On Bayesian Analysis and Computation for Functions with Monotonicity and Curvature Restrictions

William J. McCausland*
Département de sciences économiques, Université de Montréal,
CIREQ† and CIRANO‡
C.P. 6128, succursale Centre-ville
Montréal QC H3C 3J7
Canada
Email: william.j.mccausland@umontreal.ca
Phone: (514) 343-7281
Fax: (514) 343-7221
October 12, 2006

Abstract

Following many in the literature on inference for functions with monotonicity and curvature restrictions, we use a functional form consisting of finite linear combinations of basis functions. We consider Bayesian inference on the number of functions and their coefficients. Prior elicitation is difficult because of the irregular shape and infinite dimension of the parameter space. We show how to elicit priors that

*We appreciate comments by John Geweke, Narayana Kocherlakota, Dale Poirier, John Stevens and three referees. We thank William Harbaugh for providing experimental data. Software for computations reported in this paper is available from the author. We alone are responsible for any errors.
†Centre Interuniversitaire de Recherche en Économie Quantitative
‡Centre Interuniversitaire de Recherche en Analyse des Organisations
are flexible, theoretically consistent, and proper. We point out that some naively elicited priors assign very low probability to regions of the parameter space that many would consider plausible. We offer a solution to this problem. We introduce methods for prior and posterior simulation that meet challenges posed by the irregular shape of the parameter space. In an example, we analyze data from a consumer demand experiment.

Key words: Bayesian methods; Flexible functional forms; Shape restrictions
1 Introduction

There are many examples in economics where theoretically consistent choice behavior is described by multivariate functions subject to regularity constraints such as monotonicity and concavity. These include production, cost, profit, utility, expenditure and indirect utility 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 inference for such functions. 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, which 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 functions indexed by elements of the restricted parameter set are regular over their domain. If they are regular throughout the domain, we have global theoretical consistency. If they are regular at a point, we have local theoretical consistency. 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.

For example, the Constant Elasticity of Substitution (CES) class of utility functions, with non-negativity constraints on its parameters, is globally theoretically consistent but not very flexible. Popular “flexible functional form” classes of demand functions include the trans-log and Almost Ideal Demand System (AIDS) models. These classes are locally flexible in the sense that with appropriate choices of their parameters one can attain arbitrary elasticities at a given point. However, they are not globally theoretically consistent: there are values of the parameters for which the function is not everywhere regular on its domain. We cannot rule out these values without renouncing local flexibility.

There are at least three distinct approaches to approximating regular functions as linear combinations of basis functions spanning the space of continuous functions.

Gallant (1981) initiates this literature with his Fourier flexible form. Basis functions are sinusoidal, and any continuous function on a bounded subset of
$\mathbb{R}^n$ can be approximated arbitrarily closely in $L^2$ by a linear combination of a finite number of these basis functions. If the function has continuous derivatives, we can simultaneously approximate the function and these derivatives. This 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. For example, Shephard’s lemma and Roy’s identity give choices as functions of derivatives of the cost and indirect utility functions, respectively. The proximity of two functions in either $L^2$ or 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 regularity conditions can often be expressed in terms of derivatives. If we can simultaneously approximate derivatives, we can guarantee that the approximating function is regular.

Unfortunately, sinusoidal functions lie outside the space of regular functions and so it can take many terms to build up a particular function. Gallant (1981) offers a partial solution by adding a quadratic term to the expansion.

Checking for regularity is not straightforward. Gallant (1981) gives 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.

Barnett and Jonas (1983) use a multivariate Müntz-Szasz expansion. The set of basis functions is

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

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) 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 the basis functions are themselves regular. However, as with the Fourier flexible form, there are no known conditions that are easy to verify, necessary and sufficient for regularity. Non-negativity of all coefficients is easy to verify and sufficient for regularity, but not necessary. In fact, the constraint imposes substitutability on all inputs. Terrell (1994) uses simulations to show that even when inputs are
substitutes, much flexibility is lost. See also Koop, Osiewalski and Steel (1994) on this point.

In unpublished work, Geweke and Petrella have shown that the following set of basis functions also spans the space of continuous functions on a compact set in $\mathbb{R}^n_{++}$:

$$\left\{ \prod_{i=1}^{n} p_i^{b_i} : t \in \mathbb{N}_0^n \right\}, \quad (1)$$

where $b > 0$. The basis functions $\prod_{i=1}^{n} p_i^{b_i}$ satisfying $t_i < b^{-1}$, $i = 1, \ldots, n$, are themselves regular, which is convenient for constructing regular approximations with a small number of terms. As with the Fourier flexible form and the Müntz-Szasz expansion, 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).

In this paper, we also use the result of Evard and Jafari (1994), and we generalize the approximation result of Geweke and Petrella to a wide variety of sets of basis functions.

Inference for regular functions involves 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 much more easily handled using a Bayesian, rather than frequentist analysis. Inference on parameter values, number of terms and even competing functional forms\(^1\) is done 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 regular functions compatible with the observed data, individual functions may differ considerably on regions outside the sample. Averaging over these functions leads 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. We can put priors on certain features of functions and leave the prior

\(^1\)See Gordon, (1996) on Bayesian comparison of competing functional forms.
otherwise flat, but we should not confuse flatness for un informativeness: we will see that such priors can assign very low probability to regions of the parameter space that many will find plausible.

We believe that priors have received too little attention in the literature on Bayesian inference on flexible functional forms. In this paper, we discuss prior elicitation in some depth, identifying some of the associated problems and offering a solution.

Section 2 concerns the approximation of regular functions by linear combinations of a finite number of basis functions. Following Geweke and Petrella (unpublished), we apply a result from Evard and Jafari (1994) to show that any twice continuously differentiable regular function can be arbitrarily well approximated on a compact subset $X$ of its theoretical domain $X$ by a function that is regular on $\bar{X}$. The approximation is a simultaneous approximation of the function, its gradient and its Hessian.

In Section 3, we discuss prior distributions over flexible functional forms. We give sufficient conditions for a prior to be flexible and theoretically consistent. We say a prior is flexible if for any twice continuously differentiable function $u$ that is regular on $X$, there is a constant $c$ such that the prior assigns positive probability to every neighbourhood of $u - c$. It is theoretically consistent if it assigns zero probability to the set of functions that are irregular on $\bar{X}$. We offer an approach to prior elicitation that involves eliciting prior distributions over economically relevant quantities and not directly over 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.

In Section 4, we discuss the problem of prior and posterior simulation for parameters of the functional form. The shape of the parameter space presents special problems, and we propose simulation methods to meet this challenge.

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), and find that out-of-sample
predictions are better on average than any model based only on numbers of observed violations of the Generalized Axiom of Revealed Preference.

We conclude in Section 6.

2 Parametric Classes of Functions

Our objective is Bayesian learning, from choice data, about utility, indirect utility, expenditure, production, cost or profit functions, all of which have well-established theoretical properties. These properties can be classified as monotonicity properties (utility is increasing in quantities, indirect utility is decreasing in prices), curvature properties (utility is quasi-concave, profit is convex), homogeneity properties (indirect utility is homogeneous degree zero, expenditure is homogeneous degree one in prices), and continuity properties.

Each of these functions is defined on its theoretical domain $X$. It is often convenient to define equivalent functions on alternative domains, in order not to have to deal with homogeneity properties. For example, the indirect utility function is a homogeneous degree zero function $v(w, m)$ of prices $w$ and income $m$ on the domain $X = \mathbb{R}^{n+1}_{++}$. One can instead define the indirect utility function as a function of income-normalized prices $w_i/m$, $i = 1, \ldots, n$ on the domain $X = \mathbb{R}^{n}_{++}$.

In the case of homogeneous degree one functions, such as the profit function, we can restrict the domain to a subset such as $X = \{x \in \mathbb{R}^n_{++} : v'x = 1\}$, where $v$ is a positive vector. We can do this because the values of the function at points in the larger domain are uniquely determined by its values on $X$.

We need to 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. Parameters are the coefficients of the basis functions and their number.

Using results known to us, we can guarantee approximation only on compact sets, not unbounded sets such as the classical consumption set $X = \mathbb{R}^n_{++}$. We settle for approximation on a compact restricted domain $\tilde{X} \subseteq X$, 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 will also impose concavity or convexity on functions that theory tells us are quasi-convex or quasi-concave. This is constraining, since there are increasing quasi-concave functions for which there are no increasing concave functions with the same level curves. See de Finetti (1949) or the document “A Pedagogical Example of Non-concavifiable Preferences” by James Schummer, available at the author’s website. However, a result by Kannai (1973) on the approximate representation of complete, continuous, monotone and convex preferences using concave functions will reassure many that the restriction is not very severe. We impose concavity or convexity to simplify the shape of the parameter space, which is useful for prior elicitation and for simulation of the prior and posterior distributions.

Finally, we also 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.

For definiteness, we now define a function as regular if it is continuous, non-decreasing and concave. Other forms of regularity can be accommodated with obvious modifications.

### 2.1 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.

If we restrict the domain of homogeneous degree one functions in the way we described in the introduction to this section, then it is already compact. In other cases, \( X \) is the positive orthant of \( \mathbb{R}^n \), a non-compact set. One natural choice for \( \bar{X} \) is a hyper-rectangle of the form \([0, \bar{x}_1] \times \ldots \times [0, \bar{x}_n] \), for some \( \bar{x} \in \mathbb{R}^n_+ \). Another is a simplex of the form \( \{ x \in X : w^t x \leq m \} \), 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, trade 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.

2.2 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 \( I \), \( \sum_{\iota \in I} \lambda_i x^\iota \) 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 \( X \) with that of approximating \( u \circ \phi^{-1} \) on \( \phi(X) \).

In the simulations and empirical example of this paper, we use the following example of a transformation \( \phi \):

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

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 make \( \phi(\bar{X}) \) large, which adversely affects flexibility at other scales. The constant \( x^* \) establishes a reference value around which the monomials of first and second order are nearly “orthogonal”.

For the particular choice of \( \phi(x) \) in equation (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 \( \lambda_i / \sum_{j=1}^{n} \lambda_j \). Higher order monomials in \( \phi(x) \) allow the shares to change with prices and income.

Geweke and Petrella’s basis functions, given in equation (1), can be interpreted as monomials in \( \phi \) for the following alternative choice of \( \phi \):

\[
\phi(x) = (x_1^b, \ldots, x_n^b) \quad \forall x \in X. \tag{3}
\]
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 \( \mathbb{N}_0 \setminus \{0\} \). Excluding the multi-index 0 means excluding the constant term of \( u(\cdot; \lambda) \). We do this to make it easier to elicit proper priors. The exclusion means that we can only approximate a regular function \( u \) up to an additive constant, but this is innocuous. The set of polynomials \( \sum_{\epsilon \in I_k} \lambda_\epsilon [\phi(x)]^\epsilon \) becomes more flexible as \( k \) increases.

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

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

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

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

\[
\Lambda^k_X \equiv \{ \lambda \in \mathbb{R}^{|I_k|} : u(\cdot; \lambda) \text{ is regular on } \bar{X} \},
\]

We also define \( \Lambda_X \equiv \cup_{k=1}^{\infty} \Lambda^k_X \), the complete set of vectors \( \lambda \) associated with functions that are regular on \( \bar{X} \).

### 2.3 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 \( k \in \{1, 2, \ldots\} \), \( \Lambda^k_X \) is a convex cone.

**Proof.** The set of functions regular on \( \bar{X} \) is closed under addition and positive scalar multiplication. Therefore \( \Lambda^k_X \) is a convex cone. \( \Box \)

---

\(^2\)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 \( \cup_{k=0}^{\infty} I_k = I \).

\(^3\)For any set \( A \), we denote by \(|A|\) the cardinality of \( A \).
The following alternate proof yields some important insights that are relevant for later sections. We can express the regularity conditions as follows. For all \( x \in \tilde{X} \),
\[
\frac{\partial u(x; \lambda)}{\partial x} = \sum_{i \in I_k} \lambda_i \frac{\partial [\phi(x)]^i}{\partial x} \geq 0,
\]
and for all \( x \in X \) and \( v \in \mathbb{R}^n \),
\[
v^i \frac{\partial^2 u(x; \lambda)}{\partial x \partial x^i} v = \sum_{i \in I_k} \lambda_i v^i \frac{\partial^2 [\phi(x)]^i}{\partial x \partial x^i} v \leq 0.
\]

For every choice of \( x \in \tilde{X} \), equation (5) gives \( n \) linear inequalities in \( \lambda \), one for each component of the gradient. For every choice of \( x \in X \) and \( v \in \mathbb{R}^n \), equation (6) gives another linear inequality in \( \lambda \). We see, therefore, that the parameter space \( \Lambda^k_X \) is the intersection of the half spaces defined by the inequalities above. The half spaces are convex, so their intersection is as well.

The following approximation result tell us that we can simultaneously approximate any twice continuously differentiable function \( u \), regular on \( \tilde{X} \), together with its gradient and Hessian, arbitrarily closely on \( \tilde{X} \), 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 \( \tilde{X} \). In practice, it is much easier to check that the approximating function is regular on \( \tilde{X} \) than to verify its proximity to some function regular on \( \tilde{X} \). For notational convenience, we define the following norm for twice continuously differentiable functions:
\[
\| f \| \equiv \max \left[ \sup_{x \in \mathcal{X}} |f(x)|, \sup_{x \in \mathcal{X}, i \in \{1, \ldots, n\}} \left| \frac{\partial f(x)}{\partial x_i} \right|, \sup_{x \in \mathcal{X}, i, j \in \{1, \ldots, n\}} \left| \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right| \right].
\]

The result is a generalization of an unpublished result by Geweke and Petrella. 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 \( \phi: \bar{X} \to \mathbb{R} \) is such that \( \phi(X) \) is compact and that the inverse \( \phi^{-1} \) on \( \phi(X) \) exists and is twice continuously differentiable. Then for every twice continuously differentiable \( u: X \to \mathbb{R} \) regular on \( \bar{X} \), and every \( \epsilon > 0 \), there exists a \( \lambda \in \Lambda_X \) and a constant \( c \) such that

\[
\|c + u(\cdot; \lambda) - u(\cdot)\| < \epsilon
\]

(7)

Proof. Let \( u \) be a twice continuously differentiable regular function and let \( \epsilon > 0 \).

Rather than approximating \( u \) directly, which may be on the boundary of the regular region, we will approximate a nearby function \( \hat{u}: \bar{X} \to \mathbb{R} \) 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

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

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,

\[
\|\hat{u}(\cdot) - u(\cdot)\| \leq \frac{\epsilon}{2}. \tag{8}
\]

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

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

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 \( \phi(X) \). Furthermore, \( \phi(X) \) is compact. Therefore the corollary implies that for all \( \epsilon' > 0 \), there exists a \( \lambda \in \bigcup_{k=1}^{\infty} \mathbb{R}^{[1k]} \) and a \( c \) such that for all \( i, j \in \{1, \ldots, n\} \) and all \( z \in \phi(X) \),

\[
\left| c + \sum_{i \in I_k} \lambda_i z^i - (\hat{u} \circ \phi^{-1})(z) \right| < \epsilon'.
\]
\[
\left| \frac{\partial}{\partial z_i} \sum_{i \in I_k} \lambda_i z^i - \frac{\partial}{\partial z_i} (\hat{u} \circ \phi^{-1})(z) \right| < \epsilon'
\]

and
\[
\left| \frac{\partial^2}{\partial z_i \partial z_j} \sum_{i \in I_k} \lambda_i z^i - \frac{\partial^2}{\partial z_i \partial z_j} (\hat{u} \circ \phi^{-1})(z) \right| < \epsilon'.
\]

The function $\phi$ maps $\bar{X}$ to $\phi(\bar{X})$, and therefore for all $i, j \in \{1, \ldots, n\}$ and $x \in \bar{X}$,
\[
\left| c + \sum_{i \in I_k} \lambda_i [\phi(x)]^i - \hat{u}(x) \right| = |c + u(x; \lambda) - \hat{u}(x)| < \epsilon',
\]
\[
\left| \frac{\partial}{\partial x_i} u(x; \lambda) - \frac{\partial}{\partial x_i} \hat{u}(x) \right| < \epsilon'M_1
\]
and
\[
\left| \frac{\partial^2}{\partial x_i \partial x_j} u(x; \lambda) - \frac{\partial^2}{\partial x_i \partial x_j} \hat{u}(x) \right| < \epsilon'M_2,
\]
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 $\epsilon'$ such that
\[
\|c + u(x; \lambda) - \hat{u}(x)\| < \frac{\epsilon}{2}, \quad (9)
\]
and for all $x \in \bar{X}$ and all $i, j \in \{1, \ldots, n\},$
\[
\frac{\partial u(x; \lambda)}{\partial x} \geq 0, \quad (10)
\]
and
\[
v \frac{\partial^2 u(x; \lambda)}{\partial x \partial x'} v \leq 0 \quad \forall v \in \mathbb{R}^n. \quad (11)
\]

Equations (8), (9), and the triangle inequality guarantee that (7) holds. Equations (10) and (11) guarantee that $\lambda \in \Lambda_{\bar{X}}$. $\square$
3 Priors

We have just described a parametric class 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 $\lambda = (\lambda_i)_{i \in I_k}$. 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 used in 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 achieving both. We then propose a basic approach for eliciting prior distributions over $\Lambda_X^k$ for fixed $k$, involving the elicitation of priors for economically relevant quantities depending on the parameters. We give conditions for these priors to be proper. We identify an inconvenient property of these priors and offer a solution.

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 $\Lambda_X^k$, then the prior is flexible and theoretically consistent.

Proof. Let $u: X \rightarrow \mathbb{R}$ be twice continuously differentiable and regular on $\bar{X}$, and let $\epsilon > 0$. By Result 2.2, we can find a $k \in \mathbb{N}$ and a $\lambda^* \in \Lambda_X^k$ such that $\|c + u(x; \lambda^*) - u(x)\| < \epsilon/2$. The conditional distribution $\lambda|k$, whose support is $\Lambda_X^k$, assigns positive probability to the set

$$\Lambda^* \equiv \left\{ \lambda \in \Lambda_X^k : |\lambda^*_i - \lambda_i| < \frac{\epsilon}{2\|I_k\|\|\phi(x)^i\|} \quad \forall i \in I_k \right\}.$$

For all $\lambda \in \Lambda^*$, $\|u(\cdot; \lambda) - u(\cdot; \lambda^*)\| < \frac{\epsilon}{2}$, and therefore $\|c + u(\cdot; \lambda) - u(\cdot)\| < \epsilon$. 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 the \( \| \cdot \| \)-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 \( \Lambda_X \). □

3.2 Conditional Prior Distributions on \( \Lambda_X^k \)

The conditional priors \( \lambda \mid k \) are distributions on the \( \Lambda_X^k \). Except for being convex cones, the \( \Lambda_X^k \) are irregularly shaped, and this makes elicitation difficult. We look for conditional priors with the following properties. They should be proper, as this guarantees that the conditional posterior distribution is proper and allows us to do inference on \( k \) and to compare models using Bayes factors. They should assign sufficient probability to all plausible regions of \( \Lambda_X^k \). Finally, we would like to be able elicit priors on economically relevant quantities rather than the parameters themselves.

Our first approach to elicitation takes advantage of the fact that \( \Lambda_X^k \) is a convex cone. We choose a vector \( v \in \mathbb{R}^{\lvert I_k \rvert} \) such that for all \( c > 0 \), the intersection of the hyper-plane \( v' \lambda = c \) with \( \Lambda_X^k \) is non-empty and has a finite volume. For some choices of \( v \), \( v' \lambda \) has an economically relevant interpretation. For example, we can express the difference \( u(\bar{x}) - u(0) \) as \( v' \lambda \), where \( v \) is the vector with elements \( v_i = [\phi(\bar{x})] - [\phi(0)]^t \), \( i \in I_k \). For other choices of \( v \), \( v' \lambda \) is a linear combination of differences and/or gradient components.

While the irregular shape of the intersection of \( v' \lambda \) with \( \Lambda_X^k \) makes computation of its volume intractable, the conic nature of \( \Lambda_X^k \) means that this volume is proportional to \( c^{\lvert I_k \rvert - 1} \). Thus any prior density \( f \) defined up to a multiplicative constant by

\[
f(\lambda) \propto \begin{cases} 
(v' \lambda)^{-\lvert I_k \rvert - 1} \cdot h(v' \lambda) & \lambda \in \Lambda_X^k, \\
0 & \lambda \notin \Lambda_X^k,
\end{cases}
\]

where \( h \) is any proper density on \( \mathbb{R}^{++} \), is a proper density on \( \Lambda_X^k \). The density \( f \) is constant on each hyperplane \( v' \lambda = c \), and the implied distribution of \( v' \lambda \) has density \( h \).

Recall that we require a vector \( v \in \mathbb{R}^{\lvert I_k \rvert} \) such that for \( c > 0 \), the intersection of the hyper-plane \( v' \lambda = c \) with \( \Lambda_X^k \) has a finite volume. For the purpose of establishing the finiteness of this volume, we prove in the appendix that
the set \( \{ \lambda \in \Lambda_X^k : u(\bar{x}; \lambda) \leq 1 \} \) is bounded. Thus, to show that the intersection has finite volume, it suffices to find a upper bound for \( u(\bar{x}; \lambda) \) on the intersection of the hyper-plane \( v' \lambda = c \) with \( \Lambda_X^k \). For example, let \( v \) be the vector such that \( v' \lambda = u(\bar{x}; \lambda) - u(0; \lambda) \). If \( \lambda \in \Lambda_X^k \), then \( u \) is non-increasing, and so \( u(\bar{x}; \lambda) \leq u(0; \lambda) + c \leq u(x^*; \lambda) + c = c \). Result A.1 implies that \( \{ \lambda \in \Lambda_X^k : u(\bar{x}; \lambda) \leq c \} \) is bounded, and so the intersection of the hyper-plane \( v' \lambda = c \) with \( \Lambda_X^k \) is bounded and thus has finite volume.

For the examples of this paper, we choose \( v \) such that

\[
v' \lambda = \sum_{i=1}^{n} \frac{\Delta u_i(\lambda)}{\beta_i},
\]

where \( \Delta u_i(\lambda) \) is the increase in \( u \) as \( x_i \) goes from 0 to \( \bar{x}_i \), with all other \( x_j \) set to \( x_j^* \). 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 choose a gamma density with shape parameter \( \alpha \) and scale parameter 1 for the prior density \( h \) on \( v' \lambda \). The \( \beta_i \) determine the scale, for each \( i \) independently.

The first panel of Figure 1 shows a scatterplot of the joint prior distribution of \( \Delta u_1 \) and \( \Delta u_2 \) for the following choices of fixed parameters: \( n = 2, \xi = (0.1, 0.1), x^* = (1.0, 1.0), \bar{x} = (12.0, 12.0), \alpha = 4 \) and \( \beta = (25.0, 25.0) \).

We see that despite being flat on the the truncated hyperplanes \( \{ v \in \Lambda_X^k : \Delta u_1(\lambda)/\beta_1 + \Delta u_2(\lambda)/\beta_2 = c \}, c > 0 \), the prior is quite informative about the relative magnitudes of \( \Delta u_1 \) and \( \Delta u_2 \). Specifically, it puts very low probability on the regions where either \( \Delta u_1 \) or \( \Delta u_2 \) are close to zero. Many will find these regions quite plausible for certain applications and will thus find the prior inadequate. For example, if \( u \) is a utility function, then the probability that two goods have very different marginal utilities is very low.

Given the flatness of the prior on these truncated hyperplanes, the assignment of low probability to the regions where either \( \Delta u_1 \) or \( \Delta u_2 \) is close to zero must be because these regions are tight corners of \( \Lambda_X^k \) with low volume. Intuitively, as \( \Delta u_1 \) approaches zero for fixed \( c \), it becomes more and more difficult to choose \( \lambda \) to satisfy monotonicity and concavity.

If we have prior information suggesting that \( \Delta u_1 \) might be considerably lower than \( \Delta u_2 \), then we can always adjust the \( \beta_i \) to tilt the distribution to favor low values of \( \Delta u_1 \). The second panel of Figure 1 shows the same scatterplot for a prior with \( \beta = (5.0, 25.0) \) but otherwise identical. The
Figure 1: Scatterplots of $(\Delta u_1, \Delta u_2)$ for Three Different Priors
alternate prior indeed tilts the distribution as desired. However, the prior is still very informative about the distribution of \((\Delta u_1, \Delta u_2)\).

We suggest that the problem arises not because of the particular choice of \(v\), but because the contours of the prior density are flat: a prior whose contours lack curvature will put low probability in the vicinity of any given bounding hyper-plane of the form \(v'\lambda = 0\), such as \(\Delta u_i = 0\). Simulations we have done suggest that as \(k\) increases and \(u(\cdot; \lambda)\) becomes more flexible, the prior, paradoxically, puts less and less probability in the vicinity of the bounding hyper-planes \(\Delta u_i = 0\). Intuitively, as \(k\) increases and the dimension of \(\Lambda^k_X\) increases with it, the curvature of its boundary increases as well, reducing the relative volume close to the hyper-plane \(\Delta u_i = 0\).

If we want to make the prior distribution of \((\Delta u_1(\lambda), \ldots, \Delta u_n(\lambda))\) more diffuse, we can do so by putting more probability mass near the boundary of \(\Lambda^k_X\). Suppose we multiply the prior by the factor

\[
\prod_{i=1}^n \left( \frac{\Delta u_i}{\sum_{j=1}^n \Delta u_j/\beta_j} + \delta \right)^{-p},
\]

where \(\delta\) and \(p\) are positive constants. The factor is bounded on the truncated hyper-plane \(\{ \lambda \in \Lambda^k_X : \Delta u_1(\lambda)/\beta_1 + \Delta u_2(\lambda)/\beta_2 = c \}\), so the prior remains proper. It is homogeneous of degree zero in the \(\Delta u_i/\beta_i\) and therefore in \(\lambda\), so it does not change the distribution of \(\sum_{i=1}^n \Delta u_i/\beta_i\), which we specify. It does, however, put more mass in the tight corners of \(\Lambda^k_X\), as desired.

Choosing \(\delta = 0.001\) and \(p = -3.8\) makes \((\Delta u_1/\beta_1)/(\Delta u_1/\beta_1 + \Delta u_2/\beta_2)\) nearly uniform. The third panel of Figure 1 shows the scatterplot of \((\Delta u_1, \Delta u_2)\) for this modified prior. We see that although the distribution of \(\Delta u_1/\beta_1 + \Delta u_2/\beta_2\) does not change, the joint distribution of \((\Delta u_1, \Delta u_2)\) is more diffuse. The three examples we have just seen suggest that we can center the prior over whatever ratio \(\Delta u_1/\Delta u_2\) is appropriate for a given application and also independently choose the degree of diffuseness of the prior.

Figure 2 gives an idea of the implied joint prior distribution of various economic choices. Recall that for the transformation \(\phi\) given by (2), the partial sum \(\lambda_{(1,0)}\phi(x)^{(1,0)} + \lambda_{(0,1)}\phi(x)^{(0,1)}\) is approximately a Cobb-Douglas utility with expenditure shares \(\lambda_{(1,0)}/(\lambda_{(1,0)} + \lambda_{(0,1)})\) and \(\lambda_{(0,1)}/(\lambda_{(1,0)} + \lambda_{(0,1)})\). Higher order terms allow changes in expenditure share with prices and income. In the figure, the expenditure share of the first good is given as a function of the ratio \(w_1/w_2\) of income-normalized prices \(w_1\) and \(w_2\) adding to one. The 20 curves are drawn from the prior distribution whose density includes the
factor in (12). The prior seems to favor smoothness, but looking ahead to Figure 4, we see that after 11 observations, the posterior distribution can feature very abrupt changes in expenditure share with high probability.

4 Prior and Posterior Simulation

In the previous section, we describe conditional prior densities \( f(\lambda | k) \) on the convex cones \( \Lambda^k_X \) whose union is the regular parameter space. In this section we describe methods for Markov chain Monte Carlo (MCMC) simulation of distributions on \( \Lambda^k_X \).

We saw in the previous section that the parameter space \( \Lambda^k_X \) has tight corners where one or more of the \( \Delta u_i \) is much smaller than the rest. 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 density in these tight corners. An efficient chain for simulating the target distribution must be able to spend a lot of 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 \( \Lambda^k_X \), 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 state-dependant 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 distri-
Figure 2: Prior Scattergraph of the Expenditure Share of Good 1 versus $w_1/w_2$, where $w_1$ and $w_2$ are income-normalized prices adding to 1.
butions efficiently on the parameter subspace $\Lambda_X^k$.

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 $\Lambda_X^k$. This means that large jumps across $\Lambda_X^k$ have reasonably high probability.

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

### 4.1 Definitions

We introduce a few preliminary definitions that will be important in the next sections. We define, for each $i$, the univariate function $u_i : [0, x_i] \rightarrow \mathbb{R}$ as the restriction of $u(x; \lambda)$ to the line segment $\{x \in X: x_j = x_j^*, j \neq i\}$, 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 $[\phi(x^* + (x_i - x_i^*)e_i)]^\prime$ is non-zero only for $i = \kappa e_i$, $\kappa = 1, \ldots, k - 1$. Thus the univariate $u_i$ depends only the elements $\lambda_i$ such that $i = \kappa e_i$, $\kappa = 1, \ldots, k - 1$.

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

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

where $\lambda^{(i)}$ consists of the elements $\lambda_i$ for which $\nu_j = 0$ 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 $\lambda^o$ consists of all remaining elements of $\lambda$.

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

21
4.2 A Convex Cone Enclosing $\Lambda^k_X$

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

It is easy to construct matrices \( V \) such that \( V\lambda \geq 0 \) contains \( \Lambda^k_X \). Equations (5) and (6), which give various necessary conditions for regularity, are inequalities of the form \( v\lambda \geq 0 \). Constructing \( V \) involves vertically stacking \( N_v \) row vectors \( v \) satisfying \( v\lambda \geq 0 \).

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 \((i, t) \in \{1, \ldots, n\} \times I_J\). For each \( i \) and \( t \), we generate a row of \( V \) using the necessary conditions
\[
\begin{align*}
\frac{\partial u(x^{(i)}; \lambda)}{\partial x_i} - \frac{\partial u(x^{(i+t)}; \lambda)}{\partial x_i} &\geq 0 & t_i &\leq J \\
\frac{\partial u(x^{(i)}; \lambda)}{\partial x_i} &\geq 0 & t_i &= J,
\end{align*}
\]
where \( \{x^{(i)} : t \in I_J\} \) is a grid of points. Each \( x^{(i)} \) is defined by
\[
\phi(x^{(i)}) = \left( z_1^{lo} + \frac{t_1}{J-1}(z_1^{hi} - z_1^{lo}), \ldots, z_n^{lo} + \frac{t_n}{J-1}(z_n^{hi} - z_n^{lo}) \right),
\]
where \( z_i^{lo} \equiv \phi(0) \) and \( z_i^{hi} \equiv \phi(\bar{x}) \).

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 \( \pi_- \leq 0 \) and \( \pi_+ \geq 0 \) such that \( \lambda + (\pi_-)w \) and \( \lambda + (\pi_+)w \) are on the boundary of the cone \( V\lambda \geq 0 \) enclosing \( \Lambda^k_X \). 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 \( \lambda^* \in \Lambda^k_X \). We now describe some of these steps in more detail.
4.3.1 Computing $\pi_- \text{ and } \pi_+$

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

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

Similarly,

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

4.3.2 Drawing Random Directions

For the simulations reported in this paper, we use the following set of directions. There is a direction $w^{(i)} \in \mathbb{R}^{|I_k|}$ for each multi-index $i \in I_k \setminus \{e_1, \ldots, e_n\}$. For multi-indices $i$ and $\kappa$, $w^{(i)}_\kappa$ is element $\kappa$ of the direction vector $w^{(i)}$ and is given by the coefficient of $z^\kappa$ in the $n$-variate polynomial $\prod_{i=1}^n P_i(a_i z_i + b_i)$, 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,

$$w^{(i)}_\kappa = \prod_{i=1}^n \sum_{j=\kappa_i}^{\kappa_i} L_{ij,i} \binom{j}{\kappa_i} a_i^{\kappa_i} b_i^{-\kappa_i},$$

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 $\sum_{\kappa \in I_k} w^{(i)}_\kappa z^\kappa$ are orthogonal on $\phi(X)$, 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, $V w^{(i)}_\kappa$ 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 $\Lambda_X^k$, typically involves verifying $\lambda^* \in \Lambda_X^k$ 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 checks the regularity of $u(\cdot; \lambda^*)$ without error. We use an algorithm that
consists of a battery of regularity tests involving verification of 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 point out that $V\lambda^* \geq 0$ by construction, so $\lambda^*$ has already survived all of the regularity testing implied by this condition.

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

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

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 its value 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 $(\lambda^{(1)}, \ldots, \lambda^{(i^*-1)}, \pi\lambda^{(i^*)}, \lambda^{(i^*+1)}, \ldots, \lambda^{(n)}, \lambda^o)$ is on the boundary of the cone $V\lambda \geq 0$. Finally, we draw $\pi^*$ from a log-normal distribution truncated to $[\pi_{\min}, \infty)$ and construct the proposal

$$\lambda^* = (\lambda^{(1)}, \ldots, \lambda^{(i^*-1)}, \pi^*\lambda^{(i^*)}, \lambda^{(i^*+1)}, \ldots, \lambda^{(n)}, \lambda^o). \quad (13)$$

The advantage of the second proposal is that the chain can move quickly out of the tight corners of $\Lambda^k_X$.

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

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

We point out that there is no $\pi_{\max}$. That is, there is no $\pi > 1$ such that $V(\lambda + (\pi - 1)w) \not\geq 0$. This is because $w$ and $\lambda$ are both in $\Lambda^k_X$ and $\Lambda^k_X \subseteq \{\lambda : V\lambda \geq 0\}$. 

24
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 $\sigma \equiv 0.25$.

We then construct $\lambda^*$ as in equation (13). 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 $\Lambda_X^k$: 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 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 $\lambda^o$ 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

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

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

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

The advantage of the third proposal is that the chain can move quickly into the tight corners of $\Lambda_X^k$.

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

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

Regularity of the $u_{it}$, $i \neq i^*$ 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 $V(\lambda + (\pi - 1)w) \not\geq 0$. 

25
Once we have $\pi_{\text{max}}$, we draw $\pi^*$ from a log-normal distribution truncated to $[0, \pi_{\text{max}}]$. For the simulations described in this paper, $\log \pi^*$ is normal with mean 0 and standard deviation $\sigma \equiv 0.25$.

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

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

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

## 5 An Empirical Application

We present a consumer demand application to illustrate our methods. Utility functions are supposed to 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 $u: X \rightarrow \mathbb{R}$ is the utility function, 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. Figure 3 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.

Using the theory of Theil (1974) or that of McCausland (2004), we obtain the following distribution for the t’th choice \( x_t \).

\[
\Pr(x_t = x) = \begin{cases} 
\frac{\exp u(x)}{\sum_{y \in B_t} \exp u(y)} & x_t \in B_t \\
0 & \text{otherwise},
\end{cases}
\]

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.
Figure 3: Budgets for the “GARP for Kids” experiment.
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 (15).

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 $X = [0,12]^2$, 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

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

The multi-index constellations are rectangular lattices given by

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

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 $k \geq 2$. Thus we have three terms with probability $1/2$, eight with probability $1/4$, 15 with probability $1/8$, and so on. As in the first example of Section 3, we choose the prior parameters $\alpha = 4$ and $\beta = (25.0,25.0)$, which means that $\Delta u_1/\beta_1 + \Delta u_2/\beta_2$ is gamma with shape parameter 4 and scale parameter 1. We modify the prior as described in Section 3 by multiplying the prior density by the factor in (12). We choose $\delta = 0.001$ 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 the first 30 subjects 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 $k \in \{2, 3, 4, 5\}$. 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 reported in Table 2 are all less than 0.15. The average log marginal likelihood is –12.49. Subject 1088 always spent his income on the more expensive good. This exuberant irrationality earned him a log marginal likelihood of -30.36, by far the lowest. Subject 1105 always spent his income on the cheaper good, and choose 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 a low –21.43. However, we can see in Figure 4 that the posterior distribution puts high probability on these abrupt changes. The 20 curves of that figure are constructed from 20 simulated values from the conditional posterior distribution of $\lambda$ given $k = 4$. 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 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
Table 1: Posterior Probabilities of \( k \) for First 30 Subjects in “GARP for Kids” Experiment

| Subject | \( \Pr [k = 2|X] \) | \( \Pr [k = 3|X] \) | \( \Pr [k = 4|X] \) | \( \Pr [k = 5|X] \) |
|---------|-----------------|-----------------|-----------------|-----------------|
| 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        | 0.253905        |
| 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 “GARP for Kids” Experiment

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-16.713183</td>
<td>-4.233398</td>
<td>-17.822427</td>
<td>-10.557720</td>
<td>-30.356065</td>
</tr>
<tr>
<td>-10.407859</td>
<td>-12.575744</td>
<td>-4.160754</td>
<td>-4.224852</td>
<td>-8.894845</td>
</tr>
</tbody>
</table>

−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 \( X \), we can approximate \( u \circ \phi^{-1} \) on \( \phi(X) \), where \( \phi \) is a function that can be chosen to be monotone and concave, facilitating the approximation of functions with monotonicity and curvature restrictions. 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. Despite an infinite dimensional and irregularly shaped parameter space, we have shown how to
Table 3: Counts of Numbers of GARP Violations

<table>
<thead>
<tr>
<th>Number of violations</th>
<th>Experimental subjects</th>
<th>All sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>36</td>
<td>108,846</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>140,788</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>171,718</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>272,978</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>438,074</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>646,288</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>928,790</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1,567,246</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2,081,452</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2,555,030</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>3,184,790</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>55</strong></td>
<td><strong>12,096,000</strong></td>
</tr>
</tbody>
</table>
Figure 4: Posterior Scattergraph of Modal Expenditure Share of Good One for Subject 1105.
elicit exact priors on economically relevant quantities and thereby induce proper priors on the parameters. We have remarked that there are tight corners of the parameter space which, despite their low volume, may be quite plausible in many applications. We have shown how to concentrate prior probability in these corners without sacrificing propriety or modifying the prior distribution of the economically relevant quantities.

We have pointed out 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. We find that out-of-sample predictions are better on average than any model based only on numbers of observed violations of the Generalized Axiom of Revealed Preference. This result provides some support for the theory of random consumer demand in McCausland (2004), which we used to construct the likelihood functions.

A On the Boundedness of Cone Truncations

This appendix is on the boundedness of a truncation of the cone $\Lambda^k_X$ defined in equation (4). This boundedness is relevant for the construction of proper prior distributions on $\Lambda^k_X$ as described in Section 3.2.

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

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

**Claim A.1** For every $x \in [x^*_1, \bar{x}_1] \times \ldots \times [x^*_n, \bar{x}_n]$ and every $\lambda \in \Lambda^k_X : u(\bar{x}; \lambda) \leq 1$

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

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

Let $J = |I_k|$, and order the multi-indices $i \in I_k$ as $i^{(1)}, \ldots, i^{(J)}$. 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. $[\phi_i(\frac{x^*+\bar{x}}{2})/\phi_i(\bar{x})]^{1/J} \leq q_i \leq 1$.

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

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

For all $\lambda \in \Lambda_k^X$,

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

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

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

**Claim A.2** $C$ is non-singular.
Proof. $C$ can be written as

\[
\begin{bmatrix}
(q^{(1)})^1[\phi(\bar{x})]^{(1)} & \cdots & (q^{(J)})^1[\phi(\bar{x})]^{(J)} \\
\vdots & \ddots & \vdots \\
(q^{(1)})^J[\phi(\bar{x})]^{(1)} & \cdots & (q^{(J)})^J[\phi(\bar{x})]^{(J)}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(q^{(1)})^1 & \cdots & (q^{(J)})^1 \\
\vdots & \ddots & \vdots \\
(q^{(1)})^J & \cdots & (q^{(J)})^J
\end{bmatrix}
\cdot \text{diag} \left( [\phi(\bar{x})]^{(1)}, \ldots, [\phi(\bar{x})]^{(J)} \right)
\]

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 $j, l \in \{1, \ldots, J\}$, $j \neq l \Rightarrow q^{(j)} \neq q^{(l)}$. 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 $C$. □

References


