possible probability.

An MEU financial environment is a pair $\mathcal{A} = (A, \mathcal{P})$ where $\mathcal{P}$ is an MEU uncertainty. The asset demand function for an MEU financial environment is defined according to (5). The following theorem characterizes Giffen behavior for an agent with MEU preferences.

Theorem 2. *An agent with MEU preferences and Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$ does not display Giffen behavior on any MEU financial environment if and only if u displays constant relative risk aversion.*

The proof of the theorem relies on a characterization of Bernoulli utility functions which display constant relative risk aversion. The following lemma captures this characterization.

Lemma 2. *Let $u : \mathbb{R}_{++} \to \mathbb{R}$ be a Bernoulli utility function and let $S \geq 3$. The function u displays constant relative risk aversion if and only if for all $c \in \mathbb{R}^S_{++}$ and $y \in \mathbb{R}^S$*

$$\sum_{s=1}^{S} y_s = 0 \quad \textit{and} \quad \sum_{s=1}^{S} y_s u(c_s) = 0 \quad \Longrightarrow \quad \sum_{s=1}^{S} y_s u'(c_s) c_s = 0 \tag{12}$$

We now prove Theorem 2.

Proof of Theorem 2. It is straightforward to show that MEU preferences over contingent consumption are strictly convex. This in turn guarantees that the demand function for the agent exists. It is also easy to show that if u displays constant relative risk aversion then preferences are homothetic regardless of $\mathcal{P}$. It is well known that when preferences are homothetic then demand satisfies monotonicity and so there can be no Giffen assets.

Next, suppose that u does not display constant relative risk aversion. Let $S \geq 3$. From Lemma 2 we see that there must be some $c \in \mathbb{R}^S_{++}$ and some $y \in \mathbb{R}^S$ where

$$\sum_{s=1}^{S} y_s = 0, \qquad \sum_{s=1}^{S} u(c_s) y_s = 0, \qquad \text{and} \qquad \sum_{s=1}^{S} u'(c_s) c_s y_s \neq 0 \tag{13}$$

Let $\tilde{\pi}$ and $\pi \in \Delta_+^{S-1}$ be any two probabilities such that $\tilde{\pi} - \pi$ is collinear with y. Let $V(c)$ be the $S \times S$ diagonal matrix with diagonal s equal to $u'(c_s)$. Equation (13) shows that $c'V(c)\tilde{\pi} \neq c'V(c)\pi$. This establishes that

$$V(c)(\tilde{\pi} - \pi) \quad \text{and} \quad V(c)\left(\frac{\tilde{\pi}}{c'V(c)\tilde{\pi}} - \frac{\pi}{c'V(c)\pi}\right) \quad \text{are not collinear.}$$

and neither of these vectors are 0. The separating hyperplane theorem shows that there must be a $z \in \mathbb{R}^S$ such that

$$z'V(c)(\tilde{\pi} - \pi) < 0 < z'V(c)\left(\frac{\tilde{\pi}}{c'V(c)\tilde{\pi}} - \frac{\pi}{c'V(c)\pi}\right) \tag{14}$$

Note that (13) shows that $\sum_{s=1}^S \pi_s u(c_s) = \sum_{s=1}^S \tilde{\pi}_s u(c_s)$. This allows us along with (14) to exploit the continuity of u to find some $\tilde{c}$ near c such that $\tilde{c} - c = kz$ for some $k > 0$ such that

$$\sum_{s=1}^S \tilde{\pi}_s u(\tilde{c}_s) < \sum_{s=1}^S \pi_s u(\tilde{c}_s) \quad \text{and} \quad z'\left(\frac{V(\tilde{c})\tilde{\pi}}{\tilde{c}'V(\tilde{c})\tilde{\pi}} - \frac{V(c)\pi}{c'V(c)\pi}\right) > 0 \tag{15}$$

Let $\mathcal{P} = [\pi, \tilde{\pi}]$. Define state prices p and $p' \in \mathbb{R}_{++}^S$ by

$$p = \frac{V(\tilde{c})\tilde{\pi}}{\tilde{c}'V(\tilde{c})\tilde{\pi}} \quad \text{and} \quad \tilde{p} = \frac{V(c)\pi}{c'V(c)\pi}$$

Let x denote the demand for contingent consumption. Utility maximization implies $x(p,1) = c$ and $x(\tilde{p},1) = \tilde{c}$. Now, applying (15) we see

$$(\tilde{p} - p) \cdot (x(\tilde{p},1) - x(p,1)) = (\tilde{p} - p) \cdot (\tilde{c} - c) = k(\tilde{p} - p) \cdot z > 0$$

So, we see that monotonicity fails for contingent consumption. Proposition 1 shows that there are giffen assets. $\square$

Note that Lemma 2 relies on there being at least three states. This cannot be relaxed and when there are merely two states then the agent will display Giffen behavior only if (7) fails.

4.2 Three Paradoxes

Theorem 2 allows us to discuss a connection between three behavioral paradoxes. These paradoxes are known as the Ellsberg, St. Petersburg, and Giffen paradoxes. The Giffen paradox, as we are now well-familiar, refers to the fact that utility maximization is consistent with the behavior of demanding more of a good as its price rises.

The St. Petersburg paradox refers to the fact that there are some gambles which have incredibly large, indeed infinite, expected payouts but which are close to worthless. The canonical version specifies a lottery which pays out 2 with probability 1/2, pays out 4 with probability 1/4, pays out 8 with probability 1/8, etc. It is clear that the expected payout of the lottery is $1 + 1 + 1 + \ldots = \infty$. Yet, it is extremely unlikely that the lottery pays out more than a couple of dollars. People will not want to pay much for such a gamble yet a risk neutral agent should pay out any arbitrarily large amount of money to partake in the lottery. This is often used to motivate the use of expected utility maximization with a concave utility function. However, it has been known for a long time (see Bassett (1987)) the St. Petersburg paradox is only really addressed when u is *bounded*. That is, if u is not bounded one may construct a modified version of the St. Petersburg lottery which will have infinite positive or negative payout. This point is discussed at length in Arrow (1971).

Finally, there is the Ellsberg paradox. Imagine there are two urns. Urn 1 has exactly 50 red balls and 50 black balls. Urn 2 also has 100 balls which are either red or black however the exact number of each color is unknown. Consider four bets. Bet 1: draw a ball from urn 1 and receive a dollar if the ball is red. Bet 2: draw a ball from urn 1 and receive a dollar if the ball is black. Bet 3: draw a ball from urn 2 and receive a dollar if the ball is red. Bet 4: draw a ball from urn 2 and receive a dollar if the ball is black. People are in general indifferent between bets 1 and 2 and are also indifferent between bets 3 and 4. However, people prefer bets 1 and 2 to 3 and 4. This sort of behavior is inconsistent with expected utility maximization regardless of what the subject believes is the true mixture of balls in urn 2. Models of ambiguity aversion are able to explain this type of preference.

Suppose we wish to model an agent who is impervious to these paradoxes. That is, suppose we wish the agent to have (i) ordinary demand, (ii) does not pay arbitrarily large amounts of money to play or avoid St. Petersburg type lotteries, (iii) will display ambiguity aversion in the Ellsberg experiments. If this is our goal then Theorem 2 shows that MEU is not the answer. Indeed, Theorem 2 shows that to satisfy (i) and (iii) the agent must have constant relative risk aversion. However, it is easy to see that a Bernoulli utility function which displays constant relative risk aversion is not bounded and so the agent will not satisfy property (ii).

The next subsection reveals that α-MEU also does not allow an agent to satisfy properties (i), (ii), and (iii).

4.3 α-MEU

The α-MEU model is a generalization of MEU preferences. An agent has α-MEU preferences if there is a Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$ and $\alpha \in [0,1]$ such that a contingent consumption bundle $c \in \mathbb{R}^S_{++}$ and MEU uncertainty $\mathcal{P}$ is awarded utility

$$U(c, \mathcal{P}) = \alpha \min_{\pi \in \mathcal{P}} \ \mathbf{E}_\pi[u(c)] + (1-\alpha) \max_{\pi \in \mathcal{P}} \ \mathbf{E}_\pi[u(c)]$$

Notice that when $\alpha = 1$ then α-MEU preferences coincide with MEU preferences. The coefficient α can be interpreted as an agent's attitude towards ambiguity (see Ghirardato, Maccheroni, and Marinacci (2004)). It is important to note that when $\alpha < 1$ then the function $U(\cdot, \mathcal{P})$ may not be strictly quasi-concave and thus the demand for contingent consumption is in general a correspondence. For this reason the definition of Giffen behavior must be amended slightly. We say that an agent with α-MEU preferences and demand correspondence y displays Giffen behavior if there is an MEU financial environment $\mathcal{A} = (A, \mathcal{P})$, an asset index k, prices q, $\tilde{q} \in \mathcal{Q}_A$ where $q_j = \tilde{q}_j$ for all $j \neq k$, wealth level $w \in \mathbb{R}_{++}$, and demands $\theta \in y(q, w, \mathcal{A})$, $\tilde{\theta} \in y(\tilde{q}, w, \mathcal{A})$ such that

$$(\tilde{q}_k - q_k)(\tilde{\theta}_k - \theta_k) > 0$$

Clearly this new definition of Giffen behavior coincides with the old one when demand is a function.

We show that the conclusion of Theorem 2 is unchanged under this generalization when the agent has ($\alpha > 0.5$).

Theorem 3. *An agent with α-MEU preferences with $\alpha \in (0.5, 1]$ and Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$ does not display Giffen behavior on any MEU financial environment if and only if u displays constant relative risk aversion.*

The proof of the theorem proceeds in a fashion similar to Theorem 2. The main trick is to show that when $\alpha > 0.5$ and $\mathcal{P} = [\pi, \tilde{\pi}]$ for some π, $\tilde{\pi} \in \Delta(S)$ then the agent acts as if she has MEU preferences with a different set of probabilities $\mathcal{P}$.

Lemma 3. *Let u be a Bernoulli utility function and let $\alpha \in (0.5, 1]$. Let $S \in \mathbb{N}$, π, $\tilde{\pi} \in \Delta(S)$, and $\mathcal{P} = [\pi, \tilde{\pi}]$. Let*

$$\beta = \alpha\pi + (1-\alpha)\tilde{\pi} \qquad \textit{and} \qquad \tilde{\beta} = \alpha\tilde{\pi} + (1-\alpha)\pi \tag{16}$$

Let $\bar{\mathcal{P}} = [\beta, \tilde{\beta}]$. Then

$$\min_{\bar{\beta} \in \bar{\mathcal{P}}} \mathbf{E}_{\bar{\beta}}[u(c)] = \alpha \min_{\bar{\pi} \in \mathcal{P}} \mathbf{E}_{\bar{\pi}}[u(c)] + (1-\alpha) \max_{\bar{\pi} \in \bar{\mathcal{P}}} \mathbf{E}_{\bar{\pi}}[u(c)]$$

Importantly this lemma shows us a condition which makes α-MEU utility functions strictly concave.

Corollary 2. *If $\alpha \in (0.5, 1]$, U is an α-MEU utility function, and $\mathcal{P} = [\pi, \tilde{\pi}]$ for some π, $\tilde{\pi} \in \Delta(S)$ then $U(\cdot, \mathcal{P})$ is strictly concave.*

Proof. This follows from the fact that MEU preferences are strictly concave. $\square$

The proof of Theorem 3 now mimics the proof of Theorem 2 with a few minor tweaks. So, we shall leave it to the appendix.

4.4 Choquet Expected Utility

Another popular approach to modeling ambiguity aversion is now discussed. In this model an uncertainty is a non-additive probability measure P. A non-additive probability P is a map from 2^S to $\mathbb{R}$ (where 2^S denotes the power set of S) which satisfies (i) $P(\varnothing) = 0$, (ii) $P(\{1, \ldots, S\}) = 1$, and (iii) A, $B \in 2^S$ where $A \subseteq B$ implies $P(B) \geq P(A)$. We say that non-additive probability P is convex if $P(A \cup B) + P(A \cap B) \geq P(A) + P(B)$ for all A, $B \in 2^S$. A Choquet Expected Utility (CEU) uncertainty is a convex non-additive probability P.

Choquet Expected Utility preferences rely on the notion of Choquet integration which we now define. Let $X : \mathbb{R}^S \to \mathbb{R}$ and let P be a non-additive probability. Suppose X takes values $x_1 > x_2 > \ldots > x_n$. Also, let $x_0 = -\infty$ and $\pi_i = P(\{s \in S : X(s) \leq x_i\}) - P(\{s \in S : X(s) \leq x_{i+1}\})$ for each i. The *Choquet integral* is defined as

$$\mathbf{CE}_P[X] = \int X dP = \sum_{i=1}^{n} \pi_i x_i$$

We say that an agent has CEU preferences if there is some Bernoulli utility function u such that a contingent consumption bundle $c \in \mathbb{R}^S_{++}$ and CEU uncertainty P is given a utility

$$U(c, P) = \mathbf{CE}_P[u(c)] = \int u(c) dP$$

where the integral is the Choquet integral. CEU preferences are a special case of MEU preferences. Let $\operatorname{core} P = \{\pi \in \Delta(S) : \pi(A) \geq P(A) \text{ for all } A \in 2^S\}$. So, $\operatorname{core} P$ is the set of all probabilities which weigh events higher than P. Schmeidler (1989) shows that

$$\mathbf{CE}[u(c)] = \min_{\pi \in \operatorname{core}(P)} \mathbf{E}_\pi[u(c)] \tag{17}$$

So, we see that CEU preferences can be expressed as MEU preferences where the set of probabilities $\mathcal{P}$ is confined to be a set of probabilities which majorize a convex non-additive probability. As the next theorem shows this additional structure placed on the set of probabilities allows far less restrictive assumptions on u to ensure ordinary behavior.

A *CEU financial environment* is a pair $\mathcal{A} = (A, P)$ where A is an asset matrix and P is a CEU uncertainty.

Theorem 4. *Suppose an agent has CEU preferences with Bernoulli utility function $u : \mathbb{R}_{++} \to \mathbb{R}$. The agent does not display Giffen behavior for any CEU financial environment if and only if* (7) *holds.*

To prove the theorem we introduce a few concepts. For X, $Y \in \mathbb{R}^S$ let $[X, Y] = \{\tilde{X} \in \mathbb{R}^S : \exists \alpha \in [0, 1] \text{ s.t. } \tilde{X} = (1-t)X + tY\}$. We refer to $[X, Y]$ as an interval. We say an interval $[X, Y]$ is *co-monotonic* if for all $s, s' \in S$ and $\tilde{X}, \tilde{Y} \in [X, Y]$ we never have $\tilde{X}_{s'} > \tilde{X}_s$ and $\tilde{Y}'_{s'} < \tilde{Y}'_s$.

Lemma 4. *Let $c, \tilde{c} \in \mathbb{R}^S_{++}$. There exists $c^1, \ldots, c^N \in [c, \tilde{c}]$ such that*

1. *$c = c^1$ and $\tilde{c} = c^N$.*

2. *For each integer n such that $N \geq n > 1$ the interval $[c^{n-1}, c^n]$ is co-monotonic.*

Proof. Let s, $s' \leq S$. Clearly, if $c_s < \tilde{c}_s$ and $\tilde{c}_{s'} > c_{s'}$ then there is one and only one $\bar{c} \in [c, \tilde{c}]$ such that $\bar{c}_s = \bar{c}_{s'}$. Collecting all such $\bar{c}$ for each pair s, $s' \leq S$ we are able to assemble the desired $c^1, \ldots, c^N$. □

To prove Theorem 4 we shall have to pay close attention to those points in contingent consumption space where indifference curves are kinked. For each $c \in \mathbb{R}^S_{++}$ let

$$\mathcal{P}(c) = \{\pi \in \text{core}(P) : U(c, \pi) = U(c, \mathcal{P})\}$$

where $U(c, \pi) = \mathbf{E}_\pi[u(c)]$. From (17) it is clear that $\mathcal{P}(c)$ is non-empty. For any two contingent consumption bundles c, $\tilde{c} \in \mathbb{R}^S_{++}$ we let

$$\mathcal{P}([c, \tilde{c}]) = \bigcap_{\bar{c} \in [c, \tilde{c}]} \mathcal{P}(\bar{c})$$

Lemma 5. *Suppose $[c, \tilde{c}]$ is co-monotonic. Then $\mathcal{P}([c, \tilde{c}]) \neq \varnothing$.*

Proof of Lemma 5. It is well-known that CEU preferences will satisfy the independence axiom over co-monotonic acts which implies the result. □

For the proof of Theorem 4 we shall make use of the indirect demand correspondence. That is, we shall use the map which takes a contingent consumption bundle and return prices at which that bundle is demanded. Let $\rho : \mathbb{R}^S_{++} \to \mathbb{R}^S_{++}$ be the correspondence so that $\rho(c)$ is the set of prices p such that $c = x(p, 1)$. Let $\rho : \mathbb{R}^S_{++} \times \Delta^{S-1}_+$ be defined so that $\rho(\bar{c}, \pi)$ returns the price $\bar{p}$ so that $\bar{c} = x(\bar{p}, 1; \pi)$ where $x(\cdot, 1; \pi)$ is the demand function for the expected utility function $U(\cdot; \pi)$. We now prove Theorem 4.

Proof of Theorem 4. The necessary part of the statement follows as EU is a special case of CEU. Next, let $\sigma^M = \sup \sigma(a)$ and $\sigma^m = \inf \sigma(a)$. Suppose that $\sqrt{\sigma^M} - \sqrt{\sigma^m} \leq 2$. We shall show that monotonicity will hold for contingent consumption. From there Proposition 1 establishes the result. Let P be some CEU uncertainty. It is easy to show that CEU utility functions are strictly concave and so there is a demand function $x : \mathbb{R}^S_{++} \times \mathbb{R}_{++} \to \mathbb{R}^S_{++}$ for $U(\cdot; P)$. Let $\tilde{p}, p \in \mathbb{R}^S_{++}$ be any two prices and let $\tilde{c} = x(\tilde{p}, 1)$ and $c = x(p, 1)$. Let $c^1, \ldots, c^N$ be as in Lemma 4. To each interval $[c^{n-1}, c^n]$ associate a probability $\pi^n \in \mathcal{P}([c^{n-1}, c^n])$ (Lemma 5 shows that these sets are non-empty). So, we have selected probabilities $\pi^2, \ldots, \pi^N$.

Let π^1 and π^{N+1} be probabilities such that $p = \rho(c, \pi^1)$ and $\tilde{p} = \rho(\tilde{c}, \pi^{N+1})$. Let $z = \tilde{c} - c$ and put $\rho^{n,m} = \rho(c^n, \pi^m)$. Now,

$$\begin{aligned}(\tilde{p} - p) \cdot (\tilde{c} - c) &= (\rho^{N,N+1} - \rho^{1,1}) \cdot z \\ &= \left[\sum_{n=1}^{N} (\rho^{n,n+1} - \rho^{n,n}) + \sum_{n=2}^{N-1} (\rho^{n+1,n+1} - \rho^{n,n+1})\right] \cdot z\end{aligned}$$

We shall show that $(\rho^{n,n+1} - \rho^{n,n}) \cdot z \leq 0$ and $(\rho^{n+1,n+1} - \rho^{n,n+1}) \cdot z \leq 0$. Notice that for each n

$$(\rho^{n+1,n+1} - \rho^{n,n+1}) \cdot z = \big(\rho(c^{n+1}, \pi^{n+1}) - \rho(c^n, \pi^{n+1})\big) \cdot z \leq 0$$

where the last inequality follows from the fact that $\rho(\cdot; \pi^{n+1})$ is the indirect demand function for an expected utility function which satisfies $\sqrt{\sigma^M} - \sqrt{\sigma^m} \leq 2$ and thus monotonicity must hold. Next, let n be some integer between 1 and N and define equivalence classes of integers so that for each k we have $[k] = \{s : c^N_s = c^N_k\}$. Then, clearly for each k we have $\sum_{s \in [k]} \pi^n_k = \sum_{s \in [k]} \pi^{n+1}_k$. This shows that

$$\nabla_c U(c^n; \pi^{n+1}) = \nabla_c U(c^n; \pi^n) \quad \text{and} \quad \nabla_z U(c^n; \pi^{n+1}) \leq \nabla_z U(c^n; \pi^n) \tag{18}$$

where the second relation follows from (17). The usual first order conditions for utility maximization shows that

$$\rho^{n,m} \cdot z = \frac{\nabla_z U(c^n, \pi^m)}{\nabla_c U(c^n, \pi^m)}$$

and so using 18 we see

$$\begin{aligned}(\rho^{n,n+1} - \rho^{n,n}) \cdot z &= \frac{\nabla_z U(c^n, \pi^{n+1})}{\nabla_c U(c^n, \pi^{n+1})} - \frac{\nabla_z U(c^n, \pi^n)}{\nabla_c U(c^n, \pi^n)} \\ &= \frac{\nabla_z U(c^n, \pi^{n+1}) - \nabla_z U(c^n, \pi^n)}{\nabla_c U(c^n, \pi^n)} \leq 0\end{aligned}$$

which completes the proof. □

4.5 Smooth Ambiguity Aversion

In the model of Smooth Ambiguity Aversion (SAA) the sample space is equipped with a set of probability measures $\mathcal{P}$ just as in the MEU case. However, instead of evaluating a contingent consumption bundle c according to the worst measure in $\mathcal{P}$ the agent has a probability η on $\mathcal{P}$ and an additional Bernoulli utility function v used to evaluate utility across different ambiguous situations.

An SAA uncertainty is a pair $(\mathcal{P}, \eta)$ where $\mathcal{P}$ is a finite set of probability in $\Delta_+(S)$ and η is a probability on $\mathcal{P}$. The interpretation is that elements of $\mathcal{P}$ represent different probabilities which the agent believes are possible. For $\pi \in \mathcal{P}$ the value $\eta(\pi)$ represents how strongly the agent feels that π is the the probability.

We say an agent has SAA preferences if there are two Bernoulli utility function u and v where v is more concave than u (meaning there is some increasing concave function f mapping the range of u to $\mathbb{R}$ such that $v = f \circ u$) and function $\phi = v \circ u^{-1}$ such that a contingent consumption bundle $c \in \mathbb{R}^S_{++}$ and SAA uncertainty $\Lambda = (\mathcal{P}, \eta)$ is given utility

$$U(c, \Lambda) = \mathbf{E}_\eta \Big[\phi\Big(\mathbf{E}_\pi \big[u(c)\big]\Big)\Big] = \int_{\mathcal{P}} \phi\Big(\int_S u(c) d\pi\Big) d\eta$$

We refer to u as the risk Bernoulli utility function and to v as the ambiguity Bernoulli utility function. We are interested in introducing a measure of the curvature of v similar to the role which the coefficient of relative risk aversion σ plays for u. To this end, we define the *net coefficient of relative ambiguity aversion* $\gamma : \mathbb{R}_{++} \to \mathbb{R}$ by

$$\gamma(a) = -\Big(\frac{v''(a)}{v'(a)} - \frac{u''(a)}{u'(a)}\Big)a \tag{19}$$

The form in equation (4.5) turns out to be more difficult to work with than the previous models considered. For this reason we are only able to develop a sufficient condition for monotonicity to hold under specific assumptions.

An SAA financial environment is a pair (A, Λ) where A is an asset matrix and Λ is an SAA uncertainty.

Theorem 5. *Suppose an agent has SAA preferences with risk Bernoulli utility function u and ambiguity Bernoulli utility function v. The agent does not display Giffen behavior on any SAA financial environment if u displays increasing relative risk aversion and*

$$\gamma(d) + \sigma(a) \leq 4, \qquad \textit{for all } d,\ a \in \mathbb{R}_{++} \tag{20}$$

The proof of the theorem relies on the following two lemmas.

Lemma 6. *Let $u : \mathbb{R}_{++} \to \mathbb{R}$ be a Bernoulli utility function which displays increasing relative risk aversion. Let c be a contingent consumption bundle on some probability space (S, π). Let $\bar{c} \in \mathbb{R}_{++}$ be defined by $\bar{c} = u^{-1}(\mathbf{E}[u(c)])$. Then*

$$\mathbf{E}\Big[u'(c)c\Big] \leq u'\big(\bar{c}\big)\bar{c}$$

Lemma 7. *Let $u : \mathbb{R}_{++} \to \mathbb{R}$ and $v : \mathbb{R}_{++} \to \mathbb{R}$ be Bernoulli utility functions where v is more concave than u. Let $\phi = v \circ u^{-1}$ and let γ be defined by 19. Then for all $a \in \mathbb{R}_{++}$*

$$\gamma(a) = -\frac{\phi''(u(a))}{\phi'(u(a))}u'(a)a$$

Proof of Theorem 5. Let $\sigma^M = \sup \sigma(a)$ and $\gamma^M = \sup \gamma(a)$. Let $c \in \mathbb{R}^S_{++}$ be a contingent consumption bundle and $\Lambda = (\mathcal{P}, \eta)$ be an SAA uncertainty. Suppose 20 holds. We shall show that

$$-\frac{c'\partial^2 U(c, \Lambda)c}{\partial U(c, \Lambda) \cdot c} \leq 4 \tag{21}$$

As $\partial^2 U(c, \Lambda)$ is negative semi-definite we see that if (21) holds then so does (9) and thus demand is monotonic. From there Corollary 1 establishes the result.

Suppose $\mathcal{P}$ has R elements. Let $\mathcal{P} = \{\pi_1, \ldots, \pi_R\}$. For $s \in S$ and $r \in R$ let $\pi_{r,s}$ be element s of π_r. For each $r \in R$ let $\bar{u}_r = \sum_{s=1}^S \pi_{r,s} u(c_s)$. By differentiating U in (4.5) twice it is straightforward to show that

$$\partial U(c, \Lambda) \cdot c = \sum_{s=1}^{S} \sum_{r=1}^{R} \eta_r \phi'(\bar{u}_r)\pi_{r,s}u'(c_s)c_s$$

$$c'\partial^2 U(c, \Lambda)c = \sum_{k=1}^{S} \sum_{s=1}^{S} \sum_{r=1}^{R} \eta_r \phi''(\bar{u}_r)\pi_{r,s}u'(c_s)c_s\pi_{r,k}u'(c_k)c_k + \sum_{s=1}^{S} \sum_{r=1}^{R} \eta_r \phi'(\bar{u}_r)\pi_{r,s}u''(c_s)c_s^2$$

For $t \in R$ and $k \in S$ let $\alpha_{t,k}$ be the value

$$\alpha_{t,k} = \frac{\eta_t \phi'(\bar{u}_t)\pi_{t,k} u'(c_k)c_k}{\sum_{r=1}^{R}\sum_{s=1}^{S} \eta_r \phi'(\bar{u}_r)\pi_{r,s} u'(c_s)c_s}$$

Clearly, $\sum_{r=1}^{R}\sum_{s=1}^{S} \alpha_{r,s} = 1$. For each $r \in R$ let $\bar{c}_r = u^{-1}(\bar{u}_r)$ and $\tau_r = -(\phi''(\bar{u}_r)/\phi'(\bar{u}_r))$ and for each $s \in S$ let $\sigma_s = \sigma(c_s)$. Now,

$$\begin{aligned}
-\frac{c'\partial^2 U(c,\Lambda)c}{\partial U(c,\Lambda)\cdot c} &= \sum_{k=1}^{S}\sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\tau_r \pi_{r,k} u'(c_k)c_k + \sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\sigma_s \\
&\leq \sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\tau_r u'(\bar{c}_r)\bar{c}_r + \sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\sigma_s && \text{by Lemma 6} \\
&= \sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\gamma(\bar{c}_r) + \sum_{s=1}^{S}\sum_{r=1}^{R} \alpha_{r,s}\sigma_s && \text{by Lemma 7} \\
&\leq \gamma^M + \sigma^M \leq 4
\end{aligned}$$

So, by our previous remarks the agent will not display Giffen behavior. □

5 Which Asset is Giffen?

We have provided many conditions showing when an agent will display Giffen behavior for *some* financial environment. However, we have not said anything about which assets can become Giffen assets (an asset whose price rises and quantity demanded rises). The next proposition says that once one asset can become a Giffen asset then any asset can become a Giffen asset. Further, only 3 assets are required to attain this result.

Proposition 2. *Let $S \in \mathbb{N}$ and $U : \mathbb{R}^S_{++} \to \mathbb{R}$ be a continuous, strictly concave, and increasing utility function. Let $X \in \mathbb{R}^S$ be any asset such that $X \neq 0$ and let $\bar{A}$ be an asset matrix. If demand for assets in $\bar{A}$ displays Giffen behavior then there is an asset matrix A where X is Giffen.*

Proof. If demand over $\bar{A}$ displays Giffen behavior then Proposition 1 shows that demand for contingent consumption $x : \mathbb{R}^S_{++} \times \mathbb{R}_{++} \to \mathbb{R}^S_{++}$ is not monotonic. So, there are prices $\tilde{p}$ and $p \in \mathbb{R}^S_{++}$ and wealth level $w \in \mathbb{R}_{++}$ so that

$$(\tilde{p} - p)\cdot(\tilde{c} - c) > 0 \tag{22}$$

where $\tilde{c} = x(\tilde{p}, w)$ and $c = x(p, w)$. Without loss of generality we may assume that $(\tilde{p} - p)\cdot X \neq 0$. We may make this assumption because if it does not hold then we may perturb p slightly to attain the desired relation and (22) will still hold as x is continuous. We now construct an asset matrix A where X is a giffen good. We define two assets X_1 and X_2. Let

$$X_1 = c - \frac{(\tilde{p} - p) \cdot c}{(\tilde{p} - p) \cdot X} X \qquad \text{and} \qquad X_2 = \tilde{c} - \frac{(\tilde{p} - p) \cdot \tilde{c}}{(\tilde{p} - p) \cdot X} X$$

Let $A = [X \quad X_1 \quad X_2]$. Let q, $\tilde{q} \in \mathbb{R}^3$ be defined by $q = A'p$ and $\tilde{q} = A'\tilde{p}$. Let θ and $\tilde{\theta} \in \mathbb{R}^3$ be the quantities demanded of the assets when prices are q and $\tilde{q}$ and wealth is w. It is straightforward to see that the span of the assets in A include c and $\tilde{c}$. From how we chose prices q and $\tilde{q}$ we see that $c = A\theta$ and $\tilde{c} = A\tilde{\theta}$. From (22) we see

$$(\tilde{q} - q) \cdot (\tilde{\theta} - \theta) = (\tilde{p} - p) \cdot A(\tilde{\theta} - \theta) = (\tilde{p} - p) \cdot (\tilde{c} - c) > 0 \tag{23}$$

It is straightforward to confirm that $\tilde{q}_2 = q_2$ and $\tilde{q}_3 = q_3$ (just note that $(\tilde{p}-p)\cdot X_1 = 0$ and $(\tilde{p}-p)\cdot X_2 = 0$). This fact along with (23) shows that $(\tilde{q}-q)\cdot(\tilde{\theta}-\theta) = (\tilde{q}_1-q_1)(\tilde{\theta}_1-\theta_1) > 0$ and so we see that X is a Giffen asset. $\square$

A corollary of this proposition is that if there is an asset matrix $\bar{A}$ where Giffen behavior is displayed then there is an asset matrix A in which the risk-free asset is Giffen. It should be remarked that there is no requirement in this proposition that the risk-free asset be held long. Kubler, Selden, and Wei (2013), in the expected utility context, show that if relative risk aversion is increasing then the risk free asset can only be an inferior good when it is held short (the opposite holds when risk aversion is decreasing). As inferiority is a necessary condition for Giffen behavior we see that in the context considered in Kubler et al. (2013) the risk-free asset will not be Giffen when it is held long and relative risk aversion is increasing. It should be noted that Kubler et al. work in a framework with 2 assets while proposition 2 requires at least 3 to work.

Appendix

To prove Proposition 1 we shall use the following lemma.

Lemma 8. *Let U, x, and y be as in Proposition 1. Suppose A is an $S \times S$ full rank asset matrix and let $q \in \mathcal{Q}_A$, $w \in \mathbb{R}_{++}$, and $p \in \mathbb{R}^S_{++}$ satisfy $q = A'p$. Then*

$$x(p, w) = Ay(q, w, A)$$

Proof of Lemma 8. To simplify notation let $c = x(p, w)$ and $\theta = y(q, w, A)$. The fact that $q = A'p$ means that $q \cdot \tilde{\theta} = w$ implies $p'A\tilde{\theta} = w$. Thus, $U(c) \geq U(A\theta)$. But, $q = A'p$

also means that $p \cdot \tilde{c} = w$ implies $q' A^{-1} c = w$ and so $U(A\theta) \geq U(c)$. So, $U(A\theta) = U(c)$. The fact that U is strictly concave shows that $c = A\theta$. $\square$

Proof of Proposition 1. Suppose x does not satisfy monotonicity. Then there exists $w \in \mathbb{R}_{++}$ and price pairs p, $\tilde{p} \in \mathbb{R}^S_{++}$ such that

$$(\tilde{p} - p) \cdot (x(\tilde{p}, w) - x(p, w)) > 0$$

Clearly there must exist $S-1$ assets $a_1, \ldots, a_{S-1}$ which are linearly independent and satisfy $(\tilde{p} - p) \cdot a_s = 0$ for all $s \in (S-1)$. Let a_S be an asset which is orthogonal to $a_1, \ldots a_{S-1}$ (this also means that $(\tilde{p} - p) \cdot a_S \neq 0$). Let $A = [a_1, \ldots, a_S]$. Define $q' = p'A$ and $\tilde{q}' = \tilde{p}'A$. As A is full rank we may use Lemma 8 to show

$$\begin{aligned}
0 &< (\tilde{p} - p) \cdot (x(\tilde{p}, w) - x(p, w)) \\
&= (\tilde{p} - p)' A (y(\tilde{q}, w; A) - y(q, w; A)) \\
&= (\tilde{q} - q) \cdot (y(\tilde{q}, w; A) - y(q, w; A))
\end{aligned}$$

From how A was defined we see that for all $s \neq S$ we have $\tilde{q}_s = q_s$. This shows that the agent displays Giffen behavior.

Next, suppose there is an asset matrix $A = [a_1, \ldots, a_J]$ such that $y(\cdot, \cdot; A)$ displays Giffen behavior. This means there is some $w \in \mathbb{R}_{++}$ and price pair $\tilde{q}$, $q \in \mathcal{Q}_A$ such that

$$(\tilde{q} - q) \cdot (y(\tilde{q}, w; A) - y(q, w; A) > 0$$

where there exists a $k \in S$ such that for all $s \neq k$ we have $q_s = \tilde{q}_k$. Construct an asset matrix $B = [\tilde{a}_1, \ldots, \tilde{a}_S]$ which has full rank and whose first J assets are the same as the assets in A. Let q^n and $\tilde{q}^n$ be price sequences in $\mathcal{Q}_B$ which satisfy $\tilde{q}_s = q_s$ for $s > J$ and

$$q^n \to (q_1, \ldots, q_J, \infty, \ldots, \infty), \qquad \text{and} \qquad \tilde{q}^n \to (\tilde{q}_1, \ldots, \tilde{q}_J, \infty, \ldots, \infty)$$

That is the sequence q^n converges to the price q on its first J elements and all other elements diverge to infinity. It is easy to show that the continuity of U implies that

$$By(q^n, w; B) \to Ay(q, w; A), \qquad \text{and} \qquad By(\tilde{q}^n, w; B) \to Ay(\tilde{q}, w; A) \tag{24}$$

As B is full rank we know B^{-1} exists. Let $p^n = B^{-1} q^n$ and $\tilde{p}^n = B^{-1} \tilde{q}^n$. Using Lemma

8 we see

$$\begin{aligned}(\tilde{q}^n - q^n)\cdot(y(\tilde{q}^n, w; B) - y(q^n, w; B)) &= (\tilde{p}^n - p^n)' B(y(\tilde{q}^n, w; B) - y(q^n, w; B)) \\ &= (\tilde{p}^n - p^n)\cdot(x(\tilde{p}^n, w) - x(p^n, w))\end{aligned}$$

From (24) the first expression above tends to $(\tilde{q} - q)\cdot(y(\tilde{q}, w; A) - y(q, w; A) > 0$. Thus, x does not satisfy monotonicity. □

Proof of Lemma 1. This proof consists of nothing but plugging in definitions.

$$\begin{aligned}\mathrm{MMP}(c, \pi) &= \frac{\mathbf{E}_\pi[u'(c)c]}{\mathbf{E}_\pi[u'(c)^2/u''(c)]} - \frac{\mathbf{E}_\pi[u''(c)c^2]}{\mathbf{E}_\pi[u'(c)c]} = -\frac{\mathbf{E}_\pi[u'(c)c]}{\mathbf{E}_\pi[u'(c)c/\sigma]} + \frac{\mathbf{E}_\pi[\sigma u'(c)c]}{\mathbf{E}_\pi[u'(c)c]} \\ &= \mathbf{E}_b[\sigma] - 1/\mathbf{E}_b[1/\sigma] = \psi(\sigma, b)\end{aligned}$$ □

Proof of Lemma 2. Let $\mu : \mathbb{R}^S_{++} \to \mathbb{R}$ be defined by $\mu(c) = (u(c_1), \ldots, u(c_S))$ and let $\nu(c) = (u'(c_1)c_1, \ldots, u'(c_S)c_S)$. Let $\iota \in \mathbb{R}^S$ be the vector of all ones and let I be the $S \times S$ identity matrix. Suppose that (12) holds for all $c \in \mathbb{R}^S_{++}$ and all $y \in \mathbb{R}^S$. It is straightforward to see that for all $c \in \mathbb{R}^S_{++}$ and $y \in \mathbb{R}^S$

$$y'\left(I - \tfrac{\iota\iota'}{S}\right)\mu(c) = 0 \quad\implies\quad y'\left(I - \tfrac{\iota\iota'}{S}\right)\nu(c) = 0$$

This clearly means that either $\left(I - \frac{\iota\iota'}{S}\right)\mu(c)$ and $\left(I - \frac{\iota\iota'}{S}\right)\nu(c)$ are collinear or $\left(I - \frac{\iota\iota'}{S}\right)\nu(c) = 0$. In either case there must be some $k(c) \in \mathbb{R}$ such that $\left(I - \frac{\iota\iota'}{S}\right)\nu(c) = k(c)\left(I - \frac{\iota\iota'}{S}\right)\mu(c)$. Rearranging terms we see that there must be some $\beta(c) \in \mathbb{R}$ such that

$$\nu(c) = k(c)\mu(c) + \beta(c)\iota \tag{25}$$

Subtracting element 1 from element 2 in equation (25) we see that $u'(c_2)c_2 - u'(c_1)c_1 = k(c)(u(c_2) - u(c_1))$. This shows that if c and $\tilde{c}'$ have the same values for elements 1 and 2 then $k(c) = k(\tilde{c})$. That is $k(c)$ will not change when elements 3 and higher of c are changed. But, the same arguments show that the same property holds for elements 1 and 2 and so the function k is equal to a constant and so we write $k(c) = k \in \mathbb{R}$. Again referring to (25) we see $u'(c_1)c_1 = ku(c_1) + \beta(c)$ which shows that $\beta(c)$ cannot depend on the values that c takes on elements 2 and higher. The same property holds for element 1 and so β is a constant. Thus for all $a \in \mathbb{R}_{++}$ we have $u'(a)a = ku(a) + \beta$. Differentiating we see that $u''(a)a + u'(a) = ku'(a)$. After rearranging we see that u displays constant relative risk aversion of $1 - k$.

Now, we tackle the "only if" part of the proof. It is well know that if u displays constant relative risk aversion then either u is a positive affine transformation of the function f defined by $f(a) = a^{1-\delta}/(1-\delta)$ for some $\delta > 0$ or a positive affine transformation of the log function. Suppose the first case holds. Then we may write u as $u(a) = ka^{1-\delta}/(1-\delta) + \beta$ for some $k \in$. Notice $u'(a)a = ka^{1-\delta}$. So, for any $c \in \mathbb{R}^S_{++}$ and $y \in \mathbb{R}^S$ such that $\sum_{s=1}^S y_s = 0$ and $\sum_{s=1}^S u(c_s)y_s = 0$ we have

$$0 = \sum_{s=1}^S u(c_s)y_s = \sum_{s=1}^S \left(k\frac{c_s^{1-\delta}}{1-\delta} + \beta\right) y_s = \frac{k}{1-\delta}\sum_{s=1}^S c_s^{1-\delta} y_s = \frac{1}{1-\delta}\sum_{s=1}^S u'(c_s)c_s y_s$$

So, we see that (12) holds. Next, suppose that u is an affine transformation of the log function. It is easy to see that if $\sum_{s=1}^S y_s = 0$ then $\sum_{s=1}^S y_s u'(c_s)c_s = 0$ and so (12) holds. □

Proof of Lemma 3. Let $c \in \mathbb{R}^S_{++}$. Without loss of generality suppose that $\mathbf{E}_{\tilde{\beta}}[u(c)] \geq \mathbf{E}_{\beta}[u(c)]$. Some rearrangement yields $\mathbf{E}_{\tilde{\pi}}[u(c)] \geq \mathbf{E}_{\pi}[u(c)]$. So,

$$\begin{aligned} \alpha \min_{\bar{\pi}\in\mathcal{P}} \mathbf{E}_{\bar{\pi}}[u(c)] + (1-\alpha)\max_{\bar{\pi}\in\mathcal{P}} \mathbf{E}_{\bar{\pi}}[u(c)] &= \alpha\mathbf{E}_{\pi}[u(c)] + (1-\alpha)\mathbf{E}_{\tilde{\pi}}[u(c)] \\ &= \mathbf{E}_{\beta}[u(c)] = \min_{\bar{\beta}\in[\beta,\tilde{\beta}]} \mathbf{E}_{\bar{\beta}}[u(c)] \end{aligned}$$ □

The lemma says that when the MEU uncertainty is an interval of probabilities and $\alpha \in (0.5, 1]$ then α-MEU preferences agree with MEU preferences for a different interval of probabilities.

Proof of Theorem 3. Generalize the definition of monotonic demand (8) for demand correspondences in the same way that the definition of Giffen behavior was generalized. Proposition 1 goes through in exactly the same way for correspondences.

If u displays constant relative risk aversion then preferences are homothetic. It is well known that when preferences are homothetic then demand is monotonic. Appealing to Proposition 1 we see there can be no Giffen behavior.

Suppose u does not display constant relative risk aversion and let $S \geq 3$. Use Lemma 2 to find $c \in \mathbb{R}^S_{++}$ and $y \in \mathbb{R}^S$ which satisfy (13). Let $\tilde{\pi}$, $\pi \in \Delta(S)$ be any two probabilities such that $\tilde{\pi} - \pi$ is collinear with y. Define β, $\tilde{\beta} \in \Delta(S)$ by (16). Let $V(c)$ be the $S \times S$ diagonal matrix with diagonal s equal to $u'(c_s)$. Equation (13) shows $c'V(c)\tilde{|pi} \neq c'V(c)\pi$. Some rearranging shows that $c'V(c)\tilde{\beta} \neq c'V(c)\beta$. Using Lemma 3 and Corollary 2 we may follow the steps in the proof of Theorem 2 only using probabilities β, $\tilde{\beta}$ instead of π, $\tilde{\pi}$ to complete the proof. □

Proof of lemma 6. Let $\mathcal{U}$ be the range of u. Let $g = u^{-1}$ and define $h : \mathcal{U} \to \mathbb{R}_{++}$ by

$h(b) = u'(g(b))g(b)$. Differentiating h and using the fact that $u' = (g')^{-1}$ we see

$$h'(b) = \frac{u''(b)}{u'(b)} g(b) + 1 = \sigma(g(b)) + 1$$

As g is a strictly increasing function we see that h is concave as we have assumed that σ is increasing. Applying Jensen's inequality we see

$$\mathbf{E}\Big[u'(c)c\Big] = \mathbf{E}\Big[h(u(c))\Big] \leq h\Big(\mathbf{E}[u(c)]\Big) = u'(\bar{c})\bar{c}$$

which completes the proof. □

Proof of Lemma 7. Let $a \in \mathbb{R}_{++}$, $b = u(a)$, and $g = u^{-1}$. Using the definition of g is it straightforward to show that

$$\frac{g''(u(a))}{g'(u(a))} = -\frac{u''(a)}{u'(a)^2}$$

Using the chain rule we see

$$\phi'\big(u(a)\big) = v'(a)g'\big(u(a)\big) \qquad \text{and} \qquad \phi''\big(u(a)\big) = v''(a)g'\big(u(a)\big)^2 + v'(a)g''\big(u(a)\big)$$

Using the fact that $g'(u(a)) = u'(a)^{-1}$ and the preceding results we see

$$\begin{aligned} -\frac{\phi''(u(a))}{\phi'(u(a))} u'(a)a &= -\left(\frac{v''(a)}{v'(a)} g'\Big(u(a)\Big) + \frac{g''\Big(u(a)\Big)}{g'\Big(u(a)\Big)}\right) u'(a)a \\ &= -\left(\frac{v''(a)}{v'(a)} - \frac{u''(a)}{u\text{''}(a)}\right) a = \gamma(a) \end{aligned}$$

which completes the proof. □

References

Kenneth Arrow. Essays in the theory of risk-bearing. 1971.

Gilbert W. Jr. Bassett. The St. Petersburg Paradox and bounded utility. *History of Political Economy*, 19(4):517–523, 1987.

Pierre-Andre Chiappori and Monica Paiella. Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, 9(6):1021–1052, 2011.

Daniel Ellsberg. Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.

Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci. Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118:133–173, 2004.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

Robert T. Jensen and Nolan H. Miller. Giffen behavior and subsistence consumption. *The American Economic Review*, 98(4):1553–1577, 2008.

Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73:1849–1892, November 2005.

Felix Kubler, Larry Selden, and Xiao Wei. Inferior good and Giffen behavior for investing and borrowing. *The American Economic Review*, 103(2):1034–1053, April 2013.

Massimo Marinacci. Probabilistic sophistication and multiple priors. *Econometrica*, 70 (2):755–764, 2002.

Andreu Mas-Colell. On the uniqueness of equilibrium once again. In Equilibrium Theory and Application, editors, *Barnett, W., Cornet, B., D'Aspremont, C., Gabszewicz, J. and Mas-Colell, A.*, chapter 12, pages 275–296. Cambridge University Press, 1991.

Jean-Claude Milleron. Unicite et stabilite de l'equilibre en economie de distribution. *Seminaire d'Econometric Roy-Malinvaud*, 1974.

Leonid Mityushin and Victor Polterovich. Criteria for monotonicity of demand functions. *Ekonomika i Matematicheskie Metody*, (14):122–128, 1978.

John K.-H. Quah. The law of demand and risk aversion. *Econometrica*, 71(2):719–721, March 2003.

David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.

O. Shisha and B. Mond. Difference of means. *Bulletin of the American Mathematical Society*, 73:328–333, 1967.

Chapter 3

Weak Separability: A Revealed Preference Test

Weak Separability: A Revealed Preference Test

Abstract

We develop a revealed preference test for weakly separable preferences that is distinct from the test of Varian (1983). Our test disposes of the concavity requirements upon which Varian's test is built. This allows us to work with non-linear pricing. It also allows us to calculate Afriat's critical cost efficiency index as a measure of goodness-of-fit. Further, implementing our test involves searching a finite solution space whereas implementing Varian's test involves solving a non-convex programming problem.

1 Introduction

Suppose we observe a consumer making purchases from L goods, with a typical observation t consisting of the bundle $x^t \in \mathbb{R}^L_+$ chosen by the consumer and the price vector $p^t \in \mathbb{R}^L_{++}$ that the consumer faces. A finite dataset $\mathcal{O} = \{(x^1, p^1), \ldots, (x^T, p^T)\}$ is said to be *rationalized* by the function $U : \mathbb{R}^L_+ \to \mathbb{R}$ if, for all t, the observed bundle x^t maximizes $U(x)$ over the set $\{x \in \mathbb{R}^L_+ : p^t \cdot x \leqslant p^t \cdot x^t\}$. Afriat's Theorem gives us the precise condition under which a data set can be rationalized by a *well-behaved*, i.e., strictly increasing[1] and continuous, utility function U. It says that this is possible if and only if the data set obeys the *generalized axiom of revealed preference* or GARP, for short (see Afriat (1967) and Varian (1982)). This property requires that there be no strict revealed preference cycles on the set of observed consumption bundles $\mathcal{X} = \{x^1, \ldots, x^T\}$, where a bundle x^t is said to be revealed preferred (revealed strictly preferred) to another bundle x^s if $p^t \cdot x^t \geqslant (>)\, p^t \cdot x^s$. GARP is an easy to check property, either directly or via a linear program, so Afriat's Theorem has become the cornerstone of a large empirical literature on consumer demand.

It is common in empirical and theoretical work to impose more conditions on the utility function, in addition to requiring it to be well-behaved. A particularly common and convenient property is *weak separability*. For this reason, it is useful to develop a characterization of datasets that could be rationalized by utility functions with this

[1] This means that $U(x') > U(x)$ whenever $x' > x$.

added feature. The objective of this paper is to solve this problem. Let us consider what characterizes a weakly separable utility function.

Suppose we have a desired way of dividing the L goods an agent purchases into categories. Suppose category j has ℓ_j goods and so the consumption bundle in this category may be represented by a vector $x_j \in \mathbb{R}_+^{\ell_j}$. We can interpret these categories in different ways. For example, category j could represent a particular class of goods, say clothing items or food items, it could represent goods consumed in period j in an intertemporal model, or goods consumed in state j in a model with uncertainty. In these cases and in many others, one may wish to check whether the agent has a preference over bundles in category j that is independent of her consumption of goods outside that set. We denote the agent's consumption bundle by $x = (x_1, x_2, ..., x_J)$, where the subvector $x_j \in \mathbb{R}_+^{\ell_j}$ gives his consumption of the goods in category j. The data set $\mathcal{O} = \{(x^1, p^1), \ldots, (x^T, p^T)\}$ is said to be *rationalized* by a weakly separable utility function if there are well-behaved sub-utility functions $U_j : \mathbb{R}_+^{\ell_j} \to \mathbb{R}_+$ and aggregator function $F : \mathbb{R}_+^J \to \mathbb{R}$ such that the utility function $G : X \to R$, given by

$$G(x) = F\Big(U_1(x_1), U_2(x_2), ..., U_J(x_J)\Big), \tag{1}$$

rationalizes $\mathcal{O}$. Let us consider what problem weak separability addresses.

When presented with time series data for an individual's consumption most analysis will at least implicitly assume some sort of separability in the individual's preferences. That is, if one only has data on a person's grocery consumption over a period of one year and we would like to say something meaningful about the individual's preferences, it will be essential that the agent's preferences for groceries are not affected by the individual's consumption of something unobserved, say furniture. If it were the case that every time the agent purchased a new lawn chair the consumer experienced a sudden aversion to potatoes then having only data on grocery consumption would be insufficient. In this case we would need to have data on the person's consumption of furniture. Indeed it seems that whenever we do not see the entirety of an individual's consumption we must make some reasonable assumption of separability in order to say anything meaningful about the preferences of the individual. This paper allows us to test this critical assumption.

The form introduced in (1) allows us to speak of preferences over bundles in category j without making reference to any other goods being consumed. Indeed if x_{-j} and x'_{-j} are two bundles of goods that have every category of good represented except for j and if x_j and x'_j are two bundles of goods of category j then clearly

$$G(x_j, x_{-j}) \geqslant G(x'_j, x_{-j}) \iff G(x_j, x'_{-j}) \geqslant G(x'_j, x'_{-j}) \iff U_j(x_j) \geqslant U_j(x'_j)$$

While we have just presented a dataset $\mathcal{O}$ with prices $p^t \in \mathbb{R}^L_{++}$ in the body of the paper we shall work with more general constraint sets. Following the approach in Forges and Minelli (2009) we shall introduce a set property called regularity which shall characterize the sets we are considering. We shall speak momentarily about the advantages of this approach.

While this paper does not introduce the first revealed preference test for weakly separable preferences, we believe our approach offers some advantages. The work of Varian (1983), which was later generalized by Diewert and Parkan (1985), presents a revealed preference test for weakly separable preferences using an approach that is distinct from ours. Varian's approach is to test for a concave aggregator function F and concave sub-utility functions $U_1, \ldots, U_J$ such that $G(x) = F(U(x_1), \ldots, U(x_J))$ rationalizes the data. Their approach assumes linear pricing. This allows one to exploit the usual first order conditions for utility maximization to derive necessary conditions for maximization with a weakly separable utility function. These necessary conditions turn out to be sufficient to guarantee the existence of a weakly separable utility function that rationalizes the data.

Our approach makes no such concavity assumptions on aggregator F or on sub-utility functions $U_1, \ldots, U_J$. Our test then imposes fewer restrictions on the utility function and can be thought of as a way of testing for weakly separable preferences without imposing anything additional.

Our approach also allows for greater diversity in the constraint sets. As Varian's approach relies on the first order conditions for utility maximization it requires the constraint sets to be convex. Clearly this is no problem when each good has a linear price in each period. However, there are contexts where we would like the constraint sets to not be characterized in this way. Examples are provided in Matzkin (1991) and Forges and Minelli (2009). Examples include households with production, households facing progressive or regressive taxation, and what Forges and Minelli (2009) call strategic market games. Our general budget sets also allow us to calculate the popular goodness-of-fit measure developed in Afriat (1972), the critical cost efficiency index (CCEI).

Our test is guaranteed to terminate in finite time as our search space has a finite number of elements. This seems a significant improvement over other tests which have to be implemented as non-convex programming problems as these non-convex programming problems provide no guarantee that they can be solved in finite time.

The condition we develop is easy to interpret and is similar to Afriat's condition, GARP. GARP: the requirement that there are no strict revealed preference cycles in the data, is an obvious necessary condition for the data to be rationalized by well-behaved preferences. We require there to be a complete preorder on each of the categories where these preorders do not induce any strict revealed preference cycles in the data. Notice that each sub-utility function U_j will induce a complete ordering of $\mathbb{R}_+^{\ell_j}$. So, the group of sub-utility functions $U(x) = (U_1, \ldots, U_J)$ will induce an order on $\mathbb{R}_+^L$ which extends the usual Euclidean order. The challenge will be to ensure that the this U does not induce an order that contradicts the revealed preference relations on $\mathcal{O}$.

We organize this paper as follows. We begin by providing some definitions from the revealed preference literature and introduce a version of Afriat's Theorem. Afterwards we present our main theorem which presents necessary and sufficient conditions for a dataset to be rationalized with weakly separable preferences. This is followed by a presentation of the proof. The last section shows how the CCEI can be incorporated into our approach.

2 Preliminary Definitions and Results

2.1 Consumption Space

Let $\mathbb{N}$ to denote the set of natural numbers. For $n \in \mathbb{N}$ we write $\mathbb{N}_n$ to denote the set of natural number $\{1, \ldots, n\}$. $X = \mathbb{R}_+^L$ will denote the consumption space. We assume that the $L \in \mathbb{N}$ goods can be divided into $J \leqslant L$ categories. Category $j \in \mathbb{N}$ has $\ell_j \in \mathbb{N}$ goods. For each $j \in \mathbb{N}_J$ let $X_j = \mathbb{R}_+^{\ell_j}$ be the set of possible bundles that can be consumed of goods in category j. For $x \in X$ we shall often write $x_j \in X_j$ to denote the bundle of goods in category $j \in \mathbb{N}_j$ which are in x. So, we have $x = (x_1, \ldots, x_J)$.

We shall equip the positive orthants $\mathbb{R}_+^n$ for all $n \in \mathbb{N}$ with the usual topology relative to the positive orthants. So, a set $E \subseteq \mathbb{R}_+^n$ is open if there is a set $E' \subseteq \mathbb{R}^n$ that is open in the usual topology on $\mathbb{R}^n$ such that $E = \mathbb{R}_+^n \cap E'$. For $B \subseteq \mathbb{R}_+^n$ let B^c be the complement of B relative to $\mathbb{R}_+^n$. That is $B^c = \mathbb{R}_+^n \backslash B$. We say that a set B is *co-convex* if B^c is convex. For $B \subseteq \mathbb{R}_+^n$ let $\bar{B}$ denote the closure of B and let B^o denote the interior of B. That is $\bar{B}$ is the intersection of all closed sets containing B and B^o is the union of all open sets contained in B. We denote the boundary of B by $\partial B = \bar{B} \cap \overline{B^c}$.

2.2 Datasets

Suppose $T \in \mathbb{N}$. We suppose we have a collection of observations of a consumer's decisions $\mathcal{O} = \{(x^1, B^1), \ldots, (x^T, B^T)\}$ where $B^t \subseteq X$ represents the possible bundles that could have been consumed in period t while $x^t \in B^t$ represent the bundle which the consumer actually chose during period t. Our goal will be to test if this collection of observations could be generated by a consumer with weakly separable utility. We shall assume that the sets B^t satisfy a property we call *regularity* which is defined in the next section. We refer to any collection $\mathcal{O} = \{(x^1, B^1), \ldots, (x^T, B^T)\}$ where all the sets B^t are regular and the element x^t is on the boundary of B^t as a *dataset*.[2] For any $t \in \mathbb{N}_T$ we call the ordered pair (x^t, B^t) an *observation*. The set B^t is called a *constraint set*. The bundle x^t is called the bundle chosen in period t.

2.3 Regular Sets

We have mentioned that a dataset, by definition, has regular constraint sets. We have yet to define this concept.

Definition 1. A set $B \subseteq \mathbb{R}^n_+$ is *regular* if it is compact, satisfies $B = \overline{B^o}$, and for $y \in B$ and $x \in \mathbb{R}^n_+$ we have

$$x < y \implies x \in B^o \tag{2}$$

Let us provide some intuition about how these sets behave. Because they are compact we know that they are not "too" large. The fact that $B = \overline{B^o}$ means that they have no isolated points. In particular this means that $B \neq \{0\}$. Equation (2) shows that B is downward closed in the Euclidean order. In particular if $b \in B$ then the set $\{b' \in \mathbb{R}^n_+ : b' \leqslant b\}$ is contained in B. The fact that anything lower than $b \in B$ must be in the interior of B means that a set of the form $D(b) = \{b' \in \mathbb{R}^n_+ : b' \leqslant b\}$ is not regular.

An important special case of a regular set is the familiar linear budget set. For a price vector $p \in \mathbb{R}^n_{++}$ and wealth level $w \in \mathbb{R}_{++}$ let us write

$$B(p, w) = \{c \in \mathbb{R}^n_+ : p \cdot c \leqslant w\} \tag{3}$$

So long as we impose the requirement that wealth is positive (i.e. $w > 0$) the set in (3) is regular.

[2] Recall that we are working with the topology relative to the positive orthant. This will mean that 0 is in general not on the boundary of $B^t \subseteq \mathbb{R}^t_+$. In fact, the definition of regularity will preclude $0 \in \partial B^t$.

2.4 Rationalizing Datasets

Suppose $\mathcal{O} = \{(x^1, B^1), \ldots, (x^T, B^T)\}$ is a dataset. We say that $\mathcal{O}$ is rationalizeable by a function $H : \mathbb{R}^n_+ \to \mathbb{R}$ if

$$H(x^t) = \sup_{x \in B^t} H(x)$$

A function $H : \mathbb{R}^n_+ \to \mathbb{R}$ is *well-behaved* if it is continuous and strictly increasing. Strictly increasing means $H(y) > H(x)$ when $y > x$. If $\mathcal{O}$ is rationalizeable by some well-behaved H we say that $\mathcal{O}$ can be rationalized by well-behaved preferences. We say that a function $G : X \to \mathbb{R}$ is weakly separable if there exists well-behaved functions $F : \mathbb{R}^J \to \mathbb{R}$ and $U_j : X_j \to \mathbb{R}$ for each $j \in \mathbb{N}_J$ such that

$$G(x) = F\Big(U_1(x_1), \ldots, U_J(x_J)\Big)$$

Clearly, G will be well-behaved if F and U_j are well-behaved for all $j \in \mathbb{N}_J$. We shall refer to F as the *aggregator* function and to $U_1, \ldots, U_J$ as *sub-utility* functions. We say that a dataset is *rationalizeable by weakly separable preferences* if there exists a weakly separable G that rationalizes the dataset. Our goal is to find necessary and sufficient conditions for a dataset $\mathcal{O}$ to be rationalizeable by weakly separable preferences.

2.5 Revealed Preferences and Order Theory

Here we introduce some key concepts in revealed preference theory and order theory.

Definition 2. Let $\succeq$ be a binary relation on a subset Y of $\mathbb{R}^n_+$. We say that $\succeq$ is a *preorder* if it is

1. (transitive) For $x, y, z \in Y$ if $x \succeq y$, $y \succeq z$ then $x \succeq z$.
2. (reflexive) For $x \in Y$ we have $x \succeq x$.

A binary relation is said to be *complete* if for $x, y \in Y$ we have either $x \succeq y$ or $y \succeq x$.

For a binary relation $\succeq$ on Y we write $\succ$ to denote the asymmetric part of $\succeq$. That is $y \succ x$ means $y \succeq x$ and *not* $x \succeq y$. We shall use $\sim$ to denote the symmetric part of $\succeq$. That is $y \sim x$ if $y \succeq x$ and $x \succeq y$. Suppose that $\succeq$ and $\succeq'$ are binary relations on $Y \subseteq \mathbb{R}^n_+$. We say that $\succeq'$ *extends* $\succeq$ if $y \succeq x$ implies $y \succeq' x$. We say that a function $f : Y \to \mathbb{R}$ *extends* a relation $\succeq$ if $y \succeq x$ implies $f(y) \geqslant f(x)$ and if $y \succ x$ implies $f(y) > f(x)$. For a binary relation $\succeq$ on Y we say that $\succeq'$ is the *transitive closure* of $\succeq$ if we have $y \succeq' x$ if and only if we may find $n \in \mathbb{N}$ elements $x^1, \ldots, x^n \in Y$ such that $y \succeq x^1 \succeq x^2 \succeq \ldots \succeq x^{n-1} \succeq x^n \succeq x$.

For a dataset $\mathcal{O} = \{(x^1, B^1), \ldots, (x^T, B^T)\}$ we shall write

$$\mathcal{X} = \{x^1, \ldots, x^T\}, \qquad \mathcal{X}_j = \{x_j^1, \ldots, x_j^T\} \text{ for all } j \in \mathbb{N}_J, \qquad \bar{\mathcal{X}} = \prod_{j=1}^{J} \mathcal{X}_j$$

So, $\mathcal{X}$ is the set of bundles that are consumed in some time period $t \in \mathbb{N}_T$. The set $\mathcal{X}_j$ is the set of bundles of goods in category j which are consumed in some time period $t \in \mathbb{N}_T$. Finally, $\bar{x} \in \bar{\mathcal{X}}$ is a bundle such that for each $j \in \mathbb{N}_T$ the bundle $\bar{x}_j \in X_j$ was consumed in some time period. So, the set $\bar{\mathcal{X}}$ is in general considerably bigger than $\mathcal{X}$. In general $\mathcal{X}$ will have around T elements while $\bar{\mathcal{X}}$ has around T^J elements.

The following definition presents a key concept in revealed preference theory.

Definition 3. The *direct revealed preference relation* for $\mathcal{O}$ is the binary relation $\succeq^*$ on $\bar{\mathcal{X}}$ where $y \succeq^* x$ if $y = x^t$ for some $t \in \mathbb{N}_T$ and $x \in B^t$. Similarly, the *strict direct revealed preference relation* for $\mathcal{O}$ is the binary relation $\succ\!\!\succ^*$ on $\bar{\mathcal{X}}$ where $y \succ\!\!\succ^* x$ if $y = x^t$ for some $t \in \mathbb{N}_T$ and $x \in (B^t)^o$. Let $\succeq^{**}$ be the transitive closure of $\succeq^*$. We call $\succeq^{**}$ the *revealed preference relation* for $\mathcal{O}$. We say that a preorder $\succeq$ *extends the revealed preference relations on* $\mathcal{O}$ if $y \succeq^* x$ implies $y \succeq x$ and $y \succ\!\!\succ^* x$ implies $y \succ x$.

2.6 GARP

We now define an important concept for determining if a dataset may be rationalized by a well-behaved utility function.

Definition 4. A dataset $\mathcal{O}$ satisfies the *generalized axiom of revealed preferences* (GARP) if $x^t \succeq^{**} x^s$ implies not $x^s \succ\!\!\succ^* x^t$ for all $s, t \in \mathbb{N}_T$.

Afriat (1967) and Forges and Minelli (2009) show that a dataset may be rationalized by a continuous and increasing utility function if and only if the dataset satisfies GARP.

The following is a slightly more delicate version of Forges and Minelli (2009) which is itself a generalization of Afriat's Theorem and so we shall refer to it as the Afriat-Forges-Minelli (AFM) Theorem.

Theorem 0 (AFM)**.** *Suppose $\mathcal{O}$ is a dataset. If $\succeq$ is a complete preorder on $\mathcal{X}$ that extends the revealed preference relations on $\mathcal{O}$ then the data can be rationalized by a well-behaved utility function $H : X \to \mathbb{R}_+$ that extends $\succeq$ and satisfies $H(0) = 0$. Further, if the constraint sets $B^1, \ldots, B^T$ are co-convex then H can be chosen to be concave.*

The proof of this theorem appears in Quah (2014) and so we omit it. This result gives us more control over the utility function we use to rationalize datasets.

3 Testing for Weak Separability

3.1 R Revealed Preferred

We present appropriate modifications to the revealed preference relations, as expressed in definition 3, which allow us to state our main theorem.

Suppose $R = (\succeq_1, \ldots, \succeq_J)$ is a collection of pre-orders on $\mathcal{X}_1, \ldots, \mathcal{X}_J$ respectively. We define a binary relation $\succeq_R$ on $\bar{\mathcal{X}}$ that aggregates the $\succeq_j$ relations onto $\bar{\mathcal{X}}$ in a natural way. For $x, y \in \bar{\mathcal{X}}$ we let

$$y \succeq_R x \quad \text{if} \quad \forall j \in \mathbb{N}_J \quad \text{we have } y_j \succeq_j x_j$$

We now introduce the two relations which serve as the main revealed preference relations in this paper. The relation $\succeq_R^*$ is the composition of $\succeq^*$ and $\succeq_R$ where the relation is restricted to $\mathcal{X}$. The relation $\ggcurly_R^*$ is the union of the composition of $\succ^*$ with $\succeq_R$ and the composition of $\succeq^*$ with $\succ_R$ where the relation is restricted to $\mathcal{X}$. That is

$$\begin{aligned} &x^t \succeq_R^* x^s \quad \text{if} \quad \exists \bar{x} \in \bar{\mathcal{X}} \text{ s.t.} && x^t \succeq^* \bar{x} \text{ and } \bar{x} \succeq_R x^s \\ &x^t \ggcurly_R^* x^s \quad \text{if} \quad \exists \bar{x} \in \bar{\mathcal{X}} \text{ s.t. either} && \begin{cases} \text{(i)} & x^t \ggcurly^* \bar{x} \text{ and } \bar{x} \succeq_R x^s \text{ or} \\ \text{(ii)} & x^t \succeq^* \bar{x} \text{ and } \bar{x} \succ_R x^s \end{cases} \end{aligned} \tag{4}$$

where $\succeq_R^*$ and $\ggcurly_R^*$ are defined on $\mathcal{X}$. When $x^t \succeq_R^* x^s$ we shall say that x^t is *R direct revealed preferred* to x^s and when $x^t \ggcurly_R^* x^s$ we shall say that x^t is *strictly R direct revealed preferred* to x^s. Let $\succeq_R^{**}$ be the transitive closure of $\succeq_R^*$. We say that x^t is *R revealed preferred to* x^s if $x^t \succeq_R^{**} x^s$. We now introduce the condition which will prove to be necessary and sufficient to rationalize $\mathcal{O}$ with weakly separable preferences.

Definition 5. Suppose $R = (\succeq_1, \ldots, \succeq_J)$ where $\succeq_j$ is a preorder on $\mathcal{X}_j$ for all $j \in \mathbb{N}_J$. If $x^t \succeq_R^{**} x^s$ implies $x^s \not\ggcurly_R^* x^t$ for all $s, t \in \mathbb{N}_T$ then we say that $\mathcal{O}$ satisfies *R-GARP*. Further, if for all $j \in \mathbb{N}_J$ the preorder $\succeq_j$ is complete on $\mathcal{X}_j$ then we say that $\mathcal{O}$ satisfies *complete R-GARP*.

We now present our main theorem.

Theorem 1. *A dataset $\mathcal{O} = \{(x^1, B^1), \ldots, (x^T, B^T)\}$ is rationalizable by a weakly separable utility function if and only if there is a collection of preorders $R = (\succeq_1, \ldots, \succeq_J)$ such that $\mathcal{O}$ satisfies complete R-GARP. Also, if $\mathcal{O}$ satisfies complete R-GARP then for each $j \in \mathbb{N}_J$ we can choose U_j to extend $\succeq_j$. If the constraint sets $B^1, \ldots, B^T$ are co-convex then the sub-utility functions can be chosen to be concave.*

Before going into the proof a few remarks on complete R-GARP and weak separability will help build some understanding of the issues involved. For sub-utility functions $U_1, \ldots, U_J$ let $U : X \to \mathbb{R}^J$ be defined by $U(x) = (U_1(x_1), \ldots, U_J(x_J))$. Let us consider the canonical case where the constraint sets are linear. For all $j \in \mathbb{N}_J$ define

$$\mathcal{O}_j = \left\{ \left(x_j^1, B(p_j^1, p_j^1 \cdot x_j^1)\right), \ldots, \left(x_j^T, B(p_j^T, p_j^T \cdot x_j^T)\right) \right\}$$

Let $\succeq_j^*$ and $\ggcurly_j^*$ be the revealed preference relations for this dataset. It is a well known property of weak separable utility that if $F(U(y)) = \max_{x \in B(p,w)} F(U(x))$ then it must also be the case that $U_j(y_j) = \max_{x_j \in B(p_j, p_j \cdot y_j)} U_j(x_j)$. That is, a utility maximizing agent must allocate their consumption optimally within each category.

From this discussion it is reasonable to suspect that any weak separable function G that rationalizes $\mathcal{O}$ also must have its sub-utility functions rationalize $\mathcal{O}_1, \ldots, \mathcal{O}_J$. This is in fact the case. Notice that R-GARP does not explicitly require that the datasets $\mathcal{O}_1, \ldots, \mathcal{O}_J$ satisfy GARP even though this would seem to be an obvious necessary condition for $\mathcal{O}$ to be rationalized by a weakly separable utility function. As one might suspect, this is in fact implied by complete R-GARP although not explicitly stated. What is actually required by complete R-GARP is that the relations in R agree with the revealed preference relations on $\mathcal{O}_1, \ldots, \mathcal{O}_J$ which of course implies that these datasets satisfy GARP. To see this suppose that $x_j^t \succeq_j^* x_j^s$ but $x_j^s \succ_j x_j^t$. This means that $x^t \succeq^* (x_j^s, x_{-j}^t)$ and $(x_j^s, x_{-j}^t) \succ_R x^t$ which shows that $x^t \ggcurly_R^* x^t$ and so the data does not satisfy complete R-GARP.

We also remark that if $\mathcal{O}$ satisfies complete R-GARP then each preference $\succeq_j$ must agree with the usual Euclidean order on X_j. To see this, suppose that $x_j^t > x_j^s$. As B^t is regular we see that $(x_j^s, x_{-j}^t) \in (B^t)^o$ which means $x_j^t \succ_j^* x_j^s$ and we have seen that if we have complete R-GARP then $\succeq_j$ must agree with $\succeq_j^*$. So, we see that each $\succeq_j$ must agree with the usual Euclidean order for complete R-GARP to hold.

We now prove that a dataset that is rationalizable by a weakly separable utility function satisfies complete R-GARP.

Proof of the Necessary Part of Theorem 1. Suppose $G = F \circ U$ rationalizes $\mathcal{O}$ where $U(x) = (U_1(x_1), \ldots, U_J(x_J))$. For each $j \in \mathbb{N}_J$ define $\succeq_j$ on $\mathcal{X}_j$ so that $y_j \succeq_j x_j$ if and only if $U_j(y_j) \geqslant U_j(x_j)$. Let $R = (\succeq_1, \ldots, \succeq_J)$. We see that $x^t \succeq_R^* x^s$ means $G(x^t) \geqslant G(x^s)$ and $x^t \succ_R^* x^s$ means $G(x^t) > G(x^s)$. So, if $x^t \succeq_R^{**} x^s$ and $x^s \ggcurly_R^* x^t$ then we would have $G(x^t) \geqslant G(x^s) > G(x^t)$ which is a contradiction. Thus, $\mathcal{O}$ satisfies complete R-GARP. □

The sufficiency part is much more challenging to show.

3.2 Constructing the Rationalizing Function

We prove the sufficiency part of our theorem by using the AFM Theorem to construct the sub-utility functions $U_1, \ldots, U_J$ and also to define the aggregator function F. Suppose we have some sub-utility functions $U_1, \ldots, U_J$ and let $U : X \to \mathbb{R}^J$ be defined by

$$U(x) = \Big(U_1(x_1), \ldots, U_J(x_J)\Big) \tag{5}$$

Define a new collection[3] $\mathcal{O}_U$ by

$$\mathcal{O}_U = \Big\{\Big(U(x^1), U(B^1)\Big), \ldots, \Big(U(x^T), U(B^T)\Big)\Big\} \tag{6}$$

where $U(B^t) = \{u \in \mathbb{R}^J : \exists x \in B^t \text{ s.t. } U(x) = u\}$. Let $\mathcal{X}_U = \{U(x^1), \ldots, U(x^T)\}$. We would like to use Theorem 0 on $\mathcal{O}_U$ to generate an aggregator function F but in order to do this we require the constraint sets in $\mathcal{O}_U$ to be regular. Clearly, this will require U_j to satisfy $U_j(0) = 0$. In fact, as the next lemma shows, if U_j is well-behaved this is all that is required.

Lemma 1. *Suppose $B \subseteq X$ is regular and $U_j : X_j \to \mathbb{R}_+$ is continuous, strictly increasing, and satisfies $U_j(0) = 0$ for all $j \in \mathbb{N}_J$. Define $U : X \to \mathbb{R}^J_+$ by equation (5). Then $U(B) = \{u \in \mathbb{R}^J_+ : \exists x \in B \text{ s.t. } u = U(x)\}$ is regular.*

Proof. $U(B)$ is compact as U is continuous and B is compact. If B is empty the proof is trivial so assume B is non-empty. The proof relies on two claims. Claim 1: if $E \subseteq X$ is open then so is $U(E)$. To see this, note that if $E_j \subseteq X_j$ is open then $U_j(E_j)$ is open as U_j is strictly increasing, continuous, and maps to $\mathbb{R}_+$. This means that for any $x \in E$ there are open sets $E_1, \ldots, E_J$ such that $x \in E_1 \times \ldots \times E_J \subseteq E$ and so $U(x) \in U(E_1 \times \ldots \times E_J)$ and $U(E_1 \times \ldots \times E_J)$ is open. So, any point in $U(E)$ is in an open set contained in $U(E)$ which establishes the claim. Next, for $x \in \mathbb{R}^n_+$ let $D(x) = \{y \in \mathbb{R}^n_+ : y \leqslant x\}$. Claim 2: $U(D(x)) = D(U(x))$. $U(D(x)) \subseteq D(U(x))$ follows from U being strictly increasing. Next, notice that $D(U_j(x_j)) \subseteq U_j(D(x_j))$ as $U_j(D(x_j)$ is merely the closed interval between 0 and $U_j(x_j)$. This means $D(U(x)) \subseteq U(D(x))$ which establishes the claim.

Suppose, $u \in U(B)$ and $u' < u$. As $u \in U(B)$ there is an $x \in B$ such that $U(x) = u$. Now, claim 2 shows us that there is some $x' \in B$ such that $x' < x$ and $U(x') = u'$. As B is regular we see that $x' \in B^o$. Now, claim 1 shows that $U(x') \in U(B)^o$. So, we see that (2) holds. Next, we show that $U(B) = \overline{U(B)^o}$. As $U(B)$ is closed we have $\overline{U(B)^o} \subseteq U(B)$. Clearly, $0 \in U(B)$ and $\overline{U(B)^o}$ so suppose that $u \in U(B)$ where $u \neq 0$.

[3] We do not refer to this new collection as a dataset as it is not clear that the constraint sets are regular.

Let $K = D(u)\backslash\{u\}$. As we have shown that $U(B)$ satisfies (2) we see that $K \subseteq U(B)^o$ and clearly $u \in \overline{K}$ so $u \in \overline{K} \subseteq \overline{U(B)^o}$ which completes the proof. $\square$

So, if we have sub-utility functions satisfying the assumptions of Lemma 1 then $\mathcal{O}_U$ defined in (6) is a dataset. In this case we may apply Theorem 0 to obtain an aggregator function $F : \mathbb{R}^J_+ \to \mathbb{R}$ that rationalizes $\mathcal{O}_U$. The following lemma demonstrates that the function $G = F \circ U$ will rationalize the original dataset.

Lemma 2. *Let $U : X \to \mathbb{R}^J_+$ be a continuous function. Then $\mathcal{O}_U$ defined by (5) is rationalized by continuous function F if and only if $\mathcal{O}$ is rationalized by $G = F \circ U$.*

Proof. Fix $t \in \mathbb{N}_T$ and let $u^t = U(x^t)$ and $K^t = U(B^t)$. The result follows from

$$\begin{aligned} G(x^t) \geqslant G(x), \quad \forall x \in B^t &\iff F(U(x^t)) \geqslant F(U(x)), \quad \forall x \in B^t \\ &\iff F(u^t) \geqslant F(u), \qquad\qquad \forall u \in K^t \end{aligned}$$

which completes the proof. $\square$

Lemma 2 shows that if we can find an aggregator function F that rationalizes $\mathcal{O}_U$ defined by (6) for some sub-utility functions $U_1, \ldots, U_J$ then we can rationalize $\mathcal{O}$ with weakly separable preferences.

The challenging part is to define these sub-utility functions $U_1, \ldots, U_J$ in a way that guarantees $\mathcal{O}_U$ satisfies GARP. The first thing to note is that the most natural way to define the sub-utility functions does not work. Let $B^t_j \subseteq X_j$ be the set of bundles in X_j that are available when the bundle x^t_{-j} is purchased and the overall constraint set is B^t. That is $x_j \in B^t_j$ implies $(x_j, x^t_{-j}) \in B^t$. Notice that when the constraint sets are described by price and wealth levels as in equation (3) then $B^t_j = \{x_j : p_j \cdot x_j \leqslant p_j \cdot x^t_j\}$. Define a dataset by

$$\hat{\mathcal{O}}_j = \{(x^1_j, B^1_j), \ldots, (x^T_j, B^T_j)\}$$

From our previous discussion it is clear that U_j must be chosen to rationalize $\mathcal{O}_j$. Unfortunately, this is not sufficient to ensure that $\mathcal{O}_U$ defined in (6) will satisfy GARP.

To gain some intuition into why this is, suppose we have constructed sub-utility functions $U_1, \ldots, U_J$ to rationalize the sub-datasets $\hat{\mathcal{O}}_1, \ldots, \hat{\mathcal{O}}_J$. It is entirely possible that $\mathcal{O}$ satisfied GARP but $\mathcal{O}_U$ does not. The issue, in short, is that there may be revealed preference relations in the dataset $\mathcal{O}_U$ that do not exist in $\mathcal{O}$. Or, more precisely, there may be some x^t and x^s purchased in $\mathcal{O}$ where $x^s \notin B^t$ *but* $U(x^s) \in U(B^t)$. This means that $x^t \not\succeq^* x^s$ *but* $U(x^t) \succeq^*_U U(x^s)$ (where $\succeq^*_U$ is the direct revealed preference relation on the dataset $\mathcal{O}_U$). That is, when transforming $\mathcal{O}$ to $\mathcal{O}_U$ it is likely that we introduce many revealed preference relations that were not in the original dataset. This can cause $\mathcal{O}_U$ to not satisfy GARP even if $\mathcal{O}$ does. The key then is to be parsimonious in the construction of $U_1, \ldots, U_J$.

4 Proving the Sufficiency Part of the Main Theorem

We now provide a brief summary of the approach we will use to complete the sufficiency part of the proof. The basic idea is to first assume that we have an $R = (\succeq_1, \ldots, \succeq_J)$ such that $\mathcal{O}$ satisfies complete R-GARP. Then we construct artificial datasets $\mathcal{O}_1, \ldots, \mathcal{O}_J$ where $\mathcal{O}_j = \{(x_j^1, C_j^1), \ldots, (x_j^T, C_j^T)\}$.[4] We rationalize these datasets with sub-utility functions $U_1, \ldots, U_J$ so that U_j extends $\succeq_j$. From the way we construct these artificial datasets we guarantee that the dataset $\mathcal{O}_U$ defined in (6) satisfies GARP which will allow us to obtain the appropriate aggregator function using the AFM Theorem.

Let us consider properties which the artificial datasets $\mathcal{O}_1, \ldots, \mathcal{O}_J$ should satisfy to guarantee that the dataset $\mathcal{O}_U$ satisfies GARP. First, each dataset $\mathcal{O}_j$ must be constructed to ensure that $\succeq_j$ extends its revealed preferences. This is because we would like the sub-utility function U_j to extend $\succeq_j$ which is only possible if $\succeq_j$ is not contradicted by the revealed preference relations on $\mathcal{O}_j$. A more subtle requirement is that we would like U defined by (5) to not map an observation x^t into the image of another observation's constraint set unless this is required by the preorders $(\succeq_1, \ldots, \succeq_J)$. That is, we would like to prevent $U(x^t) \in U(B^s)$ whenever possible. The trick to doing this will be to define the artificial constraint sets $(C_1^t, \ldots, C_J^t)$ so that the rationalizing utility functions ensure $U(x^t)$ is not dominated (in the Euclidean order) by $U(x)$ where $x \in B^s$. We may think of the problem then in terms of a trade-off between choosing the artificial constraint sets to be small enough so that they do not contradict our specified preorders $(\succeq_1, \ldots, \succeq_J)$ but large enough to ensure that the rationalizing utility functions do not map $U(x^t)$ into $U(B^s)$ when this can be prevented. Our first task will be to take a closer look at how we would like sub-utility functions to behave.

Suppose $U_1, \ldots, U_J$ are well-behaved sub-utility functions and U is defined according to (5). Suppose $\mathcal{O}_U$ is defined by (6). Let $\mathcal{X}_U = \{U(x^1), \ldots, U(x^T)\}$. Let $\succeq_U^*$ and $\ggcurly_U^*$ denote the directed revealed preferences on $\mathcal{O}_U$. The key to proving our main result is to show that if a dataset satisfies complete R-GARP then we may guarantee that

$$U(x^s) \succeq_U^* U(x^t) \quad \Longleftrightarrow \quad x^s \succeq_R^* x^t \tag{7}$$

$$U(x^s) \ggcurly_U^* U(x^t) \quad \Longleftrightarrow \quad x^s \ggcurly_R^* x^t \tag{8}$$

[4]These artificial datasets will be distinct from the datasets $\hat{\mathcal{O}}_1, \ldots, \hat{\mathcal{O}}_J$ defined in the previous section.

One way of understanding equations (7) and (8) is to note that when restricting U to be a map from $\mathcal{X}$ onto $\mathcal{X}_U$ then U is an order isomorphism from $(\mathcal{X}, \succeq_R^*)$ to $(\mathcal{X}_U, \succeq_U^*)$ and from $(\mathcal{X}, \succ\!\!\succ_R^*)$ to $(\mathcal{X}_U, \succ\!\!\succ_U^*)$. This means that the R revealed preference and the revealed preference $\succeq_U^*$ are essentially the same relation. We shall show that if (7) and (8) hold then $\mathcal{O}$ will satisfy complete R-GARP if and only if $\mathcal{O}_U$ satisfies GARP. Noting that Lemma 1 shows that the constraint sets in $\mathcal{O}_U$ are regular we see that if $\mathcal{O}_U$ satisfies GARP then we can use Theorem 0 to rationalize $\mathcal{O}_U$.

That we may define sub-utility functions $U_1, \ldots, U_J$ that satisfy the conditions (7) and (8) will follow from a series of lemmas that we present later. Once we are able to guarantee the existence of such utility functions the remainder of the proof of Theorem 1 is straightforward and we present it now.

Proof of Theorem 1. Suppose $\mathcal{O}$ satisfies complete R-GARP. The lemmas that follow guarantee that we may construct well-behaved sub-utility functions $U_1, \ldots, U_J$ and U defined by (5) such that $U(0) = 0$ and (7) and (8) hold. For a contradiction suppose that $\mathcal{O}_U$ does not satisfy GARP. This means that there is some $t, s \in \mathbb{N}_T$ such that $U(x^t) \succeq_U^{**} U(x^s)$ while $U(x^s) \succ\!\!\succ_U^* U(x^t)$. But, equations (7) and (8) show that $x^t \succeq_R^{**} x^s$ and $x^s \succ\!\!\succ_R^* x^t$ which contradicts complete R-GARP. So, we see that $\mathcal{O}_U$ obeys GARP. Now, Theorem 0 shows that we may find an aggregator function $F : \mathbb{R}_+^J \to \mathbb{R}$ that rationalizes $\mathcal{O}_U$. Lemma 2 shows that $G = F \circ U$ rationalizes $\mathcal{O}$. $\square$

What remains is to show that if $\mathcal{O}$ satisfies complete R-GARP then we may find $U_1, \ldots, U_J$ and U defined by (5) such that $U(0) = 0$ and (7) and (8) hold. To do this we apply Theorem 0 to specially constructed artificial datasets. For each $j \in \mathbb{N}_J$ and $t \in \mathbb{N}_T$ we shall construct a constraint set $C_j^t \subseteq X_j$ such that $x_j^t \in \partial(C_j^t)$ and define a dataset

$$\mathcal{O}_j = \{(x_j^1, C_j^1), \ldots, (x_j^T, C_j^T)\} \tag{9}$$

The trick will be to select the constraint sets in such a way that an application of Theorem 0 to each of the datasets $\mathcal{O}_j$ guarantees that (7) and (8) hold. We shall show that it suffices to find regular constraint sets $C_1^1, \ldots, C_1^T, \ldots, C_J^1, \ldots, C_J^T$ that satisfy the following properties:

*A*1 For each $j \in \mathbb{N}_J$ the relation $\succeq_j$ extends the revealed preference relations on $\mathcal{O}_j$.

*A*2 If $x^s \not\succeq_R^* x^t$ then for any $x \in B^s$ there exists a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^t)^o$.

*A*3 If $x^s \not\succ\!\!\succ_R^* x^t$ then for any $x \in B^s$ either (i) there exists a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^t)^o$ or (ii) for all $j \in \mathbb{N}_J$ we have $x_j \in C_j^t$.

*A*4 If B^t is co-convex for all $t \in \mathbb{N}_T$ then all the constraint sets are co-convex.

Recall our discussion of the trade-off faced when constructing these artificial constraint sets. We noted that the challenge was to make the constraint sets small enough to conform to the relations in R while being large enough to prevent the rationalizing sub-utility functions from mapping some bundle x^t into B^s in the sense that $U(x^t) \in U(B^s)$ when this was not required. Put in the terms of this discussion, item $A1$ ensures that the constraint sets are small enough while items $A2$ and $A3$ makes sure our constraint sets are large enough.

Let us call any collection of datasets $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$ whose constraint sets satisfy these items an *R-suitable collection of datasets*. We shall now give a brief description on how these items shall be used to prove that an R-suitable collection of datasets allows us to produce sub-utility functions with the desired properties. Item $A1$ will ensure the sub-utility functions $U_1, \ldots, U_J$ agree with the relations in R. $A2$ ensures that (7) holds and $A3$ ensures that (8) holds. $A4$ will allow us to make the sub-utility functions concave when the constraint sets are co-convex.

Lemma 3. *Suppose we have an R-suitable collection of datasets* $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$ *then there are sub-utility functions* $U_1, \ldots, U_J$ *and* U *defined by* (5) *that satisfy* (7) *and* (8). *Further, if the constraint sets* $B^1, \ldots, B^T$ *are co-convex then the sub-utility functions can be chosen to be concave.*

Proof. As $\succeq_j$ extends the revealed preference relations on $\mathcal{O}_j$ (by item $A1$) we may use Theorem 0 to generate a sub-utility function that rationalizes $\mathcal{O}_j$ and agrees with $\succeq_j$. Further, if the constraint sets in $\mathcal{O}$ are co-convex then item $A4$ allows U_j to be concave. Proceed to define $U_1, \ldots, U_J$ in this way. We now show that (7) and (8) hold.

Suppose $x^s \nsucceq_R^* x^t$. Notice that if $U(x^s) \succeq_U^* U(x^t)$ then by definition there must be some $u \in U(B^s)$ such that $u = U(x^t)$. This would mean that there is some $x \in B^s$ such that $u = U(x) = U(x^t)$. But, item $A2$ states that there must be a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^t)^o$ and as U_j rationalizes $\mathcal{O}_j$, we obtain $U_j(x_j^t) > U_j(x_j)$ and thus $U(x^t) \neq U(x)$. So $U(x^s) \nsucceq_U^* U(x^t)$. Next, suppose $x^t \succeq_R^* x^s$. This means there is an $\bar{x} \in \bar{\mathcal{X}}$ such that $x^t \succeq^* \bar{x}$ and $\bar{x} \succeq_R x^s$. As $\bar{x} \in B^t$ we see that $U(\bar{x}) \in U(B^t)$ and so $U(x^t) \succeq_U^* U(\bar{x})$. As U_j was defined to agree with $\succeq_j$ for all $j \in \mathbb{N}_J$, we see that $U(\bar{x}) \geqslant U(x^s)$. By Lemma 1 the set $U(B^t)$ is regular and so $U(x^t) \succeq_U^* U(x^s)$. So we have shown that (7) holds.

Suppose $x^s \not\ggcurly_R^* x^t$. If $U(x^s) \ggcurly_U^* U(x^t)$ then by definition there would be some $u' \in U(B^s)^o$ such that $u' = U(x^t)$ and thus there would be some $u \in U(B^s)^o$ such that $u > U(x^t)$. This would mean that there is some $x \in B^s$ such that $u = U(x) > U(x^t)$. Item $A3$ states that either (i) there is a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^t)^o$ and as U_j rationalizes $\mathcal{O}_j$ this would mean that $U_j(x_j^t) > U_j(x_j)$ which contradicts $U(x) > U(x^t)$ or (ii) for all $j \in \mathbb{N}_J$ we have $x_j \in C_j^t$ which implies $U(x_j^t) \geqslant U_j(x_j)$ and therefore contradicts $U(x) > U(x^t)$. We conclude that $U(x^s) \not\ggcurly_U^* U(x^t)$. Next, suppose $x^t \ggcurly_R^* x^s$. This means either (i) there exists $\bar{x} \in \bar{\mathcal{X}}$ such that $x^t \gg^* \bar{x}$ and $\bar{x} \succeq_R x^s$ or (ii) there exists $\bar{x} \in \bar{\mathcal{X}}$ such that $x^t \succeq^* \bar{x}$ and $\bar{x} \succ_R x^s$. If (i) holds then we have $\bar{x} \in (B^t)^o$ and so there is some $\bar{x}' > \bar{x}$ such that $\bar{x}' \in B^t$. As U is strictly increasing we see that $U(\bar{x}') > U(\bar{x})$. The fact that $U(\bar{x}') \in U(B^t)$ and $U(B^t)$ is regular shows us that $U(\bar{x}) \in U(B)^o$ and thus $U(x^t) \gg^*_U U(\bar{x})$. Because U_j was defined to extend the relation $\succeq_j$ for all $j \in \mathbb{N}_J$, we see that $\bar{x} \succeq_R x^s$ implies $U(\bar{x}) \geqslant U(x^t)$. This shows that $U(x^t) \gg^*_U U(x^s)$. Suppose (ii) holds. We have already shown that $x^t \succeq^* \bar{x}$ implies $U(x^t) \succeq^*_U U(\bar{x})$. Because U_j was defined to extend the relation $\succeq_j$ for all $j \in \mathbb{N}_J$, we see that $\bar{x} \succ_R x^s$ implies $U(\bar{x}) > U(x^t)$. This shows that $U(x^t) \gg^*_U U(x^s)$ and so we see that (8) holds. □

Our remaining task is to show how to construct an R-suitable collection of datasets $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$ when $\mathcal{O}$ satisfies complete R-GARP. This task we leave to the appendix.

5 Goodness-of-Fit

Revealed preference tests are often stated as binary outcome tests. Either the dataset satisfies the tested property or the dataset does not. In the case that the property is found not to hold we would often like a measure of the degree of failure. This is the idea behind Afriat's critical cost efficiency index (CCEI) presented in Afriat (1972) and Varian (1993). The CCEI is used in many studies including Choi, Fisman, Gale, and Kariv (2007a), Choi, Fisman, Gale, and Kariv (2007b), Andreoni and Miller (2002), and Choi, Kariv, Wieland, and Silverman (2014).

We hypothesize that a consumer has preferences satisfying certain properties (for example preferences characterized by a well-behaved utility function) but, through lack of attention or some other mechanism, the consumer does not always choose optimally. We hypothesize that, while they may misallocate their consumption, they shall never behave so inefficiently that they waste more than a certain portion of their choice set. More concretely, suppose we have a dataset $\mathcal{O}$ and let $e \in [0, 1]$. We can hypothesize that there is a well-behaved utility function $H : X \to \mathbb{R}$ where

$$H(x^t) \geqslant \max_{x \in e \cdot B^t} H(x) \tag{10}$$

When the constraint set is characterized by prices and wealth this has a nice interpretation. It says that no more than $(1 - e)$ of the consumer's wealth can be wasted. That is, (10) states that the consumer with wealth w does no worse than a truly optimizing consumer with only ew to spend.

The CCEI is very easy to calculate when we are merely testing for the existence of a well-behaved utility function to rationalize our data. In fact, it only requires defining a new revealed preference relation $\succeq^{*,e}$ and $\succ\!\!\succ^{*,e}$ where

$$\begin{aligned} x^t \succeq^{*,e} x \quad &\implies \quad x \in eB^t \text{ or } x^t \geqslant x \\ x^t \succ\!\!\succ^{*,e} x \quad &\implies \quad x \in (eB^t)^o \text{ or } x^t > x \end{aligned} \tag{11}$$

Let $\succeq^{**,e}$ be the transitive closure of $\succeq^{*,e}$. If $x^t \succeq^{**,e} x^s$ means not $x^s \succ\!\!\succ^{*,e} x^t$ (i.e. something akin to GARP is satisfied) then there is a utility function H that rationalizes the data for CCEI level e. In other words, there is some well-behaved H that satisfies (10). We shall proceed to apply this concept to the case of weak separability.

Definition 6. For a dataset $\mathcal{O}$ and $e \in [0,1]$ we say that the dataset is *rationalizeable by weakly separable utility for CCEI level* e if there is a well-behaved weakly separable utility function $G : X \to \mathbb{R}$ such that

$$G(x^t) \geqslant \max_{x \in e \cdot B^t} G(x), \qquad \forall t \in \mathbb{N}_T$$

We proceed to identify the conditions on $\mathcal{O}$ which allow us to rationalize this dataset with weakly separable utility for a specific e.

5.1 R Revealed Preferences with CCEI

Let us modify the R revealed preference relations introduced in (4) in a way that allows us to rationalize a dataset with weakly separable utility for a given CCEI level. Suppose $R = (\succeq_1, \ldots, \succeq_J)$ is a collection of pre-orders. We define the relations $\succeq_R^{*,e}$ and $\succ\!\!\succ_R^{*,e}$ by

$$\begin{aligned} &x^t \succeq_R^{*,e} x^s \;\; \text{if} \;\; \exists \bar{x} \in \bar{\mathcal{X}} \text{ s.t.} && x^t \succeq^{*,e} \bar{x} \text{ and } \bar{x} \succeq_R x^s \\ &x^t \succ\!\!\succ_R^{*,e} x^s \;\; \text{if} \;\; \exists \bar{x} \in \bar{\mathcal{X}} \text{ s.t. either} && \begin{cases} \text{(i)} & x^t \succ\!\!\succ^{*,e} \bar{x} \text{ and } \bar{x} \succeq_R x^s \text{ or} \\ \text{(ii)} & x^t \succeq^{*,e} \bar{x} \text{ and } \bar{x} \succ_R x^s \end{cases} \end{aligned} \tag{12}$$

Let $\succeq_R^{**,e}$ be the transitive closure of $\succeq_R^{*,e}$. The following is the analogue to complete R-GARP in this context.

Definition 7. Let $e \in [0,1]$ and suppose $R = (\succeq_1, \ldots, \succeq_J)$ where $\succeq_j$ is a preorder on $\mathcal{X}_j$ for all $j \in \mathbb{N}_J$. If $x^t \succeq_R^{**,e} x^s$ implies $x^s \not\succ\!\!\succ_R^{*,e} x^t$ for all $s, t \in \mathbb{N}_T$ then we say that $\mathcal{O}$ satisfies *R-GARP for CCEI level* e. Further, if for all $j \in \mathbb{N}_J$ the preorder $\succeq_j$ is complete on $\mathcal{X}_j$ then we say that $\mathcal{O}$ satisfies *complete R-GARP for CCEI level* e.

As might be expected, this condition is necessary and sufficient to rationalize a dataset for a certain CCEI level e with weakly separable preferences.

Theorem 2. *$\mathcal{O}$ can be rationalized with weakly separable preferences for CCEI level e if and only if $\mathcal{O}$ satisfies Complete R-GARP for CCEI level e.*

To prove this theorem our approach will be to produce a dataset $\mathcal{O}^e$ such that if this dataset is rationalized with weakly separable preferences then the same rationalizing function G will rationalize $\mathcal{O}$ for CCEI level e. This approach is appealing as all of the heavy lifting has already been performed when Theorem 1 was proved.

5.2 Modifying Constraint Sets for CCEI

The most straightforward way of producing $\mathcal{O}^e$ would be to define $D(y) = \{x \in X : x \leqslant y\}$ and replace each constraint set B^t with

$$B^{t,e} = eB^t \cup D(x^t) \tag{13}$$

so that we have a collection $\{(x^1, B^{1,e}), \ldots, (x^T, B^{T,e})\}$ which we may attempt to rationalize. The issue here is that the constraint sets $B^{t,e}$ are *not* regular. The problem is that there is an element $x \in D(x^t)$ such that $x < x^t$ but $x \notin D(x^t)^o$ (although of course $x \in D(x^t)$). The following lemma shows that we can define appropriate constraint sets.

Lemma 4. *Suppose $\mathcal{O}$ is a dataset and $e \in [0,1]$. For any $t \in \mathbb{N}_T$ there exists a regular set $\hat{B}^{t,e}$ such that $B^{t,e} \subseteq \hat{B}^{t,e}$ and for any $\bar{x} \in \bar{\mathcal{X}}$ we have $\bar{x} \in \hat{B}^{t,e}$ if and only if $\bar{x} \in B^{t,e}$.*

We put the proof of this lemma in the appendix. This lemma allows us to prove our CCEI theorem.

Proof of Theorem 2. Suppose $G = F \circ U$ rationalizes $\mathcal{O}$ for CCEI level e where $U(x) = (U_1(x_1), \ldots, U_J(x_J))$. For each $j \in \mathbb{N}_J$ define $\succeq_j$ on $\mathcal{X}_j$ so that $y_j \succeq_j x_j$ if and only if $U_j(y_j) \geqslant U_j(x_j)$. Let $R = (\succeq_1, \ldots, \succeq_J)$. We see that $x^t \succeq_R^{*,e} x^s$ means $G(x^t) \geqslant G(x^s)$ and $x^t \succ_R^{*,e} x^s$ means $G(x^t) > G(x^s)$. So, if $x^t \succeq_R^{**} x^s$ and $x^t \ggcurly_R^{*} x^s$ then we would have $G(x^t) \geqslant G(x^s) > G(x^t)$ which is a contradiction. Thus, $\mathcal{O}$ satisfies complete R-GARP for CCEI level e.

Next, suppose that $\mathcal{O}$ satisfies complete R-GARP for CCEI level e. Lemma 4 allows us to find regular sets $\hat{B}^{t,e}$ for all $t \in \mathbb{N}_T$ such that (i) $B^{t,e} \subseteq \hat{B}^{t,e}$ and (ii) for any $\bar{x} \in \bar{\mathcal{X}}$ we have $\bar{x} \in B^{t,e}$ if and only if $\bar{x} \in \hat{B}^{t,e}$. This will mean that $\bar{x} \in \hat{B}^t$ if and only if $x^t \succeq^{*,e} \bar{x}$ and $\bar{x} \in (\hat{B}^t)^o$ if and only if $x^t \ggcurly^{*,e} \bar{x}$. Define $\mathcal{O}^e = \{(x^1, \hat{B}^1), \ldots, (x^T, \hat{B}^T)\}$ and let $\succeq_R^*$ and $\ggcurly_R^*$ be the R revealed preferences on this dataset. Because we have $x^t \succeq_R^{*,e} x^s$ if and only if $x^t \succeq_R^* x^s$ and $x^t \ggcurly_R^{*,e} x^s$ if and only if $x^t \ggcurly_R^{*,e} x^s$ we see that $\mathcal{O}^e$ satisfies complete R-GARP. Thus, there is a weakly separable utility function G that rationalizes $\mathcal{O}^e$. Clearly this G rationalizes $\mathcal{O}$ for CCEI level e. $\square$

Appendix

A Proof of the Sufficiency Part of the Main Theorem

Our remaining task to prove the sufficiency part of the main theorem is to show how to construct an R-suitable collection of datasets $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$ when $\mathcal{O}$ satisfies complete R-GARP.

A.1 Creating a Suitable Collection of Datasets

We now construct the suitable collection of datasets. Our approach will be to construct regular sets $C_j^{t,s} \subseteq X_j$ for all $j \in \mathbb{N}_J$ and $t, s \in \mathbb{N}_T$ such that the sets

$$C_j^t = \bigcup_{s \in \mathbb{N}_T} C_j^{t,s} \tag{14}$$

will form the desired constraint sets in $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$. We require the sets $C_j^{t,s}$ to satisfy the following properties:

$B1$ (i) $x_j^r \succ_j x_j^t \implies x_j^r \notin C_j^{t,s}$ and (ii) $x_j^r \succeq_j x_j^t \implies x_j^r \notin (C_j^{t,s})^o$.

$B2$ If $x^s \not\trianglerighteq_R^* x^t$ then for any $x \in B^s$ there is a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^{t,s})^o$.

$B3$ If $x^s \not\gg_R^* x^t$ then for any $x \in B^s$ either (i) there is a $j \in \mathbb{N}_J$ such that $x_j \in (C_j^{t,s})^o$ or (ii) for all $j \in \mathbb{N}_J$ we have $x_j \in C_j^{t,s}$.

$B4$ If B^s is co-convex then so is $C_j^{t,s}$.

Notice that whenever we have $s, t \in \mathbb{N}_T$ such that $x^s \gg_R^* x^t$ we may set $C_j^{t,s} = \varnothing$ and items $B1$ - $B4$ will be satisfied. However, C_j^t will never be empty as it will always be the case that $x_j^t \in C_j^t$. To see this note that $x^t \not\gg_R^* x^t$ and so item $B3$ shows that either (i) there is a $j \in \mathbb{N}_J$ such that $x_j^t \in (C_j^{t,t})^o$ or (ii) for all $j \in \mathbb{N}_J$ we have $x_j^t \in C_j^{t,t}$. But, item $B1$ states that $x_j^t \notin (C_j^{t,t})^o$ as we have $x_j^t \succeq_j x_j^t$ and so it must be that for all $j \in \mathbb{N}_J$ we have $x_j^t \in C_j^{t,t}$. As $C_j^{t,t} \subseteq C_j^t$ we have $x_j^t \in C_j^t$. We now show that defining the constraint sets according to (14) allows us to construct an R-suitable collection of datasets $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$.

Lemma 5. *Suppose for all $j \in \mathbb{N}_J$ and $s, t \in \mathbb{N}_T$ we have a set $C_j^{t,s}$ that satisfies items* B*1*-B*4. Then, datasets $(\mathcal{O}_1, \ldots, \mathcal{O}_J)$ with constraint sets defined by* (14) *form an R-suitable collection of datasets.*

Proof. First, let us show that $A1$ holds. This entails showing that $\succeq_j$ extends the revealed preference relations on $\mathcal{O}_j$. This means showing that $x_j^t \succeq_j^* x_j^r$ implies $x_j^t \succeq_j x_j^r$ and $x_j^t \ggcurly_j^* x_j^r$ implies $x_j^t \succ_j x_j^r$. This is the same as showing that $x_j^r \succ_j x_j^t$ implies $x_j^t \not\succeq_j^* x_j^r$ and $x_j^r \succeq_j x_j^t$ implies $x_j^t \not\ggcurly_j^* x_j^r$. If $x_j^r \succ_j x_j^t$ then $B1$ states that for all $s \in \mathbb{N}_T$ we have $x_j^r \notin C_j^{t,s}$ and so $x_j^r \notin C_j^t$ which shows that $x_j^r \not\succeq_j^* x_j^t$. If $x_j^r \succeq_j x_j^t$ then $B1$ states that for all $s \in \mathbb{N}_T$ we have $x_j^r \notin (C_j^{t,s})^o$. If $x_j^r \in (C_j^t)^o$ then there is some $x_j > x_j^r$ such that $x_j \in C_j^t$. But then x_j must be in some $C_j^{t,s}$ and as $C_j^{t,s}$ is regular, $x_j^r \in (C_j^{t,s})^o$ is not the case. Thus, $x_j^r \notin (C_j^t)^o$. So, we have shown that $\succeq_j$ extends the revealed preference relations on $\mathcal{O}_j$.

It is clear that $A2$ follows from item $B2$ and item $A3$ follows from $B3$. Finally, if $B4$ holds then $A4$ holds since the union of co-convex sets is co-convex. $\square$

All that is left to do is to show that we can create sets $C_j^{t,r}$ satisfying $B1$-$B4$. To do this we introduce a new definition and then provide two lemmas.

Definition 8. A set $A \subseteq \mathbb{R}_+^n$ is *co-regular* if A is closed, A^c is bounded, and for $x \in A$ and $y \in \mathbb{R}_+^n$

$$x < y \implies y \in A^o$$

Now, we provide two lemmas which allow us to complete the proof.

Lemma 6. *Let $A \subseteq X$ be co-regular, $D_1 \subseteq X_1, \ldots, D_J \subseteq X_J$ be finite sets, $D = D_1 \times \ldots \times D_J$, and assume $D \subseteq A$. Then there are co-regular sets $H_1 \subseteq X_1, \ldots, H_J \subseteq X_J$ and $H = H_1 \times \ldots \times H_J$ satisfying*

E*1* $D \subseteq H \subseteq A$.

E*2* *If $x_j \in D_j$ and $\{x_j\} \times D_{-j} \subseteq A^o$ then $x_j \in H_j^o$.*

Further, if A is convex then H can be chosen to be convex.

We shall postpone the proof of this lemma and instead show how to create the desired sets $C_j^{t,s}$ using the lemma. First we develop some facts about regular sets.

A.2 Regular Sets

Lemma 7. *Suppose that $B \subseteq \mathbb{R}_+^n$ is regular (co-regular). Then*

R*1 :* $\overline{B^c}$ *is co-regular (regular).*

R*2 :* $B^o = \left(\overline{B^c}\right)^c$

R*3 :* $B^c = \left(\overline{B^c}\right)^o$

Proof. It is well known that for *any* set $F \subseteq \mathbb{R}_+^n$

$$\overline{F^c} = (F^o)^c \tag{15}$$

This shows that $(\overline{B^c})^c = B^o$ so $R2$ holds. Suppose B is regular. If B is empty the results hold trivially so let us assume that B is not empty. Using the fact that $\overline{B^o}$ and (15) we see that $\overline{\left[\overline{B^c}\right]^c} = \overline{\left[(B^o)^c\right]^c} = \overline{B^o} = B$. Using this result and (15) we see that $B = \overline{\left[\overline{B^c}\right]^c} = \left[\left(\overline{B^c}\right)^o\right]^c$. Taking complements of both sides of this equation shows that $R3$ holds. Suppose that $x \in \overline{B^c}$ and let y satisfy $y > x$. If $y \in B$ then as B is regular $x \in B^o$. But then $R2$ shows that $x \notin \overline{B^c}$ which is a contradiction. Thus, $y \in B^c$ and $R3$ shows that $y \in (\overline{B^c})^o$. Clearly, $(\overline{B^c})^c$ is bounded and so we conclude that $\overline{B^c}$ is co-regular. The proof for the case when B is co-regular can proceed in the same fashion so we are done. $\square$

We shall refer to the results of this lemma by their item number.

A.3 Returning to the Proof of the Sufficiency Part of the Main Theorem

We now show that if Lemma 6 is true then we can complete the proof.

Lemma 8. *If $\mathcal{O}$ satisfies complete R-GARP then we may construct sets $C_j^{t,s}$ for each $j \in \mathbb{N}_J$ and $t, s \in \mathbb{N}_T$ such that items* B*1*-B*4 hold.*

Proof. If $x^s \succ\!\!\succ_R^* x^t$ then we have already commented that setting $C_j^{t,s} = \varnothing$ for all $j \in \mathbb{N}_J$ will satisfy $B1$-$B4$. So, let us assume that $x^s \not\succ\!\!\succ_R^* x^t$. For all $j \in \mathbb{N}_J$ let $D_j = \{x_j \in \mathcal{X}_j : x_j \succeq_j x_j^t\}$ and $D = D_1 \times \ldots \times D_J$. Notice that $x^s \not\succ\!\!\succ_R^* x^t$ implies that $(B^s)^o \cap D = \varnothing$. Let $\alpha = \sup_{a \in \mathbb{R}} \left\{\left[(1+a)B^s\right] \cap D = \varnothing\right\}$. Notice that $\alpha = 0$ if and only if $B^s \cap D \neq \varnothing$. Let $B = (1 + 0.5\alpha)B^s$. Clearly B is regular and $B^o \cap D = \varnothing$. Let $A = \overline{B^c}$. Item $R1$ shows that A is co-regular. Item $R2$ shows that $B^o = A^c$ and so $A^c \cap D = \varnothing$ which means $D \subseteq A$. So, we may apply Lemma 6 to obtain co-regular sets $H_1, \ldots, H_J$ and $H = H_1 \times \ldots \times H_J$ satisfying $E1$ and $E2$. For each $j \in \mathbb{N}_J$ let $C_j^{t,s} = \overline{(H_j)^c}$. As H_j is co-regular item $R1$ shows that $C_j^{t,s}$ is regular.

To show $B1$ holds suppose that $x_j^r \succ\!\!\succ_j x_j^t$. This means $x_j^r \in D_j$. If there is an $\bar{x}_{-j} \in D_{-j}$ such that $(x_j^r, \bar{x}_{-j}) \in B$ then our earlier remark shows that $\alpha = 0$ and thus $(x_j^r, \bar{x}_{-j}) \in B^s = B$ which implies $x^s \succ\!\!\succ_R^* x^t$. However, we have assumed that $x^s \not\succ\!\!\succ_R^* x^t$ and thus $\{x_j^r\} \times D_{-j} \cap B = \varnothing$. Item $R3$ shows that $B^c = A^o$ and so $\{x_j^r\} \times D_{-j} \subseteq A^o$ and now item $E2$ states that $x_j^r \in H_j^o$ and so item $R2$ states that $x_j^r \notin C_j^{t,s}$ which is the desired result. Next, suppose that $x_j^r \succeq_j x_j^t$. This means $x_j^r \in D_j$. As $D \subseteq H$ we see that $D_j \subseteq H_j$ and so $x_j^r \in H_j$. Now, item $R3$ states that $H_j^c = (C_j^{t,s})^o$ and thus $x_j^r \notin (C_j^{t,s})^o$ which is what we endeavoured to show.

To establish $B2$ suppose that $x^s \not\succeq_R^* x^t$. Our previous remark about α demonstrates that $\alpha > 0$. We establish the proposition that for any $x \in X$ such that $x_j \notin (C_j^{t,s})^o$ for all $j \in \mathcal{J}$ then $x \notin B^s$ which is equivalent to $B2$. So, suppose we have some $x \in X$ such that $x_j \notin (C_j^{t,s})^o$ for all $j \in \mathbb{N}_J$. Item $R3$ shows that $(C_j^{t,s})^o = H_j^c$ and so $x_j \in H_j$ for all $j \in \mathbb{N}_J$. Thus, $x \in H$. Item $R2$ and the fact that $H \subseteq A$ show that $x \notin B^o$. But, as $\alpha > 0$ we see that $B^s \subseteq B^o$ and so $x \notin B^s$ which establishes $B2$.

To establish $B3$ suppose that $x^s \not\succ_R^* x^t$. Our previous remark about α demonstrates that $\alpha = 0$. We establish the proposition that if $x \in X$ such that for all $j \in \mathbb{N}_J$ we have $x_j \notin (C_j^{t,s})^o$ and for some $k \in \mathbb{N}_J$ we have $x_k \notin C_k^{t,s}$ then $x \notin B^s$ which is equivalent to $B3$. So, suppose we have such an $x \in X$. As in the proof that $B2$ holds we can show that $x \in H$. Further, as $x_k \notin C_k^{t,s}$ item $R2$ establishes that $x_k \in H_k^o$. As $x_k \in H_k^o$ we may choose a $y_k < x_k$ such that $y_k \in H_k$. As $H \subseteq A$ we see that $(y_k, x_{-k}) \in A$. As A is co-regular and $(y_k, x_{-k}) < x$ we see that $x \in A^o$. Now, item $R3$ states that $x \notin B$ and as $\alpha = 0$ we know that $B^s = B$ and so we have established the result.

Next, suppose that B^s is co-convex. Clearly B will also be co-convex. This means that A will be convex. Lemma 6 allows us to choose an H that is convex. Clearly this means that $H_1, \ldots, H_J$ are convex. Now, each set $C_j^{t,s} = \overline{H_j^c}$ will be co-convex which establishes $B4$. □

We still have to prove Lemma 6.

A.4 Proof of Lemma 6

For a set $B \subseteq X$ and $F_{-j} \subseteq X_{-j}$ let define a set $M_j(B; F_{-j})$ by $M_j(B; F_{-j}) = \{x_j \in X_j : \forall y_{-j} \in F_{-j} \text{ we have } (x_j, y_{-j}) \in B\}$. This set is the unique largest set (ordered by set inclusion) $Z_j \subseteq X_j$ such that $Z_j \times F_{-j} \subseteq B$. It is straightforward to verify that

$$Z_j \subseteq M_j(B; F_{-j}) \text{ if and only if } Z_j \times F_{-j} \subseteq B \tag{16}$$

Lemma 9. *If A is co-regular and F_{-j} is closed then $A_j = M_j(A, F_{-j})$ is co-regular. Also, $A_j^o = M_j(A^o, F_{-j})$.*

Proof. We claim that if $z_j \in X_j$ and $\{z_j\} \times F_{-j} \subseteq A^o$ then there is an open set $E_j \subseteq X_j$ such that

$$z_j \in E_j \quad \text{and} \quad E_j \times F_{-j} \subseteq A^o \tag{17}$$

Define a sequence $(\rho^m) = (1 - 1/m)$. Suppose for a contradiction that for each $m \in \mathbb{N}$ there is some $f_{-j}^m \in F_{-j}$ such that $(\rho^m \cdot z_j, f_{-j}^m) \in (A^o)^c$. From Lemma 7 we know that $(A^o)^c = \overline{A^c}$ and is regular and thus compact. So there must be some sub-sequence (m_k) such that $(\rho^{m_k} \cdot z_j, f_{-j}^{m_k})$ converges to some element $y = (y_j, y_{-j}) \in (A^o)^c$. From how the sequence was defined it is clear that $y_j = z_j$. As F_{-j} is closed we see that $y_{-j} \in F_{-j}$.

But, this means that $(y_j, y_{-j}) \in A^o$ which contradicts $(y_j, y_{-j}) \in (A^o)^c$. So, there is some $m \in \mathbb{N}$ such that $(\rho^m \cdot z_j, f_{-j}) \in A^o$ for all $f_{-j} \in F_{-j}$. This means $\{\rho^m \cdot z_j\} \times F_{-j} \in A^o$. Let E_j be the interior of the set $\{x_j \in X_j : x_j \geqslant \rho \cdot z_j\}$. Notice that each element in $\rho \cdot z_j$ is either strictly below z_j or is 0. So $z_j \in E_j$ and so E_j satisfies (17).

Clearly A_j^c is bounded and so $\overline{A_j^c}$ is compact and A_j^c is not empty. Let $y_j \in A_j$ and let $z_j > y_j$. As $y_j \in A_j$ we see that $\{y_j\} \times F_{-j} \in A$ and because A is co-regular we see that $\{z_j\} \times F_{-j} \subseteq A^o$. So, we may select an open set E_j that satisfies (17). As $E_j \times F_{-j} \subseteq A$ we have $E_j \subseteq A_j$ and so $z_j \in A_j^o$. This shows that A_j is co-regular.

Suppose that $z_j \in M_j(A^o; F_{-j})$. This means that $\{z_j\} \times F_{-j} \subseteq A^o$. So we may select an open set E_j to satisfy (17). We now obtain $z_j \in E_j \subseteq A_j$ and so $z_j \in A_j^o$. So, we see that $M_j(A^o; F_{-j}) \subseteq A_j^o$. Now, suppose that $z_j' \in A_j^o$. Select an element $z_j \in A_j$ such that $z_j < z_j'$. As $z_j \in A_j$ we have $\{z_j\} \times F_{-j} \subseteq A$. As A is co-regular we see that $\{z_j'\} \times F_{-j} \subseteq A^o$. This means that $z_j' \in M_j(A^o; F_{-j})$. We see that $A_j^o \subseteq M_j(A^o; F_{-j})$ and so $A_j^o = M_j(A^o; F_{-j})$. This completes the proof. □

Suppose $B \subseteq \mathbb{R}_+^n$ is regular. Define a function $g_B : \mathbb{R}_+^n \to \mathbb{R}_+$ by

$$g_B(x) = \inf\{k \in (0, \infty) : x \in kB\} \tag{18}$$

We refer to a function defined by (18) as a *gauge function*. From the regularity of B it is immediate that

$$x \in B^o \iff g_B(x) < 1, \qquad x \in \partial B \iff g_B(x) = 1, \qquad x \notin B \iff g_B(x) > 1$$

Notice that a linear constraint set $B = B(p, w)$ has a linear gauge function where $g_B(x) = (p \cdot x)/w$. If A is co-regular we define its gauge function to be the gauge function for $\overline{A^c}$. We now define a function which allows a handy deformation of a space X into the border of a co-regular set A.

Definition 9. For a co-regular set $A \subseteq X$ with gauge function $g_A(\cdot)$ define its *regular deformation* $\Gamma_A : X \times [0, 1] \to X$ by $\Gamma_A(x, \rho) = x/[g_A(x)^\rho]$

It is straightforward to verify that for any $x \in A$ the function $\Gamma_A(x, \cdot)$ is a continuous decreasing function and

$$\Gamma_A(x, 0) = x \qquad \text{and} \qquad \Gamma_A(x, 1) \in \partial A \tag{19}$$

Further, if $x \in \partial A$ then $\Gamma_A(x, \rho) = x$ for all $\rho \in [0, 1]$ and if $x \in A^o$ then $\Gamma_A(x, \cdot)$ is strictly decreasing. We may understand $\Gamma_A(x, \rho)$ as an operator that pulls x closer to the boundary of A as ρ increases.

We are now ready to prove Lemma 6.

Proof of Lemma 6. Let $A_j = M_j(A; D_{-j})$. Lemma 9 shows that A_j is co-regular. Let $\Gamma_j : X_j \times [0,1] \to X_j$ be the regular deformation of A_j. Define a function $\Gamma : X \times [0,1] \to X$ by $\Gamma(x,\rho) = (\Gamma_1(x_1,\rho), \ldots, \Gamma_J(x_J,\rho))$. For each $d \in D$ such that $d \in A^o$ select a $\rho_d \in (0,1)$ such that $\Gamma(d;\rho_d) \in A^o$. That way may do this follows from $\Gamma(d;\cdot)$ being continuous and satisfying $\Gamma(d;0) = d$. Let ρ be the minimum of all the ρ_d. This minimum exists as D is finite. By our remark following the definition of a regular deformation we see that for all $d \in D$ such that $d \in \partial A$ we have $\Gamma(d;\rho) = d$. So, we have found a $\rho > 0$ such that $\Gamma(d,\rho) \in A$ for all $d \in D$. Set $F = \Gamma(D,\rho)$ and let $F_1, \ldots, F_J$ satisfy $F = F_1 \times \ldots \times F_J$. Let $H_1 = M_1(A; F_{-1})$ and $\hat{H}_{-1} = F_{-1}$. Define H_j and $\hat{H}_{-j}$ recursively by

$$H_j = M_j(A; \hat{H}_{-j}) \qquad \text{and} \qquad \hat{H}_{-j} = H_1 \times \ldots \times H_{j-1} \times F_{j+1} \times \ldots \times F_J \tag{20}$$

Lemma 9 shows H_j is co-regular. Let $H = H_1 \times \ldots \times H_J$. Notice that by definition $\hat{H}_{-J} = H_1 \times H_2 \times \ldots \times H_{J-1}$ and so $H = H_J \times \hat{H}_{-J}$. By (16) and the definition of $\hat{H}_{-j}$ we have $H = H_J \times \hat{H}_{-J} \subseteq A$. Next, we show that for all $j \in \mathcal{J}$ we have $F_j \subseteq H_j$. To do this, we show that (i) $F_1 \subseteq H_1$ and (ii) for all $j \neq J$ we have $F_j \subseteq H_j \implies F_{j+1} \subseteq H_{j+1}$. To see (i) note that $F_1 \times \hat{H}_{-1} = F \subseteq A$ and so (16) shows $F_1 \subseteq \hat{H}_1$. Next, suppose that $F_j \subseteq H_j$. This means $F_j \times \hat{H}_{-j} \subseteq H_j \times \hat{H}_{-j} \subseteq A$. Now, using $H_j \times \hat{H}_{-j} = F_{j+1} \times \hat{H}_{-(j+1)} \subseteq A$ and (16) we see that $F_{j+1} \subseteq H_{j+1}$ and so (ii) holds. So, using (i) and (ii) we see that $F_j \subseteq H_j$ for all $j \in \mathcal{J}$. Let $d_j \in D_j$. Because $\Gamma_j(d_j,\cdot)$ is decreasing, $\Gamma(d_j,\rho) \in F_j \subseteq H_j$, and H_j is co-regular we have $d_j \in H_j$. We see that $D_j \subseteq H_j$ which completes the proof of $E1$. Next, suppose that $d_j \in D_j$ satisfies $\{d_j\} \times D_{-j} \subseteq A^o$. Equation (16) shows that $d_j \in M_j(A^o; F_{-j})$ and so Lemma 9 shows that $d_j \in A_j^o$. This means that $\Gamma_j(d_j,\cdot)$ is a strictly decreasing. Because $\Gamma_j(d_j,\rho) \in F_j \subseteq H_j$ and H_j is co-regular we see that $d_j \in H_j^o$. This shows that $E2$ holds.

Now, suppose that A is convex. Let $x_j, x_j' \in H_j$ and let x_j'' be a convex combination of x_j and x_j'. From the definition H_j and equation (16) we see that $\{x_j\} \times \hat{H}_{-j} \subseteq A$ and $\{x_j'\} \times \hat{H}_{-j} \subseteq A$. As A is convex we see that $\{x_j''\} \times \hat{H}_{-j} \subseteq A$ which means $x_j'' \in H_j$. Thus, H_j is convex. □

B Proof of Lemma 4

We show how to construct the set $\hat{B}^t$.

Proof of Lemma 4. Our approach is to deform the set $\{x \in X : x \leqslant x^t\}$ so that it becomes regular. Let $\iota \in \mathbb{R}^L_{++}$ be the vector with a 1 for each element. For $y \in \mathbb{R}^L$ let $H(y) = \{x \in \mathbb{R}^L : (x - y) \cdot \iota = 0\}$ and let $\rho(\cdot, y) : \mathbb{R}^L \to H$ be the orthogonal projection onto $H(y)$. Now, let $\alpha : \mathbb{R}^L \times \mathbb{R}^L \times [0, 1] \to \mathbb{R}^L$ be defined by

$$\alpha(x, y, \delta) = (1 - \delta)x + \delta\rho(x, y)$$

Clearly, $\alpha(x, y, 0) = x$ and $\alpha(x, y, 1) \in H(y)$. We may think of $\alpha(x, y, \cdot)$ as a function which sends elements in $\mathbb{R}^L$ to $H(y)$ as δ goes from 0 to 1. Let $d(y) = \{x \in \mathbb{R}^L : x \leqslant y\}$ and let $d(y, \delta) = \{x \in \mathbb{R}^L : \exists x' \in d(y) \text{ s.t. } x = \rho(x', y, \delta)\}$.

We claim that $d(y, \delta)$ is regular for any $\delta \in (0, 1)$. To see this suppose $x \in d(y, \delta)$ and $x' < x$. From the definition of α we see that $x' \in d(y, \delta)$. It is straightforward to confirm that for any two points $z', z \in \partial d(y, \delta)$ we have $z \not> z'$. This means that $x' \in d(y, \delta)$. The rest of the properties required for $d(y, \delta) \cap X$ to be regular are easy to prove.

Set $\delta > 0$ small enough to ensure that

$$\bar{x} \notin d(x^t, \delta) \iff \bar{x} \notin d(x^t, 0), \qquad \forall \bar{x} \in \bar{\mathcal{X}}$$

Define $\hat{B}^t = eB^t \cup d(x^t, \delta)$. This set is clearly regular (it is the union of two regular sets) and δ was chosen so that we have $\bar{x} \in eB^t \cup D(x^t)$ if and only if $\bar{x} \in \hat{B}^t$. $\square$

References

Sydney Afriat. The construction of a utility function from expenditure data. *International Economic Review*, 8, 1967.

Sydney Afriat. Efficiency estimation of production functions. *International Economic Review*, 13:568–598, 1972.

James Andreoni and John Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Revealing preferences graphically: An old method gets a new tool kit. *American Economic Review*, 97(2):153–158, 2007a.

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5):1921–1938, 2007b.

Syngjoo Choi, Shachar Kariv, Muller Wieland, and Dan Silverman. Revealing preferences graphically: An old method gets a new tool kit. *American Economic Review*, 104(6):1518–1550, 2014.

W. Erwin Diewert and Celik Parkan. Tests for the consistency of consumer data. *Journal of Econometrics*, 30:127–147, 1985.

Francoise Forges and Enrico Minelli. Afriat's theorem for general budget sets. *Journal of Economic Theory*, 144:135–145, 2009.

Rosa L. Matzkin. Axioms of revealed preference for nonlinear choice sets. *Econometrica*, 59(6):1779–1786, 1991.

John K.-H. Quah. A test for weakly separable preferences. *University of Oxford, Department of Economics, Discussion Paper*, May 2014.

Hal R. Varian. The non-parametric approach to demand analysis. *Econometrica*, 50(4): 945–973, July 1982.

Hal R. Varian. Non-parametric tests of consumer behaviour. *The Review of Economic Studies*, 50(1):99–110, January 1983.

Hal R. Varian. Goodness-of-fit for revealed preference tests. *manuscript*, 1993.