Formulation and solution strategies for nonparametric nonlinear stochastic programmes with an application in finance

Bastin, F.; Cirillo, C.; Toint, Philippe

Published in:
Optimization

DOI:
10.1080/02331931003700707

Publication date:
2010

Document Version
Early version, also known as pre-print

FORMULATION AND SOLUTION STRATEGIES FOR NONPARAMETRIC NONLINEAR STOCHASTIC PROGRAMS, WITH AN APPLICATION IN FINANCE

by F. Bastin[1], C. Cirillo[2] and Ph. L. Toint[3]

Report 07/16 December 19, 2007

[1] Department of Computing Science and Operational Research,
University of Montreal, Canada
email: bastin@iro.umontreal.ca

[2] Department of Civil and Environmental Engineering,
University of Maryland, USA
email: ccirillo@umd.edu

[3] Department of Mathematics,
University of Namur,
61, rue de Bruxelles, B-5000 Namur, Belgium,
Email: philippe.toint@fundp.ac.be

Formulation and solution strategies for nonparametric nonlinear stochastic programs, with an application in finance.

Fabian Bastin[*], Cinzia Cirillo[†], Philippe L. Toint[‡]

December 19, 2007

Abstract

We consider a class of stochastic programming models where the uncertainty is classically represented using parametric distributions families. The parameters are then usually estimated together with the optimal value of the problem. However, misspecification of the underlying random variables often leads to irrealistic results when little is known about their true distributions. We propose to overcome this difficulty by introducing a nonparametric approach where we replace the estimation of the distribution parameters by that of cumulative distribution functions (CDF). A practical algorithm is described which achieves this goal by using a monotonic spline representation of the inverse marginal CDF's and a projection-based trust-region globalization. Applications of the new algorithm to discrete choice theory are finally discussed, both with simulated data and in the context of a practical financial application related to interventions of the Bank of Japan in the foreign exchange market.

1 Introduction

Stochastic programming has an significant place in the mathematical programming field, where the central role of uncertainty gains even wider recognition. Depending of the influence and nature of this uncertainty in the optimization problem, various solution approaches have been proposed, ranging from expected-value minimization to robust optimization. However, the quality of the stochastic models considered remains crucially dependent on an adequate choice of the distributions of the involved random variables. This is especially true for some classes of estimation problems where one aims at calibrating the heterogeneity present in the data. Although it is typically possible to estimate parameters of predefined distributions in such cases, the very choice of

[*]Department of Computing Science and Operational Research, University of Montreal, Canada.
[†]Department of Civil and Environmental Engineering, University of Maryland, USA.
[‡]Department of Mathematics, University of Namur, Belgium.

these distributions often remains difficult or arbitrary, and it may be influenced more by classical assumptions and ease of computation than by real knowledge of the problem.

The purpose of this paper is to discuss a nonparametric approach to resolve the difficulties associated with the identification of underlying unknown base random distributions. Our proposal is characterized by the explicit estimation of the shape of the unknown distributions, expressed via their cumulative distribution functions, as a part of the complete calibration procedure.

Nonparametric techniques are an important topic is statistics and simulation (see for instance Chapter 6 in Law [31], Dong and Wets [18] or Wasserman [44]), allowing to deal with uncertainty when the underlying distributions are difficult to recover by approximating the cumulative distribution functions and probability densities. One of the most popular tool is the empirical distribution, which has also received attention in stochastic programming (Pflug [36]), where an approximation of the cumulative distribution function is constructed from observed realisations of the random variable, even if the true distribution is unknown. As for standard stochastic programs for which random variables are completely specified with standard parametric distributions, it is also assumed in this case that underlying distributions of the involved random variables can be estimated during a preliminary phase, prior the optimization. This assumption appears to be reasonable for numerous applications, but, as we will illustrate in this paper, not all. We will indeed focus our attention on problems in which the estimation of the random variables is itself part of the optimization process, as is for instance the case for the calibration of mixed logit models, a recent and popular technique in econometrics and transportation (see Bastin, Cirillo and Hess [5], Bhat [13] or Train [41], amongst others). The methodology proposed here, based on spline approximations of inverse cumulative distribution functions, has also been explored in a less general framework and successfully applied to transportation problems by Bastin, Cirillo, and Toint [7], who show its value as a new tool for studying the value of time. The present paper pursues the same idea and casts it in the general context of nonconvex stochastic programming. It also describes its application to a problem in finance for which traditional approaches have been found to be limited (Bernal and Gnabo [11], Beine, Bernal, Gnabo, and Lecourt [8]).

The paper is organized as follows. In Section 2 we present the class of considered problems and motivate our nonparametric approach. We develop the related optimization technique in Section 3, and report numerical experiments in Section 4. The methodology is tested on simulated data, exhibiting the danger of misspecification, as well as on a real econometric model derived from financial markets. Some conclusions and perspectives are finally given in Section 5.

2 Problem formulation

We consider the general stochastic program (SP)

$$\min_{x \in \mathcal{X}} g\left(E[f(x, \boldsymbol{\xi})]\right), \tag{1}$$

where $\mathcal{X}$ is a compact set in $\mathbb{R}^n$, $\boldsymbol{\xi}$ is a random vector of size m defined on the probability space $(\Xi, \mathcal{F}, P)$, and g is a function from $\mathbb{R}$ to $\mathbb{R}$. For simplicity, we here assume

that $\mathcal{X}$ is deterministic. Crucially, we also assume that a computationally efficient projection on this feasible set is available (a typical example is when $\mathcal{X}$ defines simple bounds on the components of x). Smoothness and regularity assumptions are made below, but we do not require convexity of the objective function. Since (1) cannot usually be solved analytically, we consider a solution process based on the sample average approximation (SAA), which is constructed by sampling over $\boldsymbol{\xi}$:

$$\min_{x \in \mathcal{X}} \hat{g}(x) \stackrel{\text{def}}{=} g\left(\hat{f}_R(x)\right), \tag{2}$$

where $\hat{f}_R(x)$ is defined as

$$\hat{f}_R(x) = \frac{1}{R} \sum_{r=1}^{R} f(x, \xi_r)$$

for R random draws. In order to ensure consistency of this formulation, we make the following assumptions.

A.0 The random draws $\{\xi_q\}_{q=1}^{\infty}$ are independently and identically distributed.

A.1 For P-almost every ξ, the function $f(\cdot, \xi)$ is continuously differentiable on S.

A.2 The family $f(x, \xi)$, $x \in \mathcal{X}$, is dominated by a P-integrable function $K(\xi)$, i.e. $E_P[K]$ is finite and $|f(x, \xi)| \leq K(\xi)$ for all $x \in \mathcal{X}$ and P-almost every ξ.

A.3 Each gradient component $\frac{\partial}{\partial [x]_l} f(x, \xi)$ ($l = 1, \ldots, n$), $x \in \mathcal{X}$, is dominated by a P-integrable function.

A.4 The function g is twice continously differentiable in its argument.

Note that **A.3** allows us to apply the results of Rubinstein and Shapiro [39], page 71, and deduce that the expected value function $E[f(x, \boldsymbol{\xi})]$ is continuously differentiable over $\mathcal{X}$, and that the expectation and gradient operator can be interchanged in the expression of the gradient. Also observe that **A.0**–**A.2** together imply the existence a uniform law of large numbers (ULLN) on S, for the approximation $\hat{f}_R(x)$ of $f(x)$:

$$\sup_{x \in \mathcal{X}} \left| \hat{f}_R(x) - f(x) \right| \longrightarrow 0 \quad \text{almost surely as } R \to \infty,$$

which in turn allows us to deduce the following property.

Proposition 1. *Under Assumptions* **A.0–A.2**, **A.4**, *we have the uniform law of large numbers*

$$\sup_{x \in \mathcal{X}} \left| g\left(\hat{f}_R(x)\right) - g(f(x)) \right| \xrightarrow{a.s.} 0 \quad \textit{as } R \to \infty.$$

Proof. Let $\epsilon > 0$. By the continuity of $g(\cdot)$, we have that there exists some $\delta > 0$ such that $|z_1 - z_2| < \delta$ implies that $|g(z_1) - g(z_2)| < \epsilon$. From the ULLN on $\hat{f}_R(\cdot)$ with

respect to $f(\cdot)$, we have that there exists some R_δ such for all $R \geq R_\delta$, for all x in $\mathcal{X}$, $\left|\hat{f}_R(x) - f(x)\right| < \delta$ almost surely, and therefore $\left|g\left(\hat{f}_R(x)\right) - g(f(x))\right| < \epsilon$ almost surely. □

From **A.3** and **A.4**, we can similarly establish ULLN's between partial derivatives of the SAA and the true objective. First-order convergence (i.e. convergence of a sequence of first-order solutions of (2) to a first-order solution of (1) can then be derived from stochastic variational inequalities, as presented in Gurkan, Özge, and Robinson [24] and Shapiro [40]. Consider a mapping $\Phi : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$ and a multifunction $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Suppose that we have a well-defined function $\phi(x) := h\left(E_{P_\Pi}[\Phi(x, \boldsymbol{\xi})]\right)$. We refer now to

$$\phi(x) \in \Gamma(x) \tag{3}$$

as the true, or expected value, generalized equation and say that a point $x^* \in \mathbb{R}^m$ is a solution of (3) if $\phi(x^*) \in \Gamma(x^*)$. If $\{\xi_1, \ldots, \xi_R\}$ is a random sample, we refer to

$$\hat{\phi}_R(x) \in \Gamma(z) \tag{4}$$

as the SAA generalized equation, where

$$\hat{\phi}_R(x) = h\left(R^{-1} \sum_{i=1}^{R} \Phi(x, \xi_i)\right).$$

We denote by S^* and S_R^* the sets of (all) solutions of the true (3) and SAA (4) generalized equations, respectively. Let us denote by $d(x, A) := \inf_{x' \in A} \|x - x'\|$, the distance from $x \in \mathbb{R}^n$ to A, and $D(A, B) := \sup_{x \in A} D(x, B)$, the deviation of the set A from the set B. We then have the following result (Shapiro [40]), whose proof does not depend on the explicit forms of $\phi(x)$ and $\hat{\phi}_R(x)$:

Theorème 1. *Let $\mathcal{S}$ be a compact subset of $\mathbb{R}^m$ such that $S^* \subseteq \mathcal{S}$. Assume that*

(a) the multifunction $\Gamma(x)$ is closed, that is if $x^{(k)} \to x$, $y^{(k)} \in \Gamma(x^k)$ and $y^{(k)} \to y$, then $y \in \Gamma(x)$,

(b) the mapping $\phi(x)$ is continuous on $\mathcal{S}$,

(c) almost surely, $\emptyset \neq S_R^* \subseteq \mathcal{S}$ for sufficiently large R, and

(d) $\hat{\phi}_R(x)$ converges to $\phi(x)$ almost surely uniformly on $\mathcal{S}$ as $R \to \infty$.

Then $D(S_R^*, S^*) \to 0$ almost surely as $R \to \infty$.

When $\mathcal{X}$ is convex, we can define the first-order criticality conditions for some point x^* as the requirement that $-\nabla_x g(x^*)$ belongs to the normal cone to $\mathcal{X}$ at x^*, denoted by $\mathcal{N}_\mathcal{X}(x^*)$. Theorem 1 then allows an easy proof of almost sure first-order convergence. Consider the choice $\Gamma(\cdot) = \mathcal{N}_\mathcal{X}(\cdot)$; $\phi(x^*)$ belongs to $\Gamma(x^*)$ if and only if

$$\langle \phi(x^*), u - x^* \rangle \leq 0, \ \forall u \in \mathcal{X}.$$

Following Shapiro [40], we refer to such variational inequalities as stochastic variational inequalities and note that the assumption (a) of Theorem 1 always holds in this case. Let S^* and S_R^* represent the set of first-order critical points of the true (3) and SAA (4) generalized equations, respectively. Then under **A.0**–**A.4**, we have that $\phi(x) = -\nabla_x g(x)$, and that $\phi(x)$ is a continuous random vector on $\mathcal{X}$, yielding assumption (b). Assumption (d) results from the ULLN, while **A.1** and the compacity of $\mathcal{X}$ ensure assumption (c) by setting $\mathcal{S} = \mathcal{X}$. Thus Theorem 1 guarantees first-order criticality in the limit as $R \to \infty$, almost surely. Second-order convergence is considerably more difficult to establish, but can nevertheless be deduced under additional assumptions as in Bastin et al. [6].

2.1 A nonparametric approach

When dealing with problems (1) and (2), one often implicitely makes suppositions on the distribution of the random vector $\boldsymbol{\xi}$ to generate the draws needed to construct the SAA. In practice, this may lead to problems, as we will demonstrate below. If we nevertheless continue to represent uncertainty in the problem at hand by using random variables in our model, we therefore wish to do with minimal assumptions on their distributions, avoiding in particular specifics on their shape.

A first observation is that we may consider each of the m components of $\boldsymbol{\xi}$ separately, at the price of assuming independence between them. As a consequence, we only have to draw from univariate random variables, which is considerably simpler than handling the multivariate case. If X is an univariate (known) random variable, a well-known technique to generate draws from its distribution consist to sample a uniform distribution on $[0, 1]$, hereafter denoted by $U[0, 1]$, and to apply the inverse cumulative distribution function F_X^{-1} to these draws:

$$S_X = \{F_X^{-1}(U),\ U \sim U[0, 1]\},$$

where S_X represents the sample set drawn from the random variable X. It is usually assumed that F_X^{-1} is available (or at least can be well approximated), if the distribution of X is known. This method is known as the inversion technique in the random numbers generation litterature (Devroye [17], Law [31]), and is also popular in the context of variance reduction methods (see for instance L'Ecuyer [32]). The use of the inverse cumulative distribution function has also previously been proposed in standard nonparametric estimation by Hora [27] and Avramidis and Wilson [2]. In order to capitalize on this idea, we introduce the assumption that

A.5 the components of $\boldsymbol{\xi}$ are independent.

The discussion above then implies that we may obtain the necessary draws from the distribution of the random variable $\boldsymbol{\xi}$ if we are able to estimate for each component X an inverse cumulative distribution function F_X^{-1} with the properties that

- $F_X^{-1} : [0, 1] \to \mathbb{R}$,
- F_X^{-1} is monotonically increasing,

and if we restrict ourselves to continuous variates,

- F_X^{-1} is continuous.

In other terms, we have to estimate a continuous real function whose domain is $[0,1]$, and which is monotonically increasing. One may argue that this approach has advantages in generality and efficiency, since estimating the density instead of F_X^{-1} would only make sense for continuous distributions, and since it is usually easier to estimate a function rather than its derivative[1]. In order to further limit our choices, we also make the following additional assumption:

A.6 Each component of the random vector $\boldsymbol{\xi}$ is continuous and has a bounded support.

This assumption is often realistic for practical data sets and has the advantage of explicitly avoiding the presense of tails which are often difficult to interpret.

Functions approximation is a large field of mathematics, and various techniques can be considered for the problem of estimating F_X^{-1}. In our case, we choose to represent the desired inverse cumulative distribution function as a finite linear combination of some basis functions $\{l_q,\ q = 0,\ldots,v\}$, which are continuous on the interval $[0,1]$ for continuous variables, in which case we belive that a basis choice suitable for our purposes is that of B-splines. In general, a B-spline function $C(u)$ of degree p is a piecewise polynomial of degree p, defined on the interval $[a,b]$, which can be expressed as a linear combination of $n+1$ basis functions $N_{i,p}(u)$, as follows:

$$C(u) = \sum_{i=0}^{v} P_i N_{i,p}(u).$$

The coefficients $P_0, P_1, \ldots, P_v$ are called the control points, and u is the knot vector $(u_0 = a, u_1, \ldots, u_m = b)$. The basis functions can be constructed by recurrence (on the degree p) as follows:

$$N_{i,0} = \begin{cases} 1 & \text{if } u \in [u_i, u_{i+1}), \\ 0 & \text{otherwise.} \end{cases}$$

and

$$N_{i,p} = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u),$$

so that v is equal to $m-p-1$. There are several types of knot vectors, but one especially convenient for our purposes is the nonperiodic (or clamped or open) knot vector, which has the form

$$U = \{\underbrace{a,\ldots,a}_{p+1}, u_{p+1},\ldots,u_{m-p-1}, \underbrace{b,\ldots,b}_{p+1}\}, \tag{5}$$

that is the first and last knots have multiplicity $p+1$. It is possible to show that the function $C(u)$ is $p-1$ times continuously differentiable. In this paper, we will consider

[1] In some situations, one can however prefer to estimate densities as approximations like empirical distributions can produce asymptotically unbounded densities as the number of observations grows to infinity (L'Ecuyer, Cordeau and Simard [33]).

cubic B-splines, i.e. we will set p to 3, giving twice continuously differentiable inverse cumulative distribution functions. But the most crucial property[2] of B-splines in our context is that, with these basis and knots choices, $C(u)$ is monotonically increasing if the control points have the same property, that is if

$$P_0 \leq P_1 \leq \ldots \leq P_v.$$

As we will describe in the next Section, this property can be algorithmically guaranteed in our estimation procedure.

For simplicity of presentation, we will assume that all random variates are non-parametric, while in practice (as in our real application), we can mix parametric and nonparametric distributions. Our nonparametric estimation problem is then to solve, for some fixed R,

$$\min_{\substack{x \in \mathcal{X} \\ P_{i,j}}} \; g\left(\frac{1}{R}\sum_{r=1}^{R} f(x, \xi_r)\right), \tag{6}$$

under the additional constraints that, for $j = 1, \ldots, m$,

$$P_{0j} \leq P_{1j} \leq \ldots \leq P_{vj} \tag{7}$$

and

$$[\xi_r]_j = \sum_{i=0}^{v} P_{ij} N_{i,p}(\zeta_r), \tag{8}$$

where ζ_r are draws from the uniform distribution of $[0, 1]$.

2.2 A trust-region method for efficient estimation

In order to solve the program (2), under the constraints that the control points describing our inverse cumulative distribution functions are monotonic, we first substitute (8) in (6), which yields a (potentially nonconvex) objective function $f(w) \stackrel{\text{def}}{=} \hat{g}(x, P)$ to minimize with respect to $w \stackrel{\text{def}}{=} (x, P)$ on the feasible region $\mathcal{X} \times \mathcal{C}$ where $\mathcal{C}$ is defined by

$$\mathcal{C} = \prod_{i=1}^{m} \{(P_{0j}, \ldots P_{vj}) \text{ such that } P_{0j} \leq P_{1j} \leq \ldots \leq P_{vj}\}, \tag{9}$$

Fortunately, it is easy to verify that $\mathcal{C}$ is a convex set, so that consistency results, especially Theorem 1, can be applied, as long as we also assume that the iterates remain in a compact set..

If there is only one nonparametric coefficient, $\mathcal{C}$ defines v ordered variables. $\mathcal{C}$ is then called the order-simplex. The key of our algorithm is that projection onto the order-simplex can be performed easily and efficiently, since several efficient algorithms of $O(n)$ complexity have been designed for this task (Best and Chakravarti [12]), and their extension to Cartesian products is easy since they can be applied independently

[2]For a more complete review of B-splines properties, we refer the reader to Piegl and Tiller [37].

on each subset of monotonic variables. Since an efficient projection operator onto $\mathcal{X}$ is also assumed to exist, an efficient projection operator onto $\mathcal{X} \times \mathcal{C}$ is easily constructed and can then be exploited successfully in the solution of the nonlinear optimization problem defined by (6) (with fixed R), (7) and (9). This is achieved by a computationally efficient specialized trust-region algorithm (see Chapter 12 of Conn, Gould, and Toint [16]) which we now state.

Algorithm 1: Projected trust-region algorithm

Step 0. Initialization. An initial point $w_0 \in \mathcal{C}$ and an initial trust-region radius Δ_0 are given. The constants η_1, η_2, γ_1, and γ_2 are also given and satisfy

$$0 < \eta_1 \leq \eta_2 < 1 \text{ and } 0 < \gamma_1 \leq \gamma_2 < 1. \tag{10}$$

Compute $f(w_0)$ and set $k = 0$.

Step 1. Model definition. Construct a model m_k in the trust-region $\mathcal{B}_k$, defined as

$$\mathcal{B}_k = \{w \text{ such that } \|w - w_k\| \leq \Delta_k\}.$$

Step 2. Step calculation. Compute a step $s^{(k)}$ that "sufficiently reduces the model" $m^{(k)}$ and such that $w^{(k)} + s^{(k)} \in (\mathcal{X} \times \mathcal{C}) \cap \mathcal{B}_k$.

Step 3. Acceptance of the trial point. Compute $f(w^{(k)} + s^{(k)})$ and define

$$\rho_k = \frac{f(w^{(k)}) - f(w^{(k)} + s^{(k)})}{m_k(w^{(k)}) - m_k(w^{(k)} + s^{(k)})}. \tag{11}$$

If $\rho_k \geq \eta_1$, then define $w^{(k+1)} = w^{(k)} + s^{(k)}$; otherwise define $w^{(k+1)} = w^{(k)}$.

Step 4. Trust-region radius update.

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{if } \rho_k \geq \eta_2, \\ [\gamma_2 \Delta_k, \Delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{if } \rho_k < \eta_1. \end{cases}$$

Increment k by 1 and go to Step 1.

In this description, reasonable values for the constants of (10) are for instance given by

$$\eta_1 = 0.01, \ \eta_2 = 0.9, \ \text{ and } \gamma_1 = \gamma_2 = 0.5,$$

but other values may be selected. We chose the Euclidean norm in the definition of $\mathcal{B}_k$. We finally followed the usual practice and defined the model m_k to be a quadratic function of the type

$$m_k(w^{(k)} + s) = f(w^{(k)}) + \nabla_w f(w^{(k)})^T s + \frac{1}{2} s^T H_k s,$$

where H_k is either the Hessian $\nabla^2_{ww} f(w^{(k)})$ or some approximation of it. In our experiments, we have used either the SR1 approximation (Conn, Gould and Toint [15]), either the BFGS approximation (see for instance Chapter 5 in Nocedal and Wright [35]), with similar success.

If $\rho_k \geq \eta_1$ in Step 1, the iteration k is said to be successful since the candidate point $w^{(k)} + s^{(k)}$ is accepted; otherwise the iteration is declared unsuccessful and the new point is rejected. If $\rho_k \geq \eta_2$, the agreement between the model and the function is particularly good, and the iteration is said to be very successful. This then suggests increasing the trust-region radius, as in Step 4, in order to allow a longer step at the next iteration.

The step $s^{(k)}$ is computed by first attempting to identify the active constraints by minimizing the model along the projected gradient path (using the complete projection onto $\mathcal{X} \times \mathcal{C}$). Note that, at the iteration k, a monotonicity constraint $P_{ij} \leq P_{i+1,j}$ will be said active if

$$P_{ij}^{(k)} = P_{i+1,j}^{(k)} \text{ and } \left.\frac{\partial \hat{g}(x,P)}{\partial P_{ij}}\right|_{(x^{(k)},P^{(k)})} \leq \left.\frac{\partial \hat{g}(x,P)}{\partial P_{i+1,j}}\right|_{(x^{(k)},P^{(k)})}.$$

Once the active set has been identified by minimizing the model along this path, further reduction of the model can then be obtained by applying a (possibly restarted) conjugate-gradients algorithm within $(\mathcal{X} \times \mathcal{C}) \cap \mathcal{B}_k$.

Convergence results for this algorithm can directly be deduced from the theory presented in Chapter 12 [16], together with more details on the algorithm. In particular, one obtains that all limit points of the sequence of iterates are first- and second-order stationary points of problem (2) for fixed R under the constraints (9). Moreover, the active constraints (in our case the confluent control points of the spline defining the inverse cumulative distribution functions F_X^{-1}) are identified after a finite number of iterations. As already mentioned above, the consistency results of Section 2 then apply provided the control points iterates remain in a compact set that contains $\mathcal{C}$.

3 Numerical experiments

We have experimented the proposed algorithm in the context of discrete choice theory (Ben Akiva and Lerman [9], Train [41]), in particular in the field of mixed logit problems. These constitute a recent development in the theory and are used today in a variety of contexts, e.g. politics (Glasgow [23]), economics (Rigby and Burton [38]), marketing, transportation (Brownstone, Bunch and Train [14]), and finance [4], in order to explain the behaviour of individuals/households/companies who express their choices amongst a finite set of alternatives. In this framework, taste heterogeneity across the population is captured using parametric models whose random variables have distributions with specific functional forms. In a large majority of the applications published up to now, the distributions are chosen to be normal. However, the use of unbounded distributions (such as the normal) appears inappropriate in a number of cases, especially when certain attributes are assumed to be positively (or negatively) valued by all individuals. In order to circumvent these difficulties, more recent models use bounded distributions, often obtained as simple transformations of normals.

Train and Sonnier [42] specify mixed logit models with lognormal, censored normal and Johnson Sb distributions bounded on both sides. Less restrictive nonparametric or semi-parametric approaches have also been proposed, such as mass point mixed logit models (Dong and Koppelman [19], discrete mixture of GEV models over a finite set of distinctive support points (Hess, Bierlaire, and Polak [26], nonparametric density techniques based on Klein and Spady estimators (Fosgerau [20]) and more recently nonparametric models based on Legendre polynomials (Fosgerau and Bierlaire [21]). The bounded support assumption **A.6** can therefore be considered appropriate, since extreme behaviours, corresponding to very large (absolute) values of X, are usually not welcome. It should also be noted that if the use of normal distributions is known to facilitate the estimation process, failures in convergence have nevertheless been reported for unbounded or nonparametric distributions, as well as difficulties created by the presence of numerous local maxima (Fosgerau and Bierlaire [21], Fosgerau and Hess [22]).

In our application context, we consider a set of I individuals, each individual i having to choose one alternative within a finite set $\mathcal{A}_i$. We associate an utility U_{ij} to each alternative A_j in $\mathcal{A}_i$, as perceived by individual i. In line with accepted econometric theory, we also assume that individuals aim at maximizing their utility, but this utility is not fully observable. The standard technique is then to decompose the utility U_{ij} in the sum of a deterministic measurable part $V_{ij}(\beta)$, where β is a vector to be estimated, and a random, unobserved part ϵ_{ij}. The probability that individual i chooses alternative j is then given by

$$P_{ij}(\beta) = P[V_{ij}(\beta) + \epsilon_{ij} \geq V_{ik}(\beta) + \epsilon_{ik},\ \forall A_k \in \mathcal{A}_i].$$

The probability expression is of course dependent of the distribution choice for ϵ_{ij}. When the ϵ_{ij} are assumed to be independent and identically Gumbel distributed amongst the individuals and alternatives, we obtain the traditional logit probability formula

$$P_{ij} = \frac{e^{V_{ij}(\beta)}}{\sum_{k=1}^{\mathcal{A}_i} e^{V_{ik}(\beta)}} \stackrel{\text{def}}{=} L_{i,j}(\beta). \tag{12}$$

In the mixed logit framework, we relax the assumption that β is a constant vector, but instead consider it as a random vector with cumulative distribution function $F_B(\beta)$ so that the probability on the left-hand side of (12) is now conditional on the realization of β, and the unconditional probability becomes

$$P_{ij} = E_B[L_{ij}(\beta)] = \int L_{ij}(\beta) dP_B(\beta). \tag{13}$$

As we typically cannot directly estimate B, we assume that it can be described as $B = B(\Gamma, \theta)$, where Γ is some random vector, and θ some constant parameters vector, again to be estimated. In other terms, we assume some distribution family for B, parametrized by θ. If additionally the vector β is continuous, we can rewrite (13) as

$$P_{ij}(\theta) = \int L_{ij}(\gamma, \theta)\phi(\gamma, \theta) d\gamma,$$

where $\phi(\gamma, \theta)$ is the density of B, with parameters vector θ.

In the case when the same individual can express several choices, we observe for each individual the sequence of choices $y_i = (j_{i1}, \ldots, j_{iT_i})$, which can be assumed to be correlated. (Such cases are referred to as "panel data".) A simple way to accomodate this situation is to assume the heterogeneity is present at the population level only, but not at the level of individuals. The probability to observe the individuals' choices is then given by the product of logit probabilities $L_{ij_{it}}$ (Train [41]):

$$P_{iy_i}(\theta) = \int \left(\prod_{t=1}^{T_i} L_{ij_{it}}(\gamma, \theta) \right) \phi(\gamma, \theta) d\gamma.$$

The vector of unknown parameters θ is then estimated by maximizing the log-likelihood function, i.e. by solving the problem

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln P_{iy_i}(\theta), \tag{14}$$

where y_i is the vector of alternative choices made by the individual i. As pointed in Bastin et al. [6], (14) can be viewed as an extension of (1), and the corresponding SAA is

$$\max_{\theta} SLL^R(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln SP^R_{iy_i}(\theta), \tag{15}$$

where

$$SP^R_{iy_i} = \frac{1}{R} \sum_{r_i=1}^{R} \prod_{t=1}^{T_i} L_{ij_{it}}(\gamma_{r_i}, \theta),$$

and R is the number of random draws γ_{r_i}. The SAA program (15) can be solved directly by Algorithm 1, using the equivalent form of minimizing $-SLL^R(\theta)$. We use the nul vector as the starting point in our experiments.

3.1 Simulations

We first validate our estimation procedure on simulated data. We wish to verify that it approximately recovers the underlying distributions, and to compare it to traditional parametric estimation. In our simulated experiments we create a synthetic population of 2000 individuals, each one delivering one observation. The design contains four alternatives, normally distributed with parameters $N(0.5, 1)$, and three independent variables, one normal, one lognormal and one spline. The parameters used for these distributions are given in Table 1. We then estimate four models, one using correct distributions families, one constructed with normals only, one made with lognormals only, one using splines only. In our estimation procedure, $\mathcal{X} = \mathbb{R}^n$, so that in order to be consistent with the developed theory, we assume that the complete vectors of iterates remain in a compact set. We chose 5 equally spaced knot points per spline approximation, that is, taking account of the repetitions in the form (5), we choose $U = \{0, 0, 0, 0, 0.25, 0.5, 0.75, 1, 1, 1, 1\}$.

The worst fit is obtained when all distributions are assumed to be lognormal, which is not surprising since the lognormal is unable to capture negative coefficients. The use of normals provides a good fit, even if it is unable to capture the fact that the coefficient corresponding to the lognormal in the original specification should be positive. This is in line with other reported results, especially in Train and Weeks [43]. The model specification built only on splines outperforms all models in terms of log-likelihood optimal value, and is close to the one obtained with the correct specification. The difference is however of the order of the sampling accuracy, so no strong conclusion can be reached, but we cannot exclude a slight overfit on the sample data. The numerical convergence is slower with the splines-only specification, but again, this should not be a surprise since the two first parameters violate Assumption **A.6**.

Distribution	Real coeff.	Correct model	Normals	Lognormals	Splines
Normal	0.0 (μ)	0.025	0.013	-26.369	-16.783
	2.0 (σ)	2.200	2.069	16.648	-1.125
					-1.125
					-0.600
					1.439
					1.439
					7.163
Lognormal	0.0 (μ)	0.0322	1.262	-0.600	-2.520
	1.0 (σ)	0.880	0.924	0.884	1.331
					1.331
					1.331
					1.356
					2.038
					3.088
Spline	-15.0	-6.157	0.135	-28.532	-6.086
	-3.0	-6.157	2.4780942	22.13822	-6.086
	-0.5	0.134			0.0184
	0.0	0.6091			0.644
	0.5	0.6091			0.644
	3.0	0.6091			0.644
	15.0	27.076			25.865
Log-likelihood	-	-1.213	-1.216	-1.268	-1.212

Table 1: Calibration on simulated data

The same observations can be made in Figure 1, illustrating the agreement between the true cumulative distribution functions, and the different estimations. The all-lognormal model performs the worst, even for the second variable, while the best model is, as expected, the correctly specified one, which would typically be unavailable in practice. The all-normal model achieves good agreement, except on the last variable. The agreement for the lognormally distributed parameter is better than expected, the only problem being the infinite tail on the left, although the true spline specification is bounded. The all-splines model can be seen as a compromise when we have no information on the real distributions. The agreement with the first and third parameter are particularly good, while the model captures the essence of the second parameter, in a better way than with the all-lognormal model.

These findings however have to be taken with care, as by construction, two of our

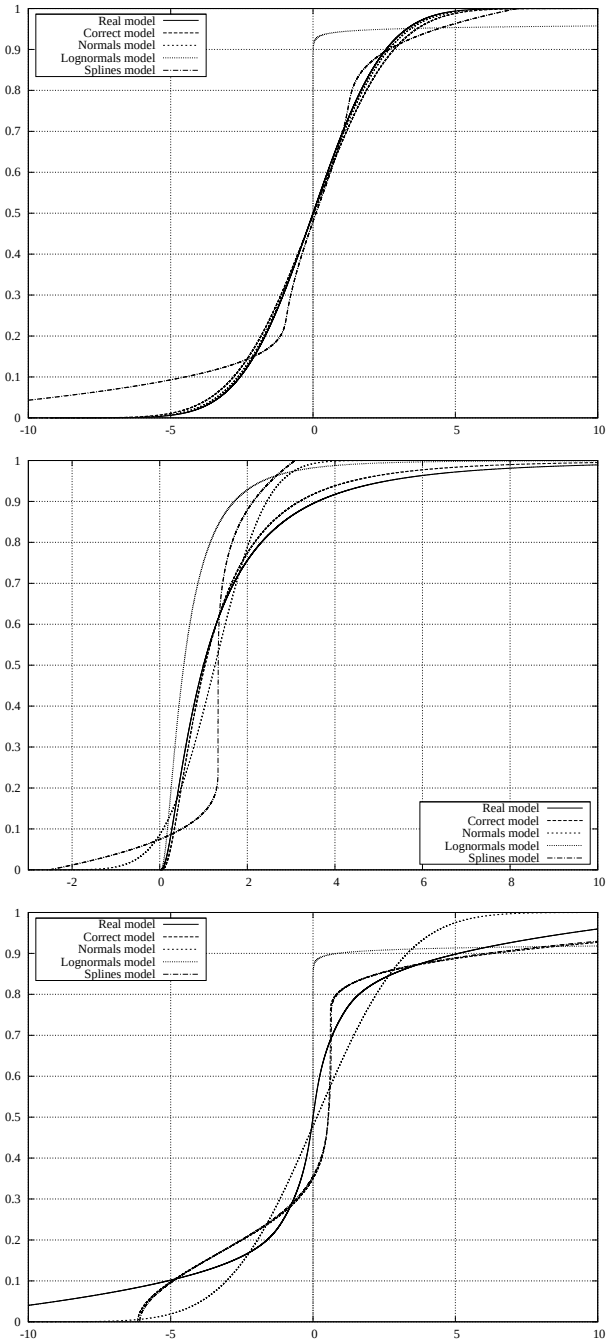

Figure 1: First simulated case: normal, lognormal and spline variates.

parameters are symmetric, as is the normal distribution. Such conditions favor the normal distribution, but we expect the opposite to happen when the distribution are highly unsymmetric.

This led us to design a second experiment, where we now use two independent variables, one lognormal, and one spline, for which the selected parameters are chosen to ensure unsymmetry. As illustrated in Table 2 and Figure 2, the all-normal model now perfoms very poorly when compared to the other formulations, while the all-lognormal model is acceptable. The true specification is obviously better, but the all-spline model again provides a good compromise.

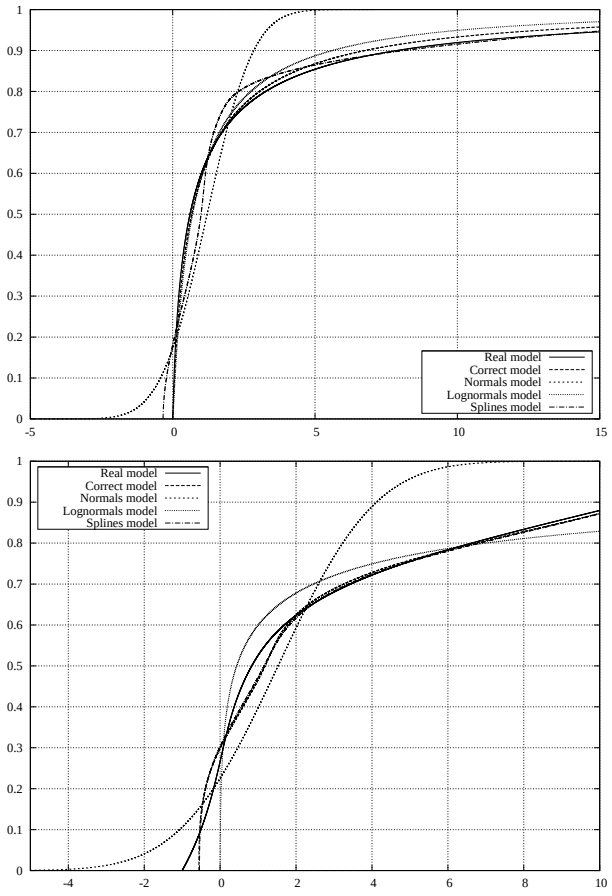

Figure 2: Second simulated case: normal, lognormal and spline variates.

These two examples indicate that a parametric model can behave well when information about the distributions families is available, but also that a misspecified model can lead to a very poor estimation. The nonparametric approach allows us to capture the randomness of the model, even in the absence of any knowledge of the underlying

Distribution	Real coeff.	Correct model	Normals	Lognormals	Splines
Lognormal	-0.5 (μ)	-0.40642264	1.209	-0.360	-0.337
	2.0 (σ)	1.8074374	1.278	1.625	-0.337
					0.222
					1.151
					1.151
					3.446
					27.123
Spline	-1.0	-0.564	1.526	-0.831	-0.543
	-0.5	-0.564	2.024	3.2983643	-0.543
	0.0	-0.564			-0.543
	0.5	1.456			1.317
	3.0	1.863			2.246
	12.0	13.454			13.240
	15.0	13.454			13.240
Log-likelihood	-	-1.100	-1.108	-1.103	-1.099

Table 2: Calibration on simulated data, second case.

distributions.

3.2 An application in finance

The algorithm developed has been applied to a financial problem concerning central bank interventions and dynamics in the foreign exchange (FX) market. Intervening in the FX market is a complex process (Beine et al. [8]); recently, several research papers have explored the determinants of interventions (Almekinders [1], Baillie and Osterberg [3], Ito [28], Ito and Yabu [29], Kearns and Rigobon [30], Bernal [10], and Bernal and Gnabo [11]). The main conclusions are that interventions tend to be conducted to counteract large deviations of the exchange rate from past levels. The existence of non-intervention bands have also been outlined, suggesting that monetary authorities incur significant costs when intervening on the foreign exchange markets. Our application concerns the Bank of Japan (BoJ), which has intervened in the FX market more than 300 times since the beginning of the 1990s, and has played a major role by conducting very large-scaled operations (Bernal [10]). The data used for our analysis have been collected from the Japanese Ministry of Finance's website (where they are publicly available) for the period April, 1991 to September, 2004. For each intervention, the exact date, amount and currencies involved are provided; our database contains data for a total of 3497 official trading days. In a given situation where intervention is possible, there are four possible outcomes of the central bank's decision: no intervention, public intervention, secret intervention detected by the market, secret intervention not detected by the market[3] (see Table 3 for the details about BoJ decisions in the period 1991-2004).

[3] "Pure secret interventions" are secret interventions which were not detected by the market, i.e. for which there were no news reports of the intervention at all. Conversely, "detected secret interventions" are secret interventions which were detected by the market: that is for which there was a highly speculative news report about the intervention (Beine et al. [8]).

Trading days	Intervention	Public	Undetected	Detected
3497	342	212	74	56

Table 3: Bank of Japan decisions 1991-2004

We again have that $\mathcal{X} = \mathbb{R}^n$ in this case. As for the simulated data, we chose 5 equally spaced knot points per spline approximation.

Discrete choice models are used to analyze BoJ decisions and to determine the factors that affect interventions on the FX market and their types. The four described outcomes constitute the set of alternatives available to BoJ. The independent variables are listed in Table 4; the relative coefficients are estimated as specific to the alternatives. For an exhaustive description of the procedure adopted to calculate these variables we refer the reader to Beine et al. [8]. The results of our model estimations are reported in Table 3, where we compare three model formulations. The estimated coefficients and their level of significance are reported for (a) a logit model, (b) a mixed logit model with error component and random parameter on the "amount" variable (assumed to be normally distributed), (c) a mixed logit with error component and random parameter on "amount" with nonparametric distribution (B-Spline). The error component is specified here to capture the correlation across the alternatives that share the secrecy in the intervention of the central bank. The fit of the model improves significantly from the logit formulation to the mixed logit formulation; the use of B-Spline also produces an improvement of the final value of the log-likelihood. We note that significant variables keep the same sign on the three different specifications and that they are consistent with the expectations of financial analysts. The error component is reported to be significant, confirming the hypothesis of correlation amongst the alternatives that share this error component. For a detailed economic interpretation of the results obtained we refer again to Beine et al. [8], and focus our attention in what follows on the variable "amount". In particular, central bankers reports that large interventions are much more likely to be detected (Neely [34]). The coefficient of the variable "amount" which plays a significant role in each intervention decision is found to be positive in the logit formulation (as expected). When the same coefficient is supposed to be normally distributed a large share (53%) of the intervention decisions are characterized by a negative value of the "amount" parameter. To overcome this problem a nonparametric distribution has been adopted, which produces a positively bounded coefficient in the interval $[0.01, 1.88]$.

The spline model also informs us that most of the time (70%), the invested amount is associated to a low parameter, as already suggested by the logit model, while this factor can be more important in approximately one third of the time. In view of this low "amount" parameter, we could reasonably wonder if the mixed logit model performs well only because of the error component. We therefore calibrated one additional model, fixing the "amount" parameter to a constant while using the error component term. The log-likelihood optimal value is then -596.84, suggesting that there is a significative heterogeneity in the influence of the invested amount, and that a random parameter helps to capture these variations. We conclude this section by observing that the use of mixed multinomial logit improves the fit of the model, but that the classi-

W (no intervention decision)	
Short term	Absolute level of short-term exchange rate deviation (%)
Medium term	Absolute level of medium-term exchange rate deviation (%)
Long term	Absolute level of long-term exchange rate deviation (%)
Misalignment	Absolute level of exchange rate misalignment (%)
Statement	1 if authorities made a statement expressing some discomfort with the exchange rate or confirming/discussing the intervention on the day of the operation
$\text{Intervention}_{t-1}$	1 if there was an official intervention the day before
RV_{t-1}	Exchange rate realized volatility of preceding day, estimated at the end of the day
Z (public process)	
Leaning	1 if the intervention tries to reverse recent exchange rate trend
Previous report	1 if the last detected intervention was a success
Inconsistency	1 if the intervention direction is inconsistent with the reduction of the exchange
Statement sum	Number of verbal interventions from the authorities signaling a discomfort with the exchange rate in the 5 days before the intervention
X (detection process)	
Amount	Amount invested in the daily intervention
Concertation	1 if intervention is concerted
Success	1 if the intervention moves the exchange rate in the desired direction
Cluster	1 if there is at least one detected intervention over the last 5 preceding days

Table 4: Variables description.

cal normal distribution fails to detect the real shape of the coefficient associated to the variable. Furthermore the nonparametric B-Spline model is able to detect the nonsymmetric nature of the random parameter and reveals that this coefficient is very different from zero in just one third of the intervention cases.

4 Conclusion

We have studied stochastic programming problems for which some probability distribution parameters are estimated along with the optimization, in which the determination of the true distribution or even of an empirical distribution prior the optimization is either difficult or impossible.

To alleviate this difficulty, we propose a new nonparametric approach based on approximations of the inverse cumulative distribution functions, in particular using cubic B-splines. We apply the proposed methodology in the context of discrete choice mod-

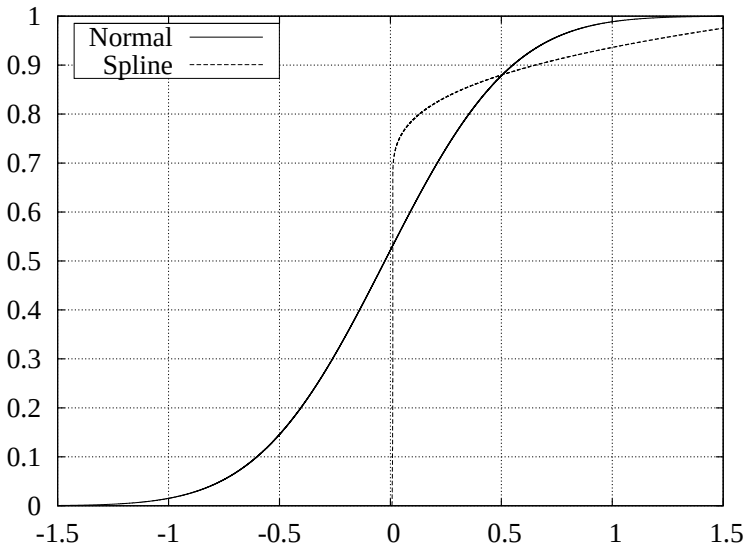

Figure 3: investment distributions.

eling. On simulated data, we conclude that misspecification of the underlying random variables often leads to irrealistic results while the nonparametric approach allows us to capture the randomness inside the population. We also validate the method on a financial model, dealing with bank interventions in the FX market, and obtain encouraging results.

Various questions however remain open. One may first wonder at how to choose the number of knots and their position in order to increase the estimation efficiency and quality, since too many knots would produce overfitting, and too few can lead to a poor approximation. We could possibly allow the knots to be chosen dynamically, along with the optimization process, but this problem is known to be very hard (see for instance Harashima, Ferrari, and Sanka [25]). Other heuristics are therefore of interest. A second question arises from the observation that most methods for generating multivariate distributions rely on the marginal functions and some treatment of the dependency between these marginals. These methods usually exploit the inversion technique to generate the marginals, and a natural extension of the method proposed here would concern its application to such multivariate random vector generation techniques.

The results obtained so far with our new nonparametric model appear to be encouraging. However, the authors are well aware that only continued experience with a broader variety of models and stochastic dependences will ultimately reveal the true protential of the ideas exposed in the present paper.

References

[1] G. J. Almekinders. *Foreign exchange intervention: theory and evidence*. Edward Elgar Publishing Ltd., Cheltenham, United Kingdom, 1995.

[2] A. N. Avramidis and J. R. Wilson. A flexible method for estimating inverse distribution functions in simulation experiments. *ORSA Journal of Computing*, 6:342–355, 1994.

[3] R. Baillie and W. Osterberg. Why do central banks intervene? *Journal of International Money and Finance*, 16(6):909–919, 1997.

[4] C. P. Barros, C. Ferreira, and J. Williams. Analysing the determinants of performance of best and worst european banks: A mixed logit approach. *Journal of Banking and Finance*, 31(7):2189–2203, 2007.

[5] F. Bastin, C. Cirillo, and S. Hess. Evaluation of advanced optimisation methods for estimating mixed logit models. *Transportation Research Record*, 1921:35–43, 2005.

[6] F. Bastin, C. Cirillo, and Ph. L. Toint. Convergence theory for nonconvex stochastic programming with an application to mixed logit. *Mathematical Programming*, 108(2–3):207–234, 2006.

[7] F. Bastin, C. Cirillo, and Ph. L. Toint. Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. *Transportation Science*, Submitted.

[8] M. Beine, O. Bernal, J.-Y. Gnabo, and C. Lecourt. Intervention policy of the BoJ: a unified approach. *Journal of Banking and Finance*, Submitted.

[9] M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, 1985.

[10] O. Bernal. Do interactions between political authorities and central banks influence FX interventions? Evidence from Japan. DULBEA working paper n. 06-03.RS.29, Université Libre de Bruxelles, Belgium, 2006.

[11] O. Bernal and J.-Y. Gnabo. Talk, financial operations or both? Generalising central banks FX reaction functions. SSRN working paper n. 932566, Université Libre de Bruxelles, Belgium, 2006.

[12] M. J. Best and N. Chakravarti. Active set algorithms for isotonic regression: a unifying framework. *Mathematical Programming*, 47:425–439, 1990.

[13] C. R. Bhat. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research*, 35B(7):677–693, Aug. 2001.

[14] D. Brownstone, D. S. Bunch, and K. Train. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research B*, 34(5):315–338, 2000.

[15] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Mathematical Programming*, 50(2):177–196, 1991.

[16] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, USA, 2000.

[17] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA, 1986.

[18] M. X. Dong and R. J.-B. Wets. Estimating density functions: A constrained maximum likelihood function. *Journal of Nonparametric Statistics*, 12(4):549–595, 2000.

[19] X. Dong and F. S. Koppelman. Mass point mixed logit model: Development and application. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, 2003.

[20] M. Fosgerau. Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, 40:688–707, 2006.

[21] M. Fosgerau and M. Bierlaire. A practical test for the choice of mixing distribution in discrete choice models. *Transportation Research Part B*, 41:784–794, 2007.

[22] M. Fosgerau and S. Hess. Competing methods for representing random taste heterogeneity in discrete choice models. Working paper, Danish Transport Research Institute, Copenhagen, Denmark, 2007.

[23] G. Glasgow. Mixed logit models for multiparty elections. *Political Analysis*, 9(1):116–136, 2001.

[24] G. Gürkan, A. Özge, and S. M. Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 84(2):313–333, 1999.

[25] M. Harashima, L. A. Ferrari, and P. V. Sankar. Spline approximation using kalman filter state estimation. *IEEE Transactions on Circuits and Systems. 2, Analog and digital signal processing*, 44(5):421–424, 1997.

[26] S. Hess, M. Bierlaire, and J. W. Polak. Discrete mixture of GEV models. Paper presented at 5th Swiss Transport Research Conference, 2005.

[27] S. C. Hora. Estimation of the inverse function for random number generation. *Communications of the ACM*, 26:590–594, 1983.

[28] T. Ito. Is foreign exchange intervention effective? The Japanese experience in the 1990's, monetary history, exchange rates and financial markets. In P. Mitzen, editor, *Monetary History, Exchange Rates and Financial Markets, Essays in Honour of Charles Goodhart*, volume 2, pages 126–153. Edward Elgar, United Kingdom, 2003.

[29] T. Ito and T. Yabu. What promotes Japan to intervene in the foreign market? A new approach to a reaction function. *Journal of International Money and Finance*, 26(2):193–212, 2007.

[30] J. Kearns and R. Rigobon. Identifying the efficacy of central bank interventions: evidence from Australia and Japan. *Journal of International Economics*, 66(1):31–48, 2005.

[31] A. M. Law. *Simulation Modeling & Analysis*. McGraw-Hill, New York, USA, 4th edition edition, 2007.

[32] P. L'Ecuyer. Efficiency improvement and variance reduction. In *Proceedings of 1994 Winter Simulation Conference*, pages 122–132, Orlando, USA, 1994.

[33] P. L'Ecuyer, J.-F. Cordeau, and R. Simard1. Close-point spatial tests and their application to random number generators. *Operations Research*, 48(2):308–317, 2000.

[34] C. J. Neely. Authorities'beliefs about foreign exchange intervention: Getting back to the hood. Technical Report 2006-045A, Federal Reserve Bank of Saint Louis, 2006.

[35] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, USA, 1999.

[36] G. C. Pflug. Stochastic optimization and statistical inference. In A. Shapiro and A. Ruszczyński, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 427–482. Elsevier, 2003.

[37] L. A. Piegl and W. Tiller. *The NURBS Book*. Springer-Verlag, New York, NY, USA, second edition, 1996.

[38] D. Rigby and M. Burton. Modeling disinterest and dislike: A bounded bayesian mixed logit model of the uk market for gm food. *Journal Environmental and Resource Economics*, 33(4):485–509, 2006.

[39] R. Y. Rubinstein and A. Shapiro. *Discrete Event Systems*. John Wiley & Sons, Chichester, England, 1993.

[40] A. Shapiro. Monte Carlo sampling methods. In A. Shapiro and A. Ruszczyński, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 353–425. Elsevier, 2003.

[41] K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, USA, 2003.

[42] K. Train and G. Sonnier. Mixed logit with bounded distributions of correlated partworths. In A. Alberini and R. Scarpa, editors, *Applications of Simulations Methods in Environmental Resource Economics*, pages 117–134. Kluwer Academics Publisher, Dordrecht, The Netherlands, 2005.

[43] K. Train and M. Weeks. Discrete choice models in preference space and willingness-to-pay space. In A. Alberini and R. Scarpa, editors, *Applications of Simulations Methods in Environmental Resource Economics*, pages 1–16. Kluwer Academics Publisher, Dordrecht, The Netherlands, 2005.

[44] L. Wasserman. *All of Nonparametric Statistics*. Springer Series in Statistics, New York, USA, 2007.

Model	Logit	Mixed logit I	Mixed logit II
Alternative specific constants			
No intervention	4.685 (21.7)	6.129 (14.7)	7.116
Public intervention	-0.797 (3.1)	-0.359 (0.9)	0.698
Detected secret intervention	-1.493 (6.0)	-1.850 (11.1)	-1.825
Error component	-	2.536 (6.3)	4.188
W (no intervention decision)			
Short term	-0.052 (0.3)	-0.082 (0.3)	-0.148
Medium term	-0.084 (1.5)	-0.103 (0.6)	-0.104
Long term	-0.015 (0.8)	-0.196 (1.6)	-0.099
Misalignement	-0.015 (1.9)	-0.028 (1.2)	-0.030
Statement	-0.712 (4.6)	-0.825 (4.0)	-0.851
$\text{Intervention}_{t-1}$	-2.832 (15.3)	-2.34 (5.27)	-2.364
RV_{t-1}	0.257 (2.1)	0.069 (0.5)	0.056
Z (public process)			
Leaning	3.807 (9.6)	3.904 (6.3)	3.789
Previous report	3.593 (6.3)	3.485 (7.7)	3.525
Inconsistency	-0.688 (1.3)	0.081 (0.1)	0.23
Statement sum	-0.462 (0.4)	0.021 (0.2)	-1.058
X (detection process)			
Invested amount P1	0.014 (4.5)	-0.025 (5.8)	0.007
Invested amount P2	-	0.450 (7.7)	0.010
Invested amount P3	-	-	0.011
Invested amount P4	-	-	0.011
Invested amount P5	-	-	0.011
Invested amount P6	-	-	1.882
Concertation	6.772 (2.0)	0.629 (15.1)	1.466
Success	2.447 (3.4)	1.586 (6.0)	1.649
Cluster	-0.563 (0.9)	-0.499 (1.5)	-0.333
Log-likelihood	-683.43	-571.61	-563.36

Table 5: Estimated parameters.