Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 484–507
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

# On Bayesian analysis and computation for functions with monotonicity and curvature restrictions

William J. McCausland[a,b,c,*]

[a]*Département de sciences économiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, QC, Canada H3C 3J7*
[b]*Centre Interuniversitaire de Recherche en Économie Quantitative, Canada*
[c]*Centre Interuniversitaire de Recherche en Analyse des Organisations, Canada*

Available online 14 September 2007

**Abstract**

Our goal is inference for shape-restricted functions. Our functional form consists of finite linear combinations of basis functions. Prior elicitation is difficult due to the irregular shape of the parameter space. We show how to elicit priors that are *flexible*, *theoretically consistent*, and proper. We demonstrate that uniform priors over coefficients imply priors over economically relevant quantities that are quite informative and give an example of a non-uniform prior that addresses this issue. We introduce simulation methods that meet challenges posed by the shape of the parameter space. We analyze data from a consumer demand experiment.


*JEL classification:* C11; C14; C15; D12; D21

*Keywords:* Bayesian methods; Flexible functional forms; Shape restrictions; Consumer demand

## 1. Introduction

There are many examples in economics where theoretically consistent choice behavior is described by multivariate functions subject to monotonicity and curvature restrictions. These functions include utility, expenditure, indirect utility, production, cost, and profit functions.

Much empirical analysis in economics involves learning about these functions using data on the choices of consumers and firms. There is a large literature on such inference. See Deaton and Muellbauer (1980), Diewert and Wales (1987), Lau (1986), Matzkin (1994) and Terrell (1996).

Analysis typically begins with two choices: a parametric class of functions, and constraints on the parameter vector. The constraints define a restricted parameter set.

The literature identifies two important objectives governing these choices, theoretical consistency and flexibility. To a large extent, they are competing. Theoretical consistency refers to the extent to which the

*Département de sciences économiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, QC, Canada H3C 3J7. Tel.: +1 514 343 7281; fax: +1 514 343 7221.
*E-mail address:* william.j.mccausland@umontreal.ca

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.08.018

functions indexed by elements of the restricted parameter set satisfy the applicable monotonicity and curvature restrictions over their domain. If the functions satisfy the restrictions throughout the domain, we have *global theoretical consistency*. If they satisfy them at a point or in a region, we have *local theoretical consistency* or *regional theoretical consistency*, respectively. Flexibility refers to the variety of functions indexed by elements of the restricted parameter set, and it too may be more or less global, depending on how large is the subset of the domain where the relevant flexibility properties hold.

Consider, for example, two commonly used classes of indirect utility functions: the constant elasticity of substitution (CES) class, and the trans-log class. According to standard theory, indirect utility functions are non-increasing and quasi-convex in income-normalized prices. The CES class of indirect utility functions, with non-negativity restrictions on its parameters, is globally theoretically consistent. However, it is quite inflexible, in the sense that elasticities of substitution cannot vary with prices and income. The trans-log class (see Christensen et al. (1975)) is locally flexible in the sense that with appropriate choices of the parameters one can attain arbitrary elasticities at a given point. However, it is not globally theoretically consistent: there are values of the parameters for which the function is not everywhere on its domain non-increasing and quasi-convex. We cannot rule out these values without renouncing local flexibility.

There are at least three distinct classes of functions whose flexibility allows the simultaneous approximation of a continuous function, and any continuous derivatives it may have, on a compact subset $\bar{X}$ of its theoretical domain $X$. We call $\bar{X}$ the *restricted domain* and note that it can be chosen to include the empirically relevant region. The three classes consist of linear combinations of basis functions.

The simultaneous approximation of a function and its derivatives is important for two reasons. First, it is desirable to approximate the behavior that a function represents, and theoretically consistent choices are often given in terms of the function's derivatives. Roy's identity, for example, gives choices as functions of derivatives of the indirect utility function. The proximity of two functions in, for example, the sup norm does not guarantee the proximity of their derivatives: the difference of the two functions may have low amplitude but high frequency ripples. A second reason is that by simultaneously approximating derivatives, we can guarantee that the approximating function satisfies the applicable monotonicity and curvature restrictions, which we can express in terms of derivatives.

Gallant (1981) launches this literature with his Fourier flexible form. Basis functions are sinusoidal, and any continuous function on $\bar{X}$ can be approximated arbitrarily closely in sup norm by a linear combination of a finite number of these basis functions. If the function has bounded derivatives up to some order, we can simultaneously approximate the function and these derivatives in sup norm.

Unfortunately, sinusoidal functions do not satisfy typical monotonicity and curvature restrictions and so it can take many terms to build up an approximation. In the context of approximating indirect utility functions, Gallant (1981) proposes adding linear and quadratic terms.

Barnett and Jonas (1983) use a multivariate Müntz–Szasz expansion to approximate a firm's unit cost function, a function of the prices $p_1,\ldots,p_n$ of $n$ input factors. The set of basis functions is

$$\left\{\prod_{i=1}^{n} p_i^{\lambda(\imath_i)} : \imath \in \mathbb{N}_0^n\right\},$$

where $\mathbb{N}_0 \equiv 0,1,2,\ldots$ and the sequence $\lambda(k)$, $k=1,2,\ldots$, satisfies $\sum_{k=1}^{\infty}(1/\lambda(k))=\infty$. Barnett and Jonas (1983) and Barnett et al. (1991a, b) take $\lambda(k)=2^{-k}$. Barnett and Yue (1988) give conditions for various modes of convergence of the function and its derivatives. An advantage of this approach is that all basis functions satisfy the appropriate monotonicity and curvature restrictions for unit cost functions: they are non-decreasing and concave.

In unpublished work, Geweke and Petrella (2000) also approximate a firm's unit cost as a function of input prices $p_1,\ldots,p_n$, but use the following set of basis functions:

$$\left\{\prod_{i=1}^{n} p_i^{b\imath_i} : \imath \in \mathbb{N}_0^n\right\}, \tag{1}$$

where $b>0$. The functions $\prod_{i=1}^{n} p_i^{b\imath_i}$ satisfying $\imath_i < b^{-1}$, $i=1,\ldots,n$, are themselves non-decreasing and concave, which is convenient for constructing approximations of non-decreasing concave unit cost functions using a

small number of terms. If the function to approximate has continuous derivatives, we can simultaneously approximate the function and these derivatives. The approximation result is based on a mathematical result on the simultaneous approximation of a function and its derivatives using polynomials due to Evard and Jafari (1994).

For the remainder of the paper, we restrict attention to a particular set of monotonicity and curvature restrictions: we define a function as *regular* if it is non-decreasing and concave. In Section 2, we will discuss the applicability of our results to other cases.

Once we have an approximation of regular functions as linear combinations of basis functions, we can do inference for functions by doing inference for the number of basis functions to include and their coefficients. We adopt a Bayesian approach, which has many advantages in this context. Inequality restrictions, which figure prominently, are more easily handled using a Bayesian, rather than frequentist analysis. Inference on parameter values, numbers of terms and even competing functional forms[1] is done in a unified approach using methods with known and desirable properties.

Bayesian predictive inference automatically takes into account *a posteriori* uncertainty about the regular function. Within a set of functions satisfying the appropriate monotonicity and curvature restrictions and compatible with observed data, individual functions may differ considerably on regions outside the sample. Averaging over these functions by taking the posterior expectation may lead to better out-of-sample prediction than using only one function, however, great its "likelihood".

However, prior elicitation is difficult. The parameter space is infinite-dimensional and irregularly shaped. Parameters have no obvious direct economic interpretation. It is difficult to think about distributions over functions.

The main contributions of the paper are the following. We describe classes of basis functions that allow one to construct a wide variety of regular functions with a small number of terms. We prove a result, similar to that of Geweke and Petrella (2000), on the simultaneous approximation of a function and its derivatives. We illustrate using simulations that for a fixed finite number of basis functions, a uniform prior on the set of coefficients associated with regular functions implies quite informative prior distributions for certain economically relevant quantities. The unsuspecting practitioner may be assigning very low probability to regions of the parameter space that are quite plausible. We give an example of a proper, non-uniform prior over coefficients that implies a more diffuse prior distribution for these economically relevant quantities. We propose methods for prior and posterior simulation that meet the challenge posed by the shape of the parameter space and made more difficult when we use the kind of non-uniform prior that we suggest in this paper.

Section 2 concerns the approximation of functions by linear combinations of a finite number of basis functions. We introduce a wide variety of sets of basis functions nesting the set of basis functions used by Geweke and Petrella (2000). Following them, we apply an approximation result by Evard and Jafari (1994). We show that using any of our sets of basis functions, any twice continuously differentiable regular function can be arbitrarily well approximated on a compact subset $\bar{X}$ of its theoretical domain $X$ by a function that is regular on $\bar{X}$. The approximation is a simultaneous approximation (in sup norm) of the function, its gradient and its Hessian.

In Section 3, we discuss prior distributions over regular functions. We define notions of flexibility and theoretical consistency for these prior distributions, and give sufficient conditions for both. We describe an approach to prior elicitation that emphasizes economically relevant quantities rather than the parameters themselves, which are difficult to interpret. We discuss the non-trivial issue of ensuring that the prior is proper, which is essential for model comparison using Bayes factors.

We simulate from a uniform prior distribution to illustrate what this prior implies about the distribution of economically relevant features of the regular function. Interpreting the regular function as a utility function, we consider the marginal rate of substitution between two goods at a typical point in the consumption set. Interpreting the function as a negative indirect utility function, we consider ratios of quantities demanded at a typical point in the space of income-normalized prices. We show that the implied prior distribution of these features of the regular function is quite informative, and argue that the uniform prior assigns low probability

[1] See Gordon (1996) on Bayesian comparison of competing functional forms.

to sets of functions that many would find plausible. We give an example of a non-uniform prior and show that it implies more diffuse prior distributions over marginal rates of substitution and ratios of quantities demanded.

In Section 4, we discuss the problem of prior and posterior simulation of the unknown coefficients. The shape of the parameter space presents special problems, and these are exacerbated when we adopt the kind of non-uniform prior we propose in Section 3. To overcome these problems we propose Markov chain Monte Carlo (MCMC) simulation methods based on three Metropolis–Hastings (see Metropolis et al., 1953 and Hastings, 1970) proposal distributions adapted to the shape of the parameter space.

In Section 5, we present an empirical application of our econometric methods. It is common in empirical applications to adopt a measurement error approach to construct a data density, whereby an error distribution accounts for discrepancies between observed choices and the optimal choices given by the regular function. Instead, we use a stochastic model for observed choices where distributions over choices are determined by the regular function itself. Theil (1974) and McCausland (2004) give very different theoretical underpinnings for the model. We analyze individual choice data from a consumer experiment described in Harbaugh et al. (2001).

We conclude in Section 6.

## 2. Parametric classes of functions

Our objective is Bayesian learning, from choice data, about a regular function representing choice behavior. For Bayesian analysis, we must be able to express prior and posterior uncertainty about these functions in terms of probability distributions. We make use of a parametric class of functions, which permits us to express this uncertainty in terms of distributions over parameters. Functions in our parametric class are linear combinations of basis functions. Unknown parameters are the coefficients of the basis functions and their number. The regularity of functions defines a restricted parameter set.

We defined, for definiteness, a particular kind of regularity, but we will see that this does not restrict the applicability of our results and methods as much as it may first appear.

We do not know of any results that allow us to approximate any regular function on an unbounded set such as the classical consumption set $X = \mathbb{R}^n_+$. We settle for approximation on a compact restricted domain $\bar{X} \subseteq X$, which we can always choose to include the empirically relevant region for any given application.

We also restrict functions to be twice continuously differentiable. This ensures that we can simultaneously approximate the function, its gradient and its Hessian.

We want to be able to generate a reasonable variety of regular functions using a small number of terms. To this end, we introduce a transformation $\phi$ on $X$. Rather than directly approximate a function $u$ on $X$, we will approximate $u \circ \phi^{-1}$ on $\phi(\bar{X})$. Choosing a $\phi$ that is regular itself makes it easier to approximate regular functions, since first order monomials of $\phi$ are then regular.

### 2.1. Regularity

Most of the results of this paper pertain to non-decreasing and concave functions, which we call regular. With simple and obvious modifications, we can replace non-decreasing with non-increasing or concave with convex.

Replacing concavity with quasi-concavity (for application to utility functions) or quasi-convexity (indirect utility functions) is a little more problematic. This is because we take advantage of the fact that the set of concave functions is closed under convex combination. Quasi-concave and quasi-convex functions do not have this property.

However, we argue that imposing concavity on utility functions is innocuous for our purposes. We grant that it is restrictive: while all complete, continuous, monotone and convex preferences have quasi-concave utility representations, not all have concave utility representations: see deFinetti (1949) or the document "A Pedagogical Example of Non-concavifiable Preferences" by James Schummer, available at the author's website. Kannai (1974) shows, however, that on any non-empty compact set, a complete, continuous,

monotone and convex preference can be arbitrarily well approximated (in a sense made precise in Lemma 2 of that article) by a complete, continuous, monotone and convex preference that has a concave utility representation. Given that we are already restricting the domain of utility functions to a compact set $\bar{X}$ to obtain an approximation result, concavity is not a severe additional restriction.

Interpreting the indirect utility function as the representation of preferences over income-normalized prices, we can make a similar argument about imposing convexity on indirect utility functions.

### 2.2. Restricted domains

The choice of a restricted domain $\bar{X} \subseteq X$ is part of the prior specification, so it should not depend on observed choices. If the decision maker is a price-taker, then prices and income are exogenous and $\bar{X}$ can depend on observed prices and income. We choose $\bar{X}$ to include the empirically relevant region. One natural choice for $\bar{X}$ is a hyper-rectangle of the form $[0, \bar{x}_1] \times \cdots \times [0, \bar{x}_n]$, for some $\bar{x} \in \mathbb{R}^n_{++}$. Another is a simplex of the form $\{x \in X : w'x \leqslant m\}$, for some $w \in \mathbb{R}^n_{++}$ and $m \in \mathbb{R}_{++}$.

Since we impose regularity on the approximating function only on $\bar{X}$, enlarging $\bar{X}$ serves the goal of theoretical consistency. However, one trades off flexibility: as $\bar{X}$ grows, the set of parameters for which the function is regular on $\bar{X}$ shrinks. Our approximation result holds for any compact $\bar{X}$, however large, but we lose flexibility for any fixed number of terms.

### 2.3. A parametric functional form

We first introduce some definitions that help simplify notation for monomials and polynomials on $\mathbb{R}^n$. A *multi-index* of length $n$ is an $\iota \in \mathbb{N}_0^n$, where $\mathbb{N}_0$ is the set of non-negative integers. For vectors $x \in \mathbb{R}^n$ and multi-indices $\iota \in \mathbb{N}_0^n$, we define the notation $x^\iota \equiv \prod_{i=1}^n x_i^{\iota_i}$. We will call a finite subset $I \subset \mathbb{N}_0^n$ a *constellation* of multi-indices. Thus for any multi-index $\iota$, $x^\iota$ is a monomial and for any constellation $I$ and vector $(\lambda_\iota)_{\iota \in I}$ of coefficients, $\sum_{\iota \in I} \lambda_\iota x^\iota$ is a polynomial.

We want to be able to construct a wide variety of regular functions using a small number of terms. To do this, we introduce a transformation $\phi : X \to \mathbb{R}^n$ of the domain and choose as our basis functions the set of monomials $\{[\phi(x)]^\iota : \iota \in \mathbb{N}_0^n\}$. In effect, we replace the problem of approximating the regular function $u$ on $\bar{X}$ with that of approximating $u \circ \phi^{-1}$ on $\phi(\bar{X})$.

In the simulations and empirical example of this paper, we use the following choice of $\phi$:

$$\phi(x) = \left( \log\left(\frac{x_1 + \xi_1}{x_1^* + \xi_1}\right), \ldots, \log\left(\frac{x_n + \xi_n}{x_n^* + \xi_n}\right) \right) \quad \forall x \in X, \tag{2}$$

where $\xi \in \mathbb{R}^n_{++}$ and $x^* \in \bar{X}$ are fixed constants that the econometrician chooses in advance for computational convenience.

The vector $\xi$ must be strictly positive, to ensure that the region $\phi(\bar{X})$ on which we approximate $u \circ \phi^{-1}$ is compact. By choosing small elements of $\xi$, we obtain flexibility at small scales, but we also enlarge $\phi(\bar{X})$, which adversely affects flexibility at other scales. The constant $x^*$ establishes the origin of the transformed choice space $\phi(\bar{X})$, since $\phi(x^*) = 0$. We find in practice that we obtain better flexibility for a fixed number of terms by choosing $\bar{X}$, $\xi$ and $x^*$ such that $\phi(x^*)$ is near the centroid of $\phi(\bar{X})$.

For the particular choice of $\phi(x)$ in Eq. (2), the set of monomials in $\phi(x)$ has some useful properties. It includes the regular functions $\log((x_i + \xi_i)/(x_i^* + \xi_i))$, $i = 1, \ldots, n$. For small values of $\xi_i$, linear combinations $\sum_{i=1}^n \lambda_i \log((x_i + \xi_i)/(x_i^* + \xi_i))$ of these regular functions approximate Cobb–Douglas utility functions with arbitrary expenditure shares $\lambda_i / \sum_{j=1}^n \lambda_j$. When we add higher order monomials in $\phi(x)$ to the utility function, we allow the shares to change with prices and income. Also for small values of the $\xi_i$, any trans-log indirect utility function can be approximated by a linear combination of a small number of terms.

Geweke and Petrella's (2000) basis functions, given in Eq. (1), can be interpreted as monomials in $\phi$ for the following alternative choice of $\phi$:

$$\phi(x) = (x_1^b, \ldots, x_n^b) \quad \forall x \in X.$$

We want to construct a sequence of families of polynomials in $\phi(x)$ with growing flexibility. We choose a sequence $\{I_k\}_{k=0}^{\infty}$ of constellations increasing[2] towards $\mathbb{N}_0^n \backslash \{0\}$. The set of polynomials $\sum_{\imath \in I_k} \lambda_\imath [\phi(x)]^\imath$ becomes more flexible as $k$ increases. Excluding the multi-index 0 means excluding the constant term of $u(\cdot;\lambda)$. We do this to simplify prior elicitation and simulation. The exclusion means that we can only approximate a regular function $u$ up to an additive constant, but this is innocuous.

One natural choice of $I_k$ is the sequence of hyper-rectangular lattices $\{\imath \in \mathbb{N}_0^n : \imath_i \leqslant k-1 \text{ for all } i\}$, which we use in this paper. Another is the sequence of simplicial lattices $\{\imath \in \mathbb{N}_0^n : \sum_{i=1}^n \imath_i \leqslant k-1\}$.

For given constellation index $k \in \mathbb{N}_0$, coefficient vector $\lambda \equiv (\lambda_\imath)_{\imath \in I_k}$, and transformation $\phi : \bar{X} \to \mathbb{R}$, we define the function $u(\cdot,\lambda) : \bar{X} \to \mathbb{R}$ by

$$u(x;\lambda) \equiv \sum_{\imath \in I_k} \lambda_\imath [\phi(x)]^\imath \quad \forall x \in X.$$

The unknown parameters of our approximation are the constellation index $k$ and the vector $\lambda \equiv (\lambda_\imath)_{\imath \in I_k}$ of monomial coefficients. For each $k \in \mathbb{N}_0$, we define $\Lambda_{\bar{X}}^k$ as the set of vectors $\lambda$ of length[3] $|I_k|$ associated with functions that are regular on $\bar{X}$:

$$\Lambda_{\bar{X}}^k \equiv \{\lambda \in \mathbb{R}^{|I_k|} : u(\cdot;\lambda) \text{ is regular on } \bar{X}\}. \tag{3}$$

We also define $\Lambda_{\bar{X}} \equiv \bigcup_{k=1}^{\infty} \Lambda_{\bar{X}}^k$, the complete set of vectors $\lambda$ associated with functions that are regular on $\bar{X}$.

### 2.4. Results

We now present two results on the parametric functional form. We will see in Section 3 that the following result is relevant for prior elicitation. In Section 4, we show that it is important for prior and posterior simulation.

**Result 2.1.** *For every $k \in \{1,2,\ldots\}$, $\Lambda_{\bar{X}}^k$ is a convex cone.*

**Proof.** The set of functions regular on $\bar{X}$ is closed under addition and positive scalar multiplication. Therefore $\Lambda_{\bar{X}}^k$ is a convex cone. □

The following alternate proof yields some important insights that are relevant for later sections. We can express the regularity conditions as follows. For all $x \in \bar{X}$,

$$\frac{\partial u(x;\lambda)}{\partial x} = \sum_{\imath \in I_k} \lambda_\imath \frac{\partial [\phi(x)]^\imath}{\partial x} \geqslant 0 \tag{4}$$

and for all $x \in X$ and $v \in \mathbb{R}^n$,

$$v' \frac{\partial^2 u(x;\lambda)}{\partial x \, \partial x'} v = \sum_{\imath \in I_k} \lambda_\imath v' \frac{\partial^2 [\phi(x)]^\imath}{\partial x \, \partial x'} v \leqslant 0. \tag{5}$$

For every choice of $x \in \bar{X}$, Eq. (4) gives $n$ linear inequalities in $\lambda$, one for each component of the gradient. For every choice of $x \in \bar{X}$ and $v \in \mathbb{R}^n$, Eq. (5) gives another linear inequality in $\lambda$. We see, therefore, that the parameter space $\Lambda_{\bar{X}}^k$ is the intersection of half spaces defined by the inequalities above. The half spaces are convex, so their intersection is as well.

The following approximation result tells us that we can simultaneously approximate any twice continuously differentiable function $u$, regular on $\bar{X}$, together with its gradient and Hessian, arbitrarily closely on $\bar{X}$, up to an additive constant. The approximation of the gradient is important because we want to approximate the behavior that the function represents, which is often given in terms of derivatives. The approximation of the gradient and Hessian is important for guaranteeing that the approximating function is regular on $\bar{X}$. In practice, it is much easier to check that the approximating function is regular on $\bar{X}$ than to verify its proximity to some function regular on $\bar{X}$. For notational convenience, we define the following norm for twice

[2] A sequence of sets $\{I_k\}_{k=0}^{\infty}$ *increases towards* set $I$, denoted $I_k \uparrow I$, if $I_k \subseteq I_{k+1}$ for all $k \in \mathbb{N}_0$ and $\bigcup_{k=0}^{\infty} I_k = I$.

[3] For any set $A$, we denote by $|A|$ the cardinality of $A$.

continuously differentiable functions:

$$\|f\| \equiv \max\left[\sup_{x\in\bar{X}}|f(x)|, \sup_{x\in\bar{X},i\in\{1,\ldots,n\}}\left|\frac{\partial f(x)}{\partial x_i}\right|, \sup_{x\in\bar{X},i,j\in\{1,\ldots,n\}}\left|\frac{\partial f(x)}{\partial x_i\,\partial x_j}\right|\right].$$

The result is a generalization of an unpublished result by Geweke and Petrella (2000). It is they who recognized the significance of a result by Evard and Jafari (1994) on the simultaneous approximation of a function and its derivatives by polynomials. The modest contributions in this section include the recognition that different transformations $\phi$ can be used to generate sets of basis functions and a complete proof of the following result.

**Result 2.2** (*Approximation*). *Suppose the transformation $\phi\colon \bar{X}\to\mathbb{R}$ is such that $\phi(\bar{X})$ is compact and that the inverse $\phi^{-1}$ on $\phi(\bar{X})$ exists and is twice continuously differentiable. Then for every twice continuously differentiable $u\colon X\to\mathbb{R}$ regular on $\bar{X}$, and every $\varepsilon>0$, there exists a $\lambda\in\Lambda_{\bar{X}}$ and a constant $c$ such that*

$$\|c+u(\cdot;\lambda)-u(\cdot)\|<\varepsilon. \tag{6}$$

**Proof.** Let $u$ be a twice continuously differentiable regular function and let $\varepsilon>0$.

Rather than approximating $u$ directly, which may be on the boundary of the regular region, we will approximate a nearby function $\hat{u}\colon\bar{X}\to\mathbb{R}$ in the interior. We choose $\hat{u}$ close enough to $u$ that the approximation of $\hat{u}$ is a sufficiently close approximation of $u$ itself. The function $\hat{u}$ is defined by

$$\hat{u}(x)\equiv u(x)+\frac{\varepsilon}{2}\prod_{i=1}^{n}(1-e^{\bar{x}_i-x_i})\quad\forall x\in\bar{X},$$

where $\bar{x}_i=\max_{x\in\bar{X}}x_i$, $i=1,\ldots,n$. Since $u$ is non-decreasing, concave and twice continuously differentiable on $\bar{X}$, $\hat{u}$ is increasing, strictly concave and twice continuously differentiable on $\bar{X}$. Also,

$$\|\hat{u}(\cdot)-u(\cdot)\|\leqslant\frac{\varepsilon}{2}. \tag{7}$$

A direct corollary of Corollary 3 of Evard and Jafari (1994) is that for every twice continuously differentiable function $f\colon\bar{X}\to\mathbb{R}$, and every $\varepsilon'>0$, there exists a polynomial $p\colon\bar{X}\to\mathbb{R}$ such that for all $i,j\in\{1,\ldots,n\}$ and all $x\in\bar{X}$,

$$|f(x)-p(x)|<\varepsilon',\quad\left|\frac{\partial f}{\partial x_i}-\frac{\partial p}{\partial x_i}\right|<\varepsilon'\quad\text{and}\quad\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}-\frac{\partial^2 p}{\partial x_i\,\partial x_j}\right|<\varepsilon'.$$

Since $\phi$ has an inverse $\phi^{-1}$ that is twice continuously differentiable on $\bar{X}$, $\hat{u}\circ\phi^{-1}$ is also twice continuously differentiable on $\phi(\bar{X})$. Furthermore, $\phi(\bar{X})$ is compact. Therefore the corollary implies that for all $\varepsilon'>0$, there exists a $\lambda\in\bigcup_{k=1}^{\infty}\mathbb{R}^{|I_k|}$ and a $c$ such that for all $i,j\in\{1,\ldots,n\}$ and all $z\in\phi(\bar{X})$,

$$\left|c+\sum_{\iota\in I_k}\lambda_\iota z^\iota-(\hat{u}\circ\phi^{-1})(z)\right|<\varepsilon',$$

$$\left|\frac{\partial}{\partial z_i}\sum_{\iota\in I_k}\lambda_\iota z^\iota-\frac{\partial}{\partial z_i}(\hat{u}\circ\phi^{-1})(z)\right|<\varepsilon'$$

and

$$\left|\frac{\partial^2}{\partial z_i\,\partial z_j}\sum_{\iota\in I_k}\lambda_\iota z^\iota-\frac{\partial^2}{\partial z_i\,\partial z_j}(\hat{u}\circ\phi^{-1})(z)\right|<\varepsilon'.$$

The function $\phi$ maps $\bar{X}$ to $\phi(\bar{X})$, and therefore for all $i,j\in\{1,\ldots,n\}$ and $x\in\bar{X}$,

$$\left|c+\sum_{\iota\in I_k}\lambda_\iota[\phi(x)]^\iota-\hat{u}(x)\right|=|c+u(x;\lambda)-\hat{u}(x)|<\varepsilon',$$

$$\left|\frac{\partial}{\partial x_i}u(x;\lambda)-\frac{\partial}{\partial x_i}\hat{u}(x)\right|<\varepsilon' M_1$$

and

$$\left|\frac{\partial^2}{\partial x_i\,\partial x_j}u(x;\lambda)-\frac{\partial^2}{\partial x_i\,\partial x_j}\hat{u}(x)\right|<\varepsilon' M_2,$$

where $M_1$ and $M_2$, derived from uniform bounds on the derivatives of $\phi$ on $\bar{X}$, do not depend on $x$.

We can choose $\varepsilon'$ such that

$$\|c+u(x;\lambda)-\hat{u}(x)\|<\frac{\varepsilon}{2} \tag{8}$$

and for all $x\in\bar{X}$ and all $i,j\in\{1,\ldots,n\}$,

$$\frac{\partial u(x;\lambda)}{\partial x}\geqslant 0, \tag{9}$$

and

$$v'\frac{\partial^2 u(x;\lambda)}{\partial x\,\partial x'}v\leqslant 0\quad\forall v\in\mathbb{R}^n. \tag{10}$$

Eqs. (7), (8), and the triangle inequality guarantee that (6) holds. Eqs. (9) and (10) guarantee that $\lambda\in\Lambda_{\bar{X}}$. □

## 3. Priors

We have just described parametric classes of regular functions. In this section, we discuss the problem of expressing prior uncertainty about regular functions by specifying a prior probability distribution for the constellation index $k$ and the coefficient vector $\lambda=(\lambda_\iota)_{\iota\in I_k}$. The ideas expressed in this section apply to various choices of the transformation $\phi$ and even to different systems of basis functions, such as those of the Fourier flexible form and the Müntz–Szasz expansion. For definiteness, we take $\bar{X}=[0,\bar{x}_1]\times\cdots\times[0,\bar{x}_n]$ and $\phi$ defined in (2).

We first define notions of flexibility and theoretical consistency for these prior distributions and give sufficient conditions for both. We then focus on the problem of eliciting prior distributions over $\Lambda^k_{\bar{X}}$ for fixed $k$. We address the issue of ensuring that these are proper, which is important for inference on $k$ and model comparison using Bayes factors.

We show that if the regular function is a utility function, a uniform prior implies prior distributions over marginal rates of substitution that many would consider too informative. When the regular function is a negative indirect utility function, the prior distributions of ratios of quantities demanded are very informative. We give an example of a non-uniform prior that implies more diffuse prior distributions over marginal rates of substitution or ratios of quantities.

### 3.1. Flexibility and theoretical consistency

We will say that a prior is *flexible* if for any twice continuously differentiable function $u$ that is regular on $\bar{X}$, there is a constant $c$ such that the prior assigns positive probability to any $\|\cdot\|$-neighborhood of $u-c$. We will say that it is *theoretically consistent* if it assigns zero probability to the set of functions that are not regular on $\bar{X}$. The following result gives sufficient conditions on the prior for flexibility and theoretical consistency.

**Result 3.1.** *If $\sum_{k=K}^{\infty}\pi_k>0$ for all $K\in\mathbb{N}$, and the conditional distributions $\lambda|k$ have support $\Lambda^k_{\bar{X}}$, then the prior is flexible and theoretically consistent.*

**Proof.** Let $u\colon X\to\mathbb{R}$ be twice continuously differentiable and regular on $\bar{X}$, and let $\varepsilon>0$. By Result 2.2, we can find a $k\in\mathbb{N}$ and a $\lambda^*\in\Lambda^k_{\bar{X}}$ such that $\|c+u(x;\lambda^*)-u(x)\|<\varepsilon/2$. The conditional distribution $\lambda|k$, whose

support is $\Lambda^k_{\bar{X}}$, assigns positive probability to the set

$$\Lambda^* \equiv \left\{ \lambda \in \Lambda^k_{\bar{X}} : |\lambda^*_\iota - \lambda_\iota| < \frac{\varepsilon}{2|I_k|\|\phi(x)^\iota\|} \quad \forall \iota \in I_k \right\}.$$

For all $\lambda \in \Lambda^*$, $\|u(\cdot;\lambda) - u(\cdot;\lambda^*)\| < \frac{\varepsilon}{2}$, and therefore $\|c + u(\cdot;\lambda) - u(\cdot)\| < \varepsilon$. Since $\sum_{k=K}^{\infty} \pi_k > 0$ for all $K \in \mathbb{N}$, the prior assigns positive probability to $\Lambda^*$, and therefore positive probability to any $\|\cdot\|$-neighborhood of $u$. The fact that the prior assigns zero probability to the set of functions that are not regular on $\bar{X}$ follows trivially from the fact that the support of the prior is $\Lambda_{\bar{X}}$. □

### 3.2. Conditional prior distributions on $\Lambda^k_{\bar{X}}$

The conditional priors $\lambda|k$ are distributions on the $\Lambda^k_{\bar{X}}$. Except for being convex cones, the $\Lambda^k_{\bar{X}}$ are irregularly shaped, and this makes elicitation difficult.

Utility functions that differ only up to the multiplication of a positive constant are observationally equivalent. The same holds for negative indirect utility functions. If the regular function is a utility function or a negative indirect utility function, then without loss of generality, we can normalize it by imposing the condition $u(\bar{x};\lambda) - u(0;\lambda) = 1$. This is a linear equality in $\lambda$, so it defines a hyperplane in $\mathbb{R}^{|I_k|}$. By Result A.1, the intersection of this hyperplane with $\Lambda^k_{\bar{X}}$ is bounded. This implies that a uniform prior on this intersection is proper.

This is not the only possible normalization. For illustrative purposes we will consider the following normalization:

$$\sum_{i=1}^{n} \beta_i \frac{\Delta u_i(\lambda)}{\bar{x}_i} = 1,$$

where $\Delta u_i(\lambda)$ is the increase in $u$ as $x_i$ goes from the lower extreme of 0 to the upper extreme of $\bar{x}_i$, with all other $x_j$ set to the reference value $x^*_j$. That is,

$$\Delta u_i(\lambda) \equiv u(x^* + (\bar{x}_i - x^*_i)e_i;\lambda) - u(x^* + (0 - x^*_i)e_i;\lambda),$$

where $e_i$ is the unit vector on the $i$th coordinate axis in $\mathbb{R}^n$. We can think of $\Delta u_i/\bar{x}_i$ as the average of the gradient component $\partial u/\partial x_i$ on the line segment from $x^* + (0 - x^*_i)e_i$ to $x^* + (\bar{x}_i - x^*_i)e_i$.

This normalization is also a linear equality in $\lambda$, and using Result A.1, we can show that the hyperplane it defines has a bounded intersection with $\Lambda^k_{\bar{X}}$. The uniform distribution on this intersection is therefore proper.

Using the methods described in Section 4, we simulate from this uniform distribution. We take $n = 2$, $\xi = (0.01, 0.01)$, $x^* = (0.1, 0.1)$, $\bar{x} = (1.0, 1.0)$, and $\beta = (0.5, 0.5)$.

The first panel of Fig. 1 shows a histogram of the prior distribution of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$. If we interpret $u$ as a utility function, this quantity is the log marginal rate of substitution between goods 1 and 2 at the commodity bundle $x^*$. If $u$ is a negative indirect utility function, it is the log ratio of the quantities demanded of goods 1 and 2 at income-normalized prices $x^*$.

We see that despite being uniform on the set

$$\left\{ \lambda \in \Lambda^k_{\bar{X}} : \sum_{i=1}^{n} \beta_i \frac{\Delta u_i(\lambda)}{\bar{x}_i} = 1 \right\},$$

the prior is quite informative about the value of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$. Specifically, it puts very low probability on the regions where $\partial u(x^*;\lambda)/\partial x_1$ is much smaller or much greater than $\partial u(x^*;\lambda)/\partial x_2$. Many will find these regions quite plausible for certain applications and will thus find the prior inappropriate.

We would like to make clear that the point we are making concerns scale, not location. We can always change the units of measurement of commodity quantities (and adjust prices accordingly) to change the location of this distribution.

To persuade the skeptical reader that the prior is not diffuse enough for some applications, we consider the following prior elicitation exercise, concerning individual consumption of two goods: gasoline and a composite good representing everything else. Hypothetical individual A earns $30,000 after tax, takes the

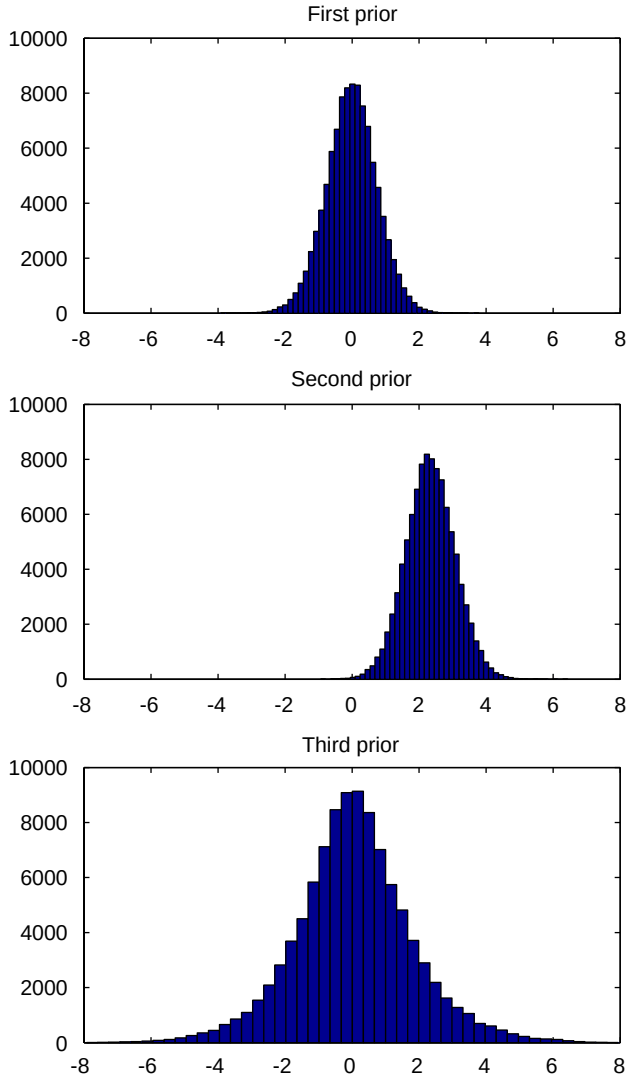


Fig. 1. Histograms of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ for three different priors.

subway to work and buys ten gallons of gasoline per month. B earns the same amount, drives to work from the suburbs and spends 100 gallons of gasoline a month. The two individuals are hardly unusual, but the ratio of quantities differs by a factor of ten between them, even though they have the same income and face the same prices. A factor of ten corresponds to slightly more than three standard deviations. It is quite plausible that if the price of gasoline doubled, A would not reduce consumption to less than 35 gallons a month. It is also plausible that if the price of gasoline dropped by a factor of five, B would not increase consumption to more than 35 gallons a month. This would imply that their marginal rates of substitution at 35 gallons per month differ by at least a factor of ten.

A sensitivity analysis in Appendix B suggests that for more than two goods, the distribution of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ becomes even more informative. For $n = 3$ goods, a factor of ten in the ratio of gradient components $\partial u(x^*;\lambda)/\partial x_1$ and $\partial u(x^*;\lambda)/\partial x_2$ corresponds to about 4.5 standard deviations.

Given the uniformity of the prior, the assignment of low probability to the regions where one gradient component is much smaller than the other must be because these regions have very low volume. Since these regions are also near the boundary of the set, we will refer to them (loosely) as the "tight corners" of $\Lambda^k_{\bar{X}}$. Intuitively, for fixed $\partial u(x^*;\lambda)/\partial x_1$ and $\partial u(x^*;\lambda)/\partial x_2$, it is easier to construct regular functions when they are

nearly equal than when one is much larger than the other: when they are nearly equal, there is more wiggle room.

We illustrate using an example that the distribution of $\log(\partial u(x;\lambda)/\partial x_1) - \log(\partial u(x;\lambda)/\partial x_2)$ is sensitive to the choice of normalization. We tilt the hyperplane defined by the normalization $\beta_1 \Delta u_1/\bar{x}_i + \beta_2 \Delta u_2 \bar{x}_i = 1$ by replacing $\beta = (\frac{1}{2}, \frac{1}{2})$ with $\beta = (\frac{1}{6}, \frac{5}{6})$. The second panel of Fig. 1 shows the prior histogram for $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ implied by a uniform prior on the intersection of $\Lambda^k_{\tilde{X}}$ with the new hyperplane. We see that the location of the distribution changes, but little else. In particular, the prior is still very informative.

We conjecture that the tight distribution of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ arises not because of the particular choice of normalization but because the prior is uniform on a truncated hyperplane: such a prior puts low probability in the vicinity of any given bounding hyper-plane of the form $v'\lambda = 0$, such as $\partial u(x^*;\lambda)/\partial x_i = 0$. The results in Appendix B suggest that as $k$ increases and $u(\cdot;\lambda)$ becomes more flexible, the prior puts less and less probability in the vicinity of the bounding hyper-plane defined by $\partial u(x^*;\lambda)/\partial x_i = 0$. Intuitively, as $k$ increases and the dimension of $\Lambda^k_{\tilde{X}}$ increases with it, the curvature of its boundary increases as well, reducing the relative volume close to the hyper-plane.

We remark that the results shown in Fig. 1 depend on fixed parameters and the choice of the point $x$ at which to evaluate $\log(\partial u(x;\lambda)/\partial x_1) - \log(\partial u(x;\lambda)/\partial x_2)$. Appendix B shows the results of a sensitivity analysis. They suggest that the informativeness of the prior is robust to variations in these quantities.

If we want to make the prior distribution of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ more diffuse, we can do so by putting more probability mass near the boundary of $\Lambda^k_{\tilde{X}}$. As an example, we multiply the prior by the factor

$$\prod_{i=1}^{n} \left( \beta_i \frac{\Delta u_i}{\bar{x}_i} + \delta \right)^{-p},$$

where $\delta$ and $p$ are positive constants. The term $\delta$ ensures that the factor is bounded on the truncated hyperplane $\{\lambda \in \Lambda^k_{\tilde{X}} : \beta_1 \Delta u_1(\lambda)/\bar{x}_i + \beta_2 \Delta u_2(\lambda)/\bar{x}_i = 1\}$, so the prior remains proper.

The third panel of Fig. 1 shows the prior histogram of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ for the modified prior, with $\delta = 0.001$ and $p = -3$. We see that it is more diffuse.

The three examples we have just seen suggest that we can center the prior over whatever value of $\log(\partial u(x^*;\lambda)/\partial x_1) - \log(\partial u(x^*;\lambda)/\partial x_2)$ is appropriate for a given application and also independently choose the degree of dispersion.

The priors in these three examples have support equal to the intersection of $\Lambda^k_{\tilde{X}}$ with a hyperplane $\{\lambda : v'\lambda = 1\}$, where $v$ is a vector defining a normalization of the regular function $u$. While utility and indirect utility functions can be normalized without affecting optimal choices, expenditure, cost and profit functions cannot. For flexibility in these cases, we would like to elicit a prior with full support $\Lambda^k_{\tilde{X}}$. Rather than normalizing the regular function, we can choose a density $f$ of the form

$$f(\lambda) \propto g(v'\lambda)(v'\lambda)^{-(|I_k|-1)}, \quad \lambda \in \Lambda^k_{\tilde{X}}, \tag{11}$$

where $g$ is a proper univariate density function. The level curves of $f$ are parallel to the hyperplane $\{\lambda : v'\lambda = 1\}$, and the distribution of $v'\lambda$ is given by the density $g$.

If we want to elicit a prior whose contours are not all parallel to the hyperplane $\{\lambda : v'\lambda = 1\}$, but has the property that the scale $v'\lambda$ and the normalized regular function $(1/v'\lambda)u(\cdot;\lambda)$ are independent, we can choose a density of the form

$$f(\lambda) \propto g(v'\lambda)(v'\lambda)^{-(|I_k|-1)} h(\lambda), \quad \lambda \in \Lambda^k_{\tilde{X}}, \tag{12}$$

where $h$ is a homogeneous degree 0 function.

For practical reasons, it turns out to be useful to elicit a prior with full support $\Lambda^k_{\tilde{X}}$ even when the regular function is a utility or negative indirect utility function. And we can always ignore the unidentified scale. This means, for example, that we can simulate the uniform distribution on the intersection of $\Lambda^k_{\tilde{X}}$ and the hyperplane $\{\lambda : v'\lambda = 1\}$ by simulating the density in (11) and normalizing draws by dividing by $v'\lambda$. The distribution of the scale $v'\lambda$ will have density given by $g$.

To simulate the third, non-uniform, prior we can simulate the density in (12) with $g$ arbitrary and

$$h(\lambda) \equiv \prod_{i=1}^{n} \left( \frac{\beta_i \Delta u_i / \bar{x}_i}{\sum_{j=1}^{n} \beta_j \Delta u_j / \bar{x}_i} + \delta \right)^{-p} \tag{13}$$

and normalize by dividing by $v'\lambda$.

## 4. Prior and posterior simulation

In the previous section, we describe conditional prior densities $f(\lambda|k)$ on the convex cones $\Lambda_{\bar{X}}^k$ whose union is the regular parameter space. In this section we describe methods for MCMC simulation of distributions on $\Lambda_{\bar{X}}^k$.

We saw in the previous section that the regions near the boundary of the parameter space $\Lambda_{\bar{X}}^k$ where one or more of the gradient components is much smaller than the rest have very low volume. Thus the problems of simulating a distribution on this space are similar to those of simulating a distribution on a cigar-shaped object with long pointy ends. The problem is exacerbated when we want to simulate a distribution, such as the third prior of the previous section, that concentrates probability in the tight corners of the parameter space. An efficient chain for simulating the target distribution must be able to spend enough time in the tight corners, but also move quickly in and out of them.

Random walk Metropolis chains with constant proposal variance matrices work poorly. If the eigenvalues of the variance are large enough to generate steps big enough for efficiently exploring the central region of $\Lambda_{\bar{X}}^k$, then the acceptance probability in the tight corners is intolerably low. We experimented with Metropolis–Hastings random walks where the proposal variance varied according to how close the current state of the chain was to various bounding planes. This is quite difficult to do, and the chain's numerical efficiency is fairly low. The problem is that it takes too many iterations to move in and out of the tight corners.

The Metropolis–Hastings proposal distributions we present here do not depend on the target distribution. So we make no distinction between prior and posterior simulation and leave open the question of how the data density is specified. For the simulations reported in this paper, we obtain satisfactory numerical efficiency for both prior and posterior simulation. We believe there is scope for improving these proposals by taking advantage of features of the prior and likelihood, but we do not explore this here.

We describe three Metropolis–Hastings updates that can be used in combination (either a mixture or a sweep) to simulate prior and posterior distributions efficiently on the parameter subspace $\Lambda_{\bar{X}}^k$.

The first update generates a line passing through the current state $\lambda$ in a random direction, then draws a random proposal $\lambda^*$ on a segment of this line containing its intersection with $\Lambda_{\bar{X}}^k$. This allows large jumps across $\Lambda_{\bar{X}}^k$ with reasonable probability.

The second and third updates draw a random proposal $\lambda^*$ on a ray emanating from a point near the boundary of $\Lambda_{\bar{X}}^k$ and passing through the current state $\lambda$. It is these two updates which ensure that the chain can quickly move in and out of the tight corners of $\Lambda_{\bar{X}}^k$.

### 4.1. Definitions

We introduce a few preliminary definitions that will be important in the next sections. We define, for each good $i$, the univariate function $u_i: [0, \bar{x}_i] \to \mathbb{R}$ as the restriction of $u(x;\lambda)$ to the line segment $\{x \in \bar{X}: x_j = x_j^*, j \neq i\}$, parallel to the $i$th coordinate axis but shifted so that the $x_j$, $j \neq i$, are fixed at $x_j^*$ rather than zero. That is,

$$u_i(x_i) = u(x^* + (x_i - x_i^*)e_i), \quad x_i \in [0, \bar{x}_i],$$

where $e_i$ is the unit length $n$-vector on the $i$th co-ordinate axis.

Note that

$$\phi(x^* + (x_i - x_i^*)e_i) = \log\left(\frac{x_i + \xi_i}{x_i^* + \xi_i}\right) e_i.$$

The only non-zero element of this vector is the $i$th, which means that the monomial $[\phi(x^* + (x_i - x_i^*)e_i]^\iota$ is non-zero only for $\iota = \kappa e_i$, $\kappa = 1, \ldots, k-1$. Thus the univariate $u_i$ depends only the elements $\lambda_\iota$ such that $\iota = \kappa e_i$, $\kappa = 1, \ldots, k-1$.

This leads to the following partition of the vector $\lambda$:

$$\lambda = (\lambda^{(1)}, \ldots, \lambda^{(n)}, \lambda^\circ),$$

where $\lambda^{(i)}$ consists of the elements $\lambda_\iota$ for which $\iota_j = 0$ for all $j \neq i$. In other terms,

$$\lambda^{(i)} \equiv (\lambda_{e_i}, \lambda_{2e_i}, \ldots, \lambda_{(k-1)e_i}), \quad i = 1, \ldots, n.$$

The subvector $\lambda^\circ$ consists of all remaining elements of $\lambda$.

Since $\phi(x^* + (x_i - x_i^*)e_i)$ is non-zero only for $\iota = \kappa e_i$, $\kappa = 1, \ldots, k-1$, $u_i$ depends only on the sub-vector $\lambda^{(i)}$.

### 4.2. A convex cone enclosing $\Lambda_{\tilde{X}}^k$

We draw all three random proposals within a convex cone $\{\lambda \in \mathbb{R}^{|I_k|} : V\lambda \geqslant 0\}$, where $V$ is an $N_v \times |I_k|$ matrix such that the cone contains $\Lambda_{\tilde{X}}^k$. The tighter the fit of $V\lambda \geqslant 0$ to $\Lambda_{\tilde{X}}^k$, the less often we draw proposals that are not in the regular set $\Lambda_{\tilde{X}}^k$.

It is easy to construct matrices $V$ such that $V\lambda \geqslant 0$ contains $\Lambda_{\tilde{X}}^k$. Eqs. (4) and (5), which give various necessary conditions for regularity, are inequalities of the form $v\lambda \geqslant 0$, where $v$ is a row vector of length $|I_k|$. Constructing $V$ involves vertically stacking a number $N_v$ of row vectors $v_j$ satisfying $v_j\lambda \geqslant 0$.

For the prior and posterior simulations reported in this paper, we use $N_v = n \cdot J^n$ vectors, where $J = 20$. Each vector is indexed by a pair $(i, \iota) \in \{1, \ldots, n\} \times I_J$. For each $i$ and $\iota$, we generate a row of $V$ using the necessary conditions

$$\begin{cases} \dfrac{\partial u(x^{(\iota)}; \lambda)}{\partial x_i} - \dfrac{\partial u(x^{(\iota + e_i)}; \lambda)}{\partial x_i} \geqslant 0, & \iota_i \leqslant J, \\ \dfrac{\partial u(x^{(\iota)}; \lambda)}{\partial x_i} \geqslant 0, & \iota_i = J, \end{cases}$$

where $\{x^{(\iota)} : \iota \in I_J\}$ is a grid of points. Each $x^{(\iota)}$ is defined by

$$\phi(x^{(\iota)}) = \left(z_1^{lo} + \frac{\iota_1}{J-1}(z_1^{hi} - z_1^{lo}), \ldots, z_n^{lo} + \frac{\iota_n}{J-1}(z_n^{hi} - z_n^{lo})\right),$$

where $z^{lo} \equiv \phi(0)$ and $z^{hi} \equiv \phi(\bar{x})$.

### 4.3. A first proposal

We start at the current state $\lambda$ and generate a random proposal $\lambda^*$.

The proposal consists of three steps. First, we draw a random direction vector $w \in \mathbb{R}^{|I_k|}$ from a discrete uniform distribution over a set of precomputed direction vectors. Then we compute $\pi_- \leqslant 0$ and $\pi_+ \geqslant 0$ such that $\lambda + (\pi_-)w$ and $\lambda + (\pi_+)w$ are on the boundary of the cone $V\lambda \geqslant 0$ enclosing $\Lambda_{\tilde{X}}^k$. Finally, we draw $\pi$ from the uniform distribution on $[\pi_-, \pi_+]$ and construct the proposal $\lambda^* = \lambda + \pi w$. We accept $\lambda^*$ with probability

$$\min\left(1, \frac{f(\lambda^*)}{f(\lambda)}\right),$$

where $f$ is the unnormalized target density. Evaluating $f(\lambda^*)$ includes determining whether $\lambda^* \in \Lambda_{\tilde{X}}^k$. We now describe some of these steps in more detail.

*4.3.1. Computing $\pi_-$ and $\pi_+$*

We compute $\pi_+$ as the smallest positive value of $\pi$ for which $V(\lambda + \pi w) \geqslant 0$. That is, we choose

$$\pi_+ = \min_{j \in \{1,\ldots,N_v\}} \left\{ -\frac{(V\lambda)_j}{(Vw)_j} : (Vw)_j \leqslant 0 \right\}.$$

Similarly,

$$\pi_- = \max_{j \in \{1,\ldots,N_v\}} \left\{ -\frac{(V\lambda)_j}{(Vw)_j} : (Vw)_j \geqslant 0 \right\}.$$

*4.3.2. Drawing random directions*

For the simulations reported in this paper, we use the following set of directions. There is a direction $w^{(\iota)} \in \mathbb{R}^{|I_k|}$ for each multi-index $\iota \in I_k \backslash \{e_1, \ldots, e_n\}$. For multi-indices $\iota$ and $\kappa$, $w_\kappa^{(\iota)}$ is element $\kappa$ of the direction vector $w^{(\iota)}$ and is given by the coefficient of $z^\kappa$ in the $n$-variate polynomial $\prod_{i=1}^n P_{\iota_i}(a_i z_i + b_i)$, where $P_j$ is the $j$th Legendre polynomial, and $a_i z_i + b_i$ is the linear transformation of $z_i \equiv \phi_i(x)$ mapping $[z_i^{lo}, z_i^{hi}]$ to $[-1, 1]$. As before, $z^{lo} \equiv \phi(0)$ and $z^{hi} \equiv \phi(\bar{x})$. In other terms,

$$w_\kappa^{(\iota)} = \prod_{i=1}^n \sum_{j=\kappa_i}^{\iota_i} L_{\iota_i, j} \binom{j}{\kappa_i} a_i^{\kappa_i} b_i^{j-\kappa_i},$$

where $a_i = 2/(z_i^{hi} - z_i^{lo})$, $b_i = -(z_i^{hi} + z_i^{lo})/(z_i^{hi} - z_i^{lo})$, and $L_{ij}$ is the coefficient of the $j$th order monomial in the $i$th Legendre polynomial.

This set of directions has two useful properties. First, the $n$-variate polynomials $\sum_{\kappa \in I_k} w_\kappa^{(\iota)} z^\kappa$ are orthogonal on $\phi(\bar{X})$, inheriting the orthogonality of the (univariate) Legendre polynomials on $[-1, 1]$. This minimizes redundancy among the directions. Second, because the coordinate vectors $e_i$ are excluded, $Vw_\kappa^{(\iota)}$ always has both positive and negative elements, which ensures that the values $\pi_-$ and $\pi_+$ exist and are always finite.

*4.3.3. Checking regularity*

Evaluating the target density, if its support is $\Lambda_{\bar{X}}^k$, typically involves verifying $\lambda^* \in \Lambda_{\bar{X}}^k$ or, equivalently, verifying that $u(\cdot; \lambda^*)$ is regular on $\bar{X}$. This is a difficult problem, and we do not know of any tractable algorithm that does this without error.

Various imperfect checks of regularity appear in the literature. Gallant (1981) proposes a simple algorithm for checking convexity or concavity, but the conditions it verifies are sufficient but not necessary. Gallant and Golub (1984) check quasi-concavity by searching for a minimum of a function whose non-negativity is necessary and sufficient for quasi-concavity, but there is no guarantee that the minimization procedure finds a global minimum.

All the basis functions of the Müntz–Szasz expansion are regular, so there is an ease to verify sufficient condition for regularity: non-negativity of all coefficients. However, the condition in not necessary. In fact, it imposes substitutability on all inputs. Terrell (1996) uses simulations to show that even when inputs are substitutes, much flexibility is lost. See also Koop et al. (1994) on this point.

To check the regularity of $u(\cdot; \lambda^*)$, we run a battery of tests verifying necessary conditions. So although we will never attribute irregularity to a regular function, we cannot guarantee that we will detect the irregularity of an irregular function.

We note that any proposal $\lambda^*$ satisfies $V\lambda^* \geqslant 0$ by construction, so it already survives all of the regularity testing implied by this condition.

For all $i \in \{1, \ldots, n\}$, the expressions

$$g(z_i) \equiv (x_i + \xi_i) \frac{\partial u_i}{\partial x_i}(x_i; \lambda^{(\iota)}) \quad \text{and} \quad h(z_i) \equiv (x_i + \xi_i)^2 \frac{\partial^2 u_i}{\partial x_i^2}(x_i; \lambda^{(\iota)})$$

are polynomials of order $k - 1$ in $z_i \equiv \phi_i(x)$. We find, numerically if necessary, all the roots of $h$ and verify that none of these roots are in $[z_i^{lo}, z_i^{hi}]$, that $h(0) < 0$, and that $g(z_i^{hi}) > 0$.

We then verify that the gradient of $u$ at $\bar{x}$ is positive. Finally, we search for the maximum value of the largest eigenvalue of the Hessian of $u$ over $\bar{X}$, using a simplex algorithm, and verify that the greatest value found is negative.

### 4.4. A second proposal

The second proposal consists of three steps. First we draw a random good $i^*$ from the discrete uniform distribution on $\{1,\ldots,n\}$. Then we find $\pi_{\min}$, the value of the multiplier $\pi$ such that $(\lambda^{(1)},\ldots,\lambda^{(i^*-1)}, \pi\lambda^{(i^*)},\lambda^{(i^*+1)},\ldots,\lambda^{(n)},\lambda^\circ)$ is on the boundary of the cone $V\lambda\geqslant 0$. Finally, we draw $\pi^*$ from a log-normal distribution truncated to $[\pi_{\min},\infty)$ and construct the proposal

$$\lambda^* = (\lambda^{(1)},\ldots,\lambda^{(i^*-1)},\pi^*\lambda^{(i^*)},\lambda^{(i^*+1)},\ldots,\lambda^{(n)},\lambda^\circ). \tag{14}$$

The advantage of the second proposal is that the chain can move quickly out of the tight corners of $\Lambda^k_{\bar{X}}$.

We compute $\pi_{\min}$ as the smallest positive value of $\pi$ for which $V(\lambda+(\pi-1)w)\geqslant 0$, where $w=(0,\ldots,0,\lambda^{(i)},0,\ldots,0)$. That is, we choose

$$\pi_{\min} = \max_{j\in\{1,\ldots,N_v\}}\left\{1-\frac{(V\lambda)_j}{(Vw)_j}\right\}.$$

We point out that there is no $\pi_{\max}$. That is, there is no $\pi>1$ such that $V(\lambda+(\pi-1)w)\ngeqslant 0$. This is because $w$ and $\lambda$ are both in $\Lambda^k_{\bar{X}}$ and $\Lambda^k_{\bar{X}}\subseteq\{\lambda : V\lambda\geqslant 0\}$.

Once we have $\pi_{\min}$, we draw $\pi^*$ from a log-normal distribution truncated to $[\pi_{\min},\infty)$. For the simulations described in this paper, we draw $\pi^*$ such that $\log\pi^*$ has mean zero and standard deviation $\sigma\equiv 0.25$ before truncation.

We then construct $\lambda^*$ as in Eq. (14). We accept with probability

$$\min\left(1,\frac{f(\lambda^*)}{f(\lambda)}\cdot(\pi^*)^{k-2}\cdot\pi^*\cdot\frac{1-\Phi((\log\pi_{\min})/\sigma)}{1-\Phi((\log\pi_{\min}-\log\pi^*)/\sigma)}\right).$$

The factor $(\pi^*)^{k-2}$ is due to the fact that this is a radial draw in a $(k-1)$-dimensional subspace of $\Lambda^k_{\bar{X}}$: the volume of the differential element increases as $(\pi^*)^{k-2}$. For more rigor on this point, and for more information on radial draws in radial co-ordinate systems, see Bauwens et al. (2004). The factor $\pi^*$ comes from the Jacobian of the exponential transformation of the truncated Gaussian draw $\log\pi^*$.

### 4.5. A third proposal

The third proposal is similar to the second proposal, except that we multiply both $\lambda^{(i^*)}$ and $\lambda^\circ$ by the same random multiplier. First we draw a random good $i^*$ from the discrete uniform distribution on $\{1,\ldots,n\}$. Then we find $\pi_{\max}$, the value of the multiplier $\pi$ such that

$$(\lambda^{(1)},\ldots,\lambda^{(i^*-1)},\pi\lambda^{(i^*)},\lambda^{(i^*+1)},\ldots,\lambda^{(n)},\pi\lambda^\circ)$$

is on the boundary of the cone $V\lambda\geqslant 0$. Finally, we draw $\pi^*$ from a log-normal distribution truncated to $[0,\pi_{\max}]$ and construct the proposal

$$\lambda^* = (\lambda^{(1)},\ldots,\lambda^{(i^*-1)},\pi^*\lambda^{(i^*)},\lambda^{(i^*+1)},\ldots,\lambda^{(n)},\pi^*\lambda^\circ). \tag{15}$$

The advantage of the third proposal is that the chain can move quickly into the tight corners of $\Lambda^k_{\bar{X}}$.

We compute $\pi_{\max}$ as the largest value of $\pi$ for which $V(\lambda+(\pi-1)w)\geqslant 0$, where $w=(0,\ldots,0,\lambda^{(i)},0,\ldots,0,\lambda^\circ)$. That is, we choose

$$\pi_{\max} = \min_{j\in\{1,\ldots,N_v\}}\left\{1-\frac{(V\lambda)_j}{(Vw)_j}\right\}.$$

Regularity of the $u_i$, $i \neq i^*$ guarantees regularity of $u(\cdot; \lambda - w)$. Since, in addition, $\Lambda_X^k$ is a cone, there is no $\pi \in (0,1)$ such that $V(\lambda + (\pi - 1)w) \not\succeq 0$.

Once we have $\pi_{\max}$, we draw $\pi^*$ from a log-normal distribution truncated to $[0, \pi_{\max}]$. For the simulations described in this paper, $\log \pi^*$ is normal with mean 0 and standard deviation $\sigma \equiv 0.25$, before truncation.

We then construct $\lambda^*$ as in Eq. (15) and accept with probability

$$\min\left(1, \frac{f(\lambda^*)}{f(\lambda)} \cdot (\pi^*)^{k^n-(n-1)(k-1)-2} \cdot \pi^* \cdot \frac{\Phi((\log \pi_{\max})/\sigma)}{\Phi((\log \pi_{\max} - \log \pi^*)/\sigma)}\right),$$

where the exponent $k^n - (n-1)(k-1) - 2$ is one less than the dimension of $(\lambda^{(i^*)}, \lambda^\circ)$.

## 5. An empirical application

We present a consumer demand application to illustrate our methods. In theory, utility or indirect utility functions represent choices exactly. Following common practice, we include a random component, or disturbance, to choices in order to obtain a data density and do likelihood-based statistical inference. However, we take an unconventional approach to the specification of the random disturbance.

Usually a measurement error approach is taken, whereby an error distribution, unrelated to preference, accounts for discrepancies between observed choices and choices which maximize utility. In contrast, we use a model for observed choices where distributions over choices are given by the utility function itself. Specifically, if $u: X \to \mathbb{R}$ is the utility function over the universe $X$ of choices, then the distribution of observed choices is proportional to $\exp(u(x))$ on the frontier of the set of feasible choices. Theil (1974) and McCausland (2004) give very different theoretical underpinnings. There are several advantages of this approach. First, it is theoretically grounded. In usual practice, distributions of disturbances are given without theoretical justification. Second, the specification is parsimonious: a single function describes not only how choices broadly respond to changes in prices and income, but also the distribution of demand on any given budget. Third, the fit of an observed choice is measured by the relative desirability of the choice and its feasible alternatives, rather than by some measure on the choice set. Varian (1990), in a paper on goodness-of-fit measures, argues for preferring the former to the latter. Finally, the theories of Theil (1974) and McCausland (2004) do not rule out violations of the axioms of revealed preference. In practice, such violations are sometimes observed. The theories are more forgiving than standard consumer theory, without being undisciplined.

We analyze data from the Harbaugh et al. (2001) "GARP for Kids" experiment, undertaken in a study of the development of rational behavior. Subjects are 31 second grade students, 42 sixth grade students and 55 undergraduates. There are two goods, chips and juice, in indivisible packages. There are no prices and income as such: subjects are offered a budget of choices directly, and the budgets do not include off-frontier bundles. Fig. 2 illustrates the eleven different budgets.

The experiment has the following features.

1. Choices are individual, rather than aggregate, so consumer theory (and in particular the theory of random consumer demand in McCausland (2004)) applies.
2. Consumers select bundles from several different budgets, in the knowledge that after all decisions are made, exactly one of the budgets will be selected at random, and the consumer will be given their choice from only that budget. We can thus plausibly consider choices as being simultaneous or static, rather than dynamic.
3. Consumers have the opportunity to go back and change earlier choices, before a budget is selected at random. This mitigates the problem of learning during the experiment.
4. Choices are recorded in a laboratory. We can be fairly confident that measurement error is not a problem.
5. The number of goods, the coarseness of the indivisibilities, and prices and income are such that the number of possible choices is small. The likelihood function is therefore easily computed.

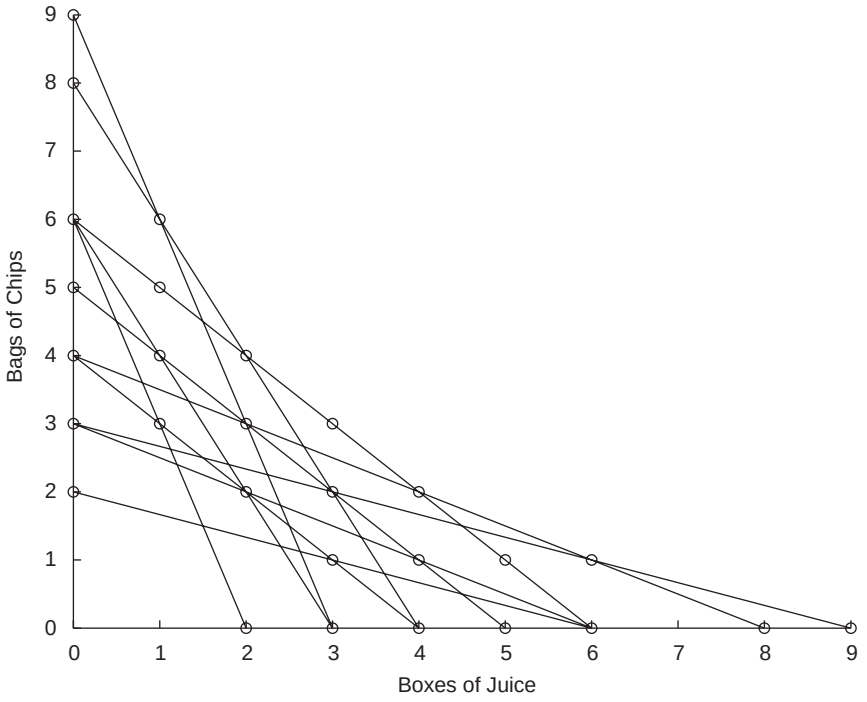


Fig. 2. Budgets for the "GARP for Kids" experiment.

Using the theory of Theil (1974) or that of McCausland (2004), we obtain the following distribution for the $t$th choice $x_t$:

$$\Pr(x_t = x) = \begin{cases} \dfrac{\exp u(x)}{\sum_{y \in B_t} \exp u(y)}, & x_t \in B_t, \\ 0 & \text{otherwise}, \end{cases} \tag{16}$$

where $B_t$ is the set of choices on the frontier of the $t$th choice set, which in this experiment is the choice set itself. We assume that the choices $x_t$ are independent and appeal to points 2 and 3 above to justify this assumption.

A few comments on the applicability of the theory by Theil (1974) are in order, as its relevance might not be immediately clear to the reader consulting it. Theil assumes that the choice set is $\mathbb{R}^n$ for some $n$, and uses a quadratic approximation for $u$ that leads to multivariate normal choice distributions. We point out that the main result of Theil (1974) also holds for finite choice sets, and that we have no need for the approximation: if we use $u$ directly, we obtain choice distributions given by (16).

### 5.1. Prior specification

The following choices define the prior distribution over the constellation index $k$ and the vector $\lambda$ of coefficients. The restricted consumption set is $\bar{X} = [0, 12]^2$, which is compact and contains all eleven budgets. The constants defining the transformation $\phi$ are $x^* = (1.0, 1.0)$ and $\xi = (0.1, 0.1)$. Thus $\phi$ is given by

$$\begin{aligned} \phi(x_1, x_2) &= \left( \log\left(\frac{x_1 + \xi_1}{x_1^* + \xi_1}\right), \log\left(\frac{x_2 + \xi_2}{x_2^* + \xi_2}\right) \right) \\ &= \left( \log\left(\frac{x_1 + 0.1}{1.1}\right), \log\left(\frac{x_2 + 0.1}{1.1}\right) \right). \end{aligned}$$

The multi-index constellations are rectangular lattices given by

$$I_k \equiv \{\imath \in \mathbb{N}_0^n : \imath \neq 0 \text{ and } 0 \leqslant \imath_i \leqslant k-1 \text{ for } i = 1, \ldots, n\}.$$

The multi-indices in $I_3$, for example, are (0,1), (0,2), (1,0), (1,1) (1,2), (2,0), (2,1) and (2,2), corresponding to the basis functions $\phi_2$, $\phi_2^2$, $\phi_1$, $\phi_1\phi_2$, $\phi_1\phi_2^2$, $\phi_1^2$, $\phi_1^2\phi_2$, and $\phi_1^2\phi_2^2$.

The prior on $k$, the constellation index, is given by $\pi_k = 2^{-(k-1)}$ for $k \geqslant 2$. Thus we have three terms with probability $\frac{1}{2}$, eight with probability $\frac{1}{4}$, 15 with probability $\frac{1}{8}$, and so on. The conditional prior density of $\lambda$ given $k$ has the form given in Eq. (12), where $v$ is the vector such that $(\Delta u_1 + \Delta u_2)/2 = v'\lambda$, $g$ is the gamma density with shape parameter 4 and scale parameter 50, and $h$ is the factor defined in (13), with $\beta_1 = \beta_2 = 0.5$, $\delta = 0.001$ and $p = k - 1$.

### 5.2. Results

We present results for the "GARP for Kids" experiment. The objective here is not to study the development of rational behavior in children, and so we report results only for the benchmark undergraduate subjects.

Table 1 shows conditional posterior probabilities for $k$ given $k \in \{2, 3, 4, 5\}$, for subjects 1074 through 1103, the first 30 of 55. We use the method of Newton and Raftery (1994) to compute $f(x_1, \ldots, x_{11}|k)$ for each value of $k$.

Table 2 lists the log marginal likelihoods for all 55 subjects, given $k \in \{2, 3, 4, 5\}$. They are in row major order, so the first row gives the marginal likelihoods for subjects 1074 through 1078. Here, the marginal likelihood is the marginal probability the model, including prior, assigns to the sequence of observed choices that a subject makes. Standard errors for the numerical approximation of the log marginal likelihoods

Table 1
Posterior probabilities of $k$ for the first 30 Subjects in the "GARP for Kids" experiment

| Subject | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
|---|---|---|---|---|
| 1074 | 0.268031 | 0.407503 | 0.219064 | 0.105402 |
| 1075 | 0.345172 | 0.324044 | 0.184888 | 0.145895 |
| 1076 | 0.258968 | 0.462476 | 0.203739 | 0.074816 |
| 1077 | 0.967457 | 0.032059 | 0.000479 | 0.000005 |
| 1078 | 0.968034 | 0.031557 | 0.000405 | 0.000004 |
| 1079 | 0.198167 | 0.339118 | 0.300826 | 0.161889 |
| 1080 | 0.025543 | 0.821864 | 0.147249 | 0.005344 |
| 1081 | 0.233941 | 0.318601 | 0.266747 | 0.180711 |
| 1082 | 0.359153 | 0.390829 | 0.172782 | 0.077236 |
| 1083 | 0.309870 | 0.333014 | 0.223047 | 0.134069 |
| 1084 | 0.363238 | 0.184935 | 0.419443 | 0.032384 |
| 1085 | 0.966052 | 0.033497 | 0.000446 | 0.000005 |
| 1086 | 0.125813 | 0.624758 | 0.222555 | 0.026874 |
| 1087 | 0.817362 | 0.100623 | 0.075123 | 0.006892 |
| 1088 | 0.414538 | 0.250175 | 0.197775 | 0.137512 |
| 1089 | 0.415736 | 0.303844 | 0.176036 | 0.104384 |
| 1090 | 0.072000 | 0.025861 | 0.900104 | 0.002035 |
| 1091 | 0.632163 | 0.222689 | 0.105932 | 0.039216 |
| 1092 | 0.968132 | 0.031479 | 0.000385 | 0.000004 |
| 1093 | 0.257544 | 0.405650 | 0.222257 | 0.114550 |
| 1094 | 0.269312 | 0.462051 | 0.203084 | 0.065554 |
| 1095 | 0.967018 | 0.032522 | 0.000457 | 0.000003 |
| 1096 | 0.101630 | 0.505159 | 0.335330 | 0.057882 |
| 1097 | 0.265254 | 0.415457 | 0.226562 | 0.092727 |
| 1098 | 0.371295 | 0.358028 | 0.181920 | 0.088757 |
| 1099 | 0.894937 | 0.088172 | 0.016743 | 0.000148 |
| 1100 | 0.118235 | 0.306723 | 0.319137 | 0.255905 |
| 1101 | 0.970699 | 0.028890 | 0.000407 | 0.000004 |
| 1102 | 0.967782 | 0.031750 | 0.000465 | 0.000004 |
| 1103 | 0.427316 | 0.360936 | 0.129459 | 0.082289 |

Table 2
Log marginal likelihoods for all subjects in the "GARP for Kids" experiment, in row major order

| | | | | |
|---|---|---|---|---|
| −7.351133 | −4.608758 | −9.940656 | −4.197768 | −4.183579 |
| −14.935639 | −6.301195 | −15.442465 | −8.161736 | −16.676400 |
| −16.713183 | −4.233398 | −17.822427 | −10.557720 | −30.356065 |
| −22.978277 | −24.813786 | −23.485596 | −4.202415 | −7.330709 |
| −9.970015 | −4.236868 | −16.950571 | −7.352623 | −8.396829 |
| −10.407859 | −12.575744 | −4.160754 | −4.224852 | −8.894845 |
| −8.242575 | −21.439450 | −11.572254 | −12.535711 | −18.998176 |
| −7.359645 | −7.946763 | −14.665421 | −18.625966 | −4.218446 |
| −9.468318 | −9.972209 | −21.650330 | −8.078603 | −13.919956 |
| −22.521666 | −15.678075 | −18.931139 | −7.500323 | −21.632655 |
| −20.188184 | −4.211171 | −16.273644 | −7.366309 | −22.680512 |

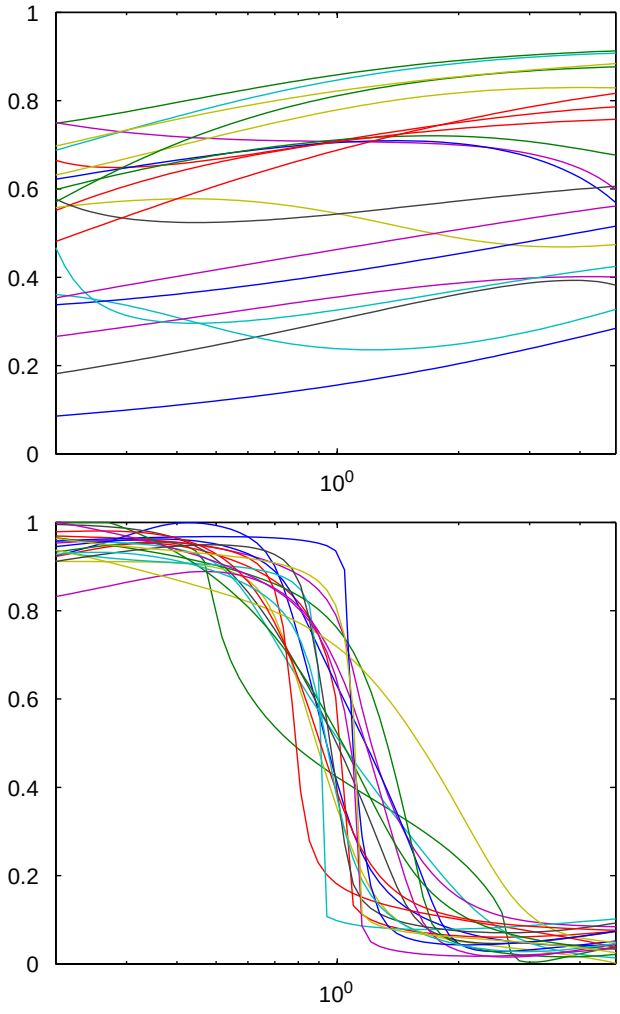


Fig. 3. Prior and posterior scattergraph, for subject 1105, of the modal expenditure share of good one versus $w_1/w_2$, where $w_1$ and $w_2$ are income-normalized prices adding to 1.

reported in Table 2 are all less than 0.15. The average log marginal likelihood is −12.49. Subject 1088 (row 3, column 5) always spent all his income on the more *expensive* good. This exuberant irrationality earned him a log marginal likelihood of −30.36, by far the lowest.

Table 3
Counts of numbers of GARP violations

| Number of violations | Experimental subjects | All sequences |
|---|---|---|
| 0 | 36 | 108,846 |
| 1 | 0 | 0 |
| 2 | 1 | 140,788 |
| 3 | 5 | 171,718 |
| 4 | 0 | 272,978 |
| 5 | 7 | 438,074 |
| 6 | 1 | 646,288 |
| 7 | 0 | 928,790 |
| 8 | 0 | 1,567,246 |
| 9 | 2 | 2,081,452 |
| 10 | 1 | 2,555,030 |
| 11 | 2 | 3,184,790 |
| Total | 55 | 12,096,000 |

Subject 1105 (row 7, column 2) always spent his income on the cheaper good, and chose equal quantities of the goods when their prices were equal. Although rational, these choices require very specific abrupt changes in expenditure shares in response to small changes in relative prices. Only a small set of $\lambda$ account for this behavior, and the marginal likelihood is low $-21.43$. However, we can see in Fig. 3 that the posterior distribution puts high probability on these abrupt changes. Both panels of Fig. 3 show 20 curves. Each curve gives the mode of the expenditure share of good one as a function of the ratio $w_1/w_2$ of income-normalized prices $w_1$ and $w_2$, adding to one. The first panel shows curves for 20 draws of the utility function from its conditional prior distribution given $k = 4$. The second panel does the same for the conditional posterior distribution given $k = 4$ and the observed choices of Subject 1105. The curves with the most abrupt changes tend to be associated with high amplitude utility functions, for which the expenditure share distribution is quite tight. The smoother curves tend to correspond to more diffuse expenditure share distributions.

To put the log marginal likelihoods in perspective, we consider the average log marginal likelihood arising from various models. The model assigning equal probability to all possible sequences of eleven choices implies a log marginal likelihood of $-16.31$ for every subject. A model which correctly and with certainty predicts the behavior of all subjects on all budgets implies a log marginal likelihood of zero for every subject. Any model that assigns probability zero to every sequence featuring at least one violation of the generalized axiom of revealed preference (GARP) gives a log marginal likelihood of negative infinity to the sequences of the 19 out of 55 subjects who violated the GARP, and therefore an average log marginal likelihood of negative infinity.

We use the data in Table 3 to derive a maximum log marginal likelihood of $-13.36$ over all models assigning equal probabilities to all sequences with the same number of GARP violations. The second column gives, for the number of GARP violations in the first column, the number of subjects having that number of violations. The third column gives the total number of distinct sequences of eleven choices having that number of violations.

## 6. Conclusions

We have pointed out that instead of approximating a function $u$ on a compact restricted domain $\bar{X}$, we can approximate $u \circ \phi^{-1}$ on $\phi(\bar{X})$, where $\phi$ is a function one can choose to be non-decreasing and concave, facilitating the approximation of non-decreasing concave functions. With modest restrictions on $\phi$, we can simultaneously approximate the function, its gradient and Hessian. We apply this idea using basis functions that are polynomials, but we note that we could easily do the same with the sinusoidal basis functions of the Fourier flexible form or the basis functions of the Müntz–Szasz expansion.

We have put considerable emphasis on prior elicitation, to which we believe the literature has not paid sufficient attention. We have described an approach to prior elicitation that puts more emphasis on economically relevant quantities than on the parameters themselves. We have shown that there are tight corners of the parameter space which, despite their low volume, may be quite plausible in many applications. We have shown that by concentrating prior probability in these corners, we can obtain more diffuse prior distributions over economically relevant quantities such as marginal rates of substitution and ratios of quantities demanded.

We believe this is relevant for applied economic research. Practitioners should be aware that flat priors on the parameter space can imply quite informative priors about economically relevant quantities. We recommend prior simulation to ensure that these quantities have reasonable prior distributions. We have provided an example of how one can modify a prior if it implies too informative a distribution for these quantities.

We have noted that the tight corners of the parameter space also pose a problem for prior and posterior simulation and that this problem is aggravated by the concentration of prior probability within them. We have introduced Metropolis–Hastings chains that are able to move quickly in and out of these tight corners, and thereby efficiently sample the prior and posterior distributions.

We have demonstrated the use of these prior distributions and simulation methods for the analysis of data from a consumer experiment.

## Acknowledgments

We appreciate comments by John Geweke, Narayana Kocherlakota, Dale Poirier, John Stevens, Walter Thurman and three referees. We thank William Harbaugh for providing experimental data. Software for computations reported in this paper is available from the author.

## Appendix A. On the boundedness of cone truncations

This appendix is on the boundedness of a truncation of the cone $\Lambda^k_{\bar{X}}$ defined in Eq. (3). This boundedness is relevant for the construction of proper prior distributions on $\Lambda^k_{\bar{X}}$ as described in Section 3.2.

**Result A.1.** *The set* $\{\lambda \in \Lambda^k_{\bar{X}} : u(\bar{x};\lambda) \leqslant 1\}$ *is bounded.*

**Proof.** We first establish bounds on the values of $u(x;\lambda)$ on the subset $[x_1^*, \bar{x}_1] \times \cdots \times [x_n^*, \bar{x}_n]$ of the restricted domain $\bar{X}$.

**Claim A.1.** *For every* $x \in [x_1^*, \bar{x}_1] \times \cdots \times [x_n^*, \bar{x}_n]$ *and every* $\lambda \in \{\lambda \in \Lambda^k_{\bar{X}} : u(\bar{x};\lambda) \leqslant 1\}$,
$0 \leqslant u(x;\lambda) \leqslant 1$.

**Proof.** The claim follows directly from the monotonicity of $u(\cdot;\lambda)$, the fact that $u(x^*;\lambda) = 0$, and the fact that $u(\bar{x};\lambda) \leqslant 1$. □

We now bound $\{\lambda \in \Lambda^k_{\bar{X}} : u(\bar{x};\lambda) \leqslant 1\}$ by enclosing it in a hyper-parallelogram defined by $\{\lambda \in \mathbb{R}^{|I_k|} : 0 \leqslant C\lambda \leqslant (1, 1, \ldots, 1)\}$, where $C$ is a non-singular matrix. The non-singularity of $C$ guarantees that the hyper-parallelogram is bounded, and thus that $\{\lambda \in \Lambda^k_{\bar{X}} : u(\bar{x};\lambda) \leqslant 1\}$ is bounded.

Let $J = |I_k|$, and order the multi-indices $\imath \in I_k$ as $\imath^{(1)}, \ldots, \imath^{(J)}$. Choose vector $q = (q_1, \ldots, q_n)$ such that for every $i \in \{1, \ldots, n\}$,

1. there exist positive integers $m_N$ and $m_D$ such that $q_i = p_{2i}^{m_N} / p_{2i-1}^{m_D}$, where $p_i$ is the $i$th prime number, and
2. $[\phi_i(\frac{x^* + \bar{x}}{2}) / \phi_i(\bar{x})]^{1/J} \leqslant q_i \leqslant 1$.

A simple modification of the proof in Rudin (1976) of the denseness of the rational numbers in the reals shows that we can do this. The inequalities $x^* < (x^* + \bar{x})/2 < \bar{x}$ ensure that we are taking the $J$th root of a positive real number strictly less than one.

Now define, for all $j \in \{1, \ldots, J\}$,

$$z_j \equiv (q_1^j \phi_1(\bar{x}), \ldots, q_n^j \phi_n(\bar{x})), \quad x_j \equiv \phi^{-1}(z_j)$$

and

$$C \equiv \begin{bmatrix} z_1^{t^{(1)}} & \cdots & z_1^{t^{(J)}} \\ \vdots & \ddots & \vdots \\ z_J^{t^{(1)}} & \cdots & z_J^{t^{(J)}} \end{bmatrix}.$$

For all $\lambda \in \Lambda_{\bar{X}}^k$,

$$C\lambda = [u(x_1; \lambda), \ldots, u(x_J; \lambda)]' = [u(\phi^{-1}(z_1); \lambda), \ldots, u(\phi^{-1}(z_J); \lambda)]'$$

and for all $j \in \{1, \ldots, J\}$, $x_j = \phi^{-1}(z_j) \in [x_1^*, \bar{x}_1] \times \cdots \times [x_n^*, \bar{x}_n]$. Claim A.1 gives us $(0, \ldots, 0)' \leqslant C\lambda \leqslant (1, \ldots, 1)'$.

We now show that $C$ is non-singular.

**Claim A.2.** *$C$ is non-singular.*

**Proof.** $C$ can be written as

$$\begin{bmatrix} (q^{t^{(1)}})^1 [\phi(\bar{x})]^{t^{(1)}} & \cdots & (q^{t^{(J)}})^1 [\phi(\bar{x})]^{t^{(J)}} \\ \vdots & \ddots & \vdots \\ (q^{t^{(1)}})^J [\phi(\bar{x})]^{t^{(1)}} & \cdots & (q^{t^{(J)}})^J [\phi(\bar{x})]^{t^{(J)}} \end{bmatrix} = \begin{bmatrix} (q^{t^{(1)}})^1 & \cdots & (q^{t^{(J)}})^1 \\ \vdots & \ddots & \vdots \\ (q^{t^{(1)}})^J & \cdots & (q^{t^{(J)}})^J \end{bmatrix} \cdot \operatorname{diag}([\phi(\bar{x})]^{t^{(1)}}, \ldots, [\phi(\bar{x})]^{t^{(J)}}).$$

We will show that both these factors are non-singular, which will then imply that $C$ is non-singular. The first factor is a Vandermonde matrix, and to establish its non-singularity, it suffices to show that for all $j, l \in \{1, \ldots, J\}$, $j \neq l \Rightarrow q^{t^{(j)}} \neq q^{t^{(l)}}$. This follows from the fact that there is a unique representation of any rational number as the ratio of two integers with no common factors, and unique prime factorizations of the two integers. The second factor is a diagonal matrix whose elements are non-zero, and so it is also non-singular. Since the two factors are non-singular, so is $C$. □

## Appendix B. Sensitivity analysis for prior simulations

We made the case in Section 3 that a uniform prior over basis function coefficients implies priors over economically relevant features of the regular function that are quite informative. We chose particular values of the fixed parameters of the functional form and focused on the value of

$$\log(\partial u(x; \lambda)/\partial x_1) - \log(\partial u(x; \lambda)/\partial x_2) \tag{17}$$

at the single point $x^*$.

In this appendix, we give the results of a sensitivity analysis that shows how the distribution of the quantity (17) depends on the point $x$ at which we evaluate (17) and values of the fixed parameters.

Table 4 shows estimates of the prior means and standard deviations of the value of (17) for seven different values of $(x_1, x_2)$. The base case corresponds to the same choice of fixed parameters as in Section 3. That is, $n = 2$, $k = 4$, $\bar{x} = (1.0, 1.0)$, $\xi = (0.01, 0.01)$ and $x^* = (0.1, 0.1)$.

The other rows illustrate the effect of various changes to the fixed parameters. All changes are relative to the base case. In the second and third rows, we decrease and increase $x^*$ for fixed $\xi$. In the fourth and fifth rows, we multiply and divide the $\xi_i$ by 4, respectively, adjusting the value of $x_i^*$ to maintain their equality to the geometric mean of $\xi_i$ and $\bar{x}_i$. In the fifth and sixth rows, the number of basis functions is first decreased to 8 (corresponding to $k = 3$) then increased to 24 (corresponding to $k = 5$).

The final row shows what happens when the number of goods increases. We take $\xi = (0.01, 0.01, 0.01)$, $x^* = (0.1, 0.1, 0.1)$ and $\bar{x} = (1.0, 1.0, 1.0)$. When evaluating (17) for the various values of $(x_1, x_2)$ we take $x_3 = x_3^*$.

Table 4
Prior means and Standard deviations of $\log(\partial u(x;\lambda)/\partial x_1) - \log(\partial u(x;\lambda)/\partial x_2)$, for various values of $(x_1, x_2)$ and fixed parameters

| Case | $(0,0)$ | $(\frac{1}{4},\frac{1}{4})$ | $(\frac{1}{2},\frac{1}{2})$ | $(1,1)$ | $(\frac{1}{2},0)$ | $(1,0)$ | $(\frac{3}{4},\frac{1}{4})$ |
|---|---|---|---|---|---|---|---|
| Base case | 0 | 0 | 0 | 0 | −0.36 | −0.22 | 0.24 |
| | (0.84) | (0.75) | (0.81) | (1.08) | (0.95) | (1.11) | (0.86) |
| $x_1^* = x_2^* = 0.025$ | 0 | 0 | 0 | 0 | −0.49 | −0.37 | 0.21 |
| | (0.75) | (0.72) | (0.83) | (1.19) | (0.93) | (1.10) | (0.87) |
| $x_1^* = x_2^* = 0.4$ | 0 | 0 | 0 | 0 | −0.20 | −0.02 | 0.28 |
| | (0.88) | (0.70) | (0.73) | (0.92) | (0.89) | (1.03) | (0.79) |
| $\xi_1 = \xi_2 = 0.04$ | 0 | 0 | 0 | 0 | −0.41 | −0.35 | 0.09 |
| $x_1^* = x_2^* = 0.2$ | (0.79) | (0.69) | (0.74) | (1.09) | (0.87) | (1.11) | (0.82) |
| $\xi_1 = \xi_2 = 0.0025$ | 0 | 0 | 0 | 0 | −0.30 | −0.15 | 0.28 |
| $x_1^* = x_2^* = 0.05$ | (0.88) | (0.80) | (0.87) | (1.09) | (1.02) | (1.14) | (0.89) |
| $k = 3$ | 0 | 0 | 0 | 0 | −0.24 | −0.42 | −0.17 |
| | (0.89) | (0.78) | (0.86) | (1.16) | (1.00) | (1.22) | (0.90) |
| $k = 5$ | 0 | 0 | 0 | 0 | −0.40 | −0.66 | 0.03 |
| | (0.84) | (0.62) | (0.62) | (1.20) | (0.82) | (1.11) | (0.69) |
| $n = 3$ | 0 | 0 | 0 | 0 | −0.32 | −0.19 | 0.20 |
| | (0.55) | (0.50) | (0.52) | (0.69) | (0.65) | (0.77) | (0.58) |

Means that we know are zero because of symmetry are given as 0. All other quantities are obtained by prior simulation, as described in Section 4. Numerical standard errors for the means in the last row are all less than 0.025. Numerical standard errors for all other means are less than 0.015.

## References


Barnett, W.A., Jonas, A., 1983. The Müntz–Szasz demand system: an application of a globally well behaved series expansion. Economics Letters 11, 337–342.

Barnett, W.A., Yue, P., 1988. Semiparametric estimation of the asymptotically ideal model: The AIM demand system. In: Rhodes, G., Fornby, T. (Eds.), Non-parametric and Robust Inference, Advances in Econometrics, vol. 7, JAI Press, Greenwich, Connecticut, pp. 229–252.

Barnett, W.A., Geweke, J.F., Wolfe, M., 1991a. Seminonparametric Bayesian estimation of the asymptotically ideal production model. Journal of Econometrics 49, 5–50.

Barnett, W.A., Geweke, J.F., Yue, P., 1991b. Seminonparametric Bayesian estimation of the asymptotically ideal model: the AIM consumer demand system. In: Barnett, W.A., Powell, J., Tauchen, G.E. (Eds.), Nonparametric and semiparametric methods in econometrics and statistics. Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, New York and Melbourne, pp. 127–173 (Chapter 6).

Bauwens, L., Bos, C.S., van Dijk, H.K., van Oest, R.D., 2004. Adaptive radial-based direction sampling: some flexible and robust Monte Carlo integration methods. Journal of Econometrics 123, 201–225.

Christensen, L.R., Jorgenson, D., Lau, L.J., 1975. Transcendental logarithmic utility functions. American Economic Review 65 (3), 367–383.

Deaton, A., Muellbauer, J., 1980. Economics and Consumer Behavior. Cambridge University Press, Cambridge.

deFinetti, B., 1949. Sulle stratificazioni convesse. Ann. de Mat. Pura Appl. 30, 173–183.

Diewert, W.E., Wales, T.J., 1987. Flexible functional forms and global curvature conditions. Econometrica 55 (1), 43–68.

Evard, J.C., Jafari, F., 1994. Direct computation of the simultaneous Stone–Weierstrass approximation of a function and its partial derivatives in banach spaces, and combination with hermite interpolation. Journal of Approximation Theory 78, 351–363.

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

Gallant, A.R., Golub, G.H., 1984. Imposing curvature restrictions on flexible functional forms. Journal of Econometrics 26, 295–321.

Geweke, J., Petrella, L., 2000. Inference for regular functions using Müntz approximations. Unpublished manuscript.

Gordon, S., 1996. Using mixtures of flexible functional forms to estimate factor demand elasticities. Canadian Journal of Economics 29 (3), 717–736.

Harbaugh, W., Krause, K., Berry, T., 2001. GARP for kids: on the development of rational choice behavior. American Economic Review 91, 1539–1545.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Kannai, Y., 1974. Approximation of convex preferences. Journal of Mathematical Economics 1, 101–106.
Koop, G., Osiewalski, J., Steel, M.F.J., 1994. Bayesian efficiency analysis with a flexible form: the AIM cost function. Journal of Business and Economic Statistics 12 (3), 339–346.
Lau, L., 1986. Functional forms in econometric model building. In: Gliliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. 3. Elsevier Science Publishers, North-Holland, Amsterdam, pp. 1515–1566 (Chapter 26).
Matzkin, R.L., 1994. Restrictions of economic theory in nonparametric methods. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. Elsevier Science Publishers, North-Holland, Amsterdam, pp. 2523–2558 (Chapter 42).
McCausland, W.J., 2004. A theory of random consumer demand. Cahiers de recherche du Département de sciences économiques, Université de Montréal, no. 2004-04.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.
Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. Journal of the Royal Statistical Society Series B 3, 3–48.
Rudin, W., 1976. Principles of Mathematical Analysis. McGraw-Hill, New York.
Terrell, D., 1996. Incorporating monotonicity and concavity conditions in flexible functional forms. Journal of Applied Econometrics 11, 179–194.
Theil, H., 1974. A theory of rational random behavior. Journal of the American Statistical Association 69, 310–314.
Varian, H.R., 1990. Goodness-of-fit in optimizing models. Journal of Econometrics 46, 125–140.