

Robust Simulation of Global Warming Policies Using the DICE Model

Zhaolin Hu

School of Economics and Management, Tongji University, 200092 Shanghai, China, huzhaolin@gmail.com

Jing Cao

School of Economics and Management, Tsinghua University, 100084 Beijing, China, caojing@sem.tsinghua.edu.cn

L. Jeff Hong

Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, hongl@ust.hk

Integrated assessment models that combine geophysics and economics features are often used to evaluate and compare global warming policies. Because there are typically profound uncertainties in these models, a simulation approach is often used. This approach requires the distribution of the uncertain parameters clearly specified. However, this is typically impossible because there is often a significant amount of ambiguity (e.g., estimation error) in specifying the distribution. 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 then show how to find the worst-case performance of a given policy for all distributions constrained by the ambiguity sets. This worst-case performance provides a robust evaluation of the policy. We test our algorithm on a famous integrated model of climate change, known as the Dynamic Integrated Model of Climate and the Economy (DICE model). We find that the DICE model is sensitive to the means and covariance of the parameters. Furthermore, we find that, based on the DICE model, moderately tight environmental policies robustly outperform the no controls policy and the famous aggressive policies proposed by Stern and Gore.

Key words: environment; global warming; programming; semidefinite; simulation; applications

History: Received September 11, 2010; accepted January 11, 2012, by Dimitris Bertsimas, optimization.

Published online in *Articles in Advance* August 3, 2012.

1. Introduction

Since the beginning of the Industrial Revolution, human activities, particularly burning of fossil fuels and clearing of forested land, have made a much “thicker” blanket of greenhouse gas around the earth, resulting in an unprecedented increase in the current level of CO₂, at around 390 ppm,¹ compared with only 280 ppm before the Industrial Revolution. Although the 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, and numerous long-term changes have already been observed, including extreme weather such as droughts, heavy precipitation, heat waves, and intensive tropical cyclones. The Intergovernmental Panel on Climate Change (IPCC 2007) projects that the

global temperature will increase 1.1 to 6.4°C over the coming century, at a magnitude that is much more rapid than any changes that have occurred in the past 10,000 years.

Facing such dramatic climate risks, the recent 194-nation United Nations–led Copenhagen Climate Summit pledged to limit global warming to 2°C along with billions of dollars in climate financing in the famous “Copenhagen Accord” (United Nations Framework Convention on Climate Change 2009). Up to now, 138 countries (representing 86.76% of global emissions) have made commitments to the accord. Even though the Copenhagen Accord does not have any legal standing itself, many major countries have announced their own national carbon reduction targets for the year 2020. For example, the United States and Canada have set a carbon emission target of 17% below the 2005 level, the European Union plans to commit a 20% reduction from the 1990 levels, China committed a 2020 carbon intensity target of 40%–45% below the 2005 level, India pledged to reduce its carbon intensity 20%–25% below the 2005 levels, and

¹The concentration of atmospheric CO₂ for April 2011 is 393.18 ppm. Readers may refer to <http://www.co2now.org/> for the trajectory of atmospheric CO₂ concentration in the past half century.

Japan plans to cut its emissions by 25% compared to the 1990 level.² However, many climate experts argue that these voluntary national targets are far from enough to avert the climate risks.

From an economic point of view, it is very difficult to evaluate (1) how much warming and climate change will occur, (2) how bad will it get, (3) when will it occur, and (4) what carbon targets or abatement efforts should be adopted to avert the likely irreversible climate catastrophic events in the future. Climate change is complicated; for many years scientists and economists typically work together and resort to numerical computer simulations, so-called “integrated assessment models” (IAM), for evaluating the economic impacts of global warming. Sir Nicholas Stern prepared a famous 700-page *Stern Review on the Economics of Climate Change* for the British government in 2005. Based on the numerical simulation using an IAM model, called the PAGE model in his document, he called for “urgent action” against global warming to prevent economic damages that could rival that of the World Wars and the Great Depression. His cost-benefit analysis also shows that spending only 1%–2% of the global gross domestic product (GDP) now could save the future potential climate loss at a magnitude of around 20% of the global GDP (Stern 2007). No matter how much praise and criticism has been attracted by this modeling exercise, this widely known Stern Review has nudged people to talk and think more seriously about climate change.

Another famous economist, William Nordhaus, provided another IAM model, called the Dynamic Integrated Model of Climate and the Economy (DICE model). The DICE model was first developed in 1974, and its current version, DICE-2007, is the outcome of the fifth major revision.³ Its basic modeling philosophy is “to incorporate the latest economic and scientific knowledge and capture the major elements of the economics of climate change in as simple and transparent a fashion as possible” (Nordhaus 2008, p. xii). The DICE model has two interrelated modules, the economic module and the geophysical module. In the economic module, it models the economy from the perspective of neoclassical economic growth theory, where the economy makes investment in capital, education, and technology, as well as emission reduction today to increase consumptions in the future. In the geophysical module, it models carbon

cycles among three reservoirs, the atmosphere, the biosphere and the deep oceans, and climate damages due to the emissions. The model is calibrated using the latest economic data as well as scientific knowledge. It conducts a cost-benefit analysis and outputs a net present value of abatement costs plus climate damages. Nordhaus (2008) used the DICE model to evaluate a number of different environmental policies, including limiting the temperature increase or the CO₂ concentration increase from the preindustrial level, the Kyoto Protocol, the Stern Review, Al Gore’s suggestion, as well as the hypothetical optimal policy, and the policy of no controls. Nordhaus then concluded that a moderately tight climate limit, e.g., a 2°C or 3°C increase in temperature or a 2× or 2.5× the preindustrial level of CO₂ concentration, is much more cost effective than aggressive climate policies, such as a 1.5°C increase in temperature, a 1.5 × CO₂ concentration, as well as the policies of Stern and Gore.

Like climate change itself, climate models also concern profound uncertainties, partly because of its intrinsic randomness and partly because of our imperfect understanding of the evolution of complex climate systems. In IAM models, these uncertainties appear to rely on a wide range of key uncertain parameters at various stages of global warming modeling. A typical approach to handling these uncertainties is to apply Monte Carlo methods (see, for instance, Mattoo et al. 2009, Webster et al. 2009). For the DICE model, Nordhaus (2008) pointed out that there are eight critical uncertain parameters in the model, including the rate of growth of total factor productivity, the rate of decarbonization, the equilibrium temperature-sensitivity coefficient, the damage parameter, the price of backstop technology, asymptotic global population, the transfer coefficient in carbon cycle, and total resources of fossil fuels. He modeled all eight parameters as normal random variables and estimated their means and variances, and he then evaluated the expected performances of different policies by generating the parameters from their distributions and running simulation experiments with these parameters.

Although such a Monte Carlo approach is easy to implement, specifying the distributions (or a joint distribution) for those uncertain parameters turns out to be a tricky issue for the modelers and policy makers. In the climate change modeling,

- there is often lack of enough data to estimate the variances of the uncertain parameters accurately;
- typically, very little information is available to estimate the correlations among the uncertain parameters;
- compared to the covariances, the means of the uncertain parameters are relatively easier to estimate

² See <http://www.usclimatenetwork.org/policy/copenhagen-accord-commitments> (last accessed August 1, 2012) for the responses of different countries to the Copenhagen Accord.

³ The DICE-2007 model as well as the related papers, documentation, and discussions are available for public use at <http://nordhaus.econ.yale.edu/DICE2007.htm>. In this paper, we focus on the DICE model because of its availability.

or specify. However, it is observed that in many well-studied random systems, the outputs are generally more sensitive to the means of random inputs than to the covariances of random inputs.⁴

It is widely known that ignoring estimation errors may lead to inaccurate simulation results 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 it (implicitly) assumes that the parameters are independent (Nordhaus 2008). However, such treatments are often problematic because the parameters generally work together to affect the outcomes of the models, and moreover, the independence has not been validated and the parameters are often believed to admit certain dependence. In the DICE model, for instance, the price of backstop technology may be correlated to a number of other parameters such as the damage parameter and the total resources of fossil fuels. It is well known in many literatures that ignoring correlations can introduce significant modeling errors; see, for instance, Clemen and Reilly (1999) and Das et al. (2001) for some recent references. In our study, we use the term “ambiguity” to describe the fact that the mean vector and the covariance matrix cannot be specified accurately, and we study how to estimate the performances robustly under the ambiguity.

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, which finds the worst-case performances among all MVN distributions constrained by the ambiguity sets to evaluate a policy. The use of worst-case analysis to account for ambiguity has a deep root in economics.

Ellsberg (1961) called the incomplete information in specifying probability distributions “ambiguity” and argued that it may be appealing to a conservative person to consider the worst of all reasonable distributions. Epstein (1999) formally defined “ambiguity aversion” and showed that it is supported by empirical evidence. Furthermore, it is worthwhile to note that we do not argue or suggest that policy makers should only consider worst-case performances when making policies. Indeed, we believe that they should take into consideration a wide range of objectives when making global warming policies. The worst-case analysis suggested in this paper provides policy makers another source of valuable information and makes them fully aware of the potential risks of different policies.

To apply the robust approach, we need to find the worst-case performance of the environmental simulation model in the ambiguity sets. However, this is not a simple task. To accurately evaluate the performance of the model at a given MVN distribution, we often need to run multiple (often a large number of) replications of the time-consuming simulation experiments. To find the worst-case performance, we often have to evaluate the performances of the model for a large number of MVN distributions with different mean vectors and covariance matrices. This requires a prohibitively large amount of computational time and is often not feasible in practice. Furthermore, even if it is feasible to conduct the performance evaluations, to find the worst-case performance requires solving a semidefinite simulation optimization problem, which is not yet studied in the current literature of simulation optimization; see, for instance, the recent reviews of Fu (2002) and Hong and Nelson (2009). To resolve this difficulty, we use a change-of-measure technique that uses the simulation results conducted at a given MVN distribution to infer simulation results for all MVN distributions constrained by the ambiguity sets. This technique has two advantages. First, it significantly reduces the computational effort required to find the worst-case performance, because simulation experiments only need to be conducted at an MVN distribution. Second, it separates the optimization part and the simulation part, and it allows us to apply mature optimization techniques to solve the newly formulated stochastic optimization problem.

To solve the stochastic optimization problem, we use a sample-average approximation that takes a sample of simulation observations and reformulates the optimization problem into a deterministic semidefinite optimization problem. However, this problem is nonconvex, because the density function of an MVN distribution is not a concave function of the mean vector and covariance matrix and therefore is

⁴For instance, in portfolio optimization literature, Chopra and Ziemba (1993) empirically showed that the impact of estimation errors in asset mean returns is approximately an order of magnitude greater than the corresponding impact of estimation errors in asset variances or covariances.

difficult to solve by available semidefinite programming (SDP) solvers. To solve this problem, we use an iterative approach. In each iteration, we solve a convex quadratic program to update the mean vector and solve a log-determinant maximization problem to update the covariance matrix. Note that both the convex quadratic programs and log-determinant maximization problems are well-known convex optimization problems that can be solved efficiently by using existing solvers (Grant and Boyd 2009). 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 types of performance measures, the net present values of the abatement costs plus the climate damages, the global temperature increases by 2105, and the probabilities that the global temperature increase by 2105 is 2°C, 3°C, or 4°C or higher than the preindustrial level. We consider five representative global warming policies, including

- no climate controls for the next 250 years, denoted as “no controls policy”;
- limiting the atmosphere CO₂ concentration to two times of the preindustrial level, denoted as “2 × CO₂ policy”;
- limiting the global temperature increase to 2°C from the preindustrial level proposed by the Copenhagen Accord, denoted as “2°C increase policy”;
- the policy recommended by the Stern Review, which is an optimal cost-benefit climate change policy with a very low discount rate (i.e., it uses the Ramsey discount formula and has an average discount rate of about 1.4%), denoted as “Stern review”; and
- Al Gore’s proposal that aims to reduce the global CO₂ by 90% by 2,050, denoted as “Gore proposal.”

We find that the various performances of the DICE model are generally robust with respect to the variances of the uncertain parameters and the worst-case performances are close to the respective average performances, given the condition that the parameters are independent. However, the ambiguities in the means and covariances appear to have significant impacts on the performances, and the worst-case performances often change significantly from the average performances. Nevertheless, these changes in general do not alter our preferences on different policies, i.e., the rankings of the policies are robust with respect to the ambiguities in both the mean vector and the covariance matrix. Coming to different policies, we find that, when the ambiguities are taken into consideration, the cost ineffectiveness of the aggressive policies, e.g., the ones of Stern and Gore, is enlarged significantly, and the risks of the no controls policy are beyond the acceptable levels. Therefore, we prefer the moderately tight policies, e.g., the 2 × CO₂ policy and

the 2°C increase policy. These policies incur moderate costs but keep the risks of global warming under control. Between the two moderately tight policies, we find that the 2 × CO₂ policy is more cost-effective, but has a large probability of resulting in a 2°C global temperature increase, and a moderate probability of resulting in a 3°C temperature increase under the worst cases.

Even though robust environmental modeling is itself a very important area, the contributions of this paper are beyond it. MVN distribution is the most widely used multivariate distribution in practice because its simple covariance structure makes it easy to model dependence among uncertain parameters. For instance, it is used in financial simulations to model correlated returns (Glasserman 2004) and in supply chain studies to model correlated demands (Netessine and Rudi 2003). Even when marginal distributions of individual parameters are not modeled as normal distributions, the dependence among the parameters may still be modeled and simulated by an MVN distribution (e.g., the normal-to-anything approach of Cario and Nelson 1997). Because the parameters of an MVN distribution are often specified with ambiguity (e.g., estimation error), our approach may be applied to these situations to improve the robustness of simulation studies.

Besides environmental modeling, our research is related to two other literatures: the simulation input modeling literature and the robust optimization literature. In the simulation input modeling literature, input-parameter ambiguity (known as a type of input-model uncertainty in the literature) has been studied; see, for instance, Henderson (2003) for a review. Among different methods proposed in this literature, Barton and Schruben (2001) discussed the resampling method, Cheng and Holland (1997) applied a first-order approximation based on sensitivity analysis, and Chick (2001) studied Bayesian model averaging. Although these approaches are conceptually applicable to the problem that we study in this paper, their implementations may not be easy. For instance, the resampling method requires input data that may not be available in our problem, the first-order approximation method requires a sensitivity analysis on the covariance matrix that is yet to be studied (Fu 2008), and the Bayesian model averaging approach requires a prior distribution on the mean vector and covariance matrix that may not be easy to specify.

The robust optimization literature considers problems where the parameters of an optimization problem are constrained in an uncertainty set and it often tries to find a solution that optimizes the objective function for the worst-case combination of parameters in the uncertainty set. For instance, Ben-Tal and Nemirovski (2000) and Bertsimas and

Sim (2004) considered linear programming problems where the uncertain parameters are bounded and symmetric, Ben-Tal and Nemirovski (1998) considered convex optimization problems where the uncertainty set is ellipsoidal, and Delage and Ye (2010) considered convex optimization problems where the uncertainty set is described by the moments of the uncertain parameters. In the robust optimization literature, robustness is often referred to as “distributional robustness,” which considers the worst-case distribution in the uncertainty set. Its basic approach is to approximate the problem as a computationally tractable deterministic optimization problem, often by using probabilistic inequalities. In our approach, however, robustness is referred to as “parameter robustness,” which considers only the ambiguity in specifying the mean vector and covariance matrix in an MVN distribution. We adopt this approach because of two reasons. First, MVN distribution is a reasonable distribution to describe the dependent uncertain parameters. Second, the distributionally robust approach requires an explicit function form of the objective function. In our problem, however, the objective function is imbedded in a complex integrated climate-economic simulation model, and it can only be estimated by running simulation experiments. Therefore, the distributionally robust approach may not be applicable to our problem. However, how to evaluate distributional robustness of environmental simulation models is certainly an interesting problem for future research.

The rest of this paper is organized as follows. In §2, we introduce the problem formulation and the change-of-measure reformulation. In §3, we develop a sequential algorithm to solve the optimization problem formulated in §2 and analyze its properties. We study the robustness of the DICE model and provide some support for global warming policy making in §4. In the electronic companion to this paper (available at <http://www.ielm.ust.hk/dfaculty/hong/>), we test the algorithm on a number of test problems and include all the proofs and extra numerical results on the DICE model.

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 an 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, the temperature increase, or the atmosphere carbon concentration. 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 experiments at ξ . In the simulation study, we are interested in estimating $E_{(\mu, \Sigma)}[c(\xi)]$, the expected value of the simulation output when the mean vector and covariance matrix of the MVN distribution are μ and Σ .

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

$$\max_{(\mu, \Sigma) \in \mathcal{M} \times \mathcal{S}} E_{(\mu, \Sigma)}[c(\xi)]. \quad (1)$$

Without loss of generality, we further assume that $c(\xi)$ is nonnegative for all $\xi \in \mathfrak{N}^d$, as in the 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).

2.2. The Ambiguity Sets

There has been extensive research on specifying the ambiguity sets of mean vector and covariance matrix for random parameters in robust optimization literature; see, e.g., Bertsimas et al. (2011) and Fabozzi et al. (2010) for a thorough review. To model the ambiguity of (μ, Σ) , we consider different scenarios that may be faced by environmental modelers. Although it is natural to consider the ambiguities of the mean vector and covariance matrix simultaneously, in this section we shall analyze them separately because of their different characteristics.

2.2.1. The Ambiguity Set of the Mean Vector.

We first consider the ambiguity set \mathcal{M} of the mean vector μ . As shall be seen, the approach we propose in §3 can handle any convex and compact ambiguity set $\mathcal{M} \subset \mathfrak{N}^d$. Here we introduce the two most common scenarios. In the first scenario, suppose there exists an independent and identically distributed (i.i.d.) sample of ξ , denoted as $\{\xi_1, \dots, \xi_m\}$. Because ξ follows an MVN distribution, based on the Hotelling’s T^2 -statistic, the $1 - \alpha$ confidence region of μ has the following ellipsoid shape (Anderson 1984):

$$\mathcal{M}_1 = \{\mu: (\mu - \bar{\mu})^T \bar{\Sigma}^{-1} (\mu - \bar{\mu}) \leq \gamma\},$$

where $\bar{\mu} = \frac{1}{m} \sum_{i=1}^m \xi_i$ and $\bar{\Sigma} = \frac{1}{m-1} \sum_{i=1}^m (\xi_i - \bar{\mu})(\xi_i - \bar{\mu})^T$ are the estimates of μ and Σ , respectively, and $\gamma > 0$ is a given constant that depends on $1 - \alpha$. Note that \mathcal{M}_1 is a convex and compact set of μ . It is also considered in Delage and Ye (2010).

In the second scenario, there is no sufficient data available to estimate μ directly. We let

$$\mathcal{M}_2 = \{\mu: \mu_l \leq \mu \leq \mu_u\},$$

where $\mu_l, \mu_u \in \mathbb{R}^d$ are the lower and upper bounds of μ . Note that \mathcal{M}_2 takes the form of a hyper-box and it is also a convex compact set of μ . The hyper-box region for mean vector is also considered in Halldórsson and Tütüncü (2003) and Goldfarb and Iyengar (2003), among others. Clearly, the box region is more suitable than the ellipsoid region when the mean vector is determined subjectively or judgementally, as in the DICE model (Nordhaus 2008).

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 semidefinite 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 \mathcal{S} is a compact subset of \mathbb{S}_{++}^d . Following the convention of optimization literature, we use $A \succeq 0$ (respectively, $A \succ 0$) to denote that $A \in \mathbb{S}_+^d$ (respectively, $A \in \mathbb{S}_{++}^d$), and for any $A, B \in \mathbb{S}^d$, use $A \preceq B$ (respectively, $A \succeq B$) to denote that $B - A \succeq 0$ (respectively, $A - B \succeq 0$).

Let $X = \Sigma^{-1}$ and $\mathcal{X} = \{X: \Sigma \in \mathcal{S}\}$. Note that X is also known as the precision matrix. Because we directly handle X instead of Σ in the algorithm that we propose in §3, we need to carefully define \mathcal{S} , the ambiguity set of Σ , so that \mathcal{X} is semidefinite representable (Ben-Tal and Nemirovski 2001) and can be solved by common matrix optimization solvers.

In this subsection, we introduce three different forms of \mathcal{S} to model the ambiguous covariance matrix under different scenarios. In the first form, we let

$$\mathcal{S}_1 = \{\Sigma: \underline{\sigma}_i^2 \leq \Sigma_{ii} \leq \bar{\sigma}_i^2, \Sigma_{ij} = 0, i \neq j, i, j = 1, \dots, d\},$$

where Σ_{ii} is the variance of the i th element of ξ . This form can be applied in scenarios where the random parameters are believed to be independent and the variances are estimated within some intervals. Under the inverse transformation, \mathcal{S}_1 becomes

$$\begin{aligned} \mathcal{X}_1 &= \{X: \text{diag}\{1/\bar{\sigma}_1^2, \dots, 1/\bar{\sigma}_d^2\} \leq X \\ &\leq \text{diag}\{1/\underline{\sigma}_1^2, \dots, 1/\underline{\sigma}_d^2\}, \\ &X_{ij} = 0, \quad i \neq j, i, j = 1, \dots, d\}, \end{aligned}$$

where $\text{diag}\{\sigma_1^2, \dots, \sigma_d^2\}$ denotes a $d \times d$ matrix with diagonal elements being $\sigma_1^2, \dots, \sigma_d^2$ and off-diagonal

elements being zero. Note that \mathcal{X}_1 is convex, compact and semidefinite representable, and thus can be handled effectively.

In the second form, we let

$$\mathcal{S}_2 = \{\Sigma: \gamma_1 \bar{\Sigma} \preceq \Sigma \preceq \gamma_2 \bar{\Sigma}\},$$

where $\gamma_1 < \gamma_2$ are positive constants and $\bar{\Sigma}$ is a positive definite matrix that can be an estimate of Σ . It is easy to see that \mathcal{S}_2 can be transformed to

$$\mathcal{X}_2 = \{X: \gamma_2^{-1} \bar{\Sigma}^{-1} \leq X \leq \gamma_1^{-1} \bar{\Sigma}^{-1}\},$$

and \mathcal{X}_2 is a convex compact subset of \mathbb{S}_{++}^d . The matrix interval form \mathcal{S}_2 can model a number of scenarios. First, when there exists an i.i.d. sample of ξ , denoted as $\{\xi_1, \xi_2, \dots, \xi_m\}$, we may let $\bar{\Sigma}$ be the sample covariance matrix. Because ξ follows an MVN distribution with mean μ and covariance matrix Σ , $\bar{\Sigma}$ follows a Wishart distribution (Anderson 1984). Then, a $1 - \alpha$ confidence region of Σ is defined as \mathcal{S}_2 with γ_1 and γ_2 being some positive constants depending on α . Second, consider a scenario where the covariance ambiguity is modeled using a matrix norm. In this scenario, $\bar{\Sigma} \succ 0$ is some nominal estimate of the covariance matrix. Goldfarb and Iyengar (2003) suggest to model the ambiguity using the set

$$\{\Sigma: \|\bar{\Sigma}^{-1/2}(\Sigma - \bar{\Sigma})\bar{\Sigma}^{-1/2}\|_2 \leq \varepsilon, \Sigma \succeq 0\},$$

where $\|A\|_2$ denotes the spectral norm of a matrix A , and ε is a parameter controlling the size of the neighborhood. This set restricts the perturbation of the covariance matrix to be within a neighborhood of $\bar{\Sigma}$ defined by the spectral norm, and it may be used to model the situation where there is little data available to estimate the covariance matrix, as frequently encountered in the global warming context. It follows from Ben-Tal and Nemirovski (2001) that the above set can be rewritten as $\{\Sigma: (1 - \varepsilon)\bar{\Sigma} \preceq \Sigma \preceq (1 + \varepsilon)\bar{\Sigma}\}$, which is also in the form of \mathcal{S}_2 , provided $\varepsilon < 1$.

In the third form, we let

$$\mathcal{S}_3 = \{\Sigma: \Sigma_{ii} \leq \sigma_i^2, i = 1, \dots, d, \Sigma \succeq \varepsilon \bar{\Sigma}\},$$

where $\sigma_i^2, i = 1, \dots, d$, and $\varepsilon > 0$ are constants and $\bar{\Sigma}$ is an estimate of the covariance matrix. Compared to \mathcal{S}_2 , \mathcal{S}_3 removes the upper bound on the entire covariance matrix while replacing it by adding upper bounds on the individual variances. It can be used in scenarios where variances may be estimated but there is little information (or data) available to estimate the correlations among the random parameters, as in the DICE model. Furthermore, it is worthwhile to note that ε is a meaningful parameter in \mathcal{S}_3 . First, it measures the deviation of the lower bound from the nominal estimate in spectral norm. Second, it also

affects the correlations among the random parameters (e.g., a smaller ε gives the correlations more freedom to change). Therefore, one may set ε small if there is no explicit information on the dependence, and vice versa. Because $\Sigma_{ii} \leq \sigma_i^2$ 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 i th column of a $d \times d$ identity matrix, \mathcal{S}_3 can be transformed to

$$\mathcal{X}_3 = \left\{ X: \begin{bmatrix} X & e_i \\ e_i^T & \sigma_i^2 \end{bmatrix} \succeq 0, i = 1, \dots, d, X \leq \varepsilon^{-1} \bar{\Sigma}^{-1} \right\},$$

which is semidefinite representable.

In some scenarios, it may be more natural to constrain all the elements of the covariance matrix, i.e., one may set the ambiguity set as $\mathcal{S} = \{\Sigma: \Sigma_i \leq \Sigma \leq \Sigma_u, \Sigma \geq 0\}$, where the inequality constraints \leq are element-wise. Unfortunately, however, this ambiguity set cannot be transformed to a convex set of X . Therefore, problem (1) with this ambiguity set on Σ may not be solved by the algorithm proposed in §3. It is worthwhile noting that this is a problem that is encountered not only by us but also by many others in analyzing MVN distributions. A recent trend is to directly parameterise MVN distributions in terms of the precision matrix X rather than the covariance matrix Σ (see, for instance, Bernardo and Smith 2000). Ravikumar et al. (2008) show that it is also often convenient to directly estimate precision matrices and build confidence regions.

2.3. A Change-of-Measure Reformulation

Problem (1) is difficult to solve because the closed form of $E_{(\mu, \Sigma)}[c(\xi)]$ is not known. Because the objective function of problem (1) may be evaluated by running simulation experiments at any $(\mu, \Sigma) \in \mathcal{M} \times \mathcal{S}$, one approach is to treat it as a simulation optimization problem (e.g., Fu 2002, Hong and Nelson 2009). 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 \mathfrak{R}^d such that $f(x) > 0$ for any $x \in \mathfrak{R}^d$, and let $E_f(\cdot)$ denote the expectation taken with respect to the distribution $f(\cdot)$. Let

$\phi_{(\mu, \Sigma)}(\cdot)$ denote the density of the MVN distribution with mean μ and covariance matrix Σ . Then,

$$\begin{aligned} E_{(\mu, \Sigma)}[c(\xi)] &= \int_{\mathfrak{R}^d} c(x) \phi_{(\mu, \Sigma)}(x) dx \\ &= \int_{\mathfrak{R}^d} c(x) \frac{\phi_{(\mu, \Sigma)}(x)}{f(x)} f(x) dx \\ &= E_f \left[c(\xi) \frac{\phi_{(\mu, \Sigma)}(\xi)}{f(\xi)} \right], \end{aligned} \tag{2}$$

where ξ has the distribution $f(\cdot)$ in the last equation. The technique used in deriving Equation (2) 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 (2), we may reformulate problem (1) as

$$\begin{aligned} \max_{(\mu, \Sigma) \in \mathcal{M} \times \mathcal{S}} E_f &\left[(2\pi)^{-d/2} \frac{c(\xi)}{f(\xi)} (\det \Sigma)^{-1/2} \right. \\ &\left. \cdot \exp \left\{ -\frac{1}{2} (\xi - \mu)^T \Sigma^{-1} (\xi - \mu) \right\} \right], \end{aligned} \tag{3}$$

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

Note that, without the change of measure, the decision variables μ and Σ are in the distribution of ξ . Therefore, to evaluate $E_{(\mu, \Sigma)}[c(\xi)]$ 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 $E_{(\mu, \Sigma)}[c(\xi)]$ for all $(\mu, \Sigma) \in \mathcal{M} \times \mathcal{S}$. This greatly reduces the required computational effort and makes it possible to solve problem (1) efficiently.

3. Solution Method

Following the notation used in §2.2, we let $X = \Sigma^{-1}$ and $\mathcal{X} = \{X \in \mathbb{S}_{++}^d: X^{-1} \in \mathcal{S}\}$, where \mathcal{S} is the ambiguity set for the covariance matrix Σ . In this paper, we only consider the cases where \mathcal{X} is convex compact in \mathbb{S}_{++}^d , as the cases discussed in §2.2. Then, problem (3) can be reformulated as follows:

$$\begin{aligned} \max_{(\mu, X) \in \mathcal{M} \times \mathcal{X}} g(\mu, X) &:= E_f \left[(2\pi)^{-d/2} \frac{c(\xi)}{f(\xi)} \right. \\ &\left. \cdot \exp \left\{ -\frac{1}{2} (\xi - \mu)^T X (\xi - \mu) + \frac{1}{2} \log \det X \right\} \right]. \end{aligned} \tag{4}$$

In the rest of this section, we develop an efficient algorithm to solve problem (4).

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 i.i.d. sample of ξ , denoted as ξ_1, \dots, ξ_n , generated from the distribution $f(\cdot)$, and the i.i.d. simulation outputs $c(\xi_1), \dots, c(\xi_n)$. We can then estimate $g(\mu, X)$ by

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

for any $(\mu, X) \in \mathcal{M} \times \mathcal{X}$. Let $a_j = (2\pi)^{-d/2} c(\xi_j) / f(\xi_j)$. Note that $a_j \geq 0$ for any $j = 1, \dots, n$, because in §2 we assumed that both $f(\xi) > 0$ and $c(\xi) \geq 0$ for all $\xi \in \mathbb{R}^d$. We suggest solving the following optimization problem:

$$\begin{aligned} & \max_{(\mu, X) \in \mathcal{M} \times \mathcal{X}} \bar{g}_n(\mu, X) \\ & := \frac{1}{n} \sum_{j=1}^n a_j \exp \left\{ -\frac{1}{2} (\xi_j - \mu)^T X (\xi_j - \mu) + \frac{1}{2} \log \det X \right\}. \end{aligned} \quad (5)$$

This approach is known as the sample-average approximation (SAA). It has been studied widely in the stochastic programming literature (see, for instance, Shapiro and Homem-de-Mello 2000, Shapiro et al. 2009).

Intuitively, if the sample size n is large enough, the solution to problem (5) should be close to the solution to problem (4). Therefore, it provides a good approximation to the solution of problem (4). In the appendix, we show that the intuition indeed holds under some mild conditions. In the rest of this section, we discuss how to solve problem (5) efficiently.

3.2. Sequential Quadratic and Maxdet Programs (SQMP)

As can be seen, problem (5) may not be a convex optimization problem. This leads to a significant difficulty in solving it. However, the problem has two special properties that are worthwhile noting. First, for the environmental problem we are interested in, there typically exists an initial feasible solution (μ_0, X_0) that may be the current estimate of (μ, Σ^{-1}) . Second, the function $-\frac{1}{2} (\xi_j - \mu)^T X (\xi_j - \mu) + \frac{1}{2} \log \det X$ is a biconcave function of μ and X , i.e., it is concave with respect to μ for every $X \in \mathcal{X}$ and concave with respect to X for every $\mu \in \mathcal{M}$. Therefore, the objective function $\bar{g}_n(\mu, X)$ in problem (5) is a summation of quasi-biconcave functions. This motivates us to consider an iterative approach to solving the problem.

In §3.2.1, we first discuss how to improve (μ_0, X_0) by changing only one of the two arguments at a time. We show that, after some transformations, the resulted two optimization problems are both standard convex optimization problems that can be solved easily by commonly used optimization solvers. In §3.2.2, we show how to iteratively solve the two problems developed in §3.2.1 to keep improving solution qualities, and we show that the resulted algorithm (called Algorithm SQMP) leads to a stationary point of problem (5).

3.2.1. Improving (μ_0, X_0) . To make use of the quasi-biconcave structure, we introduce an intermediate decision vector $y = (y_1, \dots, y_n)^T$ and reformulate problem (5) as the following problem:

$$\begin{aligned} & \max \quad \frac{1}{n} \sum_{j=1}^n a_j y_j \\ & \text{s.t.} \quad \log y_j \leq -\frac{1}{2} (\xi_j - \mu)^T X (\xi_j - \mu) \\ & \quad \quad + \frac{1}{2} \log \det X, \quad j = 1, \dots, n, \\ & \quad \quad (\mu, X, y) \in \mathcal{M} \times \mathcal{X} \times \mathbb{R}^n, \end{aligned} \quad (6)$$

where we define $\log x = -\infty$ if $x \leq 0$ to make it an extended real-valued function. Note that the newly defined $\log x$ is still a concave function of x as the epigraph of $-\log x$ defines a convex set (Rockafellar 1970). The above reformulation separates the terms in the objective function in problem (5) and embeds the quasi-biconcave structure into constraints. From the formulation, it is clear that the objective function value $(1/n) \sum_{j=1}^n a_j y_j$ of problem (6) at any feasible point (μ, X, y) is no greater than $\bar{g}_n(\mu, X)$, and moreover, the constraints defined by (7) are tight at the optimal solution of problem (6).

Suppose (μ_0, X_0) is a feasible solution of problem (5), i.e., $(\mu_0, X_0) \in \mathcal{M} \times \mathcal{X}$. It follows from constraint (7) that if we fix (μ, X) at (μ_0, X_0) , to obtain the optimal objective value of problem (6), we should set y_j as the values

$$s_j = \exp \left\{ -\frac{1}{2} (\xi_j - \mu_0)^T X_0 (\xi_j - \mu_0) + \frac{1}{2} \log \det X_0 \right\}$$

for all $j = 1, \dots, n$. It is immediately seen that $s_j > 0$ for all $j = 1, \dots, n$ and

$$\bar{g}_n(\mu_0, X_0) = \frac{1}{n} \sum_{j=1}^n a_j s_j.$$

Suppose now we fix $\mu = \mu_0$ and only change X and y to improve the objective function of problem (6). We linearize $\log y_j$ by a first-order Taylor

expansion at $y_j = s_j$ and then solve the following optimization problem:

$$\begin{aligned} \max \quad & \frac{1}{n} \sum_{j=1}^n a_j y_j \tag{8} \\ \text{s.t.} \quad & \log s_j + \frac{1}{s_j} (y_j - s_j) \leq -\frac{1}{2} (\xi_j - \mu_0)^T X (\xi_j - \mu_0) \\ & + \frac{1}{2} \log \det X, \quad j = 1, \dots, n, \\ & (X, y) \in \mathcal{X} \times \mathfrak{R}^n. \end{aligned}$$

Because $\log y_j$ is a concave function, we have

$$\log y_j \leq \log s_j + \frac{1}{s_j} (y_j - s_j) \quad \text{for any } y_j \in \mathfrak{R}.$$

Therefore, for any feasible solution (X, y) of problem (8), (μ_0, X, y) is also a feasible solution of problem (6). Furthermore, note that (X_0, s_1, \dots, s_n) is a feasible solution of problem (8). Therefore, the optimal solution of problem (8) is at least as good as (X_0, s_1, \dots, s_n) . Therefore, we may solve problem (8) to find an improved solution of problem (5).

Now we show how to solve problem (8) efficiently. Let $(X_1, y_1^1, \dots, y_n^1)$ denote an optimal solution of the problem. Then, we must have

$$\begin{aligned} y_j^1 = & \frac{1}{2} s_j (- (\xi_j - \mu_0)^T X_1 (\xi_j - \mu_0) + \log \det X_1) \\ & + s_j (1 - \log s_j), \quad j = 1, \dots, n. \end{aligned}$$

Substituting $y_j^1, j = 1, \dots, n$ into the formulation of problem (8) and noting that

$$(\xi_j - \mu_0)^T X (\xi_j - \mu_0) = \text{tr}((\xi_j - \mu_0)(\xi_j - \mu_0)^T X),$$

where $\text{tr}(A)$ denotes the trace of a matrix A , we have that X_1 is an optimal solution of the following problem:

$$\max_{X \in \mathcal{X}} \{-\text{tr}(\Pi X) + \log \det X\}, \tag{9}$$

where

$$\Pi = \frac{\sum_{j=1}^n a_j s_j (\xi_j - \mu_0)(\xi_j - \mu_0)^T}{\sum_{j=1}^n a_j s_j}.$$

Note that when \mathcal{X} is semidefinite representable, problem (9) is known as a log-determinant maximization (abbreviated as maxdet) problem, which is a well-known convex SDP 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 2011). Recently, Wang et al. (2009) developed a new method capable of solving efficiently large-scale maxdet problems where the size of the matrix can be as high as $2,000 \times 2,000$. For the DICE model we consider in this paper, the size of

the covariance matrix is only 8×8 . According to our experience, a maxdet problem of this size can typically be solved within five seconds by using CVX on a personal computer. By solving problem (9), we can easily obtain X_1 .

Following the same procedure, we may fix $X = X_0$ and solve

$$\max_{\mu \in \mathcal{M}} \{-\frac{1}{2} \mu^T X_0 \mu + V^T X_0 \mu\} \tag{10}$$

with

$$V = \frac{\sum_{j=1}^n a_j s_j \xi_j}{\sum_{j=1}^n a_j s_j}$$

to find μ_1 , and we can show that (μ_1, X_0) is at least as good as (μ_0, X_0) . Note that problem (10) is a standard convex quadratic program. It can be solved very efficiently by commonly used convex optimization solvers (Boyd and Vandenberghe 2004).

3.2.2. Algorithm SQMP and Its Properties. We have shown that we can improve any solution (μ_0, X_0) by changing one argument at a time. This motivates us to use an iterative algorithm that changes μ and X subsequently in each iteration to find better solutions of problem (5). We call this algorithm *Sequential Quadratic and Maxdet Programs* (SQMP).

Algorithm SQMP

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

Step 1. Let

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

for all $j = 1, \dots, n$. 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 $\mu_k \in \arg \max\{\mu \in \mathcal{M}: -\frac{1}{2} \mu^T X_{k-1} \mu + V_k^T X_{k-1} \mu\}$.

Step 2. Let

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

for all $j = 1, \dots, n$. 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 $X_k \in \arg \max\{X \in \mathcal{X}: -\text{tr}(\Pi_k X) + \log \det X\}$. Go to Step 1.

Note that Algorithm SQMP may start from any $(\mu_0, X_0) \in \mathcal{M} \times \mathcal{X}$. Furthermore, note that we do not specify a stopping criterion. In practice, one may stop the algorithm if the difference between $\bar{g}_n(\mu_k, X_k)$ and $\bar{g}_n(\mu_{k-1}, X_{k-1})$ is within a certain tolerance level or (μ_k, X_k) is a stationary point within a certain tolerance level.

Because Algorithm SQMP solves a sequence of standard convex quadratic programs and maxdet

Downloaded from informs.org by [144.214.42.26] on 02 September 2014, at 22:39 . For personal use only, all rights reserved.

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. The proof of the theorem is based on the proof of a similar result for DC (i.e., difference of convex functions) programs in Hong et al. (2011) and it is included in the electronic companion to this paper.

THEOREM 1. *Suppose that $\{(\mu_k, X_k), k = 1, 2, \dots\}$ is a sequence of solutions generated by Algorithm SQMP when it is applied to solve problem (5) starting from a feasible solution (μ_0, X_0) . Then, it satisfies the following properties:*

1. (μ_k, X_k) is a feasible solution of problem (5), i.e., $(\mu_k, X_k) \in \mathcal{M} \times \mathcal{X}$, for all $k = 1, 2, \dots$.
2. $\{\bar{g}_n(\mu_k, X_k), k = 0, 1, \dots\}$ 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 (5).
4. All cluster points of $\{(\mu_k, X_k), k = 0, 1, \dots\}$ are stationary points of problem (5). Furthermore, if problem (5) has only a finite number of stationary points, then $\{(\mu_k, X_k), k = 0, 1, \dots\}$ converges to a stationary point.

The first property of Theorem 1 shows that the solutions generated by Algorithm SQMP are all feasible solutions of problem (5); therefore, $\bar{g}_n(\mu_k, X_k)$ provides a lower bound on the optimal objective value for all $k = 0, 1, \dots$. The second and third properties show that Algorithm SQMP keeps improving the quality of the solutions unless the current solution is already a stationary point of problem (5). Therefore, the solutions found in later iterations are guaranteed better than the initial solution (μ_0, X_0) , unless (μ_0, X_0) is already a stationary point.

The fourth property of Theorem 1 is the most important property of the four. It shows that Algorithm SQMP indeed has the desired convergence property. Furthermore, it shows that the sequence of solutions generated by Algorithm SQMP converges to a stationary point as long as problem (5) has only a finite number of stationary points. Although problem (5) is not provably 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.

The fourth property also reveals that Algorithm SQMP may not be able to find a global optimal solution of problem (5), especially when there are multiple stationary solutions having higher objective values than the starting solution. This is a drawback of the algorithm. However, given that problem (5) is a nonconvex SDP, this hurdle appears very difficult to overcome. Note that, in environmental modeling, the starting solution is often the best estimate of (μ, Σ) .

The stationary solution found by the algorithm at least provides some information on the robustness of the model locally at the starting solution. Furthermore, if the model is not robust based on the stationary solution, it is clearly not robust based on the global optimal solution. Therefore, Algorithm SQMP may be used to screen out some no-robust models. As we point out in the introduction, environmental policy making is very complicated and often concerns a wide range of objectives. We believe that our algorithm provides information on the robustness of the model that is currently missed and ought to be included in environmental modeling. To further improve the performance of the algorithm, we may apply the algorithm on multiple well-spread starting solutions and use the one with the highest objective value to assess the robustness of the model.

Because problem (5) is a nonconvex SDP, it cannot be solved by commonly used SDP solvers such as CVX and SeDuMi. This is the reason why we develop Algorithm SQMP. Our goal is to develop an algorithm that can solve the problem efficiently using these solvers so that it may be implemented by environmental modelers and policy makers who typically have very limited knowledge on optimization. There may exist other algorithms that are capable of solving problem (5), although we have not found any in the literature. These algorithms nevertheless may require special techniques or tools that are beyond the knowledge of typical environmental modelers and policy makers.

We implement Algorithm SQMP using Matlab and call the CVX package to solve the convex quadratic program and the maxdet problem in each iteration of the algorithm. We apply our algorithm on three test problems on a laptop computer with Intel(R) Core(TM)2 Duo CPU (2.26 GHz, 2.27 GHz) and 4 GB of RAM, and report all the results in the electronic companion to this paper. For all three test problems, our algorithm works well in finding the worst-case performances when there is ambiguity in specifying the MVN distribution. The test results also show that input ambiguity is very important to simulation studies, and ignoring it may lead to unreliable conclusions.

4. Robustness of the DICE Model

In this section, we apply the approach developed in §§2 and 3 to study the robustness of the DICE model on five representative global warming policies: no controls policy, $2 \times \text{CO}_2$ policy, 2°C increase policy, Stern review, and Gore proposal, as introduced in §1. Among these five policies, the first belongs to the class of no control policies, the second and third belong to the class of moderately tight policies, and the last two belong to the class of aggressive policies. Readers

may also refer to Nordhaus (2008) for a more detailed description of these policies.

There are eight uncertain parameters in the DICE model. They are the rate of growth of total factor productivity $g(\text{TFP})$, the rate of decarbonization $g(\text{CO}_2/\text{GDP})$, the equilibrium temperature-sensitivity coefficient $T_2 \times \text{CO}_2$, the damage parameter **Dam-Coeff**, the price of backstop technology **P(back)**, asymptotic global population **Pop**, the transfer coefficient in carbon cycle **CarCyc**, and total resources of fossil fuels **Fossilim**. The means and variances of these parameters are estimated in Table 7-1 of Nordhaus (2008). As pointed out by Nordhaus (2008), these uncertain parameters have their own signs and hence need to be truncated at zero. Furthermore, our numerical experiments indicate that the model cannot handle very small values (< 0.75) of $T_2 \times \text{CO}_2$. Therefore, we truncate the MVN distribution at $(0, 0, 0.78, 0, 0, 0, 0, 0)^T$, where 0.78 is (left) two standard deviations away from the mean of $T_2 \times \text{CO}_2$. This treatment will not cause any significant problems because the probability of the related truncation is only 2.28%.

Let $\hat{\mu}$ and $\hat{\Sigma}$ denote the estimated mean vector and covariance matrix of the eight uncertain parameters provided by Nordhaus (2008). Note that the off-diagonal elements of $\hat{\Sigma}$ are set as zero because Nordhaus (2008) did not estimate the correlations. We generate 25,000 independent scenarios of the eight uncertain parameters from the MVN distribution $N(\hat{\mu}, \hat{\Sigma})$. We then call the GAMS program⁵ of the DICE model (DICE-2007.delta.v8) from Matlab, and simulate the five policies simultaneously for all the 25,000 scenarios of the uncertain parameters. These simulation results are used throughout all the experiments conducted in this section.

To investigate the robustness of the DICE model, it is important to note that the estimates of the mean vector and covariance in Nordhaus (2008) are subjective or judgmental. We quote from Nordhaus (2008, Chap. VII, p. 124) to emphasize this point.

We should pause to describe the nature of the probabilities that are used here. These are not “objective” or “frequentist” probabilities, Rather, they are “subjective” or “judgmental” probabilities, Judgmental probabilities are ones that are held by individuals and are based on formal or informal reasoning about phenomena, rather than solely on observed events.

Therefore, we believe that it is important to understand how robust the results of the DICE model are. We first analyze separately the robustness of the model with respect to the means, variances, and covariances, and see which contribute the most to the robustness

of the model. We then analyze the overall robustness by considering the mean vector and covariance matrix simultaneously, and compare the alternative policies based on their worst-case performances. It is worthwhile noting that the ambiguity sets that we use in all the studies are also subjective and judgmental. However, we believe that they are suitable for understanding the robustness of the DICE model.

In all the experiments reported in this section, we always use the simulation results of the 25,000 scenarios of the random parameters generated from $N(\hat{\mu}, \hat{\Sigma})$. We apply Algorithm SQMP to find the worst-case distributions under various ambiguity sets and resimulate the worst-case performances using 11,000 independent runs at the respective worst-case distributions. Because Nordhaus (2008) has the highest confidence believing in the estimate $(\hat{\mu}, \hat{\Sigma})$, we are especially interested in the robustness of the DICE model at this estimate. Therefore, except otherwise mentioned, we always use $(\hat{\mu}, \hat{\Sigma})$ as the starting solution for Algorithm SQMP.

4.1. Performance Measures

Making global warming policies is an extremely complicated multiobjective problem. There are a great number of performance measures that people may be interested in, and different people may emphasize different aspects. A full discussion of all these performance measures is beyond the knowledge of the authors and clearly out of the scope of this paper. In this paper, we consider only three types of performance measures: (1) the average net present values (NPV, trillions of 2005 U.S. dollars) of abatement costs and climate damages; (2) the average global temperature increases (GTI, °C) by 2,105 from the preindustrial level; and (3) the probabilities that the GTI by 2,105 exceeds 2°C, 3°C, and 4°C. Because the DICE model conducts a standard expected cost-benefit analysis, the first two performance measures are often used. In particular, NPVs are often used to compare different policies. However, a major concern or reservation about the DICE model is perhaps that such a cost-benefit analysis would fail when human beings might suffer an indefinitely large loss from abrupt climate changes, where an abrupt climate change is defined as a large-scale change in the climate system that takes place over a few decades or less, persists (or is anticipated to persist) for at least a few decades, and causes substantial disruptions in human and natural systems (U.S. Climate Change Science Program 2008). In global warming literature, there has been extensive research on the temperature change threshold that may trigger abrupt climate changes. Plentiful investigations support that 2°C above the preindustrial level can be the threshold. For instance, in a recent expert elicitation, significant probability was attached to the possibility

⁵ The GAMS code of the DICE model can be downloaded from <http://nordhaus.econ.yale.edu/DICE2007.htm>.

Table 1 Performance Measures of the DICE Model Under $N(\hat{\mu}, \hat{\Sigma})$

Policy	Avg. NPV	Avg. GTI	P(GTI > 2)	P(GTI > 3)	P(GTI > 4)
No controls	24.64	3.17	0.900	0.574	0.177
2 × CO ₂	20.89	2.37	0.739	0.144	<0.001
2°C increase	27.27	NA	0.000	0.000	0.000
Stern review	38.53	1.54	0.113	0.007	<0.001
Gore proposal	53.67	1.49	0.107	0.000	0.000

of crossing major tipping points in the climate system (that would yield specific abrupt changes) for a warming of above 2°C (Kriegler et al. 2009), and more recently, Bahn et al. (2011) provide a warming target close to 2°C by 2,100 to avoid a particular tipping point in the climate system (namely, a collapse of the Atlantic thermohaline circulation). Nevertheless, there are also considerable studies suggesting that the threshold can be as high as 4°C above the preindustrial level. For instance, the United Kingdom’s Met Office Hadley Centre prepared a map depicting some of the impacts that may occur if the temperature increases by 4°C.⁶ In our study, we report the probabilities of the GTI by 2,105 exceeding 2°C, 3°C, as well as 4°C to measure the risk of abrupt climate changes. It is worthwhile noting that these probabilities are not considered in Nordhaus (2008).

In Table 1 we report the estimated values of the performance measures for the DICE model under the MVN distribution $N(\hat{\mu}, \hat{\Sigma})$. If we only compare the NPVs of various policies, one may *falsely* conclude that, facing global warming, the no controls policy is a competitive policy because its NPV is less than not only the Stern review and Gore proposal but also the 2°C increase policy proposed by the Copenhagen Accord, and only higher than the 2 × CO₂ policy. If the risks of abrupt climate changes are taken into consideration, however, the no controls policy is clearly a dangerous policy. Its temperature increase by 2,105 is very likely to exceed 2°C and has a significant probability of exceeding 4°C, representing a risk level that is too high to be acceptable. The 2 × CO₂ policy has the lowest NPV and a negligible probability of exceeding 4°C, although its probability of exceeding 2°C is very large. The 2°C increase policy of the Copenhagen Accord has a higher NPV than the 2 × CO₂ policy, but its tight temperature control leads to a zero probability of exceeding 2°C. The NPVs of the Stern review and Gore proposal policies appear too high to be competitive compared to the 2 × CO₂ and 2°C increase policies. The results reported in Table 1 are obtained without considering the robustness of the model. They serve as benchmarks when the robustness is considered.

⁶ See <http://www.actoncopenhagen.decc.gov.uk/en/ambition/evidence/4-degrees-map/> (last accessed August 1, 2012).

4.2. Mean Robustness

We begin by studying the mean robustness for the DICE model. Because the estimates of Nordhaus (2008) are subjective or judgemental, to model the ambiguity of the means of the parameters, we first suppose there exists a 10% estimation error for the mean vector and set the ambiguity set as $\hat{\mathcal{M}} = \{\mu: 0.9\hat{\mu} \leq \mu \leq 1.1\hat{\mu}\}$. We fix the covariance matrix $\Sigma = \hat{\Sigma}$ and let the mean vector vary in $\hat{\mathcal{M}}$, and we report the worst-case outputs in Table 2.

To view the change of worst-case performances more explicitly, we define the *sensitivity ratio* (SR) for a given performance measure and a given ambiguity set as

$$SR = \frac{WEP - NEP}{NEP},$$

where WEP denotes the worst-case expected performance when the distribution varies in the given ambiguity set and NEP denotes the nominal expected performance, i.e., the expected performance when the distribution is $N(\hat{\mu}, \hat{\Sigma})$. It is worthwhile noting that the definition of SR is not new. Similar notions have been introduced in the portfolio optimization literature to study the effects of errors in random distributions on optimal portfolio choice (see, for instance, Chopra and Ziemba 1993). Together with the considered ambiguity sets, the SR helps better understand the sensitivity/robustness of the performances.

In Table 2, columns 1, 3, 5, 7, and 9 show the worst-case values of the considered performance measures, and columns 2, 4, 6, 8, and 10 show the corresponding SR values. Our first finding from the simulation study is that the DICE model is indeed sensitive to the means of the uncertain parameters. For the given 10% perturbation of the mean vector of the MVN distribution, the worst-case average NPVs are 30%–60% higher than the current average NPVs for all the policies and the worst-case probabilities also show significant changes for some policies. Specially, we note that the worst-case probability of GTI exceeding 4°C for the no controls policy has almost doubled. The sensitivity of average GTI shows a similar pattern compared to that of the average NPV, except that the changes are not as dramatic. This is to some extent reasonable as the global temperature cannot vary too much. Our second finding is that different performance measures may attain their worst cases at quite different points in the ambiguity sets. For instance, if we consider the mean of the total factor productivity **g(TFP)**, maximizing the NPV pushes it to its lower bound, whereas maximizing the GTI or probabilities pushes it to the upper bound. This is because the NPV consists of both climate damages and abatement costs.

Because the SR values for the 10% perturbation ambiguity set are large, we change the size of the ambiguity set by reducing the perturbation to 5%, and

Table 2 Worst-Case Performance for $\mu \in \hat{\mathcal{M}}$ and $\Sigma = \hat{\Sigma}$

Policy	Avg. NPV	SR (%)	Avg. GTI	SR (%)	P(GTI > 2)	SR (%)	P(GTI > 3)	SR (%)	P(GTI > 4)	SR (%)
No controls	36.58	48.46	3.61	13.88	0.949	5.44	0.752	31.01	0.351	98.31
2 × CO ₂	30.43	45.67	2.56	8.02	0.826	11.77	0.263	82.64	<0.01	NA
2°C increase	42.57	56.11	NA	NA	0.000	NA	0.000	NA	0.000	NA
Stern review	57.07	48.12	1.68	9.09	0.218	92.92	0.017	142.86	<0.001	NA
Gore proposal	71.75	33.69	1.64	10.07	0.254	137.38	0.000	NA	0.000	NA

we report the results in Table 7 in the electronic companion to this paper. Table 7 shows that even for the case of 5% error, the changes are still significant. This further confirms that the DICE model is sensitive to the estimation of the means. We also considered a larger ambiguity set where the perturbation is set as 20%. For this case, as expected, we observed changes that are more significant than the ones reported in Table 2.

4.3. Variance Robustness

As Nordhaus (2008) suggests that the eight uncertain parameters are independent, we first consider this independent case. For this case, it is natural to consider the robustness of the DICE model with respect to the variances of the uncertain parameters. To study the variance robustness, we fix the correlations at the nominal estimate zero and consider the following ambiguity set

$$\hat{\mathcal{P}}_1 = \{ \Sigma : 0.5\hat{\Sigma}_{ii} \leq \Sigma_{ii} \leq 1.2\hat{\Sigma}_{ii}, \Sigma_{ij} = 0, i \neq j, i, j = 1, \dots, 8 \}.$$

As discussed in §2.2.2, $\hat{\mathcal{P}}_1$ models the situation where the uncertain parameters are independent and their variances can vary in certain intervals. Note that, in his estimation of the variances, Nordhaus (2008) has already significantly enlarged some of the variances because he believes that “the empirical estimates of those variances are likely to underestimate the uncertainty” (p. 126). Therefore, to take his adjustments into consideration, we allow the variances to go down more than they can go up in $\hat{\mathcal{P}}_1$. Furthermore, the upper bounds used in $\hat{\mathcal{P}}_1$ may not be further enlarged because larger variances may cause certain truncation problems.

We make robustness analysis for $\hat{\mathcal{P}}_1$ by fixing $\mu = \hat{\mu}$, and we summarize the results in Table 3. We first look at the average NPV and average GTI. From

Table 3, we see that the worst-case performances of these two measures are very close to the average cases. Considering the simulation noise, it is indeed difficult to identify such differences. Note that, for the average NPV of the 2 × CO₂ policy and average GTI of the Gore proposal, we even obtain negative changes, for which an intuitive explanation is that the response surfaces are too flat with respect to the design variables and the estimation errors may sometimes dominate the small differences. The above finding suggests that if the independence assumption of Nordhaus (2008) holds, then the variances of the uncertain parameters may have very little impact on the average NPV and average GTI of the DICE model. It essentially provides a theoretical support to the observation of Nordhaus (2008, p. 136) that “the estimates in the certainty-equivalents model are very close to the estimates in the uncertainty model.”

However, this conclusion does not necessarily hold for all the performance measures. From Table 3, we observe that the probabilities may have substantial changes. An intuitive explanation of this observation is that the GTI may exhibit a certain level of symmetry, so that its expectation may not be affected by the increase of its variance but its tail probabilities may.

We also further enlarge the upper bounds of the variances in $\hat{\mathcal{P}}_1$ to 1.5 times of the nominal estimates (with the risk of some truncation problems) and summarize the results in Table 8 in the electronic companion to this paper. From the table, we find that the worst-case performances of the measures show a similar pattern. Because ±50% is a considerably large perturbation, these results reinforce the robustness assertion for the variances when the independence assumption holds.

4.4. Covariance Robustness

We have discussed the variance robustness for the independent case. However, as mentioned in the introduction, the independence has not been validated and there may exist nontrivial dependence among the

Table 3 Worst-Case Performance for $\mu = \hat{\mu}$ and $\Sigma \in \hat{\mathcal{P}}_1$

Policy	Avg. NPV	SR (%)	Avg. GTI	SR (%)	P(GTI > 2)	SR (%)	P(GTI > 3)	SR (%)	P(GTI > 4)	SR (%)
No controls	24.65	0.04	3.18	0.32	0.954	6.00	0.593	3.31	0.199	12.43
2 × CO ₂	20.73	−0.77	2.42	2.11	0.827	11.91	0.179	24.31	<0.001	NA
2°C increase	27.53	0.95	NA	NA	0.000	NA	0.000	NA	0.000	NA
Stern review	39.01	1.25	1.57	1.95	0.126	11.50	0.008	14.29	<0.001	NA
Gore proposal	55.38	3.19	1.48	−0.67	0.138	28.97	0.000	NA	0.000	NA

Downloaded from informs.org by [144.214.42.26] on 02 September 2014, at 22:39. For personal use only, all rights reserved.

Table 4 Worst-Case Performance for $\mu = \hat{\mu}$ and $\Sigma \in \hat{\mathcal{S}}_2$

Policy	Avg. NPV	SR (%)	Avg. GTI	SR (%)	P(GTI > 2)	SR (%)	P(GTI > 3)	SR (%)	P(GTI > 4)	SR (%)
No controls	26.69	8.32	3.18	0.32	0.956	6.22	0.599	4.36	0.203	14.69
2 × CO ₂	22.84	9.33	2.42	2.11	0.831	12.45	0.182	26.39	<0.001	NA
2°C increase	29.30	7.44	NA	NA	0.000	NA	0.000	NA	0.000	NA
Stern review	42.68	10.77	1.57	1.95	0.153	35.40	0.011	57.14	<0.001	NA
Gore proposal	58.72	9.41	1.49	0.00	0.134	25.23	0.000	NA	0.000	NA

parameters. In what follows we study the covariance (variances plus correlations) robustness for the DICE model. The effects of the correlations among the random parameters are more complicated, and they are often ignored by environmental modelers. As has been emphasized multiple times, in his Monte Carlo study, Nordhaus (2008) assumed that the eight parameters in the DICE model are independent. He then estimated the distributions for these parameters separately without taking into account the possible correlations. In this subsection, we investigate the robustness of the DICE model to the covariances using two ambiguity sets introduced in §2.2. The first one is

$$\hat{\mathcal{S}}_2 = \{\Sigma: 0.5\hat{\Sigma} \leq \Sigma \leq 1.2\hat{\Sigma}\},$$

whose upper and lower bounds are set based on the same logic used in setting the bounds in $\hat{\mathcal{S}}_1$. However, compared to $\hat{\mathcal{S}}_1$, $\hat{\mathcal{S}}_2$ is a bigger set and it allows the correlations to take nonzero values. The second one is

$$\hat{\mathcal{S}}_3 = \{\Sigma: \Sigma_{ii} \leq \hat{\Sigma}_{ii}, i = 1, \dots, 8, \Sigma \geq 0.5\hat{\Sigma}\}.$$

As discussed in §2.2.2, $\hat{\mathcal{S}}_3$ essentially loosens the restriction on the dependence among parameters and thus can be used to examine specifically the effects of the correlations. In $\hat{\mathcal{S}}_3$, we set the lower bound of the covariance matrix as $0.5\hat{\Sigma}$ based on the same logic used in setting the lower bounds in $\hat{\mathcal{S}}_1$.

We compute the worst-case performances for $\hat{\mathcal{S}}_2$ and $\hat{\mathcal{S}}_3$ and summarize the results in Tables 4 and 5, respectively. From the tables, we find that the covariances have a significant impact on the average NPV and the probabilities of various temperature increases. It suggests that the performance measures of the DICE model, especially the average NPV, may take a quite nonlinear form with respect to the eight parameters. Now let us take a closer look at these results. Comparing $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$, we see that the two sets have

the same range for the variances, but $\hat{\mathcal{S}}_2$ allows the correlations to have some freedom perturbing around zero. Consequently, the changes of worst-case performances for $\hat{\mathcal{S}}_2$ are much more significant than that for $\hat{\mathcal{S}}_1$. Coming to $\hat{\mathcal{S}}_2$ and $\hat{\mathcal{S}}_3$, the range of the variances for $\hat{\mathcal{S}}_3$ is smaller than that for $\hat{\mathcal{S}}_2$, whereas $\hat{\mathcal{S}}_3$ allows the correlations to have more freedom to perturb. As a result, the changes of worst-case performances for $\hat{\mathcal{S}}_3$ are much more significant than that for $\hat{\mathcal{S}}_2$. These facts suggest that the correlations play a critical role in the changes.

4.5. Overall Robustness

The above analysis has shown us that the specification of the MVN distribution has a significant impact on the simulation outputs of the DICE model. In this subsection, we examine an overall robustness by considering both the mean and covariance ambiguities simultaneously and see how it affects the simulation outputs. We let the mean vector and covariance matrix vary in $\hat{\mathcal{M}} \times \hat{\mathcal{S}}_2$. Note that the ambiguity set $\hat{\mathcal{M}} \times \hat{\mathcal{S}}_2$ requires the most computational effort for running Algorithm SQMP among all the considered situations. For this set, if we stop the algorithm when the difference between the objective values of two consecutive iterations is less than 0.01 for NPV and 0.001 for GTI and the probabilities, Algorithm SQMP typically terminates in 10–50 iterations with computational time typically varying in two to eight minutes. We summarize the worst-case performances in Table 6. From the table, we find that the worst-case performances are even more different from the average performances reported in Table 1.

Looking at all the results reported in Tables 2–8, we find that the relative preferences of the five policies do not change much in the worst cases for all the ambiguity sets that we considered, compared to the results reported in Table 1. This further supports Nordhaus' claim that moderately tight policies, e.g., the 2 × CO₂

Table 5 Worst-Case Performance for $\mu = \hat{\mu}$ and $\Sigma \in \hat{\mathcal{S}}_3$

Policy	Avg. NPV	SR (%)	Avg. GTI	SR (%)	P(GTI > 2)	SR (%)	P(GTI > 3)	SR (%)	P(GTI > 4)	SR (%)
No controls	32.13	30.40	3.26	2.84	0.959	6.56	0.600	4.53	0.214	20.90
2 × CO ₂	25.65	22.79	2.44	2.95	0.836	13.13	0.199	38.19	<0.001	NA
2°C increase	33.47	22.74	NA	NA	0.000	NA	0.000	NA	0.000	NA
Stern review	49.20	27.69	1.60	3.90	0.202	78.76	0.018	157.14	<0.001	NA
Gore proposal	68.11	26.91	1.50	0.67	0.155	44.86	0.000	NA	0.000	NA

Table 6 Worst-Case Performance for $(\mu, \Sigma) \in \hat{\mathcal{M}} \times \hat{\mathcal{P}}_2$

Policy	Avg. NPV	SR (%)	Avg. GTI	SR (%)	P(GTI > 2)	SR (%)	P(GTI > 3)	SR (%)	P(GTI > 4)	SR (%)
No controls	39.64	60.88	3.78	19.24	0.981	9.00	0.821	43.03	0.395	123.16
2 × CO ₂	32.05	53.42	2.61	10.13	0.916	23.95	0.315	118.75	<0.01	NA
2°C increase	45.96	68.54	NA	NA	0.000	NA	0.000	NA	0.000	NA
Stern review	62.12	61.23	1.74	12.99	0.252	123.01	0.016	128.57	<0.001	NA
Gore proposal	73.88	37.66	1.68	12.75	0.277	158.88	0.000	NA	0.000	NA

and 2°C increase policies, are more preferable because they incur moderate costs but have the global warming under control.

4.6. Conclusions and Suggestions

Because of the lack of data and the lack of precise understanding of the complex climate systems, there is considerable amount of ambiguity in climate modeling, which may have significant impacts on global warming studies and policy makings. In this paper, we take a robust approach to handling the ambiguity in the DICE model. Based on our investigations, we have the following suggestions:

- Whereas our primary study shows the DICE model is relatively robust to the variances when the independence assumption that is implicitly taken in Nordhaus (2008) holds, the DICE model is sensitive to the mean vector and covariance matrix (correlations play an important role). This suggests that (1) to improve the effectiveness and reliability of the model, superior effort needs to be spent on improving the estimates of the means; and (2) the independence assumption needs more scientific justifications. If there is no ground for assuming independence, then dependence among those key uncertain parameters should be considered into the global warming modeling as such dependence may have a significant impact on global warming decision making, though we believe this would be challenging.

- Although the values may change significantly for all three types of performance measures, these changes in general do not alter our preferences over different policies, i.e., the rankings of the policies are robust with respect to the ambiguities of both mean vector and covariance matrix. Therefore, we can say that the DICE model is relatively robust in comparing alternative policies.

- As emphasized in Nordhaus (2008), aggressive global warming policies such as the Stern review and Gore proposal are cost ineffective. Such ineffectiveness is enlarged significantly when the ambiguity of distribution is taken into consideration. On the other hand, no control policies may have high risks of causing abrupt climate changes, and thus facing global warming, doing nothing is by no means a rational response. Essentially, our analysis supports the policies with moderately tight climate limits, such as the 2 × CO₂ policy and 2°C increase policy.

- For the two policies with moderately tight climate limits, the 2°C increase policy is suggested in the Copenhagen Accord, whereas the 2 × CO₂ policy has not attracted that much attention. The most important merit of the 2°C increase policy may be that it can totally control the risk of global warming. Yet, according to our study, the 2°C constraint may be a little binding and will incur a moderate cost. The incurred cost may be enlarged by the ambiguity of random parameters. On the other hand, the 2 × CO₂ policy, which limits atmosphere CO₂ concentration to 560 ppm, is more cost-effective than the 2°C increase policy in both average case and worst case. If the 2 × CO₂ policy is used, it is unlikely that global temperature increase will exceed 4°C. Nevertheless, there is a significant probability that the temperature increase will exceed 2°C, and a moderate probability that the temperature increase will exceed 3°C.

- Ambiguity should be taken into consideration not only in the process of making global warming policies but also in the process of implementing these policies. Because our understanding of climate systems may change, the ambiguity in environmental modeling may also change. Policy makers and policy implementers need to constantly monitor these changes and adjust their policies accordingly.

Acknowledgments

The authors thank the associate editor and two anonymous referees for their insightful comments that improved the paper in numerous ways. A preliminary version of this paper (Hu et al. 2010) was previously published in the *Proceedings of the 2010 Winter Simulation Conference*. Zhaolin Hu is supported in part by the National Natural Science Foundation of China [Grant 70832005]. Jing Cao is supported in part by the National Natural Science Foundation of China [Grants 70803026, 71173130] and by Tsinghua University Initiative Scientific Research Program. L. Jeff Hong is supported in part by the Hong Kong Research Grants Council [Grants GRF 613410, GRF 613011, N-HKUST 626/10] and by the Hong Kong University of Science and Technology [Grant RPC10EG11].

Appendix. Convergence of Sample-Average Approximation

For simplicity of the notation, we let

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

Downloaded from informs.org by [144.214.42.26] on 02 September 2014, at 22:39. For personal use only, all rights reserved.

It is easy to verify that, for every $\xi \in \mathfrak{N}^d$, $L(\mu, X, \xi)$ is continuously differentiable on $\mathcal{M} \times \mathcal{X}$. 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 the chain rule we have

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

and

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

With a slight abuse of the notation, in what follows we let $\|\cdot\|$ denote both the Euclidean norm in \mathfrak{R}^d and the Frobenius norm in the symmetric matrix space \mathbb{S}^d . We make the following assumption on $L(\mu, X, \xi)$:

ASSUMPTION 1. *There exist measurable functions $D_i(\cdot)$, $i = 1, 2$, such that $E_f[D_i(\xi)] < \infty$, $i = 1, 2$, and*

$$\begin{aligned} \|\nabla_\mu L(\mu, X, \xi)\| &\leq D_1(\xi), \quad \forall (\mu, X) \in \mathcal{M} \times \mathcal{X}, \forall \xi \in \mathfrak{N}^d; \\ \|\nabla_X L(\mu, X, \xi)\| &\leq D_2(\xi), \quad \forall (\mu, X) \in \mathcal{M} \times \mathcal{X}, \forall \xi \in \mathfrak{N}^d. \end{aligned}$$

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

PROPOSITION 1. *Suppose there exist $\gamma > 0$, $\mu_s \in \mathfrak{N}^d$, and $\Sigma_u \in \mathbb{S}_{++}^d$, such that $\Sigma \leq \Sigma_u$ for all $\Sigma \in \mathcal{P}$, and $E_{(\mu_s, \Sigma_u)}[c^{1+\gamma}(\xi)] < \infty$. Then Assumption 1 holds for any density f .*

For Proposition 1, because \mathcal{P} is a compact set in \mathbb{S}_{++}^d , it is easy to find the matrix Σ_u . Therefore, we only need to verify the condition $E_{(\mu_s, \Sigma_u)}[c^{1+\gamma}(\xi)] < \infty$. This is, of course, a very weak condition. Note that Proposition 1 does not impose any restriction on the choice of density $f(\cdot)$. As shall be seen, in our considered situations, any density $f(\cdot)$ that satisfies $f(x) > 0$ for all $x \in \mathfrak{N}^d$ can be used to ensure the convergence of the SAA.

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

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

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

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

denote the deviation of the set $\mathcal{M}_A \times \mathcal{X}_A$ from the set $\mathcal{M}_B \times \mathcal{X}_B$. Let \bar{v}_n and \bar{S}_n denote the optimal objective value and

the set of optimal solutions, respectively, of problem (5), and let v and S denote the optimal objective value and the set of optimal solutions, respectively, of problem (4). Then, we have the following theorem on the convergence of the SAA.

THEOREM 2. *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), $\bar{v}_n \rightarrow v$ and $\mathbb{D}(\bar{S}_n, S) \rightarrow 0$ as $n \rightarrow \infty$.*

Theorem 2 shows that the optimal solutions of problem (5) provide good approximations to the optimal solutions of problem (4) when n is large. Therefore, we can solve problem (5) to approximate the optimal solutions of problem (4).

Note that problem (5) 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 (5) may be difficult. We often only find stationary points of problem (5). Then, the conclusions of Theorem 2 may not be applicable. Therefore, we need to investigate the convergence of stationary points of the SAA.

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

$$N_{\mathcal{M}}(\mu) = \{z \in \mathfrak{N}^d: z^T(\mu' - \mu) \leq 0, \forall \mu' \in \mathcal{M}\},$$

$$N_{\mathcal{X}}(X) = \{Z \in \mathbb{S}^d: \text{tr}(Z(X' - X)) \leq 0, \forall X' \in \mathcal{X}\}.$$

Recall that a point $(\mu, X) \in \mathcal{M} \times \mathcal{X}$ is a stationary point of problem (4) if $\nabla_\mu g(\mu, X) \in N_{\mathcal{M}}(\mu)$ and $\nabla_X g(\mu, X) \in N_{\mathcal{X}}(X)$, and is a stationary point of problem (5) if $\nabla_\mu \bar{g}_n(\mu, X) \in N_{\mathcal{M}}(\mu)$ and $\nabla_X \bar{g}_n(\mu, X) \in N_{\mathcal{X}}(X)$. Let K and \bar{K}_n denote the sets of stationary points of problems (4) and (5), respectively. Then, we have

$$K = \{(\mu, X) \in \mathcal{M} \times \mathcal{X}: \nabla_\mu g(\mu, X) \in N_{\mathcal{M}}(\mu), \nabla_X g(\mu, X) \in N_{\mathcal{X}}(X)\},$$

$$\bar{K}_n = \{(\mu, X) \in \mathcal{M} \times \mathcal{X}: \nabla_\mu \bar{g}_n(\mu, X) \in N_{\mathcal{M}}(\mu),$$

$$\nabla_X \bar{g}_n(\mu, X) \in N_{\mathcal{X}}(X)\}.$$

The next theorem is parallel to Theorem 2. It ensures that the stationary points of the SAA are also good approximations to those of problem (4).

THEOREM 3. *Suppose that Assumption 1 is satisfied. Then, K is nonempty and \bar{K}_n are nonempty for all n , and $\mathbb{D}(\bar{K}_n, K) \rightarrow 0$ w.p.1 as $n \rightarrow \infty$.*

REMARK 1. For nonlinear stochastic programs, the convergence of the set of Karush–Kuhn–Tucker points of the SAA is discussed by Shapiro et al. (2009) under the framework of stochastic generalized equations. In Theorem 3, we extend their result to semidefinite stochastic programs.

Because both problem (4) and its SAA are nonconvex optimization problems, we cannot guarantee to find their global optimal solutions. Therefore, we take one step back and try to find a good local optimal solution. Theorem 3 shows that, if we can find a local optimal solution of problem (5), it provides a good approximation to a local optimal solution of problem (4).

References

Anderson TW (1984) *An Introduction to Multivariate Analysis* (Wiley, New York).

- Bahn O, Edwards N, Knutti R, Stocker TF (2011) Energy policies avoiding a tipping point in the climate system. *Energy Policy* 39(1):334–348.
- Barton RR, Schruben LW (2001) Resampling methods for input modeling. *Proc. 2001 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 372–378.
- Ben-Tal A, Nemirovski A (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.
- Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* 88:411–424.
- Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (SIAM, Philadelphia).
- Bernardo JM, Smith AFM (2000) *Bayesian Theory* (Wiley, New York).
- Bertsimas D, Sim M (2004) Price of robustness. *Oper. Res.* 52(1):35–53.
- Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53(3):464–501.
- Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).
- Cario MC, Nelson BL (1997) Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.
- Cheng RCH, Holland W (1997) Sensitivity of computer simulation experiments to errors in input data. *J. Statist. Comput. Simulation* 57:219–241.
- Chick SE (2001) Input distribution selection for simulation experiments: Accounting for input uncertainty. *Oper. Res.* 49(5):744–758.
- Chopra VK, Ziemba WT (1993) The effects of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfolio Management* 19(2):6–11.
- Clemen RT, Reilly T (1999) Correlations and copulas for decision and risk analysis. *Management Sci.* 45(2):208–224.
- Das SR, Fong G, Geng G (2001) Impact of correlated default risk on credit portfolios. *J. Fixed Income* 11(3):9–19.
- Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.
- Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *Quart. J. Econom.* 75(4):643–669.
- Epstein LG (1999) A definition of uncertainty aversion. *Rev. Econom. Stud.* 66(3):579–608.
- Fabozzi FJ, Huang D, Zhou G (2010) Robust portfolios: Contributions from operations research and finance. *Ann. Oper. Res.* 176(1):191–220.
- Fu MC (2002) Optimization for simulation: Theory vs. practice. *INFORMS J. Comput.* 14(3):192–215.
- Fu MC (2008) What you should know about simulation and derivatives. *Naval Res. Logist.* 55(8):723–736.
- Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (Springer, New York).
- Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Math. Oper. Res.* 28(1):1–38.
- Grant M, Boyd S (2011) CVX: Matlab software for disciplined convex programming. Accessed October 16, 2011, <http://cvxr.com/cvx/>.
- Halldórsson BV, Tütüncü RH (2003) An interior-point method for a class of saddle point problems. *J. Optim. Theory Appl.* 116(3):559–590.
- Henderson SG (2003) Input model uncertainty: Why do we care and what should we do about it? *Proc. 2003 Winter Simulation Conf.* (IEEE Computer Society, Washington, DC), 90–100.
- Hong LJ, Nelson BL (2009) A brief introduction to optimization via simulation. *Proc. 2009 Winter Simulation Conf.* (IEEE Computer Society, Washington, DC), 75–85.
- Hong LJ, Yang Y, Zhang L (2011) Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Oper. Res.* 59(3):617–630.
- Hu Z, Cao J, Hong LJ (2010) Robust simulation of environmental policies using the DICE model. *Proc. 2010 Winter Simulation Conf.* (IEEE Computer Society, Washington, DC), 1295–1305.
- Intergovernmental Panel on Climate Change (IPCC) (2007) Summary for policymakers. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK).
- Kriegler E, Hall JW, Held H, Dawson R, Schellnhuber H-J (2009) Imprecise probability assessment of tipping points in the climate system. *Proc. Natl. Acad. Sci. USA* 106(13):5041–5046.
- Mattoo A, Subramanian A, van der Mensbrugge D, He J (2009) Reconciling climate change and trade policy. Policy Research Working Paper 5123, World Bank, Washington, DC.
- Netessine S, Rudi N (2003) Centralized and competitive inventory models with demand substitution. *Oper. Res.* 51(2):329–335.
- Nordhaus W (2008) *A Question of Balance: Weighing the Options on Global Warming Policies* (Yale University Press, New Haven, CT).
- Ravikumar P, Wainwright MJ, Raskutti G, Yu B (2008) High-dimensional covariance estimation by minimizing l_1 -penalized log-determinant divergence. Technical report, University of California, Berkeley.
- Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).
- Shapiro A, Homem-de-Mello T (2000) On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM J. Optim.* 11(1):70–86.
- Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM, Philadelphia).
- Stern N (2007) *The Economics of Climate Change: The Stern Review* (Cambridge University Press, Cambridge, UK).
- United Nations Framework Convention on Climate Change (UNFCCC) (2009) Copenhagen accord. Accessed October 16, 2011, <http://unfccc.int/resource/docs/2009/cop15/eng/107.pdf>.
- U.S. Climate Change Science Program (CCSP) (2008) Abrupt climate change. A report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research U.S. Geological Survey, Reston, VA.
- Vandenberghe L, Boyd S, Wu SP (1998) Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl.* 19(2):499–533.
- Wang C, Sun D, Toh KC (2009) Solving log-determinant optimization problems by a Newton-CG proximal point algorithm. *SIAM J. Optim.* 20(6):2994–3013.
- Webster M, Sokolov P, Reilly M, Forest E, Paltsev S, Schlosser A, Wang C, et al. (2009) Analysis of climate policy targets under uncertainty. Report No. 180, MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA.
- Wu S-P, Vandenberghe L, Boyd S (1996) MAXDET: Software for determinant maximization problems. Accessed October 16, 2011, <http://www.stanford.edu/boyd/oldsoftware/MAXDET.html>.