

# Multi-Objective Optimization using Statistical Models

Mike G. Tsionas\*

## Abstract

In this paper we consider multi-objective optimization problems (MOOP) from the point of view of Bayesian analysis. MOOP problems can be considered equivalent to certain statistical models associated with the specific objectives and constraints. MOOP that can explore accurately the Pareto frontier are Generalized Data Envelopment Analysis and Goal Programming. In turn, posterior analysis of their associated statistical models can be implemented using Markov Chain Monte Carlo (MCMC) simulation. In addition, we consider the minimax regret problem which provides robust solutions and we develop similar MCMC posterior simulators without the need to define scenarios. The new techniques are shown to work well in four examples involving non-convex and disconnected Pareto problems and to a real world portfolio optimization problem where the purpose is to optimize simultaneously average return, mean absolute deviation, positive and negative skewness of portfolio returns. Globally minimum regret can also be implemented based on post-processing of MCMC draws.

**Key Words:** Decision Analysis; Multi-objective optimization; Minimax Regret; Bayesian Analysis; Markov Chain Monte Carlo.

**Acknowledgments:** The author is grateful to two anonymous referees and Panos Xidonas for their useful comments on an earlier version.

---

\*Lancaster University Management School. LA1 4YX, U.K, email: [m.tsionas@lancaster.ac.uk](mailto:m.tsionas@lancaster.ac.uk)

# 1 Introduction

Since its inception, multi-objective optimization has found many applications in operations research and finance. See for example Florios et al. (2010), Xidonas et al. (2011, 2017), Zopounidis et al. (2015), Doumpos, and Zopounidis (2011), Doumpos et al. (2002, 2009, 2014), Zopounidis (1999), Pendaraki et al. (2005). In multi-objective optimization the problem is to optimize simultaneously several conflicting objectives subject to constraints. A promising approach is to use scalarization so that a single aggregate objective is optimized. Depending on the shape of the Pareto frontier, some aggregation function may perform better than others in tracing out the frontier.

Although it is impossible to summarize the use of meta-heuristics and other algorithms in multi-objective optimization, we should mention, the Non-dominated Sorting Genetic Algorithm (NSGA) proposed by Srinivas and Deb (1994). Refinements and advances include NSGA-II (Deb et al., 2002) and NSGA-III (Deb and Jain, 2014) and MOEA/D proposed by Zhang and Li (2007). From simulations it turns out that NSGA-II outperforms two other competing algorithms: Pareto-archived evolution strategy (Knowles and Corne, 1999) and strength Pareto evolutionary algorithm (Zitzler, 1999) in terms of finding a diverse set of solutions and convergence near the Pareto front. The two features of the evolutionary optimization approach in multiple - objective optimization are the assignment of a fitness function, and the diversity of solutions (viz. being different from each other to the greatest extent possible). To enhance convergence, “elitism” has been used widely in the literature, see for example Knowles and Corne (1999), Zitzler and Thiele (1998) and Rudolph (1998) as well as Deb and Jain (2014) for a useful summary.

In this paper we propose Bayesian analysis of multi-objective optimization programs. We show that the objectives and constraints define a sampling or statistical model which, in turn, can be used to implement posterior analysis using simulation algorithms, especially Markov Chain Monte Carlo (MCMC). In the area of evolutionary optimization and meta-heuristics, the Bayesian approach i) offers a principled way to define fitness, ii) allows incorporation of prior beliefs about the location and shape of the Pareto front when available; iii) provides access to a posterior distribution that converges to the Pareto the front using computational means; and iv) formalizes uncertainty of the statistical model as regards the Pareto front. The use of prior information is not *per se* undesirable; for example, reference points are often used (Deb and Jain, 2014, see also Jain and Deb, 2014). From another point of view, prior information may reflect the desire of the decision maker to concentrate on Pareto efficient solutions in the neighborhood of a given subset of the decision space. Moreover, formalization of uncertainty is related to how certain we are that any given point on the front or a subset of the front is Pareto non-dominated. In portfolio problems, this may offer valuable information to the decision maker, as stocks and other financial assets have inherently uncertain returns.

We rely on two techniques. First, Generalized Data Envelopment Analysis (DEA), see Yun et al. (2001) who proposed Generalized DEA to deal with possible non-convexity of Pareto sets). The second technique is applicable when the Pareto front is not only possibly non-convex but possibly disconnected as well. Dutta and Kaya (2011) have proposed techniques that can be used to deal with such disconnected sets. They have proposed a weighted Tchebychev problem that is also known as goal attainment (Mueller-Gritschneider et al., 2009) or Pascoletti–Serafini scalarization (Eichfelder, 2009a,b). The

idea explored in this paper is rather simple: Since DEA - like methods can be used to construct Pareto optima for possibly disconnected and / or non-convex Pareto sets, we can find a statistical model which is equivalent to GDEA and then perform inferences for the statistical model to approximate the solutions delivered by GDEA. In particular, access to the posterior is provided by powerful MCMC algorithms. We show that the statistical models required to approximate GDEA/GP are particularly simple (although nonlinear in the parameters of interest which are points on the Pareto front) and can be handled easily using state-of-the-art MCMC techniques.

These techniques do not require that we have prior information about the shape or other properties of the Pareto frontier, although such prior information can be used if it is available. Unsurprisingly, such techniques utilize ideas from the stochastic frontiers literature, namely that inequality constraints can be converted to equalities using one-sided error terms, plus a two-sided “noise” error term whose variance can be made arbitrarily small or can be estimated from the data, if so desired. Of course, the idea is well-known in linear programming where slack variables are used. However, in our set up these are *random* variables, not constants to be determined by the optimization.

The idea of introducing a two-sided error term, is particularly useful when the set of feasible points is the result of simulations which contain measurement errors. The number of feasible points can be very large, contrary to implementations of direct GDEA/GP. This, of course, opens up the possibility of constructing thousands of feasible points and then applying MCMC techniques (custom-made or off-the-shelve) to explore the Pareto front of large and possibly ill-behaved multi-objective optimization problems.

Moreover, we consider the problem of minimax regret (Kouvelis and Yu, 1997) and its recent implementation by Xidonas et al. (2017). The posterior of the model is explored using Markov Chain Monte Carlo and then applied to a real world portfolio optimization problem. In addition to risk and return we consider also positive and negative skewness which, to the best of our knowledge, is a novel approach to multi-objective portfolio optimization. Minimax regret is important as it can be used to deliver robust solutions to the problem. Additionally, the proposed techniques are automated in the sense that it is not necessary to define alternative scenarios: Uncertainty with respect to different states is *automatically* taken into account using posterior draws for these parameters from MCMC.

## 2 Background on multi-objective optimization and DEA approaches

### 2.1 General

In this section we provide general background material on multi-objective optimization. The reader that wants to skip this material can go directly to (12).

We consider the problem:

$$\min_{x \in X} \{f_1(x), \dots, f_M(x)\}, \tag{1}$$

where the decision variables are  $x \in X \subseteq \mathbb{R}^k$  and  $f_i : X \rightarrow \mathbb{R}$  ( $i = 1, \dots, M$ ) are  $M$  objective functions. We denote  $f(x) = [f_1(x), \dots, f_M(x)]'$ . Usually, the set  $X$  is described by nonlinear inequality, equality and bound constraints:

$$X = \{x \in \mathbb{R}^k, g_j(x) \leq 0 \forall j = 1, \dots, J, h_q(x) = 0 \forall q = 1, \dots, Q, \bar{x}_i \leq x_i \leq \underline{x}_i \forall i = 1, \dots, k\}. \quad (2)$$

For vectors  $a, b \in \mathbb{R}^M$  we use the notation:  $a \leq b \Leftrightarrow a_i \leq b_i \forall i = 1, \dots, M, \exists j \in \{1, \dots, M\} : a_j < b_j$ .

As the objectives are conflicting we use the Pareto optimality concept. The Pareto set is:

$$\mathcal{P} = \{x \in X : \exists x' \in X, f(x') \leq f(x)\}. \quad (3)$$

The Pareto frontier is the boundary of the Pareto set:

$$\mathcal{P}^* = \partial \mathcal{P} = \{y | y = f(x), x \in \mathcal{P}\}. \quad (4)$$

Locating the Pareto set is important in MOP. One way to find  $\mathcal{P}^*$  is to apply scalarization (Ehrgott, 2008). Therefore, we may consider the problem (Zadeh, 1963):

$$\min_{x \in X} \sum_{m=1}^M w_m f_m(x), \quad (5)$$

where  $w_m \geq 0 \forall m = 1, \dots, M$  and  $\sum_{m=1}^M w_m = 1$  are given weights. The method cannot locate points in a non-convex area of the Pareto front (Steuer, 1986). Kim and de Weck (2005, 2006) proposed the adaptive weighted sum method, and demonstrated that it finds solutions for non-convex regions and provide a well-distributed Pareto front. However, a Pareto-filter is applied to remove dominated solutions. Another approach is to solve problems of the form:

$$\begin{aligned} & \min_{x \in X} f_1(x) \\ & f_m(x) \leq \varepsilon_m \quad \forall m = 2, \dots, M. \end{aligned} \quad (6)$$

For the method in general see Lightner and Director (1981) and Li et al. (1999). For some developments see Laumanns et al. (2006) and Mavrotas (2009).

Given a feasible point  $x^o \in X$  another approach is to solve the problem:

$$\begin{aligned} & \min_{x \in X} f_1(x) \\ & f_m(x) \leq f_m(x^o) \quad \forall m = 2, \dots, M, \end{aligned} \quad (7)$$

which is related to the following:

$$\begin{aligned} \min_{x \in X} \sum_{m=1}^M w_m f_m(x) \\ f_m(x) \leq f_m(x^o) \quad \forall m = 1, \dots, M. \end{aligned} \quad (8)$$

In these problems, given a feasible point, one tries to determine a point that is optimal or at least better compared to a feasible solution. If we consider each problem alone we have:

$$x_{(m)}^* = \arg \min_{x \in X} f_m(x) \quad \forall m = 1, \dots, M. \quad (9)$$

In turn we can define “goals”:

$$f_m^* = f_m(x_{(m)}^*) \quad \forall m = 1, \dots, M. \quad (10)$$

Since the goals cannot be attained simultaneously, we can consider the following problem:

$$\min_{x \in X} \left\{ \sum_{m=1}^M w_m (f_m(x) - f_m^*)^p \right\}^{1/p}, \quad (11)$$

where  $p \geq 1$  is a scalar that defines an  $L_p$  norm.

## 2.2 The use of DEA methods

Next, we turn attention to DEA methods in multiple-objective optimization. DEA methods have a great potential as they target directly the envelope of a set of data and, thus, if a set of feasible points is available, the envelop approaches the Pareto frontier. The value of DEA methods has first been documented in Arakawa et al. (1998) and Yun et al. (2001). Given any point away from the Pareto front, the extent of inefficiency is a measure of the (shortest) distance of the point to the frontier. Alternatively we may consider the ray from the origin to the point. Yun et al. (2001) building on Arakawa et al. (1998) proposed solving a DEA problem. First define:

$$F_m(x) = f_m(x) + \sum_{j=1}^J \psi_j \left\{ [g_j(x)]_+ \right\}^a \quad \forall m = 1, \dots, M, \quad (12)$$

where  $\psi_j > 0$  is a penalty coefficient for the inequality restrictions,  $a$  is a parameter and  $[y]_+ = \max\{0, y\}$ . The equality and bound constraints can be treated in a similar manner and are omitted here. The DEA problem is:

$$\begin{aligned}
& \min_{\lambda, \vartheta} \vartheta \\
& [F(x^{(1)}), \dots, F(x^{(N)})] \lambda \leq \vartheta F(x^o) , \\
& \lambda \geq 0, \lambda \in \mathbb{R}^N,
\end{aligned} \tag{13}$$

to determine the efficiency of a given point (or “decision making unit”  $o \in \{1, \dots, N\}$ ). Moreover, the inputs are the values of the augmented objective function at points  $\{x^{(1)}, \dots, x^{(N)}\}$  and output is  $F(x^o)$ .

Yun et al. (2001) propose to use a Generalized DEA model to deal with possible non-convexity. The problem they propose is:

$$\begin{aligned}
& \max_{\Delta, v} \Delta \\
& \Delta \leq d_n - \alpha \sum_{m=1}^M v_m (F_m(x^o) - F_m(x^{(n)})) \quad \forall n = 1, \dots, N, \\
& v_n \geq \epsilon \quad \forall n = 1, \dots, N, \quad \sum_{n=1}^N v_n = 1,
\end{aligned} \tag{14}$$

where  $\alpha > 0$  is a constant that can be used to obtain various types of DEA problems,  $\epsilon > 0$  is a small non-Archimidean number, and  $d_n = \max_{m=1, \dots, M} \{v_m (f_m(x^{(n)}) - f_m(x^o))\}$ . In this method, taking a large  $\alpha$  can remove observations which are located far from the efficient frontier, while a small  $\alpha$  may be used in non-convex Pareto frontiers as it nests the free disposal hull in this case. Efficiency requires that the optimal value,  $\Delta = 0$ .

Since (14) can handle non-convex Pareto fronts we focus on this optimization problem. It is clear that provided  $\mathcal{X} = \{x^{(n)}, n = 1, \dots, N\}$  are all feasible, the problem simplifies to:

$$\begin{aligned}
& \max_{\Delta, v} \Delta \equiv \Delta(x^o, \mathcal{X}, \alpha) \\
& \Delta \leq d_n - \alpha \sum_{m=1}^M v_m (f_m(x^o) - f_m(x^{(n)})), n = 1, \dots, N, \\
& v_n \geq \epsilon, n = 1, \dots, N, \quad \sum_{n=1}^N v_n = 1
\end{aligned} \tag{15}$$

Therefore, given a set of feasible points one can approximate the Pareto front using the points where  $\Delta = 0$ . The notation  $\Delta(x^o, \mathcal{X}, \alpha)$  makes explicit that we evaluate observation  $o$  conditional on the set of feasible points  $\mathcal{X}$  when parameter  $\alpha$  is used.

Problem (15) is solved for each  $o \in \{1, \dots, N\}$  to obtain  $\Delta$  for each observation. **The observations corresponding to  $\Delta = 0$  can be used to approximate the Pareto front. Such methods are quite useful and provided we can cast the problem in terms of Bayesian inference, then we can exploit powerful Markov Chain Monte Carlo methods to develop posterior inferences for the Pareto frontier. We should also mention that the number of problems solved with a Goal Programming model is only defined by how many Pareto optimal solutions the decision maker wants to find. Goal Programming models are usually applied in an interactive context, so that using preferential information from the decision maker, generally only a small number of relevant solutions are derived (i.e., the ones that best suit the preferences of the decision maker).**<sup>1</sup>

In Bayesian analysis, the objective is to provide the posterior distribution of  $x^o$  conditional on the data  $\mathcal{X}$ . Therefore, given a possibly very large data set that consists of feasible solutions (called “observations”), we would like to trace out the

<sup>1</sup>The author is grateful to anonymous reviewer for making this point.

---

**Algorithm 1** Representation of basic idea

---

1. Specify a set of feasible points  $\mathcal{X} = \{x^{(i)}, i = 1, \dots, N\}$ .
  2. Focus on either the GDEA optimization problem in (15) or the Goal Attainment (GA) problem in (31).
  3. Convert the optimization problem to a statistical model with parameters  $\theta \in \Theta \subseteq \mathbb{R}^d$  and data  $\mathcal{X}$ . (How to do this precisely will be explained in detail in the main text). Points on the Pareto front are elements of  $\theta$ .
  4. Apply MCMC to explore the posterior  $p(\theta|\mathcal{X})$ :
  5. Typically, this involves drawing sequentially random numbers from the posterior conditional distributions  $p(\theta_j|\theta_{-j}, \mathcal{X}) \forall j = 1, \dots, d$ .  
Specifically, this involves drawing random numbers, sequentially, from the posterior conditional densities presented in (25)–(27) using (28) and (29).
  6. Use the posterior means of  $\{\theta_j, j = 1, \dots, d\}$  obtained from MCMC simulation to approximate the Pareto front.
- 

Pareto front by considering the posterior distribution of  $x^o$ . The method can be used in complicated and / or very large problems where i) it is critical to explore the set of restrictions, and ii) the functions can only be measured with error. Tsionas (2018) employed Bayesian analysis of the multi-objective programming problem without the use of statistical models.

Our approach can be described in terms of a pseudo-algorithm as in Algorithm 1.

The idea relies on replacing optimization by simulation. Although both are costly in terms of computing time, Bayesian simulation of a model with  $N$  observations is typical in the literature.

Moreover, we should not underestimate another advantage of the Bayesian approach. MCMC simulation delivers standard errors, probability intervals and highest-posterior-density intervals for all parameters of interest. The parameters of interest, in this paper, are points on the Pareto front so measures of uncertainty can be associated with the entire Pareto front, something that seems to be novel to the best knowledge of the author. Another issue is that implementation of this approach requires specification of a set of feasible points. As an anonymous reviewer mentioned, “if the problem is linear, then this is straightforward. But in non-convex problems (e.g., non-linear or combinatorial), it may not be easy at all to find feasible solutions.” In all artificial examples and the empirical application of this paper, we have found that a direct random search easily produces a large number of feasible solutions. In general, of course, this is not so. A general method that can be used in this context, and is closely connected to Bayesian analysis and MCMC sampling is the hit-and-run algorithm, see for example Kaufman and Smith (1994). The hit-and-run algorithm can be used to sample, uniformly, from the “posterior” defined by the constraints. For problems that involve integer-valued variables, as in combinatorial problems, these variables can be treated as parameters whose support is  $\{0, 1\}$  and the prior probability is, say,  $1/2$  (expressing ignorance over what these variables should be in an optimal solution). In turn, the hit-and-run algorithm can be augmented with a Gibbs-sampling step in which the integer-valued variables are sampled from their posterior: The posterior has the same support,  $\{0, 1\}$ , but the posterior probabilities are updated based on information from the constraints. As we do not deal with combinatorial optimization in this paper, we leave the topic for future research.

Next, we turn attention to the description of Bayesian methods.

### 3 Bayesian analysis of the problem

#### 3.1 General

Given a set of feasible points  $\mathcal{X}$  that we call “observations”, we would like to trace out the Pareto front, as best as possible, through a statistical model. To do so: *First*, we treat  $x^o$  as a parameter and we denote it by  $\tilde{x}$ .

In the constraints of (15), we add an error term  $\varepsilon_n$  with small variance  $\sigma^2$ . This allows us to have a conditionally normal distribution if the errors are normally distributed, say  $\varepsilon_n \sim iid\mathcal{N}(0, \sigma^2) \forall n = 1, \dots, N$  where  $\sigma$  is sufficiently small. However, when the functions can only be measured with error,  $\sigma$  can be treated as an unknown parameter to be inferred from the data.

*Second*, we convert the inequalities to equalities using:

$$d_n = \alpha \sum_{m=1}^M v_m \left( f_m(\tilde{x}) - f_m(x^{(n)}) \right) = \varepsilon_n + \Delta_n \forall n = 1, \dots, N, \quad (16)$$

where  $\Delta_n \geq 0$  and therefore we can assume it follows a half-normal distribution, say  $\Delta_n \sim iid\mathcal{N}_+(0, h^2) \forall n = 1, \dots, N$ . Equivalently, we obtain:

$$d_n = \sum_{m=1}^M \beta_m \left( f_m(\tilde{x}) - \bar{f}_{mn} \right) + \varepsilon_n + \Delta_n \forall n = 1, \dots, N, \quad (17)$$

where  $\bar{f}_{mn} = f_m(x^{(n)})$ ,  $\beta_m = \alpha v_m \forall m = 1, \dots, M$  so that  $\sum_{m=1}^M \beta_m = \alpha$ . Notice that  $\beta_m$ s act as *regression parameters* in this instance. Indeed the model can be interpreted as a regression where the dependent variable is  $d_n$ , the  $M$  regressors are  $f_m(\tilde{x}) - \bar{f}_{mn}$ ,  $\varepsilon_n$  is a two-sided error and  $\Delta_n$  is a one-sided error. With the new notation, we have:  $d_n = \alpha^{-1} \max_{m=1, \dots, M} \{ \beta_m (f_m(x^{(n)}) - f_m(x^o)) \}$  so the dependent variable, in fact, depends on the parameters. However, the model can be written in the form of a nonlinear regression:

$$d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m \left( f_m(\tilde{x}) - \bar{f}_{mn} \right) = \varepsilon_n + \Delta_n \forall n = 1, \dots, N, \quad (18)$$

where the dependence of  $d_n$  on  $\alpha$  and  $\beta$  is made explicit. Obviously, the  $\beta_m$ s depend on  $\alpha$  as well but we suppress this dependence to ease notation.

The posterior in this instance can be defined as:

$$p(\tilde{x}, \beta, h, \Delta | \mathcal{X}) \propto h^{-N} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N \left( d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m \left( f_m(\tilde{x}) - f_m(x^{(n)}) \right) - \Delta_n \right)^2 - \frac{1}{2h^2} \sum_{n=1}^N \Delta_n^2 \right\} p(\tilde{x}, \beta, h), \quad (19)$$

where  $\Delta = [\Delta_n, n = 1, \dots, N]$ ,  $h > 0$  is a scale parameter,  $\beta = [\beta_m, m = 1, \dots, M]$  and  $p(x, \beta, h)$  is the prior of  $x, \beta, h$ .



If we do not have information about the location of the Pareto front, a reasonable prior is:

$$p(\tilde{x}|\beta, h) \propto \mathbb{I}(\tilde{x} \in X), \quad (20)$$

where  $\mathbb{I}(\tilde{x} \in X)$  is an indicator function to enforce that we always have a feasible point. The prior on  $h$  is also non-informative:

$$p(h|\tilde{x}, \beta) \propto h^{-1}. \quad (21)$$

Finally we use a non-informative prior for the  $\beta$ s:

$$p(\beta|\tilde{x}, h) \propto \mathbb{I}(\beta \geq 0, \sum_{m=1}^M \beta_m = \alpha). \quad (22)$$

Therefore the posterior is:

$$p(\tilde{x}, \beta, h, \Delta|\mathcal{X}) \propto h^{-N-1} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N \left( d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m (f_m(\tilde{x}) - f_m(x^{(n)}) - \Delta_n) \right)^2 - \frac{1}{2h^2} \sum_{n=1}^N \Delta_n^2 \right\} \cdot \mathbb{I}(\tilde{x} \in X) \mathbb{I}(\beta \geq 0, \sum_{m=1}^M \beta_m = \alpha). \quad (23)$$

It is perhaps important to notice that if we treat  $\sigma$  as a parameter whose prior is  $p(\sigma) \propto \sigma^{-1}$  independently of other parameters, then  $\sigma$  can be integrated out of the posterior<sup>2</sup> so that it does not appear in the final problem. Moreover, we use the following prior for the scaling constant:  $p(h) \propto h^{-1}$ , to obtain:

$$p(\tilde{x}, \beta, h, \Delta|\mathcal{X}) \propto h^{-(N+1)} \left\{ \sum_{n=1}^N \left( d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m (f_m(\tilde{x}) - f_m(x^{(n)}) - \Delta_n) \right)^2 \right\}^{-(N+1)/2} \cdot \exp \left\{ -\frac{1}{2h^2} \sum_{n=1}^N \Delta_n^2 \right\} \mathbb{I}(\tilde{x} \in X) \mathbb{I}(\beta \geq 0, \sum_{m=1}^M \beta_m = \alpha). \quad (24)$$

To obtain long samples from this posterior we use a Gibbs sampler. The Gibbs sampler operates on the augmented parameter space as we consider  $\Delta$  part of the parameter vector. Notice that  $h$  is treated as a parameter so it is not necessary to assign an arbitrary value to it.

We derive the following conditional posteriors<sup>3</sup> from (24):

$$h|\tilde{x}, \beta, \Delta, \mathcal{X} : \frac{\sum_{n=1}^N \Delta_n^2}{h^2} | \tilde{x}, \beta, \Delta, \mathcal{X} \sim \chi_N^2, \quad (25)$$

<sup>2</sup>The interested reader is referred to Zellner (1971, p. 67) on how the integration can be performed analytically.

<sup>3</sup>Some background information on the derivation is provided in Appendix A.

where  $\chi_N^2$  denotes the *chi*-square distribution.

$$\Delta_n | \tilde{x}, \beta, h, \mathcal{X} \sim \mathcal{N}_+ \left( \frac{h^2 R_n}{\sigma^2 + h^2}, \frac{h^2 \sigma^2}{\sigma^2 + h^2} \right), R_n \equiv d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m \left( f_m(\tilde{x}) - f_m(x^{(n)}) \right) \quad \forall n = 1, \dots, N. \quad (26)$$

Finally, for  $\beta$  we have:

$$p(\tilde{x}, \beta, h, \Delta | \mathcal{X}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N \left( d_n(\alpha, \beta) - \sum_{m=1}^M \beta_m \left( f_m(\tilde{x}) - f_m(x^{(n)}) \right) - \Delta_n \right)^2 \right\} \mathbb{I}(\beta \geq 0, \sum_{m=1}^M \beta_m = \alpha). \quad (27)$$

The conditional posteriors in (25) and (26) are standard and drawing random numbers is straightforward. Drawing from the conditional posterior of  $\beta$  in (27) can be accomplished using a simple Metropolis-Hastings random walk step: If the current draw is  $\beta^{(s)}$  we draw a candidate  $\beta^c = \beta^{(s)} + \delta z$ , where  $z \sim \mathcal{N}(0, I_M)$ . The candidate is accepted with probability:

$$A_\beta(\beta^{(s)}, \beta^c) = \min \left\{ 1, \frac{p(\tilde{x}, \beta^c, h, \Delta | \mathcal{X})}{p(\tilde{x}, \beta^{(s)}, h, \Delta | \mathcal{X})} \right\} \quad (28)$$

else we set  $\beta^{(s+1)} = \beta^{(s)}$ . Parameter  $\delta$  is chosen so that approximately 25% of all candidates are accepted during the burn-in phase of the Gibbs sampler. For  $\tilde{x}$  we use a similar technique. If the current draw is  $\tilde{x}^{(s)}$  we draw a candidate  $\tilde{x}^c = \tilde{x}^{(s)} + \delta_{\tilde{x}} z$ , where  $z \sim \mathcal{N}(0, I_k)$  provided  $\tilde{x} \in X$ . The candidate is accepted with probability:

$$A_{\tilde{x}}(\tilde{x}^{(s)}, \tilde{x}^c) = \min \left\{ 1, \frac{p(\tilde{x}^c, \beta, h, \Delta | \mathcal{X})}{p(\tilde{x}^{(s)}, \beta, h, \Delta | \mathcal{X})} \right\} \quad (29)$$

else we set  $\tilde{x}^{(s+1)} = \tilde{x}^{(s)}$ . Parameter  $\delta_{\tilde{x}}$  is chosen so that approximately 25% of all candidates are accepted during the burn-in phase of the Gibbs sampler.

In turn we produce a long sample, say  $\{\tilde{x}^{(s)}, \beta^{(s)}, h^{(s)}, \Delta^{(s)}, s = 1, \dots, S\}$  which converges to (23) in distribution.

### 3.2 Disconnected Pareto sets in multi-objective optimization

A lot of effort has gone into devising techniques that can be used when the Pareto front is not only non-convex but possibly disconnected as well. Dutta and Kaya (2011) have proposed techniques that can be used to deal with such disconnected sets. They propose the weighted Tchebychev problem (or Tchebychev scalarization):

$$\min_{x \in X} \max_{m=1, \dots, M} w_m (f_m(x) - f_m^*) \quad (30)$$

The method has been also used in Holzmann and Smith (2018). From Miettinen (1999, Part II, Theorems 3.4.2 and 3.4.5) we know that a point is a (weak) Pareto optimum if and only if it solves problem (30), for some weights.

The method has been found quite useful in Mavrotas et al. (2015) where robustness of Pareto optima is analyzed. One can write the problem equivalently as follows:

$$\begin{aligned} & \min \gamma \\ & x \in X, \\ & w_m (f_m(x) - f_m^*) \leq \gamma \forall m = 1, \dots, M. \end{aligned} \tag{31}$$

The problem is also known as Goal Attainment (Mueller-Gritschneider et al., 2009) or Pascoletti–Serafini scalarization (Eichfelder, 2009a,b). There is an interesting statistical model associated with this approach. Given a set of feasible points<sup>4</sup>  $\mathcal{X} = \{x^{(1)}, \dots, x^{(N)}\}$  we have:

$$\beta_m (f_m(\tilde{x}) - f_m(x^{(n)})) = \varepsilon_{mn} + \gamma_n \forall m = 1, \dots, M, n = 1, \dots, N, \tag{32}$$

where  $\beta_m = w_m$ , the two-sided errors  $\varepsilon_{mn} \sim iid\mathcal{N}(0, \sigma^2) \forall m = 1, \dots, M, n = 1, \dots, N$  for a given and sufficiently small  $\sigma$ , and  $\gamma_n \sim iid\mathcal{N}_+(0, h^2) \forall n = 1, \dots, N$ .

If we denote  $\bar{f}_{mn} = f_m(x^{(n)})$  these can be treated as data and we have:

$$\beta_m (f_m(\tilde{x}) - \bar{f}_{mn}) = \varepsilon_{nm} + \gamma_n \forall m = 1, \dots, M, n = 1, \dots, N. \tag{33}$$

For all  $m = 1, \dots, M$  this is a system of  $M$  equations in the vector parameter  $\tilde{x} \in X$ , the weights appear as simple linear coefficients and the one-sided error is common for all equations / objectives. The posterior corresponding to this model is as follows:

$$p(\tilde{x}, \beta, h, \gamma | \mathcal{X}) \propto h^{-N-1} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M (\beta_m (f_m(\tilde{x}) - \bar{f}_{mn}) - \gamma_n)^2 - \frac{1}{2h^2} \sum_{n=1}^N \gamma_n^2 \right\} \mathbb{I}(\tilde{x} \in X) \mathbb{I}(\beta \geq 0), \tag{34}$$

where  $\beta = [\beta_1, \dots, \beta_M]'$ ,  $\gamma = [\gamma_1, \dots, \gamma_N]'$ . Following the same reasoning as in the previous section, the prior of  $\tilde{x}$  is uniform in the feasible set  $X$  and  $\beta$  follows an improper distribution which imposes only non-negativity of the elements of  $\beta$ . Posterior analysis using Markov Chain Monte Carlo, especially the Gibbs sampler, is again straightforward in this case. We have the

<sup>4</sup>A typical multi-objective problem has an infinite number of feasible points. Here,  $\mathcal{X}$  denotes a randomly selected set of feasible points or “observations”.

following posterior conditional distributions:

$$\frac{\sum_{n=1}^N \gamma_n^2}{h^2} | \tilde{x}, \beta, \gamma, \mathcal{X} \sim \chi_N^2, \quad (35)$$

$$\gamma_n | \tilde{x}, \beta, h, \mathcal{X} \sim \mathcal{N}_+ \left( \frac{h^2 R_n}{\sigma^2 + h^2}, \frac{h^2 \sigma^2}{\sigma^2 + h^2} \right), \quad R_n \equiv \sum_{m=1}^M (\beta_m (f_m(\tilde{x}) - \bar{f}_{mn})) \quad \forall n = 1, \dots, N. \quad (36)$$

Since

$$p(\tilde{x} | \beta, h, \gamma, \mathcal{X}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M (\beta_m (f_m(\tilde{x}) - \bar{f}_{mn}))^2 \right\} \mathbb{I}(\tilde{x} \in X), \quad (37)$$

we use a random walk Metropolis - Hastings to provide draws from the posterior conditional distributions of  $\tilde{x}$ .

From (32) we have:

$$\beta_m z_{mn} = \varepsilon_{mn} + \gamma_n \quad \forall m = 1, \dots, M, n = 1, \dots, N, \quad (38)$$

where  $z_{mn} = f_m(\tilde{x}) - f_m(x^{(n)})$ . Stacking for all observations we obtain:

$$\gamma = \beta_m z_m + \varepsilon_m, \quad m = 1, \dots, M, \quad (39)$$

where  $z_m = [z_{mn}, n = 1, \dots, N]$ ,  $\varepsilon_m = [\varepsilon_{mn}, n = 1, \dots, N]$  and  $\gamma = [\gamma_n, n = 1, \dots, N]$ . From this expression we obtain the posterior conditional distribution:

$$\beta_m | \tilde{x}, h, \gamma, X \sim \mathcal{N} \left( \frac{z'_m \gamma}{z'_m z_m}, \frac{\sigma^2}{z'_m z_m} \right) \quad \forall m = 1, \dots, M, \quad (40)$$

which involves simple operations. Since MCMC requires connected sets we can either i) restart the algorithm using random initial conditions in space  $\mathcal{X}$  or ii) apply a Pareto filter in the original space of solutions. This is not particularly expensive as MCMC converges quickly to the Pareto front from arbitrary initial conditions (Roberts and Smith, 1994). Even the first proposal is feasible as it can be parallelized easily on modern multi-core computers. Here, we follow the more conventional approach of applying a Pareto filter, although there are other MCMC techniques that can deal with the problem of disconnected Pareto sets. These methods are not described here as they would take us astray. One such prominent method is described in Hoogerheide et al. (2007) who introduced neural networks as importance or candidate generating densities for MCMC. Another technique that could be used is AdMit (Adaptive Mixtures of Student- $t$  distributions) which uses mixtures of multivariate Student- $t$  densities to craft a good proposal to investigate multimodal and / or non-elliptical posteriors. Finally, Adaptive Radial Direction Sampling (ARDS, Bauwens et al., 2004) can be used which is an easy to implement Metropolis-Hastings sampler.

Before we proceed with examples and the application, we define two models: **Model I** assumes that  $\sigma$  is fixed at

$\sigma = 10^{-3}$ . In this case we use the posterior in (23). **Model II** assumes that  $\sigma$  is unknown and, therefore, it is integrated out of the posterior as in (24). The two methods produce nearly the same results, which is not surprising as we are not dealing with noisy data.

### 3.3 Examples and Empirical Application

We apply the new techniques to some examples.<sup>5</sup> We use the posterior in (23) in all examples except example 3 where we use (34).

**Example 1** is from Yun et al. (2001), example 3:

$$\begin{aligned} \min \{f_1(x), f_2(x)\} &= \{x_1, x_2\} \\ x_1^3 - 3x_1 - x_2 &\leq 0, \\ x_1 &\geq -1, x_2 \leq 2. \end{aligned} \tag{41}$$

The Pareto front is non-convex. As we can see from Figure 1, Pareto frontiers generated by the proposed methods are smooth, even though they are non-convex. On the left we show the entire set of feasible solutions and on the right we present only the Pareto set. Both are shown in the decision space. Yun et al. (2001) compared three different methods in this instance, viz. a ranking method, DEA and GDEA. MCMC seems to perform quite well in this application, see Figure 7 in Yun et al. (2001).

**Example 2** is benchmark 1 from Chiandussi et al. (2012):

$$\begin{aligned} \min \{f_1(x), f_2(x)\} \\ f_1(x) &= 1 - \exp \left\{ - \sum_{i=1}^3 \left( x_i - \frac{1}{\sqrt{3}} \right)^2 \right\}, \\ f_2(x) &= 1 - \exp \left\{ - \sum_{i=1}^3 \left( x_i + \frac{1}{\sqrt{3}} \right)^2 \right\}, \\ -4 &\leq x_1, x_2, x_3 \leq 4. \end{aligned} \tag{42}$$

This is a bound-constrained problem with two objective functions in three variables. The Pareto front is described by:  $x_1 = x_2 = x_3$ ,  $|x_1| \leq \frac{1}{\sqrt{3}}$ . It is concave, and the results of MCMC are presented in Figure 2. The results are similar to those obtained by Figure 3.2 in Chiandussi et al. (2012).

**Example 3** is benchmark 2 from Chiandussi et al. (2012):

$$\begin{aligned} \min \{f_1(x), f_2(x)\} \\ f_1(x) = x_1, f_2(x) &= (1 + 10x_2) \left[ 1 - \left( \frac{x_1}{1+10x_2} \right)^2 - \frac{x_1}{1+10x_2} \sin(8\pi x_1) \right], \\ 0 &\leq x_1, x_2 \leq 1. \end{aligned} \tag{43}$$

---

<sup>5</sup>In all cases, including the empirical application in section 8, we use 15,000 MCMC iterations the first 5,000 of which are discarded to mitigate the possible impact of start up effects. In all cases we use  $N=10,000$  feasible points.

In this case we use the posterior in (34). It is a bound-constrained problem with two objective functions. The first is a plane and the second is a sinusoidal function with increasing amplitude. It is a difficult problem in that the Pareto front is not continuous. Our results are reported in Figure 3 and they seem to be very similar to those in Figure 3.7 of Chiandussi et al. (2012).

**Example 4** is benchmark 3 from Chiandussi et al. (2012) and Coello Coello (2001):

$$\begin{aligned}
 & \min \{f_1(x), f_2(x)\} \\
 & f_1(x) = L(2x_1 + \sqrt{2}x_2 + \sqrt{2}x_3 + x_4), \quad f_2(x) = \frac{FL}{E} \left( \frac{2}{x_1} + \frac{2\sqrt{2}}{x_2} - \frac{2\sqrt{2}}{x_3} + \frac{2}{x_4} \right), \\
 & \frac{F}{\sigma} \leq x_1 \leq 3\frac{F}{\sigma}, \\
 & \sqrt{2}\frac{F}{\sigma} \leq x_2 \leq 3\frac{F}{\sigma}, \\
 & \sqrt{2}\frac{F}{\sigma} \leq x_3 \leq 3\frac{F}{\sigma}, \\
 & \frac{F}{\sigma} \leq x_4 \leq 3\frac{F}{\sigma},
 \end{aligned} \tag{44}$$

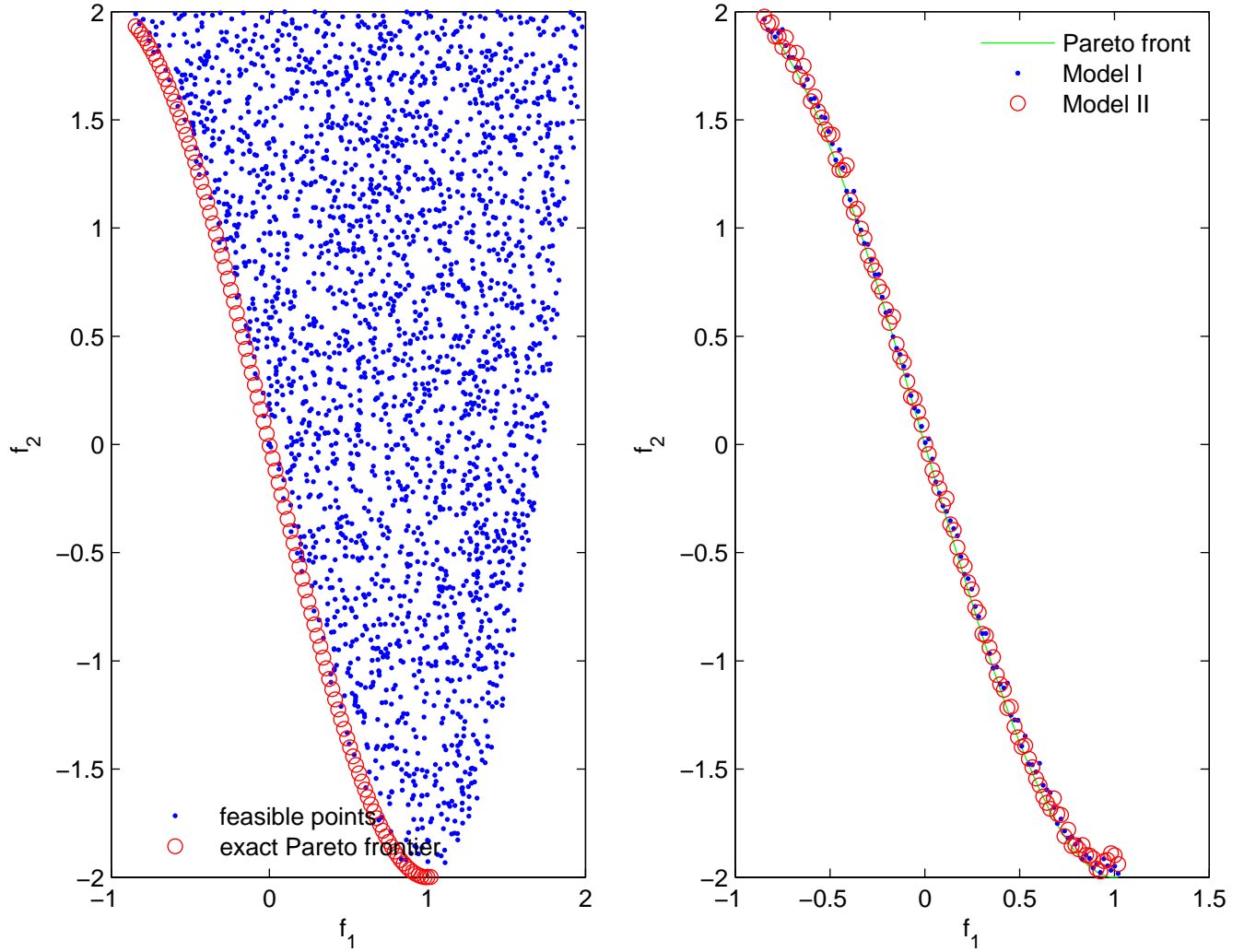
where  $F = 10$ ,  $E = 2 \times 10^5$ ,  $L = 200$ ,  $\sigma = 10$ . In this problem the objective is to minimize the mass and the compliance of a four bar plane truss respectively. In this case the Pareto front is convex. The results are presented in Figure 4 and show that the new techniques provide acceptable approximations to the Pareto front, see also Figures 3.14 - 3.17 of Chiandussi et al. (2012). As these authors mention: “The Pareto front can be fully described in detail with the drawback of a very large number of iterations. The method cannot be stopped before the conclusion of the process otherwise any optimal solution would be available for the decision maker” (p. 925-926). Therefore, to mimic this situation we decided to use only 2,000 iterations and discard the first 500 to mitigate possible start up effects. Then, we obtain points that belong to the Pareto front as in Figure 5. With 160,000 iterations the first 10,000 of which are discarded the entire Pareto front can be traced accurately (results are omitted but are available on request. Computing time was trivial on a mainframe computer and about 3 minutes on a desktop computer (Intel Core i9 7900X @ CPU 3.30 GHz). The same timings were obtained, roughly ( $\pm$  a few seconds) for Examples 1 – 3. Moreover, in all cases, including the empirical application that we present in the next section, timings are similar to a simulated annealing algorithm that uses the same number of iterations as our MCMC procedures.

The message from Figures 1 – 4, is that the techniques proposed here perform well, especially when the Pareto front is a disconnected set and we do *not* have prior knowledge about its shape. Both DEA and GA seem to perform equally well when the Pareto set is connected. When this is not the case we can use GA which is always applicable independently of the shape of the Pareto front.<sup>6</sup>

---

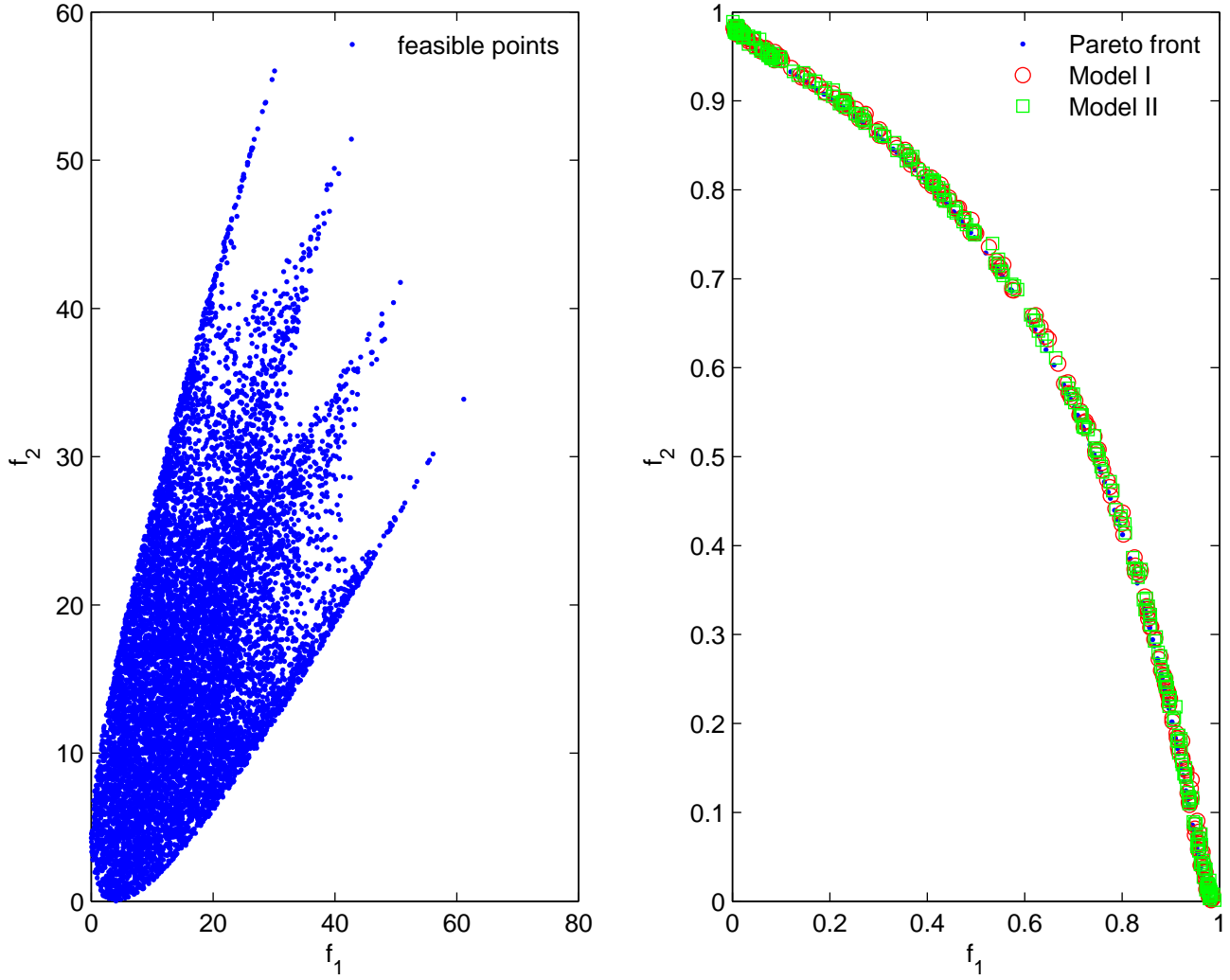
<sup>6</sup>We have used the High End Computing Cluster (HEC) at Lancaster University. This is a centrally-run service which offers over 6,500 cores, 28 TB of aggregate memory, 70TB of high performance filestore and 1.5PB of medium performance filestore. A number of nodes offer Nvidia GPU cards, which support CUDA and OpenCL applications. The cluster operating system is Scientific Linux, with job submission handled by Son of Grid Engine (SoGE). We ran the same jobs on an Intel ©i9 @ 3.4 GHZ. The first example took 7.3 seconds. The other three examples took 8.5, 8.5 and 0.6 seconds, respectively. The MCMC in actual empirical application, took 11.7 seconds. The overall timing for the application was 55.12 seconds but other methods have been used as well.

Figure 1: Example 1



Notes: On the left we show the feasible set and on the right the Pareto frontier. Model I assumes that  $\sigma$  is fixed at  $\sigma = 10^{-3}$ . In this case we use the posterior in (23). Model II assumes that  $\sigma$  is unknown and, therefore, it is integrated out of the posterior as in (24).

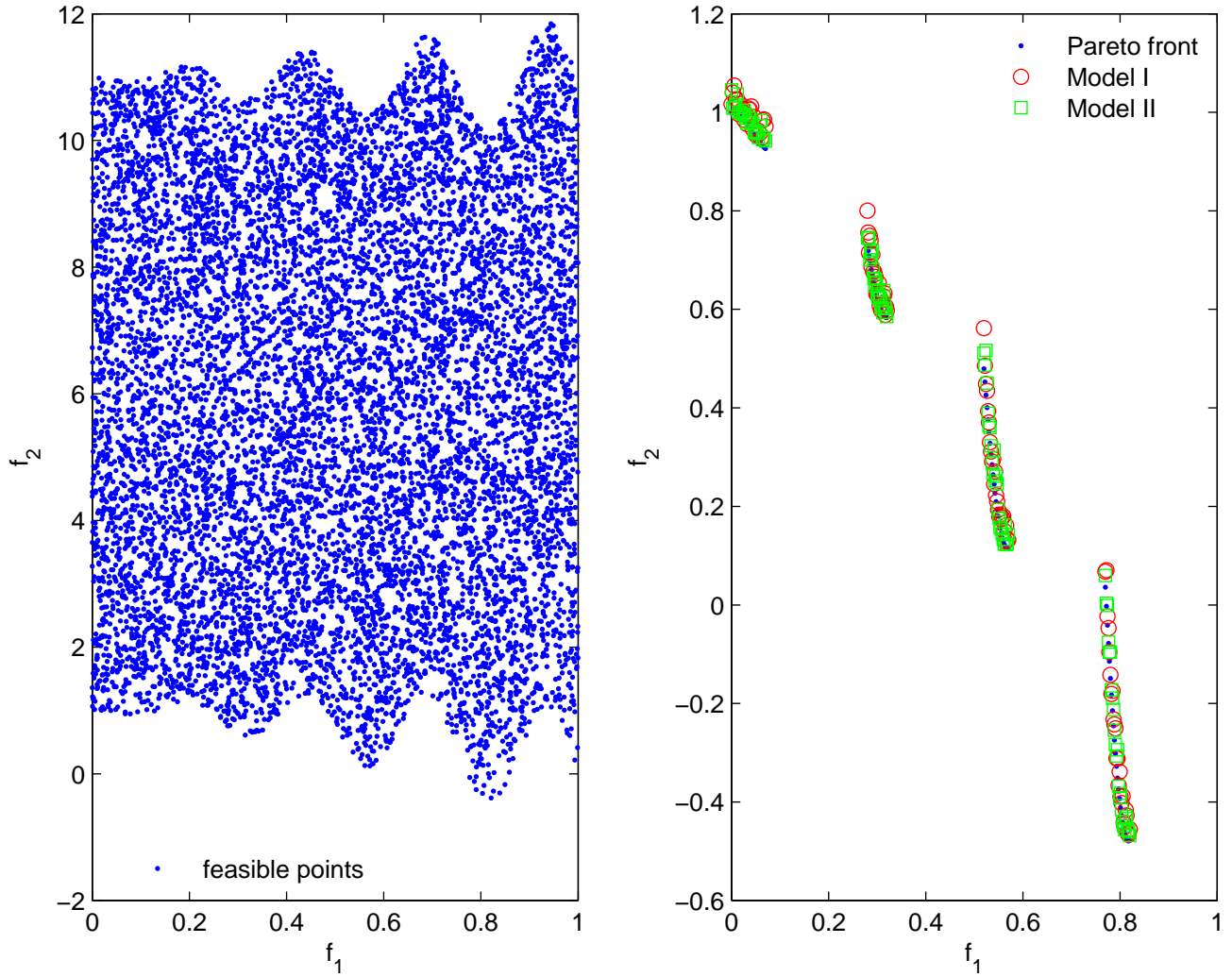
Figure 2: Example 2



Notes: On the left we show the feasible set and on the right the Pareto frontier. Model I assumes that  $\sigma$  is fixed at  $\sigma = 10^{-3}$ . In this case we use the posterior in (23). Model II assumes that  $\sigma$  is unknown and, therefore, it is integrated out of the posterior as in (24).

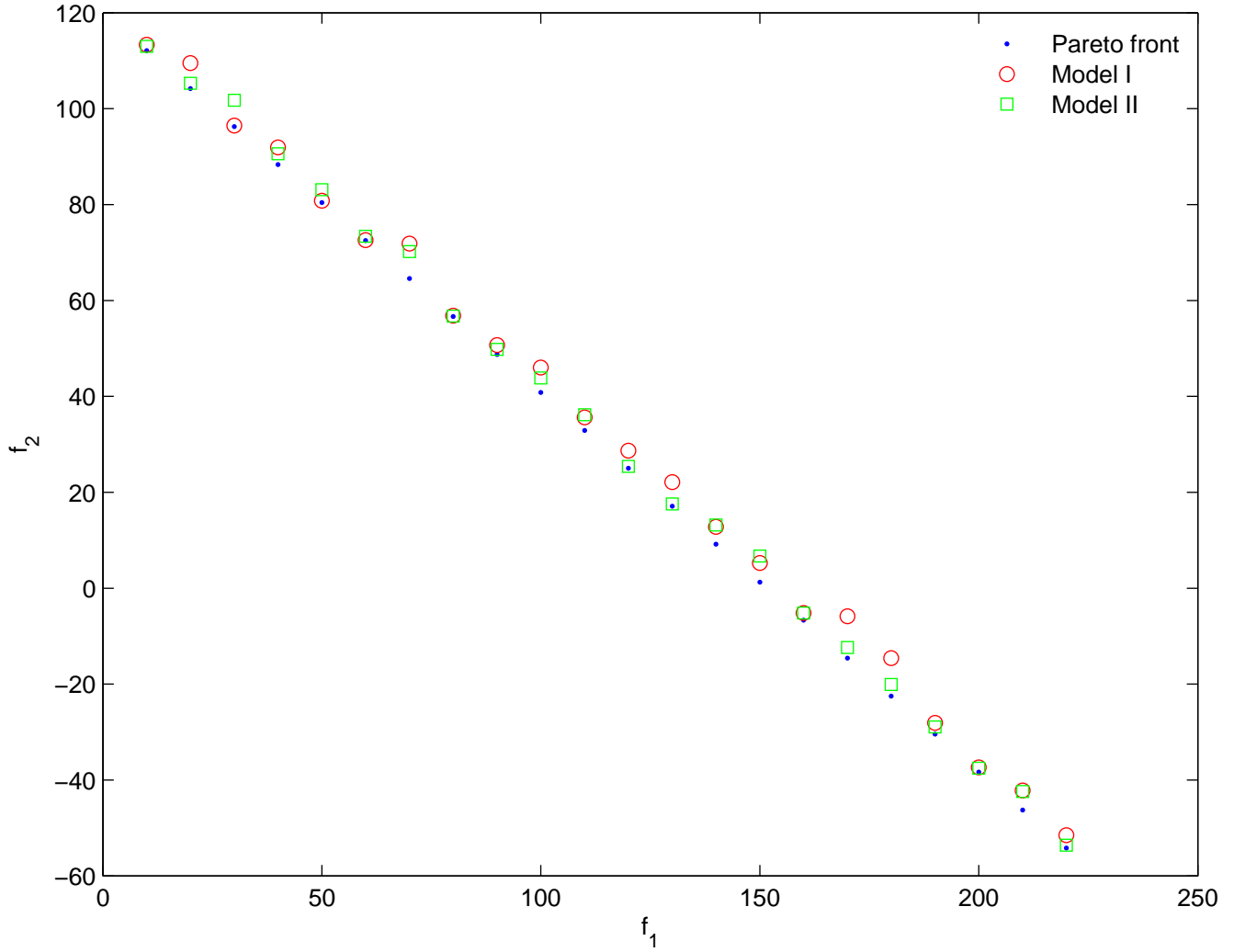


Figure 3: Example 3



Notes: On the left we show the feasible set and on the right the Pareto frontier. Model I assumes that  $\sigma$  is fixed at  $\sigma = 10^{-3}$ . In this case we use the posterior in (23). Model II assumes that  $\sigma$  is unknown and, therefore, it is integrated out of the posterior as in (24).

Figure 4: Example 4



Notes: On the left we show the feasible set and on the right the Pareto frontier. Model I assumes that  $\sigma$  is fixed at  $\sigma = 10^{-3}$ . In this case we use the

posterior in (23). Model II assumes that  $\sigma$  is unknown and, therefore, it is integrated out of the posterior as in (24). In this instance, we deliberately limit the number of MCMC simulations to mimic the procedure of Chiandussi et al. (2012), see their Figures 3.14 - 3.17. Specifically, we used only 2,000 iterations and discard the first 500 to mitigate possible start up effects.

Table 1: Comparison of methods

Problem	CPU time (in seconds)			$D$ -metric		
	NSGA-II	MOEA/D	Bayes MCMC	NSGA-II	MOEA/D	Bayes MCMC
ZDT1	1.03	0.60	1.02	0.0050	0.0055	0.0050
ZDT2	1.01	0.47	0.87	0.0049	0.0079	0.0051
ZDT3	1.02	0.55	0.90	0.0065	0.0143	0.0079
ZDT4	0.71	0.33	0.85	0.0182	0.0076	0.0081
ZDT6	0.72	0.27	0.66	0.0169	0.0042	0.0044
DTLZ1	10.31	1.25	15.32	0.0648	0.0317	0.0152
DTLZ2	8.44	1.13	13.78	0.0417	0.0389	0.0224

Notes: ZDT1–ZDT6 and DTLZ1–2 are different optimization problems defined in Zhang and Li (2007, pp. 720–721). The  $D$ -metric is defined in Zhang and Li (2007, p. 719) and “measure[s] both the diversity and convergence of in a sense. To have a low value of  $D$ , [solutions] must be very close to the PF and cannot miss any part of the whole PF.” (PF stands for Pareto front). CPU times for NSGA-II and MOEA/D are very close to those reported in Zhang and Li (2007, Tables V and VII, p. 724) but slight differences exist in certain instances. For the Bayesian MCMC approach,  $D$ -metric is computed using posterior means. Posterior standard deviations of the  $D$ -metric are, approximately, 10% of the reported metrics. “Set coverage” (Table VI of Zhang and Li, 2007) is very similar to the corresponding results reported in their paper for NSGA-II and MOEA/D. All results are obtained using a machine identical to the specification in Zhang and Li (2007), (Pentium(R) 3.2 GHZ, 1.00 GB). The author wishes to thank the IT service of Lancaster University for providing access to this machine.

### 3.4 Comparison with other approaches

We use for comparison the Non-dominated Sorting Genetic Algorithm (NSGA) proposed by Srinivas and Deb (1994). The versions we used are NSGA-II (Deb et al., 2002) and NSGA-III (Deb and Jain, 2014). We also examine Multiple Objective Evolutionary Algorithm, and in particular, MOEA/D proposed by Zhang and Li (2007). MOEA/D is interesting in that it decomposes a multiobjective optimization problem into a number of scalar optimization subproblems that can be optimized simultaneously. As they write: “MOEA/D [...] is able to generate a very uniform distribution of representative Pareto optimal solutions on the [Pareto front] on two continuous 3-objective test instances. We have also demonstrated that MOEA/D with a naive objective normalization technique can deal with disparately scaled objectives very well. These results suggest that more efficient and effective implementations of MOEA/D could be obtained if more effort is made” (Zhang and Li, 2007, p. 730).

Our results in terms of CPU time and the  $D$ -metric proposed by Zhang and Li (2007) are reported in Table 1. Although Bayes MCMC is more computationally intensive in certain difficult problems, its accuracy is remarkable and competes favorably with both NSGA-II and MOEA/D.

## 4 Minimax regret without scenarios

### 4.1 General

In the minimax regret approach we minimize the maximum regret when a number of scenarios is available for the parameters of the problem. Let us suppose that the objective function we wish to minimize is  $f(x, \delta_s)$ ,  $m = 1, \dots, M$  and  $s \in S = \{1, \dots, N_S\}$  denotes different scenarios for the parameter vector  $\delta$ . For each scenario the feasible set is  $X_s$  ( $s \in S$ ). Define  $X = \bigcap_{s \in S} X_s$ .

Then we would like to solve the following optimization problem:

$$\min_{x \in X} \max_{s \in S} \left\{ \frac{f(x, \delta_s) - f(x_s^*, \delta_s)}{f(x_s^*, \delta_s)} \right\}, \quad (45)$$

where  $x_s^*$  minimizes  $f(x, \delta_s)$  for  $x \in X_s$ . The problem can be written as follows (Kouvelis and Yu, 1997, p. 29):

$$\begin{aligned} & \min y \\ & f(x, \delta_s) \leq (1 + y)f(x_s^*, \delta_s) \quad \forall s \in S, \\ & x \in X. \end{aligned} \quad (46)$$

Xidonas et al. (2017) propose a multi-objective formulation where the purpose is to minimize simultaneously the objectives  $f_m(\delta_s)$ ,  $m = 1, \dots, M$ ,  $s \in S$ . The authors use the Pareto approach and solve the problem:

$$\begin{aligned} & \min y \\ & \sum_{m=1}^M w_m f_m(x, \delta_s) \leq (1 + y)f_s^* \quad \forall s \in S, \\ & x \in X, \end{aligned} \quad (47)$$

where  $f_s^* = \min_{x \in X_s} \sum_{m=1}^M w_m f_m(x, \delta_s)$ . Therefore, we have a single objective defined as  $f(x, \delta_s) = \sum_{m=1}^M w_m f_m(x, \delta_s)$ . As the weights are varied, the corresponding Pareto front can be traced. The assumption that a number of scenarios is available may be somewhat restrictive and an automatic procedure may be worthwhile to try in certain applications. Moreover, the Pareto approach may be relaxed to more general scalarizations using the weighted Tchebychev problem in (30). In this instance, the problem becomes:

$$\begin{aligned} & \min \gamma \\ & x \in X, \\ & \beta_m f_m(x, \delta_s) \leq (1 + \gamma)f_s^* \quad \forall m = 1, \dots, M, \quad \forall s \in S, \end{aligned} \quad (48)$$

where  $\beta_m = w_m$ ,  $m = 1, \dots, M$ . Regarding the different scenarios we assume that the parameters are allowed to vary in their domain, say  $\mathcal{D}$  with a specific distribution whose density is  $p(\delta)$ . Although one can obtain a potentially large number of random draws for  $\delta$  and solve the problem in (48) this approach may be inefficient as *the “data” (feasible set of solutions in (48)) are informative for likely values of  $\delta$* . In turn, we can interpret  $p(\delta)$  as a *prior over the possible values of the parameter*, and we can associate a posterior with (48) as in (34) if we have the statistical model in (32) and (34) when  $\delta$  is also a parameter. Indeed, only minimal modifications are required relative to (34). Our prior for this parameter is:

$$p(\delta) \propto \mathbb{I}(\delta \in \mathcal{D}), \quad (49)$$

to impose the notion that all parameter values are equally likely *a priori*. In turn, the restrictions in (48) can be written as:

$$\beta_m f_m(x, \delta) \leq (1 + \gamma) f_s^* \quad \forall m = 1, \dots, M, \quad (50)$$

for a *fixed but unknown*  $\delta \in \mathcal{D}$ . Draws from the marginal posterior density of the parameter are obtained using the conditional posterior  $p(\delta | \tilde{x}, \beta, \gamma)$  where  $\tilde{x}$  is the vector of optimal actions, treated here as a parameter, as defined in the previous sections. We use a flat prior for  $\tilde{x}, \beta, \gamma$  subject to the restrictions  $\beta_m \geq 0 \quad \forall m = 1, \dots, M, \quad \sum_{m=1}^M \beta_m = 1$ .

## 4.2 Empirical application

We follow Xidonas and Mavrotas (2014) and use the 50 stocks of the Eurostoxx-50. This is the leading blue-chip index for the Eurozone, and provides a high capitalization representation of leading sectors in the Eurozone. We use daily data for the period January 1, 2012 - May 26, 2018. With a time horizon of  $T$  periods and  $N$  assets the objectives are as follows:

$$f_1(x) = N^{-1} \sum_{i=1}^N x_i \bar{R}_i, \quad (51)$$

$$f_2(x) = T^{-1} \sum_{t=1}^T \left| \sum_{i=1}^N x_i (R_{it} - \bar{R}_i) \right|, \quad (52)$$

$$f_3(x) = \frac{T^{-1} \sum_{t=1}^T [\sum_{i=1}^N x_i (R_{it} - \bar{R}_i)]_+^3}{f_2(x)^3}, \quad (53)$$

$$f_4(x) = \frac{T^{-1} \sum_{t=1}^T [\sum_{i=1}^N x_i (R_{it} - \bar{R}_i)]_-^4}{f_2(x)^4}, \quad (54)$$

where  $R_{it}$  represents return of  $i$ th stock in  $t$ th time period,  $\bar{R}_i$  is average return,  $[x]_+ = \max(y, 0)$ ,  $[x]_- = -\min(y, 0)$ ,  $f_1(x)$  is portfolio average return,  $f_2(x)$  is portfolio's mean absolute deviation (MAD),  $f_3(x)$  is portfolio's positive skewness (PSKEW),  $f_4(x)$  is portfolio's negative skewness (NSKEW). Our objective is to maximize  $f_1(x)$  and  $f_3(x)$  and minimize  $f_2(x)$  and  $f_4(x)$  subject to the constraints:

$$x_i \geq 0 \quad \forall i = 1, \dots, N, \quad \sum_{i=1}^N x_i = 1. \quad (55)$$

We scale the objective functions using the commonly used transformation:

$$f_m(x) := \frac{f_m(x) - f_{min}}{f_{max} - f_{min}} \quad m = 1, \dots, 4. \quad (56)$$

As the data is noisy in this case, we use Model II, that is we integrate  $\sigma$  explicitly out of the posterior as in (24). The posterior distributions of weights are reported in the left panel of Figure 5. The weights favor more MAD and NSKEW, and less average return and PSKEW. In the right panel of Figure 5 we report the marginal posterior density of minimax regret under two assumptions: First, when the weights ( $w_m$ ) are drawn uniformly (straight line). Second, when they are drawn from their conditional posterior distributions. The first case corresponds to tracing out the Pareto front using random weights. The second corresponds to a *guided exploration* of the Pareto front with an aim to minimize maximum regret. Clearly, when posterior simulation is used for the weights, maximum regret is lower: Lower values of minimax regret correspond to more robust solutions from the problem formulation in Kouvelis and Yu (1997). Xidonas et al. (2017) considered two objectives (return and MAD) whereas the analysis in this paper considers four objectives. Therefore, minimax values cannot be directly compared. In addition, the time period of the analysis is different. However, as the posterior mean minimax regret is quite low in our approach, it should be useful in practical applications. Next, we turn attention to some questions:

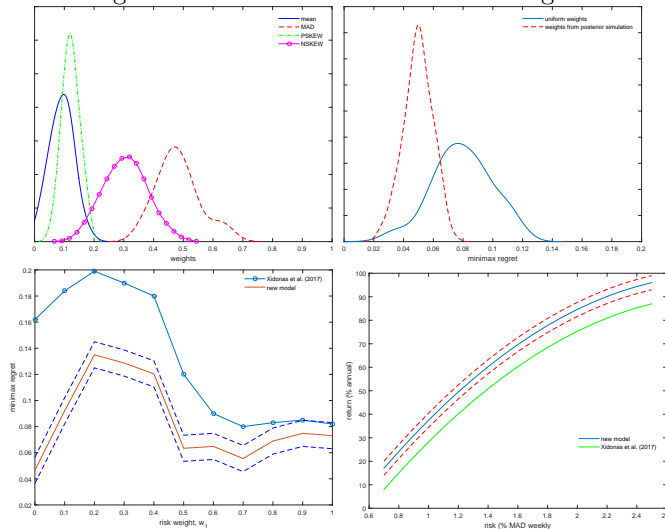
Does our approach provide more information compared to analytical approaches, such as the one of Xidonas et al. (2017)?

The current approach does provide more information in the sense that posterior uncertainty measures can be associated with optimal portfolios, minimax regret and points on the Pareto front. This (statistical) uncertainty arises from the fact that points on the Pareto front are estimated so their posterior standard errors provide a measure of robustness. **This is evidenced, for example, in Figure 5 (upper panel) where we present posterior distributions of weights of the objective functions. As these marginal posteriors are across all points in the Pareto front, they not be very informative. For this reason, in the bottom panel of Figure 5 we present a comparison of minimax regret along with 95% Bayes probability bands relative to Xidonas et al. (2017). (2017)<sup>7</sup> in the bottom left panel, and the Pareto fronts (along with 95% Bayes probability bands) of the two approaches in the bottom right panel. The Pareto fronts are presented in terms of “return” (annual percentage) versus “risk” which is measured as percentage weekly mean absolute deviation as in Xidonas et al. (2017). As the probability bands are quite narrow, the Pareto front is estimated quite accurately, in this instance. Moreover, the new techniques (which account for skewness and kurtosis of the optimal portfolios) deliver lower minimax regret and higher return for a given level of “risk”.**

Is our approach computationally more efficient? As we compare two different methods (one method relies on optimization and the other on Bayesian MCMC) computational efficiency depends on the demands of repeated optimization **as in Xidonas et al. (2017)** versus a single (possibly long) simulation. With many decision variables the advantage of MCMC is clear and relies on the same reasoning that justifies simulated annealing over local optimization algorithms: it may take longer but it provides a global optimum particularly when the Pareto front even when it is non-convex and disconnected. Despite this fact, timings on both a modern mainframe and desktop computer were very reasonable in all examples and the present application. Additionally, the direct optimization approach requires to design scenarios which may or not be feasible and / or desirable. **In terms of timing, On a modern Intel i9 personal computer the MCMC implementation took about 2**

<sup>7</sup>We compare with scenario 5 in Xidonas et al. (2017) which yields the highest return for a given level of risk, see their Figure 2, p. 303.

Figure 5: Posterior distributions of weights



minutes of clock time and negligible time on a modern mainframe computer (exact specifications were described in footnote 4).

An interesting approach to the problem is Figure 3 in Xidonas et al. (2017, p. 304) in which the authors plot minimax regret against the weight for risk for a bi-objective problem that minimizes MAD and maximizes average return. In this way they can determine weights that yield more robust portfolios. We can reproduce a similar approach if we proceed as follows:

i) From MCMC determine the weights that produce minimax regret, say  $w^*$  with corresponding minimax regret  $MMR^*$ .

ii) For each draw  $w^{(s)}$  ( $s = 1, \dots, S$ ) determine the associated regret,  $MMR^{(s)}$  ( $s = 1, \dots, S$ ).

iii) A measure of sensitivity of any given portfolio produced from MCMC draws can be measured using  $Q^{(s)} = \frac{MMR^{(s)} - MMR^*}{MMR^*}$  ( $s = 1, \dots, S$ ). Of course,  $Q^{(s)} \geq 0$ .

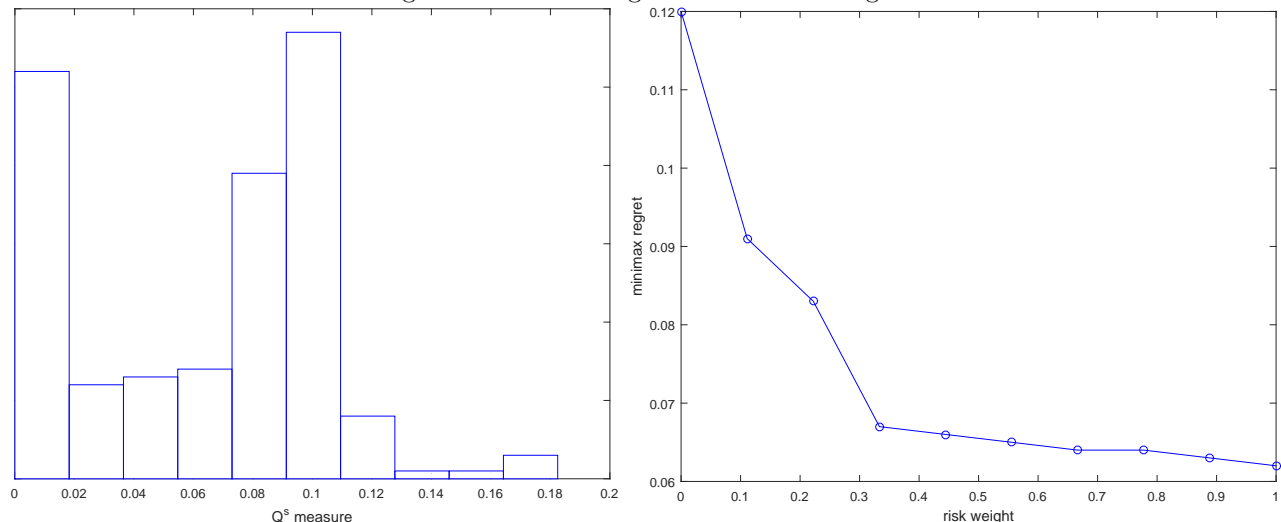
Our results are reported in the left panel of Figure 6. Evidently, most portfolios have negligible difference in terms of MMR relative to the minimax regret portfolio. This is because there is considerable accumulation of values near zero. However, many other portfolios produce regrets that are considerably higher.

The problem, in turn, is to select portfolios whose sensitivity measures ( $Q^{(s)}$ ) are “small”. Setting a target  $\varepsilon = 0.005$  say, we select portfolios with  $Q^{(s)} < \varepsilon$  during MCMC, so *robust portfolios can be selected in an automatic way*. As this requires the globally minimum regret ( $Q^*$ ) this selection can be implemented based on post-processing of MCMC draws.

Based on the universe of portfolios generated by MCMC we would also like to examine the robustness (minimax regret) in terms of certain weights, say the weight for risk. In the right panel of Figure 6 we report the relationship between MMR and the risk weight as in Figure 3 in Xidonas et al. (2017, p. 304). Weights greater than about 0.3 produce robust solutions, that is solutions whose sensitivity measure is close to zero as in the left panel of Figure 6. Again, such solutions can be produced in an automated way.

Finally, we construct a minimax regret portfolio for the period May 26, 2017 - May 26, 2018 without using the data

Figure 6: Minimax regret for different weights



for this period, so that we can estimate the required moments. The behavior of the optimal portfolio minimizing globally maximum regret is presented in Figure 7 along with the minimum variance -maximum expected return portfolio. The minimum variance-maximum expected return portfolio solves the problem:

$$\min \lambda x' \hat{\Sigma} X - (1 - \lambda) \bar{r}' x, \quad x \geq 0, \quad x' 1_N = 1, \quad (57)$$

where  $\hat{\Sigma}$  is the covariance matrix of returns. We set  $\lambda = \frac{2}{3}$ . Evidently, the mean-variance portfolio is more volatile while the minimax regret portfolio shows considerable resistance to large negative returns.

As we use Markov Chain Monte Carlo (MCMC) methods to produce (correlated) draws from the posterior it is important to know whether autocorrelations of draws for different parameters are so large (say 0.99) that prevent thorough exploration of the posterior. A MCMC whose autocorrelations are close to one, it would take forever to explore the posterior. The RNE stands for relative numerical efficiency. It was proposed by Geweke (1992) and it is equal to one if the draws are iid (so autocorrelations are zero) and approaches zero as the autocorrelation increases. Values between 0.30 and 0.70 are considered acceptable in practical applications. Typical autocorrelation functions of twenty solutions ( $x$ ) are reported in the left panel of Figure 8. Although they are evidently non-zero they are not destructively large either thus permitting efficient exploration of the posterior. Relative numerical efficiency (RNE) across all solutions is reported in the right panel of Figure 8. For iid draws from the posterior (which is, of course, never possible) RNE should be one (Geweke, 1999).

As we stated after Algorithm 1, the Bayesian approach delivers measures of uncertainty associated with the parameters (points on the Pareto front). We have omitted reporting standard errors of such parameters in the artificial examples as standard errors were orders of magnitude lower compared to the posterior means (points on the Pareto front). In the empirical application, however, it is important to show the the probability intervals for the minimax regret portfolio. Since the artificial examples do not involve noisy data it is easy to understand why standard errors are small and why confidence



Figure 7: Behavior of the optimal portfolio

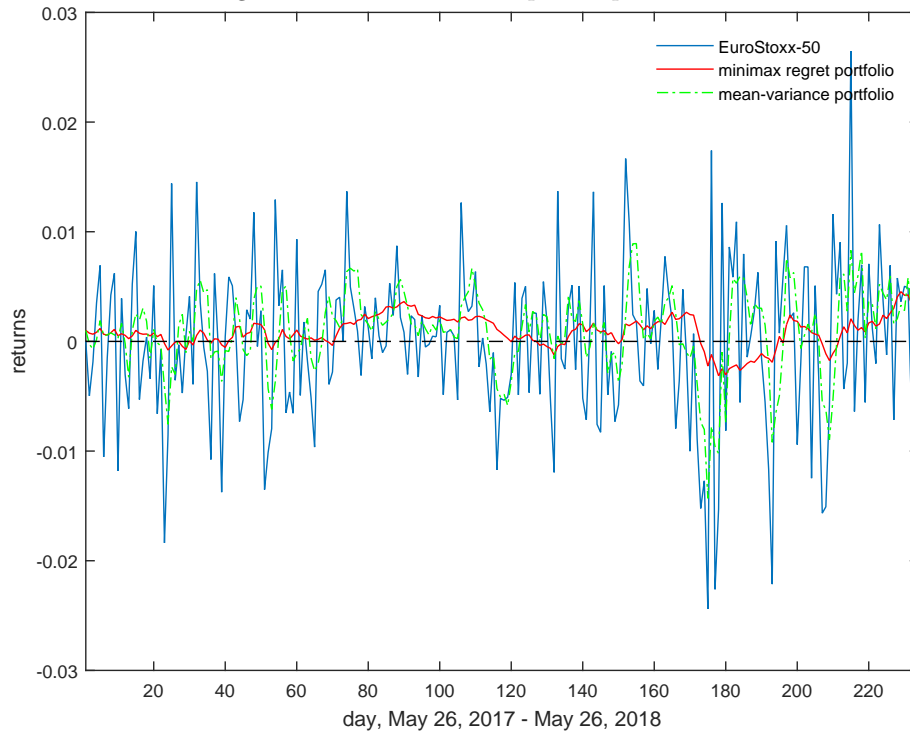


Figure 8: Autocorrelation functions

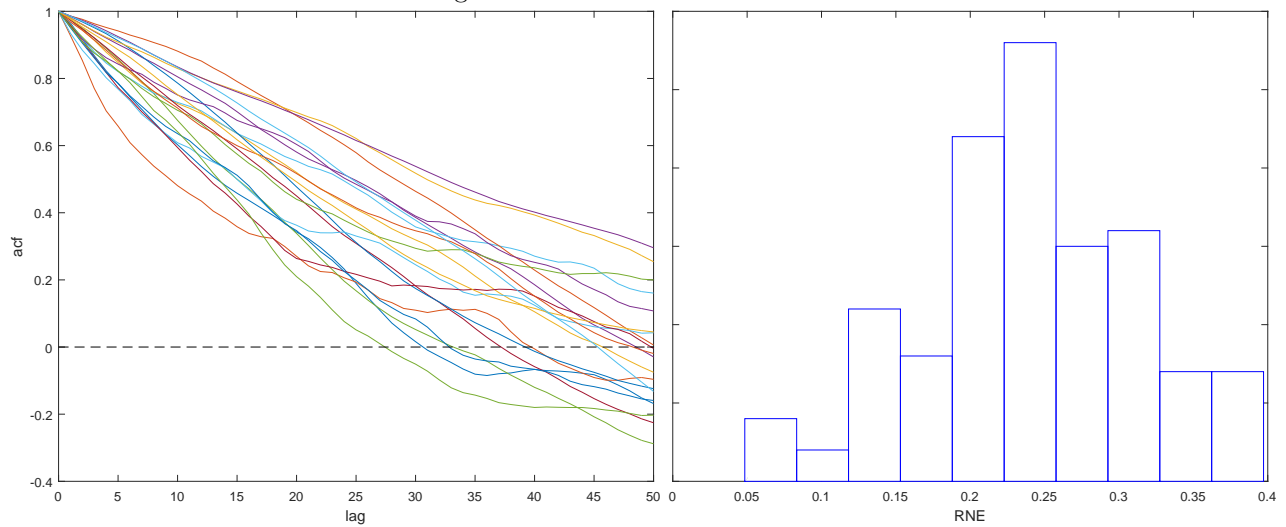
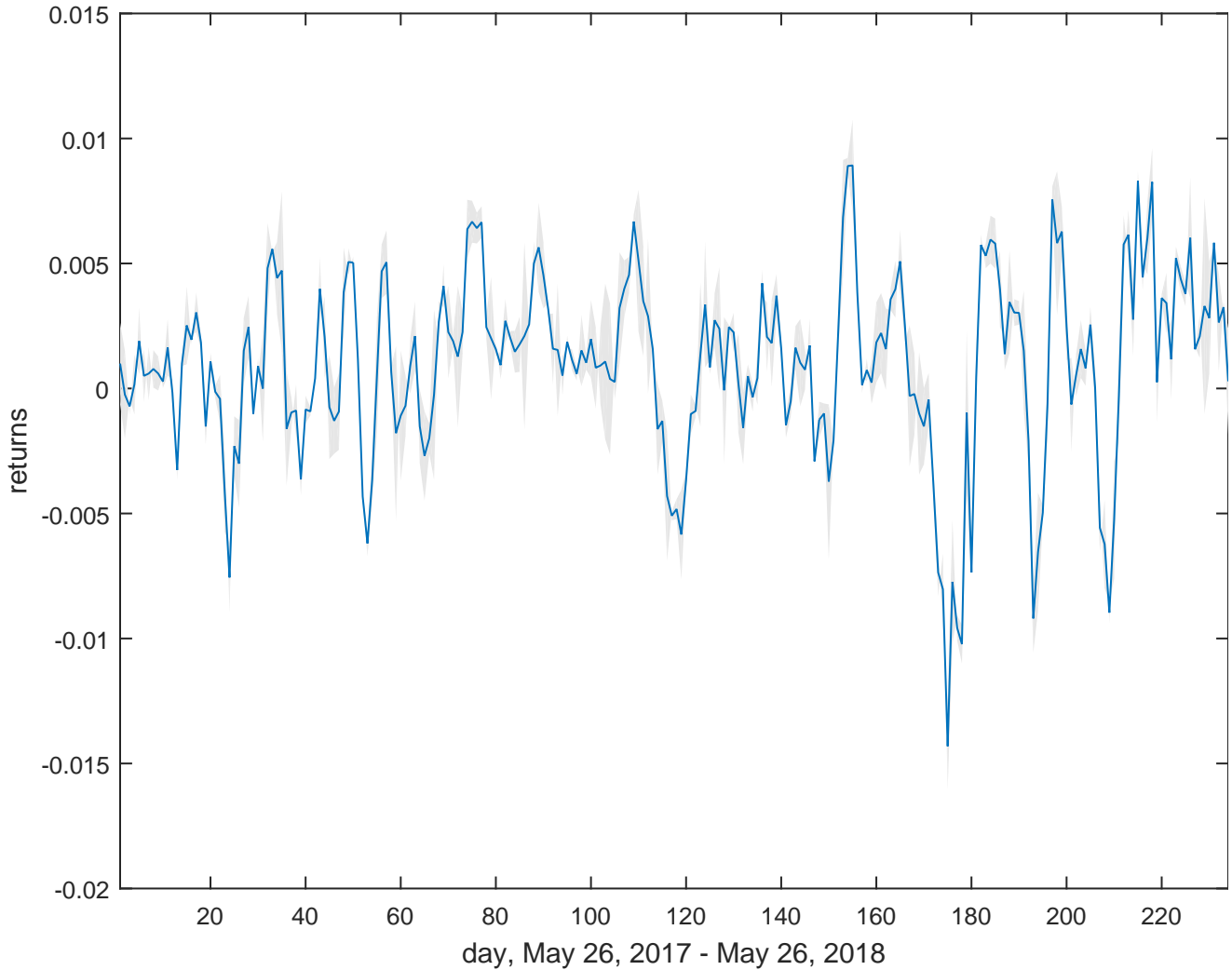


Figure 9: Minimax regret portfolio and Bayes probability interval



bands around the Pareto front are very tight (even when the front is disconnected). In **Figure 9**, we present the minimax regret portfolio as in Figure 7 along with a 95% Bayes probability interval. **Therefore, the Figure presents the returns that could have been attained if an investor had adopted this minimax regret portfolio, and it is a focused version of Figure 7.**

Of course, in this case, uncertainty is much larger although for most periods, positive and negative returns have a Bayes probability interval that stays away from zero.

## Concluding remarks

In this paper we have shown how multi-objective problems can be analyzed using Bayesian methods when appropriate statistical models are defined based on the structure of the problem (objective functions and constraints). The methods are used to analyze problems where the Pareto front is non-convex and / or disconnected. Additionally, we take up the case of

minimax regret and we show that the optimization problem can be analyzed using similar techniques: In this case uncertainty can be handled automatically without the need to specify explicitly alternative scenarios. The minimax regret approach is applied to a multi-objective portfolio problem which maximizes expected returns and positive skewness and minimizes mean absolute deviation and negative skewness of portfolio returns.

The basic idea relies on using Generalized DEA (GDEA) and Goal Attainment (GA) optimization problems which have been shown before to have good performance in multi-objective optimization. We show that such optimization problems correspond to certain statistical models whose parameters are points on the Pareto front. In particular, we propose Markov Chain Monte Carlo (MCMC) methods for Bayesian inference in the statistical models. Such methods can be implemented in a computationally efficient manner even and provide access to the implied posterior distribution of the model.

The methods can be applied whether or not the Pareto front is non-convex, disconnected, etc., as they explore the whole posterior in a thorough way. In the examples we have examined, it turns out that autocorrelation in MCMC draws is not substantial, so the posterior can be explored quickly and efficiently without sacrificing too many computing resources. This means that the Pareto front can be reconstructed efficiently using Bayesian MCMC techniques. It remains, however, to be seen whether Gibbs sampling performs equally well in other applications. Although it has been applied successfully in examples where the Pareto set is disconnected or even consists of different single points, its performance depends on GDEA or GA. If these methods fail in some respect, then MCMC will fail as well. This is left for future research.

Another avenue for future research is the exploitation of different MCMC algorithms such those in Hoogerheide et al. (2007) or Bauwens et al. (2004). These MCMC algorithms can deal with multi-modal posteriors or disconnected parameter spaces (disconnected Pareto sets in the case of multi-objective optimization). However, application of such algorithms to GDEA or GA depends on how well these methods perform in the first place. If they fail, related MCMC algorithms will fail as well. The application of the Bayesian paradigm in multi-objective optimization yields not only an approximation to the Pareto front but also delivers measures of statistical uncertainty associated with it. In future applications, this may become important, as the user may want a confidence interval associated with a particular Pareto point or a highest-posterior-interval set associated with the entire Pareto front. Based on MCMC draws, construction of such sets is trivial.

Another interesting area of research is to examine optimization (using Bayesian versions of GDEA or GA) when an explicit “utility function” can be specified in terms of function values. Most likely, this will result in an indirect prior over the parameters of the associated statistical models, so application of Bayesian analysis should be relatively straightforward. Of course, the use of complicated priors may require special MCMC techniques. It is likely, however, that simple modifications of the Gibbs sampler or off-the-shelf algorithms such as those in Hoogerheide et al. (2007) or Bauwens et al. (2004) will suffice to deliver good performance.

The techniques proposed here suggest themselves as natural candidates for computing DEA estimates (or GDEA estimates) with large data sets. The computational efficiency of Bayesian MCMC in this context relies on the replacement of repeated optimizations (linear programs in the case of DEA and nonlinear programs in the case of GDEA) by a single simulation which i) provides estimates for all unknown quantities involved in the problem, and ii) at the same time it

provides measures of statistical uncertainty associated with these estimates. With huge data sets, the need for parallelization is evident but MCMC is inherently serial. However, recent advances have made it possible to parallelize the computations, see for example Martino et al. (2016) and the references therein. Such parallelization is possible by breaking up a large data set into subsets, running MCMC on different cores, and then combining in an optimal way, see Li et al. (2017).

## References

- M. Arakawa, H. Nakayama, I. Hagiwara, & H. Yamakawa. (1998). Multiobjective optimization using adaptive range genetic algorithms with data envelopment analysis. Seventh Symposium on Multidisciplinary Analysis and Optimization (TP98-4970), Vol. 3. American Institute of Aeronautics and Astronautics, Reston, VA, 2074–2082.
- L. Bauwens, C. S. Bos, H. K. van Dijk, & R. D. van Oest (2004). Adaptive radial-based direction sampling: some flexible and robust Monte Carlo integration methods. *Journal of Econometrics* 123, 201 – 225.
- R. Benayoun R., J. de Montgolfier, J.Tergny , & O. Laritchev (1971). Linear programming with multiple objective functions: Step Method (STEM). *Mathematical Programming* 1, 366–375.
- L. M. Boychuk, & V. O. Ovchinnikov, Principal methods of solution of multicriterial optimization problems, *Sov. Automat. Control* 6 (1973) 1–4.
- G. Chiandussi, M. Codegone, S. Ferrero, & F.E. Varesio (2012). Comparison of multi-objective optimization methodologies for engineering applications. *Computers and Mathematics with Applications* 63, 912–942.
- C.A. Coello Coello (2001). A short tutorial on evolutionary multiobjective optimization, in: E. Zitzler, K. Deb, L. Thiele, C.A. Coello Coello, & D. Corne (Eds.), *First International Conference on Evolutionary Multi-Criterion Optimization*, Springer-Verlag.
- C. A. Coello Coello, C. Lamont, Gary B., & D. A. van Veldhuizen (2007). *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer.
- K. Deb (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley.
- K. Deb, & H. Jain (2014). An evolutionary many objective optimization algorithm using reference-point based non-dominated sorting approach, part I: solving problems with box constraints. *Evolutionary Computation, IEEE Transactions on*, 18 (4), 577 – 601.
- K. Deb, K., A. Pratap, S. Agarwal, & T. Meyarivan, T. (2002). A fast elitist non-dominated sorting genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6 (2), 182 – 197.
- M. Doumpos, & C. Zopounidis (2011). Preference disaggregation and statistical learning for multicriteria decision support: A review. *European Journal of Operational Research* 209 (3), 203–214.
- M. Doumpos, K. Kosmidou, G. Baourakis, & C. Zopounidis (2002). Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis. *European Journal of Operational Research* 138 (2), 392–412.
- M. Doumpos, Y. Marinakis, M. Marinaki, & C. Zopounidis (2009). An evolutionary approach to construction of

outranking models for multicriteria classification: The case of the ELECTRE TRI method. *European Journal of Operational Research* 199 (2), 496–505.

M. Doumpos, C. Zopounidis, & E. Galariotis (2014). Inferring robust decision models in multicriteria classification problems: An experimental analysis. *European Journal of Operational Research* 236 (2), 601–611.

L. Duckstein, Multiobjective optimization in structural design: the model, in: A. Atrek (Ed.), *New Directions in Optimum Structural Design*, 1984.

J. Dutta & C. Y. Kaya (2011) A new scalarization and numerical method for constructing the weak Pareto front of multi-objective optimization problems, *Optimization*, 60 (8-9), 1091–1104.

M. Ehrgott (2008). *Multiobjective Optimization*. *AI Magazine* 29(4), 47–57.

G. Eichfelder (2009a). An Adaptive scalarization method in multi-objective optimization, *SIAM J. Opt.* 19, 1694–1718.

G. Eichfelder (2009b). Scalarizations for adaptively solving multi-objective optimization problems, *Comput. Optim. Appl.* 44, 249–273.

K. Florios, G. Mavrotas, & D. Diakoulaki (2010). Solving multiobjective, multiconstraint knapsack problems using mathematical programming and evolutionary algorithms. *European Journal of Operational Research* Volume 203 (1), 14–21.

J. Geweke (1999). Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication (with discussion and rejoinder), *Econometric Reviews*, 18, 1–126.

T. Holzmann & J. C. Smith (2018). Solving discrete multi-objective optimization problems using modified augmented weighted Tchebychev scalarizations. *European Journal of Operational Research*, forthcoming, <https://doi.org/10.1016/j.ejor.2018.05.000>

L. F. Hoogerheide, J. F. Kaashoek, H. K. van Dijk (2007). On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks. *Journal of Econometrics*, 139, 154–180.

L. F. Hoogerheide and H. K. van Dijk. (2008). Possibly ill-behaved posteriors in econometric models: On the connection between model structures, non-elliptical credible sets and neural network simulation techniques. Technical Report 2008-036/4, Tinbergen Institute, Erasmus University Rotterdam, 2008. URL <http://www.tinbergen.nl/discussionpapers/08036.pdf>.

I. Hsiao, Y. Sakane, Noritaka Tsukamoto, & Yusuke Nojima (2009). Evolutionary many-objective optimization by NSGA-II and MOEA/D with large populations. 2009 IEEE International Conference on Systems, Man and Cybernetics. DOI: 10.1109/ICSMC.2009.5346628

H. Jain, & K. Deb (2014) An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point Based Nondominated Sorting Approach, Part II: Handling Constraints and Extending to an Adaptive Approach. *IEEE Transactions on Evolutionary Computation* 602–622.

D. E. Kaufman, & R. L. Smith (1994). Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research* 46, 84–95.

- I. Kim, & O. de Weck (2005). Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Structural and Multidisciplinary Optimization*, 29 (2), 149–158.
- I. Kim, & O. de Weck (2006). Adaptive weighted sum method for multiobjective optimization: a new method for Pareto front generation. *Structural and Multidisciplinary Optimization* 31 (2), 105–116.
- J. Knowles and D. Corne, “The Pareto archived evolution strategy: A new baseline algorithm for multiobjective optimization,” in *Proceedings of the 1999 Congress on Evolutionary Computation*. Piscataway, NJ: IEEE Press, 1999, pp. 98–105.
- K. Koski (1984). Multicriterion optimization in structural design, in: A. Atrek (Ed.), *New Directions in Optimum Structural Design*.
- P. Kouvelis, & Yu, G. (1997). *Robust discrete optimization and its applications*. Springer-Science Business Media.
- M. Laumanns, L. Thiele, and E. Zitzler. (2006). An efficient, adaptive parameter variation scheme for metaheuristics based on the epsilon-constraint method. *European Journal of Operational Research*, 169 (3), 932–942.
- D. Li, J. B. Yang, & M. P. Biswal (1999). Quantitative parametric connections between methods for generating noninferior solutions in multiobjective optimization. *European Journal of Operational Research* 117, 84–99.
- S. Li, G. K. F. Tso, & L. Long (2017). Powered embarrassing parallel MCMC sampling in Bayesian inference, a weighted average intuition. *Computational Statistics and Data Analysis* 115, 11–20.
- M. R. Lightner, & S. W. Director. (1981). Multiple criterion for the design of electronic systems. *Trans. Circuits Systems CAS* 28 169–179.
- L. Martino, V. Elvira, D. Luengo, J. Corander, & F. Louzada (2016). Orthogonal parallel MCMC methods for sampling and optimization. *Digital Signal Processing* 58, 64–84.
- G. Mavrotas (2009). Effective implementation of the  $\varepsilon$ -constraint method in multi-objective mathematical programming problems. *Applied Mathematics and Computation*, 213(2):455–465.
- G. Mavrotas, O. Pechak, E. Siskos, H. Doukas, & J. Psarras (2015). Robustness analysis in Multi-Objective Mathematical Programming using Monte Carlo simulation. *European Journal of Operational Research* 240, 193–201.
- K. M. Miettinen (1999). *Nonlinear Multiobjective Optimization*, Kluwer, Boston.
- K. Miettinen, & M. M. Mäkelä (2002). On scalarizing functions in multiobjective optimization. *OR Spectrum*, 24(2), 193–213.
- D. Mueller-Gritschneider, H. Graeb & U. Schlichtmann (2009). A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems, *SIAM Journal of Optimization* 20, 915–934.
- K. Pendaraki, C. Zopounidis, & M. Doumpos (2005). On the construction of mutual fund portfolios: A multicriteria methodology and an application to the Greek market of equity mutual funds. *European Journal of Operational Research* 163 (2), 462–481.
- G. O. Roberts & A. F. M. Smith (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* 49 (2), 207–216.

- G. Rudolph, “Evolutionary search under partially ordered sets,” Dept. Comput. Sci./LS11, Univ. Dortmund, Dortmund, Germany, Tech. Rep. CI-67/99, 1999.
- N. Srinivas, & K. Deb (1994). Multiobjective function optimization using non dominated sorting genetic algorithms. *Evolutionary Computation*, 2 (3), 221 – 248.
- R. Steuer (1986). Multiple criteria optimization: theory, computation and application. New York, John Wiley & Sons, Inc.
- L. Tierney (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics* 22 (4), 1701–1728.
- M. G. Tsionas (2018). A Bayesian approach to find Pareto optima in multiobjective programming problems using Sequential Monte Carlo algorithms. *Omega* 77, 73–79.
- P. Xidonas, G. Mavrotas, C. Zopounidis & J. Psarras (2011). IPSSIS: An integrated multicriteria decision support system for equity portfolio construction and selection. *European Journal of Operational Research* 210 (2) 398–409.
- P. Xidonas, G. Mavrotas, C. Hassapis, & C. Zopounidis (2017). Robust multiobjective portfolio optimization: A minimax regret approach. *European Journal of Operational Research* 262 (1), 299–305.
- L.A. Zadeh (1963). Optimality and nonscalar-valued performance criteria, *IEEE Transactions on Automatic Control* 8, 59–60.
- A. Zellner (1971). An introduction to Bayesian inference on econometrics. Wiley, New York.
- E. Zitzler (1999). *Evolutionary algorithms for multiobjective optimization: Methods and applications*, Doctoral dissertation ETH 13398, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.
- Y. B. Yun, H. Nakayama, T. Tanino, & M. Arakawa (2001). Generation of efficient frontiers in multi-objective optimization problems by generalized data envelopment analysis. *European Journal of Operational Research* 129, 586–595.
- Q. Zhang & H. Li (2007). MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11 (6), 712–731.
- E. Zitzler & L. Thiele (1999). Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation* 3, 257–271.
- C. Zopounidis (1999). Multicriteria decision aid in financial management. *European Journal of Operational Research*, 199 (2), 404–415.
- C. Zopounidis, E. Galariotis, M. Doumpos, S. Sarri, & K. Andriopoulos (2015). Multiple criteria decision aiding for finance: An updated bibliographic survey. *European Journal of Operational Research* 247 (2), 339–348.
- Xidonas, P., & Mavrotas, G. (2014). Multiobjective portfolio optimization with non-convex policy constraints: Evidence from the Eurostoxx 50. *European Journal of Finance* 20, 957–977.

## Appendix A. Background material on MCMC

Suppose we have the following statistical model:

$$y_i = f(x_i, \beta) + v_i - u_i \quad \forall i = 1, \dots, N \quad (1)$$

where  $v_i \sim iid\mathcal{N}(0, h^2)$  independently of  $x_i$  and  $u_i \sim iid\mathcal{N}_+(0, \omega^2)$  independently of  $x_i$  and  $v_i$ . Here,  $f(x_i, \beta)$  is a possibly nonlinear function depending on exogenous variables  $x_i \in \mathfrak{R}^{d_x}$  and  $\beta \in \mathfrak{R}^k$  is a parameter. If we treat  $u_i$  as part of the parameter vector along with  $\beta \in \mathfrak{R}^k$ , the augmented likelihood function:

$$L(\beta, h, \omega, \mathbf{u}; y, X) \propto h^{-N} \omega^{-N} \exp \left\{ -\frac{1}{2h^2} \sum_{i=1}^N (y_i - f(x_i, \beta) + u_i)^2 - \frac{1}{2\omega^2} \sum_{i=1}^N u_i^2 \right\}, \quad (2)$$

where  $y, X$  denotes the data (that we denote compactly as  $\mathbf{Y}$ ),  $\mathbf{u} = [u_1, \dots, u_N]'$ . Suppose our joint prior is:

$$p(\beta, h, \omega) \propto h^{-1} \omega^{-1}, \quad (3)$$

which is uninformative to reflect that we do not have specific prior information. In turn, the posterior by Bayes' theorem is:

$$p(\beta, h, \omega, \mathbf{u} | \mathbf{Y}) \propto h^{-N-1} \omega^{-N-1} \exp \left\{ -\frac{1}{2h^2} \sum_{i=1}^N (y_i - f(x_i, \beta) + u_i)^2 - \frac{1}{2\omega^2} \sum_{i=1}^N u_i^2 \right\}. \quad (4)$$

The Gibbs sampler consists of the following steps that we perform for each  $s = 1, \dots, S$  where  $S$  is the number of MCMC draws. Start with any values  $h^{(0)}, \omega^{(0)}, \mathbf{u}^{(0)}$ . Then, at iteration  $s$ :

1. Draw  $\beta^{(s)} \sim \beta | h^{(s-1)}, \omega^{(s-1)}, \mathbf{u}^{(s-1)}, \mathbf{Y}$ .
2. Draw  $h^{(s)} | \beta^{(s)}, \omega^{(s-1)}, \mathbf{u}^{(s-1)}, \mathbf{Y}$ .
3. Draw  $\omega^{(s)} | \beta^{(s)}, h^{(s)}, \mathbf{u}^{(s-1)}, \mathbf{Y}$ .
4. Draw  $\mathbf{u}^{(s)} | \beta^{(s)}, h^{(s)}, \omega^{(s)}, \mathbf{Y}$ .

Of course the blocking is arbitrary and convergence of the Gibbs sampler does not depend on the blocking. Also it does not depend on initial conditions. From (4) we have the following conditional posteriors:

$$p(\beta | h, \omega, \mathbf{u}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2h^2} \sum_{i=1}^N (\tilde{y}_i - f(x_i, \beta))^2 \right\}, \quad (5)$$



where  $\tilde{y}_i = y_i + u_i \forall i = 1, \dots, N$ .

$$p(h|\beta, \omega, \mathbf{u}, \mathbf{Y}) \propto p(\beta, h, \omega, \mathbf{u}|\mathbf{Y}) \propto h^{-N-1} \exp\left\{-\frac{1}{2h^2}Q_h\right\}, \quad (6)$$

where  $Q_h = \sum_{i=1}^N (\tilde{y}_i - f(x_i, \beta))^2$ .

$$p(\omega|\beta, h, \mathbf{u}, \mathbf{Y}) \propto \omega^{-N-1} \exp\left\{-\frac{1}{2\omega^2}Q_\omega\right\}, \quad (7)$$

where  $Q_\omega = \sum_{i=1}^N u_i^2$ . Finally, we have:

$$p(u_i|\beta, h, \omega, \mathbf{Y}) \propto \mathcal{N}_+(m_i, S^2) \quad i = 1, \dots, N, \quad (8)$$

where  $m_i = \frac{\omega^2[f(x_i, \beta) - y_i]}{h^2 + \omega^2}$ , and  $S^2 = \frac{h^2\omega^2}{h^2 + \omega^2}$ . These expressions can be derived if we complete the square in (8) in terms of  $u_i$ .

To draw  $\beta$  from (5) it is quite easy when  $f$  is linear, viz.  $f(x_i, \beta) = x_i'\beta$  but otherwise the distribution does not belong to a known family and we have to use a Metropolis - Hastings step to realize a random draw. This is quite easy but we omit the details here, see for example Tierney (1994). To draw from (8) is straightforward as there are efficient algorithms to provide draws from a truncated normal distribution. It remains to draw random numbers from (6) and (7) which are in the same family so we focus on (6) and denote  $Q = Q_h$ . Using the transformation  $H = \frac{1}{h^2}$ , it is easy to see that

$$p(H|\cdot) \propto H^{N/2-1} \exp\left\{-\frac{Q}{2}H\right\}, \quad (9)$$

which corresponds to a *gamma* distribution with shape parameter  $N/2$  and scale parameter  $Q/2$ . From this result, and using the relationship between the *gamma* and *chi*-squared distributions, we obtain:  $\frac{Q}{H} \sim \chi_N^2$ . In turn we can express (6) as follows:

$$\frac{\sum_{i=1}^N (\tilde{y}_i - f(x_i, \beta))^2}{h^2} | \beta, \omega, \mathbf{u}, \mathbf{Y} \sim \chi_N^2. \quad (10)$$