POLICY RESPONSE TO CLIMATE CHANGE IN A DYNAMIC STOCHASTIC ECONOMY

Allocation: GLCPC/0.35 Mnh
PI: Lars Hansen[1]
Senior representative: Kenneth Judd[2]
Co-PIs: William Brock[3], Thomas Hertel[4]
Collaborator: Yongyang Cai[1,2]

[1]University of Chicago
[2]Hoover Institution, Stanford University
[3]University of Wisconsin
[4]Purdue University

EXECUTIVE SUMMARY:

We extended past work on computational methods for solving dynamic programming problems by introducing an error estimation and control scheme and implementing more efficient methods for approximating multidimensional functions [1]. We extended our Integrated Assessment Model (IAM) framework, called DSICE (Dynamic Stochastic Integration of Climate and the Economy), for evaluating alternative policy responses to future climate change. The new elements allow us to use stochastic processes to model the risks in both the climate and economic systems, and to examine how robust our results are to uncertainties about the parameters of those stochastic systems [1,2].

One substantive result of DSICE is that the social cost of carbon (SCC) is substantially greater when we include economic and climate risks in analyzing the impact of climate change on the economy. Furthermore, the SCC is itself a stochastic process with significant variation. Our findings show that a single parameterization of DSICE, representing the most recent empirical macroeconomic analysis of economic growth, implies a stochastic process for carbon emissions with a ~10% chance of emissions exceeding all of the IPCC emissions scenarios.

We do not have precise knowledge about many key parameters in our models of the economic and climate systems. We performed extensive uncertainty quantification and showed that the SCC highly depends on the parameter values [1,2].

INTRODUCTION

Every economy is a complex system, but economics researchers typically ignore this by analyzing simple, stylized models of pieces of the system, and use only pencil and paper analyses (supplemented by simple computational models) of these stylized models instead of computationally intensive studies of more realistic models. We are trying to change that by creating robust and general tools that use state-of-the-art numerical methods to study economics problems. In this work, we focus on applications to climate change policy analysis.

Global warming has been recognized as a growing potential threat to economic well-being. This concern has led to an increasing number of national and international discussions on how to respond to this threat. Determining which policies should be implemented will require merging quantitative assessments of the likely economic impacts of carbon emissions with models of how the economic and climate systems interact; this is the purpose of integrated assessment models (IAMs). Most IAMs are deterministic, where economic actors have perfect knowledge of future economic and climate events. Their limitations are due to economists' aversion to modern computational tools. However, all agree that uncertainty needs to be a central part of any IAM analysis.

METHODS AND RESULTS

We developed DSICE [2], a computational, dynamic, stochastic general equilibrium framework for studying global models of both the economy and the climate. We applied it to the specific issue of how the social cost of carbon depends on stochastic features of both the climate and the economy when we applied empirically plausible specifications for the willingness to pay to reduce economic risk.

The inclusion of risks and uncertainties in DSICE makes our SCC analysis among the most computationally demanding problems ever solved in economics. We solved DSICE using numerical dynamic programming—efficient multivariate methods to approximate value functions, and reliable optimization methods to solve the dynamic stochastic problems [3,4].

We showed that economic and climate risks substantially increase SCC, and the SCC is itself a stochastic process with significant variation; for example, the basic elements of risk incorporated into our model cause the SCC in 2100 to be, with 5% probability, five times what it would be without those risks [2]. A July 2014 White House report, titled "The cost of delaying action to stem climate change," incorporated our paper's conclusion that high estimates for SCC can be justified without assuming the possibility of catastrophic events.

Fig. 1 shows four emission scenarios that the IPCC adopted as the range of plausible greenhouse gas emission paths and advocates their use as inputs for climate system models. These scenarios presumably cover the range of likely emission paths. DSICE models a stochastic process, implying that the emission paths cover a range of potential values. The shaded area in fig. 1 displays the range, 1% to 99% quantiles, of business-as-usual (BAU) emission paths implied by a single run using one set of parameters in DSICE. The dotted line shows the 90% quantile of DSICE emission paths. These highlight the importance of including uncertainty in DSICE. DSICE showed that there is a 10% chance that emissions will exceed the highest IPCC scenario. Scenarios with such high emissions are the ones that could lead to the worst outcomes. We found that the representative scenarios advocated by the IPCC ignore 10% or more of the most dangerous plausible emission paths.

DSICE has demonstrated its flexibility in other applications:

- An analysis of the impact of the carbon tax under various continuous climate tipping points showed that the costs of carbon emission used to inform policy are being underestimated and that uncertain future climate damages should be discounted at a low rate [5].
- Examining the impact of ecosystem service risk on the cost–benefit assessments of climate change policies showed that the risk of a tipping point, even if it only has non-market impacts, could substantially increase the present optimal carbon tax [6].
- Examining the implications of future uncertainty in climate impacts, climate regulation, and energy prices on the long-term trajectory of global land use and associated greenhouse gas emissions produced a World Bank Policy Research working paper [7].

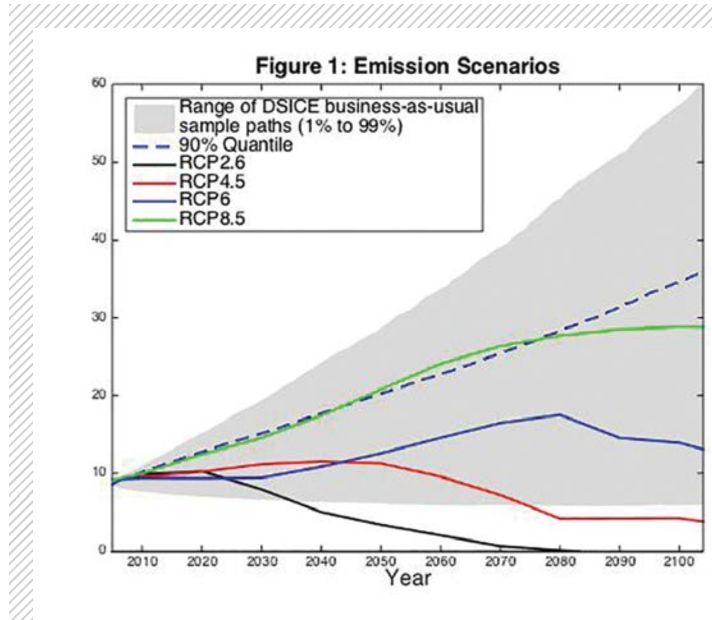

FIGURE 1: Solid lines show four emission scenarios that the IPCC adopted as the range of plausible greenhouse gas emission paths and advocates their use as inputs for climate system models. Shaded area: 1% to 99% quantiles of business-as-usual (BAU) emission paths implied by a single run using one set of parameters in DSICE. Dotted line: the 90% quantile of DSICE emission paths.

We are limited by the current state of knowledge about the critical parameters of the economics and climate system, and different decision-makers usually have different beliefs about the implications of parameter vectors. Over time, new observations will give us new information about critical parameters, such as climate sensitivity. The policy question will be whether to delay mitigation efforts until we have that new information or to proceed with aggressive interventions now.

WHY BLUE WATERS?

DSICE is a high-dimensional dynamic stochastic model that we have thus far applied to a nine-dimensional problem in [2] and a fourteen-dimensional problem in [7]. Solving these problems was very time consuming, but parallelizing the numerical dynamic programming methods allowed us to solve them efficiently [8]. For example, in the benchmark example of [2] DSICE scaled almost linearly to 84,000 cores; a serial computation would take about 77 years. Likewise, our algorithm for solving a dynamic game with multiple players and many choices [10] would take decades of serial computation. We are developing parallel computational methods to solve even higher-dimensional problems. For example, we applied a parallel algorithm to approximately solve a

400-dimensional dynamic stochastic problem in [9].

Our analyses also required extensive uncertainty quantification, which fortunately is well-suited to Blue Waters. Our applications of uncertainty quantifications used thousands of parameter specifications in [2], and hundreds in [1].

In the next year, we will study the impact of multiple interacting tipping points on the social cost of carbon in a fifteen-dimensional version of DSICE. We will also examine the effects of learning. The high-dimensional nature of both models will require Blue Waters' capabilities.

PUBLICATIONS

Lontzek, T. S., Y. Cai, K. L. Judd, and T. M. Lenton, Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy. *Nat. Clim. Change*, 5 (2015), pp. 441–444, doi:10.1038/nclimate2570.

Cai, Y., K. L. Judd, T. M. Lenton, T. S. Lontzek, and D. Narita, Environmental tipping points significantly affect the cost-benefit assessment of climate policies. *Proc. Nat. Acad. Sci.* U.S.A., 112:15 (2015), pp. 4606–4611, doi:10.1073/pnas.1503890112

COMPUTATIONAL STRATEGIES FOR APPLYING QUALITY SCORING AND ERROR MODELING STRATEGIES TO EXTREME-SCALE TEXT ARCHIVES

Allocation: Illinois/50 Knh
PI: Scott Althaus[1,2]
Co-PIs: Loretta Auvil[3], Boris Capitanu[3], David Tcheng[3], Ted Underwood[1]

[1]University of Illinois at Urbana–Champaign
[2]Cline Center for Democracy, University of Illinois at Urbana–Champaign
[3]Illinois Informatics Institute, University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY:

An important barrier to extreme-scale analysis of unstructured textual data digitized from printed copies using optical character recognition (OCR) techniques is the uncertain quality of the textual representations that have been made from scanned page images. We used Blue Waters to evaluate OCR errors on the HathiTrust Public Use dataset, which is the world's largest collection of digitized library volumes in the public domain, consisting of 3.2 million zipped files totaling nearly 3 TB. We also used Blue Waters to assess the impact of OCR errors on event-detection algorithms using a collection of 16 million articles from The New York Times. Our aim is to develop error quality scoring and correction strategies that can enhance the ability of data analytics researchers to work with digitized textual data at extreme scales.

INTRODUCTION

Researchers in the humanities and social sciences often analyze unstructured data in the form of images and text that have been scanned and digitized from non-digital sources. For this type of research, the most important barrier to conducting extreme-scale analysis of unstructured data is the uncertain quality of the textual representations of scanned images derived from optical character recognition (OCR) techniques. Our project is using Blue Waters to detect, score, and correct OCR errors in the HathiTrust Public Use dataset, which is the world's largest corpus of digitized library volumes in the public domain, consisting of over 1.2 billion scanned pages of OCR text. We are also using Blue Waters to assess the impact of OCR errors on natural language processing algorithms using a corpus of 16 million historical newspaper articles from The New York Times (NYT).

METHODS & RESULTS

In collaboration with the HathiTrust Research Center (HTRC), we performed two major computations with their data. Each of these implementations leveraged Akka [1] to provide a high-performance JVM-based framework featuring simple concurrency and distribution. The first HTRC-related computation on our Blue Waters allocation supported the evaluation of OCR errors in the HathiTrust public domain volumes (which at the time was 3.2 million volumes). We applied new error-detection algorithms produced by a Mellon-funded project called eMOP under Laura Mandell at Texas A&M. Our team's analysis of text-level quality problems will be compared and correlated with image-level quality analysis undertaken by Paul Conway of the University of Michigan on the page images from which the OCR text was extracted. Our aim is to identify how OCR errors are correlated with specific types of image distortion so that digital librarians can better anticipate quality problems from their digitization efforts.

The second HTRC-related computation on our Blue Waters allocation supported the creation of the HTRC Extracted Features (EF) dataset [2], where "features" are notable or informative characteristics of the text [3–5]. The dataset is derived from 4.8 million HathiTrust public domain volumes, totaling over 1.8 billion pages, 734 billion words, dozens of languages, and spanning multiple centuries. We processed a number of useful features at the page level including part-of-speech tagged token counts, header and footer identification, and a variety of line-level information. The EF dataset was used in the creation of the HathiTrust Bookworm [6], which is a tool that visualizes language usage trends in repositories of digitized texts in a simple and powerful way.

Our Blue Waters allocation is also supporting a third set of computations designed to clarify the impact of OCR error on natural language processing algorithms. The Cline Center for Democracy has access to the entire population of NYT articles from 1980 to 2005 in two different forms: "born-digital" content that contains pristine textual data, and ProQuest's digitized version of the same content that was derived from microfilm images using OCR. We used Blue Waters to deploy Phoenix civil unrest event identification software [7] produced by the Open Event Data Alliance (OEDA) over the combined 16 million articles in this NYT corpus. Phoenix requires making computationally intensive parse trees for the articles in order to identify events, and we have a prototype of this deployed, again using Akka, to distribute the work of creating parse trees, identifying events, and writing results. We also plan to compare 25 years of NYT OCR articles to the born-digital NYT articles using software developed for the eMOP project. This will help us evaluate OCR quality and determine the level and type of correction that most closely restores noisy OCR to resemble the born-digital content so that automated correction algorithms can be optimized for use with OCR data from ProQuest's historical newspaper collection.

WHY BLUE WATERS?

The computational demands of using natural language processing, machine learning, or rule-based scoring strategies on the scale of the HathiTrust Public Use corpus would severely tax the capabilities of other HPC platforms. Only Blue Waters offers the computational scale required to carry out the necessary quality scoring and error correction strategies in a timely fashion on an unstructured text corpus of this size.