Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

ROBUST SIMULATOIN OF ENVIRONMENTAL POLICIES USING THE DICE MODEL

Zhaolin Hu

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, CHINA

Jing Cao

Department of Economics
Tsinghua University
Beijing 100084, CHINA

L. Jeff Hong

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, CHINA

ABSTRACT

Integrated assessment models that combine geophysics and economics features are often used to evaluate environmental economic policies. In these models, there are often profound uncertainties and Monte Carlo simulations are often used to evaluate the policies. Generally, the simulation approach requires that the distribution of the uncertain parameters are clearly specified. In this paper, we adopt the widely used multivariate normal distribution to model the uncertain parameters. However, we assume that the mean vector and covariance matrix of the distribution are within some ambiguity sets. We propose a change-of-measure technique to derive the simulation results for any mean vector and covariance matrix in the sets without actually simulating them. We then show how to find the worst case performance for all mean vectors and covariance matrices in the ambiguity sets by solving a sequence of convex problems. This performance provides a robust evaluation of the policies. We test our algorithm on a famous environmental economic model, known as the DICE model, and obtain some insightful and interesting results.

1 INTRODUCTION

Since the beginning of industrial revolution, human activities particularly burning of fossil fuels and clearing forested land, has made a much "thicker" blanket of greenhouse gas around the earth, resulting in an unprecedented increase in the current level of greenhouse gases, at around 430ppm CO2 compared with only 280ppm before the Industrial Revolution. While debate on whether or not the globe is indeed warming still ensues, overwhelming scientific evidence has already suggested that climate change presents very serious global risks to human welfare. Facing such dramatic climate risks, the recent 194-nation UN-led Copenhagen Climate Summit pledged to limit global warming to two degrees Celsius along with billions of dollars in climate financing in the famous "Copenhagen Accord".

Climate change is so complicated, for many years scientists and economists typically work together and resort to numerical computer simulations, so-called "Integrated Assessment Model" (IAM) for evaluating the economic impacts of global warming. One representative IAM model is the Dynamic Integrated model of Climate and the Economy (DICE), developed by William Nordhaus (see, for instance, Nordhaus 2008). The model conducts a cost-benefit analysis and outputs a net present value of abatement costs plus climate changes. Like climate change itself, climate models seem also concern profound uncertainties, partly due to its intrinsic randomness and partly due to our imperfect understanding of the evolvement of the complex climate systems. In IAM models, these uncertainties appear to rely on a wide range of key uncertain parameters on various stages of modeling global warming. A typical approach to handling these uncertainties is to specify distributions for these parameters, and then conduct an expected cost-benefit analysis. For instance, Nordhaus (2008) pointed out there are eight critical uncertain parameters in the DICE model. He modeled all eight parameters as normal distributions and estimated their means and variances, and then evaluated the expected performances of different policies by generating the parameters from their distributions and running simulation experiments with these parameters.

However, for most of the uncertain parameters used in environmental economic models, there exists estimation error for the means and covariances. Ignoring this estimation error may lead to inaccurate estimation of performances by the simulation model, and therefore may lead to false conclusions. To resolve this difficulty, the DICE model currently only considers one parameter at a time while fixing other parameters at their means or they (implicitly) assume that the parameters are independent (Nordhaus 2008). However, this assumption is rarely true and the parameters often work together to affect the outcomes of the models. In this paper, we assume that the uncertain parameters of the

environmental simulation model follow a multivariate normal (MVN) distribution with an ambiguous mean vector and covariance matrix. The MVN distribution is used in the DICE model because of its simplicity and generality (Nordhaus 2008). To model the ambiguity in the mean vector and the covariance matrix, we assume that they are constrained in two sets, respectively, which we call "ambiguity sets". We consider different forms of ambiguity sets so that they are general enough to model practical situations, e.g., the mean vector and covariance matrix are estimated with or without data, while still maintaining the desired mathematical tractability. When there is ambiguity in specifying the MVN distribution, the results of the simulation are also ambiguous, making policy comparisons difficult. To solve this problem, we take a robust approach that is used widely (see, e.g., the robust design of Taguchi 1984 and the robust optimization of Ben-Tal and Nemirovski 1999). This approach finds the worst-case performance among all MVN distributions constrained by the ambiguity sets to evaluate a policy and compares different policies based on these worst-case performances.

To handle the above robust simulation, we implement a change-of-measure technique that uses the simulation results conducted at a given MVN distribution to infer the simulation results for all MVN distributions constrained by the ambiguity sets, and then convert the simulation problem to a stochastic optimization problem. To solve the stochastic optimization problem, we use a sample-average approximation that takes a sample of simulation replications and reformulates the optimization problem into a semi-definite optimization problem. We then take an iterative approach and solve a convex optimization problem in each iteration. We prove that the sequence of solutions generated by our algorithm is improving and converges to a stationary point of the sample-average approximation.

We apply our approach on the DICE model. We consider three performance measures, as well as five environmental policies. We find that the ambiguity of covariance matrix has very little impact on all three performance measures and, therefore, the model is robust with respect to the ambiguity covariance matrix. When the ambiguity of the mean vector is considered, however, the performances may change significantly for all three performance measures. However, these changes in general do not alter our preferences over different policies, i.e., the ranking of the policies are robust with respect to the ambiguities of both mean vector and covariance matrix. Therefore, we conclude that the DICE model is robust in comparing different policies, but its performances are only robust with respect to ambiguous covariance matrix.

The rest of this paper is organized as follows. In Section 2, we introduce the problem formulation and the change-of-measure reformulation. In section 3, we develop a sequential algorithm to solve the optimization problem formulated in Section 2 and analyze its properties. We study the robustness of the DICE model and conclude the paper in Section 4.

2 PROBLEM FORMULATION

2.1 Basic Formulation

Let ξ denote the d-dimensional vector of the key uncertain parameters in the environmental simulation model. Following the convention of environmental economics literature, we assume that ξ follows a MVN distribution with a mean vector μ and a covariance matrix Σ. Let $c(\xi)$ denote the output of the simulation model when ξ is given. In our context, $c(\xi)$ may be the cost, or the temperature increase. Because of the complexity of environmental simulation models, we assume that the closed-form expression of $c(\cdot)$ is not available and $c(\xi)$ can only be observed through running simulation experiment at ξ. In the simulation study, we are interested in estimating $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$.

Note that $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$ is a function of (μ,Σ). When there is ambiguity in the specification of (μ,Σ), there is also ambiguity in $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$. Let $\mathscr{M}$ and $\mathscr{S}$ denote the sets of possible values that μ and Σ may take, respectively, and $\mathscr{M}\times\mathscr{S}$ denote corresponding Cartesian product. We call $\mathscr{M}$, $\mathscr{S}$ and $\mathscr{M}\times\mathscr{S}$ the "ambiguity sets" and discuss in detail in Section 2.2. To handle the ambiguity in (μ,Σ), we take a robust approach and find the worst-case performance of $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$ among all (μ,Σ) in $\mathscr{M}\times\mathscr{S}$. Because $c(\xi)$ often denotes cost or temperature increase, higher values are worse. Therefore, we formulate the problem as the following maximization problem:

$$\max_{(\mu,\Sigma)\in\mathscr{M}\times\mathscr{S}} \quad \mathrm{E}_{(\mu,\Sigma)}[c(\xi)]. \tag{1}$$

Without loss of generality, we further assume that $c(\xi)$ is nonnegative for all $\xi\in\Re^d$, as in DICE model and many practical situations. If $c(\xi)$ is not always nonnegative, we can make it nonnegative by adding a positive number as long as $c(\xi)$ is bounded below, and this does not change the solution of Problem (1). For instance, if $c(\xi)\geq -M$ for all $\xi\in\Re^d$ and some constant $M>0$, then $c(\xi)+M'\geq 0$ for all $\xi\in\Re^d$ and any constant $M'\geq M$.

2.2 The Ambiguity Sets

To model the ambiguity of the mean vector μ and the covariance matrix Σ, we consider two different scenarios that are commonly faced by environmental modelers. In the first scenario, there exist some data that can be used directly

to estimate μ and Σ. Then, we model the ambiguity sets $\mathscr{M}$ and $\mathscr{S}$ by their confidence regions, denoted as $\mathscr{M}_1$ and $\mathscr{S}_1$, respectively. In the second scenario, there is no data directly available to estimate μ and Σ. The ambiguity sets are modeled more or less "subjectively" based on modeler's information or belief, and they are denoted as $\mathscr{M}_2$ and $\mathscr{S}_2$, respectively.

2.2.1 The Ambiguity Set of the Mean Vector

We first consider the ambiguity set $\mathscr{M}$ of the mean vector μ. In the first scenario, suppose there are m independent observations of ξ, denoted as $\{\xi_1,\ldots,\xi_m\}$. Because ξ follows a MVN distribution, based on the Hotelling's T^2-statistic, the $1-\alpha$ confidence region (as well of the ambiguity set) of μ has the following ellipsoid shape (Anderson 1984):

$$\mathscr{M}_1 = \left\{\mu : (\mu-\hat{\mu})^T\hat{\Sigma}^{-1}(\mu-\hat{\mu}) \le \gamma\right\}, \tag{2}$$

where $\hat{\mu} = \frac{1}{m}\sum_{i=1}^m \xi_i$ and $\hat{\Sigma} = \frac{1}{m-1}\sum_{i=1}^m (\xi_i-\hat{\mu})(\xi_i-\hat{\mu})^T$ are the estimates of μ and Σ, and $\gamma>0$ is a given constant depending on $1-\alpha$. Note that $\mathscr{M}_1$ is a convex and compact set of μ, and it is also considered in the robust optimization literature (e.g., Delage and Ye 2010).

In the second scenario, when there is no data available to estimate μ directly, we let

$$\mathscr{M}_2 = \{\mu : \mu_\ell \le \mu \le \mu_u\}, \tag{3}$$

where $\mu_\ell, \mu_u \in \Re^d$ are the lower and upper bounds of μ. Note that $\mathscr{M}_2$ takes the form of a hyper-box and it is also a convex and compact set of μ. It is more suitable when the estimations of the random parameters are "subjective" or "judgemental", as in the DICE model (Nordhaus 2008). Note that both $\mathscr{M}_1$ and $\mathscr{M}_2$ are convex and compact sets of μ. Indeed, besides these two forms, the algorithm we propose in Section 3 can handle any convex and compact ambiguity set $\mathscr{M} \subset \Re^d$.

2.2.2 The Ambiguity Set of the Covariance Matrix

Let $\mathbb{S}^d$, $\mathbb{S}^d_+$ and $\mathbb{S}^d_{++}$ denote the $d\times d$ symmetric matrix space, $d\times d$ positive semi-definite matrix space and the $d\times d$ positive definite matrix space, respectively. Because the covariance matrix Σ needs to be positive definite (we do not consider degenerate cases), we assume that the ambiguity set $\mathscr{S}$ is a compact convex subset of $\mathbb{S}^d_{++}$.

To model the ambiguity in the first scenario, we let

$$\mathscr{S}_1 = \{\Sigma : \Sigma_\ell \preccurlyeq \Sigma \preccurlyeq \Sigma_u\}, \tag{4}$$

where Σ_ℓ and Σ_u are positive definite matrices and $A \preccurlyeq B$ denotes that $B-A$ is positive semi-definite. Suppose there are m independent observations $\{\xi_1,\ldots,\xi_m\}$ of ξ. Because ξ follows a MVN with mean μ and covariance matrix Σ, the estimate $\hat{\Sigma}$ of Σ follows a Wishart distribution (Anderson 1984). Then, a $1-\alpha$ confidence region of Σ is defined as $\left\{\Sigma : \gamma_1\hat{\Sigma} \preccurlyeq \Sigma \preccurlyeq \gamma_2\hat{\Sigma}\right\}$ with $\gamma_2>\gamma_1>0$ and it is exactly in the form of $\mathscr{S}_1$. Therefore, when there are data available to estimate Σ, we may construct an ambiguity set of Σ in the form of $\mathscr{S}_1$, and the ambiguity is caused by the estimation error.

To model the ambiguity in the second scenario, we let

$$\mathscr{S}_2 = \left\{\Sigma : \Sigma_{ii} \le \sigma_i^2, i=1,\ldots,d,\ \Sigma \succeq \varepsilon I\right\} \tag{5}$$

for some $\varepsilon>0$, where I denotes a $d\times d$ identity matrix. One may set ε as a very small positive number if there is no explicit information on it. This condition is equivalent to requiring all eigenvalues of Σ be at least ε, and it is used to ensure that Σ is strictly positive definite and that $\mathscr{S}_2$ is a compact set in $\mathbb{S}^d_{++}$. Note that Σ_{ii} is the variance of the ith element of ξ. The ambiguity set in the form of $\mathscr{S}_2$ essentially constrains the variances of all random elements. In practice, when there is little information (or data) available to estimate the correlations among all random elements, $\mathscr{S}_2$ is a form that one may use to model the ambiguity set of Σ.

Let $X=\Sigma^{-1}$ and $\mathscr{X} = \{X : \Sigma \in \mathscr{S}\}$. Note that $\mathscr{X}$ is a compact subset of $\mathbb{S}^d_{++}$ if $\mathscr{S}$ is. Because the algorithm we propose in Section 3 solves for X instead of Σ, we analyze the properties of $\mathscr{X}$ in the rest of this subsection. Let

$$\mathscr{X}_1 = \left\{X : \Sigma_u^{-1} \preccurlyeq X \preccurlyeq \Sigma_\ell^{-1}\right\}.$$

It is easy to see that $\Sigma \in \mathscr{S}_1$ is equivalent to $X \in \mathscr{X}_1$. Furthermore, note that $\Sigma_{ii} \leq \sigma_i^2$ is equivalent to $e_i^T X^{-1} e_i \leq \sigma_i^2$ and, hence, is equivalent to $\begin{bmatrix} X & e_i \\ e_i^T & \sigma_i^2 \end{bmatrix} \succeq 0$, where e_i denotes the ith column of a $d \times d$ identity matrix. Let

$$\mathscr{X}_2 = \left\{ X : \begin{bmatrix} X & e_i \\ e_i^T & \sigma_i^2 \end{bmatrix} \succeq 0, i = 1, \ldots, d,\ X \preccurlyeq \varepsilon^{-1} I \right\}.$$

Then, $\Sigma \in \mathscr{S}_2$ is equivalent to $X \in \mathscr{X}_2$. Note that all constraints in $\mathscr{X}_1$ and $\mathscr{X}_2$ can be transformed into linear matrix inequalities, which can be handled efficiently by convex optimization algorithms (Boyd and Vandenberghe 2004). Indeed, besides the two forms $\mathscr{S}_1$ and $\mathscr{S}_2$, the algorithm we propose in Section 3 can handle any ambiguity sets where $\mathscr{X}$ is defined by linear matrix inequalities.

2.3 A Change-of-Measure Reformulation

Problem (1) is difficult to solve because the closed form of $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$ is not known. One approach is to treat it as an simulation optimization problem (e.g., Fu 2002 and Hong and Nelson 2009), as the objective function of Problem (1) may be evaluated by running simulation experiments at any $(\mu,\Sigma) \in \mathscr{M} \times \mathscr{S}$. However, this approach has two drawbacks. First, Σ is a positive definite matrix and, to the best of our knowledge, there are no available convergent simulation optimization algorithms to solve such problems. Second, even there exist such algorithms, this approach requires running time-consuming simulation experiments at often a large number of different (μ,Σ) values and, hence, often requires a prohibitively large amount of computational effort.

In this paper we introduce a different approach that reformulates Problem (1) based on a change-of-measure technique. Let $f(\cdot)$ denote a probability density function defined on $\Re^d$ such that $f(x) > 0$ for any $x \in \Re^d$ and let E_f denote the expectation taken with respect to the distribution $f(\cdot)$. Let $\phi_{(\mu,\Sigma)}(\cdot)$ denote the density of a MVN distribution with mean μ and covariance matrix Σ. Then,

$$\mathrm{E}_{(\mu,\Sigma)}[c(\xi)] = \int_{\Re^d} c(x)\phi_{(\mu,\Sigma)}(x)dx = \int_{\Re^d} c(x)\frac{\phi_{(\mu,\Sigma)}(x)}{f(x)}f(x)dx = \mathrm{E}_f\left[c(\xi)\frac{\phi_{(\mu,\Sigma)}(\xi)}{f(\xi)}\right], \tag{6}$$

where ξ has the density $f(\cdot)$ in the last equation. The technique used in deriving Equation (6) is known as change of measure, i.e., the probability measure that the expectation is taken with respect to is changed from $\phi_{(\mu,\Sigma)}(\cdot)$ to $f(\cdot)$, and the term $\phi_{(\mu,\Sigma)}(\cdot)/f(\cdot)$ is called a likelihood ratio or a Radon-Nikodym derivative. Plugging the function form of $\phi_{(\mu,\Sigma)}(\cdot)$ in Equation (6), we may reformulate Problem (1) as

$$\max_{(\mu,\Sigma)\in\mathscr{M}\times\mathscr{S}} \quad \mathrm{E}_f\left[(2\pi)^{-d/2}\frac{c(\xi)}{f(\xi)}(\det\Sigma)^{-1/2}\exp\left\{-\frac{1}{2}(\xi-\mu)^T\Sigma^{-1}(\xi-\mu)\right\}\right], \tag{7}$$

where $\det A$ denotes the determinant of a matrix A.

Note that, without the change of measure, the decision variables (μ,Σ) is in the distribution of ξ. Therefore, to evaluate $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$ for different (μ,Σ) values, we need to generate ξ with different means and covariance matrices and evaluate $c(\xi)$ through time-consuming simulation experiments. Then, Problem (1) is difficult to solve. With the change of measure, however, the decision variables (μ,Σ) is no longer in the distribution of ξ. This allows us to separate the decision variables (μ,Σ) and the randomness ξ. We can now use a sample of ξ and the corresponding $c(\xi)$ values to evaluate $\mathrm{E}_{(\mu,\Sigma)}[c(\xi)]$ for all $(\mu,\Sigma) \in \mathscr{M} \times \mathscr{S}$. This greatly reduces the required computational effort and makes it possible to solve Problem (7) efficiently.

3 SOLUTION METHOD

Following the notation used in Section 2.2, we let $X = \Sigma^{-1}$ and $\mathscr{X} = \{X \in \mathbb{S}^d_{++} : X^{-1} \in \mathscr{S}\}$, where $\mathscr{S}$ is the ambiguity set for covariance matrix Σ. In this paper, we only consider the cases where $\mathscr{X}$ is defined by linear matrix inequalities, as in the cases of $\mathscr{X}_1$ and $\mathscr{X}_2$ (see discussions in Section 2.2). Then, Problem (7) can be reformulated as follows:

$$\max_{(\mu,X)\in\mathscr{M}\times\mathscr{X}} \quad g(\mu,X) := \mathrm{E}_f\left[(2\pi)^{-d/2}\frac{c(\xi)}{f(\xi)}\exp\left\{-\frac{1}{2}(\xi-\mu)^T X(\xi-\mu)+\frac{1}{2}\log\det X\right\}\right]. \tag{8}$$

In the rest of this section, we develop an efficient algorithm to solve Problem (8).

3.1 Sample-Average Approximation

Because $c(\xi)$ is the output from a complex environmental simulation model, a closed-form expression of $g(\mu,X)$ is typically unavailable. Instead, we can only use simulation observations to estimate $g(\mu,X)$. Suppose that we have an independent and identically distributed (i.i.d.) sample of ξ, denoted as $\xi_1,\ldots,\xi_n$, generated from the distribution $f(\cdot)$, and the i.i.d. simulation outputs $c(\xi_1),\ldots,c(\xi_n)$. We can then estimate $g(\mu,X)$ by

$$\bar{g}_n(\mu,X)=\frac{1}{n}\sum_{j=1}^{n}(2\pi)^{-d/2}\frac{c(\xi_j)}{f(\xi_j)}\exp\left\{-\frac{1}{2}(\xi_j-\mu)^T X(\xi_j-\mu)+\frac{1}{2}\log\det X\right\}$$

for any $(\mu,X)\in\mathscr{M}\times\mathscr{X}$. Let $a_j=(2\pi)^{-d/2}c(\xi_j)/f(\xi_j)$. Note that $a_j\geq 0$ for any $j=1,\ldots,n$, because in Section 2 we assumed that both $f(\xi)>0$ and $c(\xi)\geq 0$ for all $\xi\in\Re^d$. We suggest to solve the following sample-average approximation (SAA):

$$\max_{(\mu,X)\in\mathscr{M}\times\mathscr{X}}\quad \bar{g}_n(\mu,X):=\frac{1}{n}\sum_{j=1}^{n}a_j\exp\left\{-\frac{1}{2}(\xi_j-\mu)^T X(\xi_j-\mu)+\frac{1}{2}\log\det X\right\}. \tag{9}$$

Intuitively, if the sample size n is large enough, the solution to Problem (9) should be close to the solution to Problem (8). Therefore, it provides a good approximation to the solution of Problem (8). In the rest of this subsection, we show that the intuition indeed holds under some weak conditions.

For simplicity of the notation, we let

$$L(\mu,X,\xi)=(2\pi)^{-d/2}\frac{c(\xi)}{f(\xi)}\exp\left\{-\frac{1}{2}(\xi-\mu)^T X(\xi-\mu)+\frac{1}{2}\log\det X\right\}.$$

Let "∇_μ" denote the gradient of a function with respect to μ and "∇_X" denote the derivative of a function with respect to the matrix X. Note that $\nabla_X\log\det X=X^{-1}$. Then, by chain rule we have

$$\nabla_\mu L(\mu,X,\xi)=(2\pi)^{-d/2}\tfrac{c(\xi)}{f(\xi)}\exp\left\{-\tfrac{1}{2}(\xi-\mu)^T X(\xi-\mu)+\tfrac{1}{2}\log\det X\right\}X(\xi-\mu),$$

and

$$\nabla_X L(\mu,X,\xi)=(2\pi)^{-d/2}\tfrac{c(\xi)}{f(\xi)}\exp\left\{-\tfrac{1}{2}(\xi-\mu)^T X(\xi-\mu)+\tfrac{1}{2}\log\det X\right\}\left[-\tfrac{1}{2}(\xi-\mu)(\xi-\mu)^T+\tfrac{1}{2}X^{-1}\right].$$

With some abuse of the notation we let $\|\cdot\|$ denote both the Euclidean norm in $\Re^d$ and the Frobenius norm in the symmetric matrix space $\mathbb{S}^d$ (Golub and Van Loan 1996). We make the following assumption on $L(\mu,X,\xi)$:

Assumption 1. *There exist measurable functions $D_i(\cdot), i=1,2$, such that $\mathrm{E}_f[D_i(\xi)]<\infty, i=1,2$, and*

$$\|\nabla_\mu L(\mu,X,\xi)\|\leq D_1(\xi),\quad \|\nabla_X L(\mu,X,\xi)\|\leq D_2(\xi),\quad \forall\ (\mu,X)\in\mathscr{M}\times\mathscr{X},\forall\ \xi\in\Re^d.$$

Assumption 1 states that $\{\|\nabla_\mu L(\mu,X,\xi)\|\}_{(\mu,X)\in\mathscr{M}\times\mathscr{X}}$ and $\{\|\nabla_X L(\mu,X,\xi)\|\}_{(\mu,X)\in\mathscr{M}\times\mathscr{X}}$ are dominated by some integrable functions. It is a standard assumption for the convergence of SAAs (see, e.g., Shapiro et al. 2009). In our problem, it turns out to be a very weak assumption. For instance, we have the following propositions.

Proposition 1. *Suppose the ambiguity set $\mathscr{S}$ is in the form of $\mathscr{S}_1$ and $f(\cdot)=\phi_{(\mu_s,\Sigma_u)}(\cdot)$. If there exists $\gamma>0$, such that $\mathrm{E}_{(\mu_s,\Sigma_u)}\left[c^{1+\gamma}(\xi)\right]<\infty$, then Assumption 1 holds.*

Proposition 2. *Suppose the ambiguity set $\mathscr{S}$ takes the form of $\mathscr{S}_2$ and $f(\cdot)=t(\cdot;\mu_s,\Pi,\nu)$ where $t(\cdot;\mu_s,\Pi,\nu)$ is the multivariate t distribution. If there exists $\gamma>0$, such that $\mathrm{E}_{(\mu_s,\mathbf{diag}\{d\sigma_1^2,\ldots,d\sigma_d^2\})}\left[c^{1+\gamma}(\xi)\right]<\infty$, then Assumption 1 holds.*

For any sets $\mathscr{M}_A\times\mathscr{X}_A,\mathscr{M}_B\times\mathscr{X}_B\subset\Re^d\times\mathbb{S}^d$, we let

$$\mathrm{dist}((\mu,X),\mathscr{M}_A\times\mathscr{X}_A)=\inf_{(\mu',X')\in\mathscr{M}_A\times\mathscr{X}_A}\left(\|\mu-\mu'\|^2+\|X-X'\|^2\right)^{1/2}$$

denote the distance from $(\mu,X)\in\Re^d\times\mathbb{S}^d$ to $\mathscr{M}_A\times\mathscr{X}_A$, and let

$$\mathbb{D}(\mathscr{M}_A\times\mathscr{X}_A,\mathscr{M}_B\times\mathscr{X}_B)=\sup_{(\mu,X)\in\mathscr{M}_A\times\mathscr{X}_A}\mathrm{dist}((\mu,X),\mathscr{M}_B\times\mathscr{X}_B)$$

denote the deviation of the set $\mathscr{M}_A \times \mathscr{X}_A$ from the set $\mathscr{M}_B \times \mathscr{X}_B$. Let $\bar{v}_n$ and $\bar{S}_n$ denote the optimal objective value and the set of optimal solutions, respectively, of Problem (9), and v and S denote the optimal objective value and the set of optimal solutions, respectively, of Problem (8). Then, we have the following theorem on the convergence of the SAA.

Theorem 1. *Suppose that Assumption 1 is satisfied. Then, S is nonempty and $\bar{S}_n$ are nonempty for all n, and, with probability 1 (w.p.1), $\bar{v}_n \to v$ and $\mathbb{D}(\bar{S}_n, S) \to 0$ as $n \to \infty$.*

Theorem 1 shows that the optimal solutions of the SAA Problem (9) provide good approximations to the optimal solutions of Problem (8) when n is large. Therefore, we can solve Problem (9) to approximate the optimal solutions of Problem (8). Note that Problem (9) may not be a convex optimization problem, because $\bar{g}_n(\mu, X)$ may not be concave (or quasi-concave). Then, finding an optimal solution to Problem (9) may be difficult. We often only find stationary points of Problem (9). Then, the conclusions of Theorem 1 may not be applicable. Therefore, we need to investigate the convergence of stationary points of the SAA.

Consider the convex compact subsets $\mathscr{M}$ of $\Re^d$ and $\mathscr{X}$ of $\mathbb{S}^d_{++}$. Denote by $N_{\mathscr{M}}(\mu)$ the normal cone of $\mathscr{M}$ at point $\mu \in \mathscr{M}$ and by $N_{\mathscr{X}}(X)$ the normal cone of $\mathscr{X}$ at point $X \in \mathscr{X}$, i.e.,

$$N_{\mathscr{M}}(\mu) = \left\{ z \in \Re^d : z^T(\mu' - \mu) \le 0, \forall\ \mu' \in \mathscr{M} \right\}, \quad N_{\mathscr{X}}(X) = \left\{ Z \in S^d : \mathbf{tr}(Z(X' - X)) \le 0, \forall\ X' \in \mathscr{X} \right\},$$

where $\mathbf{tr}(A)$ denotes the trace of a matrix A. Recall that a point $(\mu, X) \in \mathscr{M} \times \mathscr{X}$ is a stationary point of Problem (8) if $\nabla_\mu g(\mu, X) \in N_{\mathscr{M}}(\mu)$ and $\nabla_X g(\mu, X) \in N_{\mathscr{X}}(X)$, and is a stationary point of Problem (9) if $\nabla_\mu \bar{g}_n(\mu, X) \in N_{\mathscr{M}}(\mu)$ and $\nabla_X \bar{g}_n(\mu, X) \in N_{\mathscr{X}}(X)$. Let K and $\bar{K}_n$ denote the sets of stationary points of Problems (8) and (9), respectively. Then, we have

$$\begin{aligned} K &= \left\{ (\mu, X) \in \mathscr{M} \times \mathscr{X} : \nabla_\mu g(\mu, X) \in N_{\mathscr{M}}(\mu), \nabla_X g(\mu, X) \in N_{\mathscr{X}}(X) \right\}, \\ \bar{K}_n &= \left\{ (\mu, X) \in \mathscr{M} \times \mathscr{X} : \nabla_\mu \bar{g}_n(\mu, X) \in N_{\mathscr{M}}(\mu), \nabla_X \bar{g}_n(\mu, X) \in N_{\mathscr{X}}(X) \right\}. \end{aligned} \tag{10}$$

The next theorem is parallel to Theorem 1. It ensures that the stationary points of the SAA are also good approximations to those of Problem (8).

Theorem 2. *Suppose that Assumption 1 is satisfied. Then, K is nonempty and $\bar{K}_n$ are nonempty for all n, and $\mathbb{D}(\bar{K}_n, K) \to 0$ w.p.1 as $n \to \infty$.*

Theorem 2 shows that, if we can find a local optimal solution of Problem (9), it provides a good approximation to a local optimal solution of Problem (8). In the rest of this section, we show how to find a good local optimal solution of Problem (9).

3.2 Sequential Quadratic & Maxdet Programs

As discussed in the last subsection, Problem (9) may not be a convex optimization problem. This leads to a significant difficulty in solving it. In this subsection, we introduce an iterative convex optimization scheme that makes improvement at each iteration and leads to a stationary point of Problem (9). Note that we may introduce a set of intermediate variables $t_j, j = 1, \ldots, n$, and reformulate Problem (9) as follows:

$$\max \quad \frac{1}{n} \sum_{j=1}^{n} a_j t_j \tag{11}$$

$$\text{s.t.} \quad -\frac{1}{2}(\xi_j - \mu)^T X (\xi_j - \mu) + \frac{1}{2} \log\det X \ge \log t_j,\ \ j = 1, \ldots, n, \tag{12}$$

$$(\mu, X) \in \mathscr{M} \times \mathscr{X},\ t_j \in \Re,\ j = 1, \ldots, n$$

where we define $\log(x) = -\infty$ if $x \le 0$ to make it an extended real-valued function. Note that the newly defined $\log(x)$ is still a concave function of x as the epi-graph of $-\log(x)$ defines a convex set (Rockafellar 1970). From the formulation, it is clear to see that the constraints defined by Equation (12) are tight at the optimal solution of Problem (11).

Suppose (μ_0, X_0) is a feasible solution of Problem (9), i.e., $(\mu_0, X_0) \in \mathscr{M} \times \mathscr{X}$. For instance, (μ_0, X_0) may be the best guess for the random parameters in the DICE model. Then, we may let

$$t_j^0 = \exp\left\{ -\frac{1}{2}(\xi_j - \mu_0)^T X_0 (\xi_j - \mu_0) + \frac{1}{2} \log\det X_0 \right\},\ \ j = 1, \cdots, n.$$

Note that $t_j^0 > 0$ for all $j = 1,\ldots,n$. Then, the constraints defined by Equation (12) are tight at $(\mu_0, X_0, t_1^0, \ldots, t_n^0)$ and, thus, $(\mu_0, X_0, t_1^0, \ldots, t_n^0)$ is a feasible solution of Problem (11). Note that $\log t_j$ is concave with respect to t_j, and $-\frac{1}{2}(\xi_j - \mu)^T X (\xi_j - \mu) + \frac{1}{2}\log\det X$ is a bi-concave function of μ and X, i.e., for every $X \in \mathscr{X}$ it is concave with respect to μ, and for every $\mu \in \mathscr{M}$ it is concave with respect to X. Obviously, Equation (12) may not define a convex set and, thus, Problem (11) may not be convex. To solve the problem, we linearize $\log t_j$ by a first-order Taylor expansion at t_j^0, and fix $X = X_0$ at the same time, and then solve the following nonlinear optimization problem to update the value of μ:

$$\max \quad \frac{1}{n}\sum_{j=1}^{n} a_j t_j \tag{13}$$

$$\text{s.t.} \quad -\frac{1}{2}(\xi_j - \mu)^T X_0 (\xi_j - \mu) + \frac{1}{2}\log\det X_0 \geq \log t_j^0 + \frac{1}{t_j^0}\left(t_j - t_j^0\right),\ j = 1,\ldots,n, \tag{14}$$

$$\mu \in \mathscr{M},\ t_j \in \Re,\ j = 1,\ldots,n.$$

Note that Problem (13) is now a convex optimization problem. The linearization approach used in deriving Problem (13) is a simple approach to convexifying optimization problems where the constraint can be expressed as the difference of two convex functions. It is also used by Smola et al. (2005) and Hong et al. (2009) to solve non-convex real-valued optimization problems.

Note that the constraints defined in Equation (14) are equivalent to

$$t_j \leq \frac{1}{2}t_j^0\left(-(\xi_j - \mu)^T X_0 (\xi_j - \mu) + \log\det X_0\right) + t_j^0\left(1 - \log t_j^0\right),\ j = 1,\ldots,n.$$

Let $(\mu_1, t_1^1, \ldots, t_n^1)$ be the optimal solution of Problem (13). Then, it is clear that

$$t_j^1 = \frac{1}{2}t_j^0\left(-(\xi_j - \mu_1)^T X_0 (\xi_j - \mu_1) + \log\det X_0\right) + t_j^0\left(1 - \log t_j^0\right),\ j = 1,\ldots,n.$$

Substituting $t_j^1, j = 1,\ldots,n$, into the formulation of Problem (13), we can easily see that μ_1 is the optimal solution to the following optimization problem:

$$\max_{\mu\in\mathscr{M}} \quad \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2}a_j t_j^0\left(-(\xi_j - \mu)^T X_0 (\xi_j - \mu) + \log\det X_0\right) + \frac{1}{n}\sum_{j=1}^{n} a_j t_j^0\left(1 - \log t_j^0\right). \tag{15}$$

Let

$$V_1 = \frac{\sum_{j=1}^{n} a_j t_j^0 \xi_j}{\sum_{j=1}^{n} a_j t_j^0}.$$

Then, it is easy to verify that finding the optimal solution of Problem (15) is equivalent to finding the optimal solution to the following problem:

$$\max_{\mu\in\mathscr{M}} \quad -\frac{1}{2}\mu^T X_0 \mu + V_1^T X_0 \mu. \tag{16}$$

Problem (16) is a standard convex quadratic program and can be solved efficiently (Boyd and Vandenberghe 2004). Note that $-\frac{1}{2}\mu^T X_0 \mu + V_1^T X_0 \mu$ is strictly concave. Therefore, Problem (16) has a unique optimal solution μ_1.

Once we obtain the new mean vector $\mu = \mu_1$, we calculate

$$s_j^1 = \exp\left\{-\frac{1}{2}(\xi_j - \mu_1)^T X_0 (\xi_j - \mu_1) + \frac{1}{2}\log\det X_0\right\},\ j = 1,\cdots,n.$$

Similarly to the derivation of Problem (15), in Problem (11) we fix $\mu = \mu_1$ and linearize $\log t_j$ by a first-order Taylor expansion at s_j^1, and then solve the following semi-definite program to update the value for X:

$$\max_{X\in\mathscr{X}} \quad \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2}a_j s_j^1\left(-(\xi_j - \mu_1)^T X (\xi_j - \mu_1) + \log\det X\right) + \frac{1}{n}\sum_{j=1}^{n} a_j s_j^1\left(1 - \log s_j^1\right). \tag{17}$$

Let X_1 be the optimal solution of Problem (17). Since the objective function in Problem (17) is strictly concave, the uniqueness of X_1 can also be guaranteed. Let

$$\Pi_1 = \frac{\sum_{j=1}^n a_j s_j^1 \ (\xi_j - \mu_1)(\xi_j - \mu_1)^T}{\sum_{j=1}^n a_j s_j^1}.$$

Note that $(\xi_j - \mu_1)^T X (\xi_j - \mu_1) = \mathbf{tr}\left((\xi_j - \mu_1)(\xi_j - \mu_1)^T X\right)$. Then, X_1 is the unique optimal solution of the following problem:

$$\max_{X \in \mathscr{X}} \quad -\mathbf{tr}(\Pi_1 X) + \log\det X. \tag{18}$$

Note that, when $\mathscr{X}$ is formed by linear matrix inequalities, Problem (18) is known as a log-determinant maximization problem, which we denote as a maxdet problem. Maxdet problem is a well-known convex semi-definite program and has been studied thoroughly in the literature (Vandenberghe et al. 1998). It can be solved efficiently using MAXDET software (Wu et al. 1996) or the convex optimization package CVX (Grant and Boyd 2009). Recently, Wang et al. (2009) developed a new method capable of solving efficiently large-scale maxdet problems where the dimension of the matrix can be as high as 2000×2000. For the DICE model we consider in this paper, the dimension of the covariance matrix is only 8×8. According to our experience, a maxdet problem of this size can typically be solved within 5 seconds by using CVX on a personal computer.

Because $\log t_j$ is a concave function, we have $\log t_j \leq \log t_j^0 + \frac{1}{t_j^0}\left(t_j - t_j^0\right)$, for any $t_j \in \Re$. Then, the objective function of Problem (15) is bounded from above by the function $\bar{g}_n(\mu, X_0)$ on the feasible region $\mathscr{M}$, and at point $\mu = \mu_0$, the two functions do coincide. Furthermore, note that μ_1 is the optimal solution of Problem (15). It follows that $\bar{g}_n(\mu_1, X_0)$ is always larger than or equal to $\bar{g}_n(\mu_0, X_0)$. Using the same argument we can obtain that $\bar{g}_n(\mu_1, X_1)$ is always no less than $\bar{g}_n(\mu_1, X_0)$. Therefore, if we start with $(\mu_0, X_0) \in \mathscr{M} \times \mathscr{X}$, which is a feasible solution of Problem (9), we can solve Problems (16) and (18) to find (μ_1, X_1) that is feasible to Problem (9) and is at least as good as (μ_0, X_0). Note that, once we have (μ_1, X_1), we can then repeat this process to find even better solutions. This motivates us to develop the following iterative algorithm, which we call *Sequential Quadratic & Maxdet Programs (SQMP)*:

Algorithm SQMP

Step 0. Give initial point $(\mu_0, X_0) \in \mathscr{M} \times \mathscr{X}$, and set $k = 0$.

Step 1. Let

$$t_j^k = \exp\left\{-\frac{1}{2}(\xi_j - \mu_k)^T X_k (\xi_j - \mu_k) + \frac{1}{2}\log\det X_k\right\}$$

for all $j = 1, \ldots, n$. Set $k = k+1$, compute

$$V_k = \frac{\sum_{j=1}^n a_j t_j^{k-1} \xi_j}{\sum_{j=1}^n a_j t_j^{k-1}},$$

and let $\mu_k \in \arg\max\left\{\mu \in \mathscr{M} : -\frac{1}{2}\mu^T X_{k-1}\mu + V_k^T X_{k-1}\mu\right\}$.

Step 2. Let

$$s_j^k = \exp\left\{-\frac{1}{2}(\xi_j - \mu_k)^T X_{k-1} (\xi_j - \mu_k) + \frac{1}{2}\log\det X_{k-1}\right\}$$

for all $j = 1, \ldots, n$. Compute

$$\Pi_k = \frac{\sum_{j=1}^n a_j s_j^k (\xi_j - \mu_k)(\xi_j - \mu_k)^T}{\sum_{j=1}^n a_j s_j^k},$$

and let $X_k \in \arg\max\{X \in \mathscr{X} : -\mathbf{tr}(\Pi_k X) + \log\det X\}$.

Step 3. Stop if (μ_k, X_k) is a stationary point of Problem (9), i.e., $(\mu_k, X_k) \in \bar{K}_n$ where $\bar{K}_n$ is defined in Equation (10); otherwise, go to **Step 1**.

Because Algorithm SQMP solves a sequence of standard quadratic problems and maxdet problems, it is computationally very efficient and it can be implemented easily. Besides the ease of implementation, the algorithm also has many good theoretical properties. We summarize them in the following theorem.

Theorem 3. *Suppose that* $\{(\mu_k, X_k), k = 0, 1, \cdots\}$ *is a sequence of solutions generated by Algorithm SQMP when it is applied to solve Problem (9). Then, it satisfies the following properties:*

1. (μ_k, X_k) *is a feasible solution of Problem (9), i.e.,* $(\mu_k, X_k) \in \mathscr{M} \times \mathscr{X}$*, for all* $k = 0, 1, \ldots$*;*
2. $\{\bar{g}_n(\mu_k, X_k), k = 0, 1, \cdots\}$ *is a nondecreasing convergent sequence;*
3. *if* $(\mu_{k+1}, X_{k+1}) \neq (\mu_k, X_k)$*, then* $\bar{g}_n(\mu_{k+1}, X_{k+1}) > \bar{g}_n(\mu_k, X_k)$*; otherwise,* (μ_k, X_k) *is a stationary point of Problem (9);*
4. *all cluster points of* $\{(\mu_k, X_k), k = 0, 1, \cdots\}$ *are stationary points of Problem (9). Furthermore, if Problem (9) has a finite number of stationary points, then* $\{(\mu_k, X_k), k = 0, 1, \cdots\}$ *converges to a stationary point.*

The first property of Theorem 3 shows that the solutions generated by Algorithm SQMP are all feasible solutions of Problem (9), and, therefore, $\bar{g}_n(\mu_k, X_k)$ provides a lower bound on the optimal objective value for all $k = 0, 1, \ldots$. The second and third properties of Theorem 3 show that Algorithm SQMP keeps improving the quality of the solutions unless the current solution is already a stationary point of Problem (9). Therefore, the solutions found in later iterations are guaranteed better than the initial solution (μ_0, X_0), which is often taken to conduct simulation experiments and make decisions, unless (μ_0, X_0) is already a stationary point. The fourth property of Theorem 3 is the most important properties of the four. It shows that Algorithm SQMP indeed has the desired convergence. Furthermore, it shows that the sequence of solutions generated by Algorithm SQMP converges to a stationary point as long as Problem (9) has only a finite number of stationary points. Although Problem (9) is not provable convex, it may have only one stationary point or one stationary point that is better than the initial solution (μ_0, X_0). Then, Algorithm SQMP guarantees to converge to the point.

4 ROBUSTNESS OF THE DICE MODEL

In this section, we apply the DICE model to robustly evaluate and compare the following five representative environmental policies:

1. no controls with 250−year delay, denoted as "no controls";
2. atmosphere CO2 concentration limited to 2× the preindustrial level (i.e., 560 ppm), denoted as "2×CO2";
3. temperature increase limited to 2°C from the preindustrial level proposed by the Copenhagen Accord, denoted as "2°C increase";
4. the policy recommended by the Stern Review, denoted as "Stern Review"; and
5. Al Gore's proposal which achieves global emissions reductions of 90 percent by 2050, denoted as "Gore Proposal".

Readers may refer to Nordhaus (2008) for detailed descriptions of these policies.

We call the GAMS program[1] of the DICE model from Matlab, and simulate these five environmental policies simultaneously. The means and variances of the eight uncertain parameters, i.e., the rate of growth of total factor productivity **g(TFP)**, the rate of decarbonization **g(CO2/GDP)**, the equilibrium temperature-sensitivity coefficient $\mathbf{T_2}\times\mathbf{CO2}$, the damage parameter **DamCoeff**, the price of backstop technology **P(back)**, asymptotic global population **Pop**, the transfer coefficient in carbon cycle **CarCyc**, and total resources of fossil fuels **Fosslim** are estimated in Table 7-1 of Nordhaus (2008). We let $\hat{\mu}$ and $\hat{\Sigma}$ denote the estimated mean vector and covariance matrix of the eight uncertain parameters. Note that the off-diagonal elements of $\hat{\Sigma}$ are set as 0 because Nordhaus (2008) did not estimate the correlations. To model the ambiguity, we suppose there exists a 10% estimation error for the mean vector and set the ambiguity set as $\mathscr{M} = \{\mu : 0.9\hat{\mu} \leq \mu \leq 1.1\hat{\mu}\}$. For the covariance matrix, further enlarging the variances will cause truncation problems. Moreover, during the estimation Nordhaus (2008) has doubled some of the combined standard deviations. Therefore in this paper we focus on the left half confidence region and consider the ambiguity set $\mathscr{S} = \left\{\Sigma : 0.5\hat{\Sigma} \preccurlyeq \Sigma \preccurlyeq \hat{\Sigma}\right\}$. We set the sample size $n = 25000$, and simulate sample from the MVN distribution $N(\hat{\mu}, \hat{\Sigma})$. We implement the procedure introduced in preceding sections to find the worst case performances.

We first calculate the net present values (NPVs) of abatement costs plus climate damages for all five policies, and report them in Table 1. The first column of Table 1 reports the outputs of the DICE model when the uncertain parameters are substituted by their means. The results of this column are essentially the same as those reported in Table 5-1 of Nordhaus (2008).[2] The second column reports the average outputs of our 25000 independent simulation runs. By comparing columns 1 and 2, we find that the most likely values and the average values of all five polices are very close, which is consistent to what was reported in Nordhaus (2008).

Now we consider the robustness of the DICE model. We let the mean vector and covariance matrix vary in $\mathscr{M} \times \mathscr{S}$ and calculate the worst case outputs. In columns 3 and 4, we report the worst-case values and the percentage changes compared to the average performance, respectively. To further understand the different effects of the mean

[1]The GAMS code of the DICE model can also be downloaded at `http://nordhaus.econ.yale.edu/DICE2007.htm`.

[2]There exist some minor differences. They might be caused by the parameter settings in the GAMS code.

vector and covariance matrix, we fix the mean vector at $\hat{\mu}$ and allow the covariance matrix to vary in $\mathscr{S}$, and report the corresponding results in columns 5 and 6.

Table 1: NPVs of Abatement Costs Plus Climate Damages (Trillions of 2005 U.S. $)

Policy	Most likely	Average	Worst case	Change from Average case	Worst case (fix mean)	Change from Average case
No controls	23.12	24.64	38.99	58.21%	26.27	6.62%
2×CO2	20.00	20.89	31.88	52.62%	21.74	4.08%
2°C increase	24.72	27.27	43.75	60.42%	28.26	3.63%
Stern Review	39.96	38.53	60.28	56.42%	41.75	8.35%
Gore Proposal	45.55	53.67	71.54	33.29%	55.28	3.00%

From Table 1, we have the following observations. First, the DICE model is sensitive to the eight uncertain parameters. For the given small ambiguity set $\mathscr{M} \times \mathscr{S}$, the worst case average costs are $50\% - 60\%$ higher than the current average costs for most of the considered policies. Second, in the worst case, the order or the preference of these policies does not change compared to the most likely or the average cases (i.e., the first two columns of Table 1). Third, the cost measure seems much more sensitive to the mean vector than to the covariance matrix. From the above table we see that the 10% estimation error of the mean vector contributes to the most to the change of the costs. This suggests that, to improve the robustness of the DICE model, more effort needs to be spent on improving the estimates of the mean vector instead of the covariance matrix.

We also calculate the global temperature increases by 2105 from the pre-industrial level and report them in Table 2. The robustness of this important measure shows a similar pattern compared to the costs reported in Table 1, except that the covariance matrix of the parameters has an even smaller impact on the temperature change. Another interesting finding is that the cost measure and temperature change measure may attain their maxima at different points in the ambiguity sets. For instance, if we consider the mean of **g(TFP)**, maximizing the cost pushes it to its lower bound, while maximizing the temperature change pushes it to the upper bound. This is because the cost consists of both climate damages and abatement cost.

Table 2: Global Temperature Increases by 2105 (°C from the Pre-industrial Level)

Policy	Most likely	Average	Worst case	Change from Average case	Worst case (fix mean)	Change from Average case
No controls	3.19	3.17	3.68	16.09%	3.19	0.64%
2×CO2	2.53	2.37	2.64	11.39%	2.41	1.69%
2°C increase	2.00	2.00	2.00	0.00%	2.00	0.00%
Stern Review	1.51	1.54	1.73	12.34%	1.56	1.30%
Gore Proposal	1.49	1.49	1.68	12.75%	1.49	0.00%

DICE model is known to be a standard cost-benefit analysis. By comparing the costs and other concerned measures for different policies, it may suggest that facing global warming, "No Action" can be better than "Action". A major reservation about DICE model is perhaps that such cost-benefit analysis would fail when human beings might suffer an indefinitely large expected loss from catastrophic events. This forms the basic idea of Dismal Theory, (see e.g., Weitzman 2009). Catastrophic event is itself a vague content. Many people suggest that it would be catastrophic if the global average temperature rises by 4°C above the pre-industrial level. The United Kingdom's Met Office Hadley Centre even draws a map depicting some of the impacts that may occur if this event occurs[3]. At the end of this section we compare the probabilities of such event for different policies under current normality assumption. From the numerical simulation we estimate that the probability for no controls policy is 17.7%, with worst case probability being 36.5%. These probabilities are indeed large, indicating that there exist high risks for the no control policy. On the other hand, for 2×CO2 policy and Stern Review policy, the underlying event is identified as a rare event, with probability less than 0.1%. This means the risks of these two policies are much lower. Finally we note that for the aggressive Gore proposal we did not observe the occurrence of underlying event in these 25000 runs.

Summarizing the robustness analysis on DICE model, our conclusion is essentially consistent with Nordhaus (2008): There exist major uncertainties. Moreover, to deal with global warming, we DO need to take some MODERATE environmental policies, such as 2×CO2 policy, or 2°C increase policy. From the point of view of cost, temperature change, robustness of output, as well as risk, the 2×CO2 policy seems a competitive policy.

REFERENCE

Anderson, T. W. 1984. *An introduction to Multivariate Analysis*. New York: John Wiley and Sons, Inc.

Ben-Tal, A. and A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research*, 23:769-805.

Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.

[3]See `http://www.actoncopenhagen.decc.gov.uk/en/ambition/evidence/4- degrees-map/`.

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Tech. rep., Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Delage, E. and Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58:595-612.

Fu, M. C. 2002. Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing*, 14:192-215.

Grant, M. and S. Boyd. 2009. CVX: Matlab software for disciplined convex programming. Software available at: http://stanford.edu/boyd/cvx.

Hong, L. J. and B. L. Nelson. 2009. A brief introduction to optimization via simulation. *Proceedings of the 2009 Winter Simulation Conference*, pp.75-85.

Hong, L. J., Y. Yang, and L. Zhang. 2010. Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Operations Research*, forthcoming.

Nordhaus, W. 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press, New Haven, CT.

Shapiro, A., Dentcheva, D. and Ruszczynski, A. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.

Taguchi, G. 1984. The role of metrological control for quality control. *Proceedings of the International Symposium on Metrology for Quality Control in Production*, pp.1-7.

Smola, A.J., S.V. N.Vishwanathan, andT.Hofmann. 2005. Kernel methods for missing variables. *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, 325-332.

Wang, C., D. Sun and K. C. Toh. 2009. Solving log-determinant optimization problems by a Newton-CG proximal point algorithm. *Technical Reports*. Available at: http://www.math.nus.edu.sg/ matsundf/logdet-NAL-29Sep09.pdf.

Weitzman, Martin. 2009. On Modeling and Interpreting the Economics of Catastrophic Climate Change. *Review of Economics and Statistics*, 91(1), 1-19.

Wu, S.-P., L. Vandenberghe, and S. Boyd. 1996. MAXDET: Software for determinant maximization problems. Software available at: http://www.stanford.edu/ boyd/oldsoftware/MAXDET.html.

Vandenberghe, L., S. Boyd. and S. P. Wu. 1998. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2): 499-533.

AUTHOR BIOGRAPHIES

ZHAOLIN HU is a Ph.D. student in the Department of Industrial Engineering and Logistics Management at The Hong Kong University of Science and Technology. His research interests include simulation and stochastic optimization. His email address is <hzlshy@ust.hk>.

JING CAO is an assistant professor in the Department of Economics in the School of Economics and Management at Tsinghua University, China. She obtained her Ph.D. degree from John F. Kennedy School of Government at Harvard University in 2007. Her research interests include environmental economics, resource economics and development economics. Her email address is <caojing@sem.tsinghua.edu.cn>.

L. JEFF HONG is an associate professor in the Department of Industrial Engineering and Logistics Management and an associate director in the Logistics and Supply Chain Management Institute at The Hong Kong University of Science and Technology. His research interests include Monte Carlo method, financial engineering, and stochastic optimization. He is currently an associate editor of *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*. His email address is <jhong39@gmail.com>.