The Optimal Social Cost of Carbon & The Cost of Inaction

by

Nicole Sandra-Yaffa Dumont

A research paper
presented to the University Of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Computational Mathematics

Waterloo, Ontario, Canada, 2019

This research paper consists of material all of which I authored or co-authored: see Statement of Contributions included in the research paper. This is a true copy of the research paper, including any required final revisions, as accepted by my examiners.

I understand that my research paper may be made electronically available to the public.

Statement of contributions

Alexey Rubtsov, along with myself, Thomas Coleman, Wanqi Li, and Wenbin Liu, co-authored the working paper, "Optimal Pricing of Climate Risk" (2019). Wanqi Li created Table 4.1 of computational run times of gradient calculations. Equations from sections of the working paper typeset by Dr. Rubtsov are included in this report in sections 3.4 and 3.5. Dr. Rubtsov also edited sections I authored in the working paper which I have modified for use in this report (section 5.1).

Abstract

The EZ Climate model of Daniel et al. (2018) estimates the Social Cost of Carbon (SCC) - the societal cost of emitting one additional tonne of CO_2 assuming optimal emission reductions using asset pricing theory. Determination of the SCC is of great interest as this number underlies many government climate policies. By applying automatic differentiation and Quasi-Newton optimization to the EZ model, we significantly improve its computational performance. This improvement allows for increased scalability, parameter sensitivity analysis, robust optimization, and evaluations of suboptimal policies. We find that postponing climate change mitigation until 2030 is equivalent to a loss of $5.4 trillion (in 2015 USD).

Acknowledgments

I would like to acknowledge the contributions of my supervisor Thomas Coleman, whose advice and guidance I am deeply grateful for.

I would like thank Alexey Rubtsov and Wanqi Li for their immense contributions to this research project. In particular, I'd like to thank Dr. Rubtsov for his suggestions and editing of my contributions to the working paper, "Optimal Pricing of Climate Risk" (2019).

Additionally, this project owns thanks for the contributions from Wenbin Liu and Wei Xo, and to Bob Littermam and Kent Daniel for introducing this problem to Dr. Coleman.

I must again thank Dr. Coleman and Dr. Rubtsov for serving as my readers.

Dedication

I dedicate this paper to my partner, Kira Selby.

Contents

List of Figures

List of Tables

1 Introduction

It is well known that emissions of carbon dioxide and other greenhouse gases (GHGs) into the atmosphere are detrimental to the environment. High concentrations of GHGs in the atmosphere cause climate change which in turn is a cause of heat waves, flooding, wildfires, loss of coastal land, reduced crop yields, health problems in human populations, and extreme weather events. The effect that climate change will have on the global economy and the quality of life around the world should not be understated. The Fourth National Climate Assessment predicts the U.S. economy will shrink by up to 10% in 100 years if emissions continue on their current trajectory (Wuebbles et al. (2017)). This is a loss of billions for the U.S alone. Producers of carbon emissions do not pay for this future cost to society and thus there is no market mechanisms to limit emissions in a way that accounts for these future damages. This makes GHG emissions a negative production externality. The preferred solution to this problem by many economists is to assign a price to such activities equal to the monetary value of the damages they cause. In the case of GHG emssions, such a price is called the social cost of carbon (SCC), the monetary value of damages caused by each additional ton of CO_2-equivalent emissions (this price would be used in practice as a taxation rate or used to design a cap-and-trade program).

Calculating the SCC is of great interest. A review (Wang et al. (2018)) found that estimates for the SCC in peer-reviewed literature range from -13.36 to 2386.91\$ per ton of CO_2, with a mean value of 54.70\$/tCO_2. Estimates of the SCC are already used by government bodies around the globe to decide climate policy. As with any tractable modelling of a large, complex systems, models of the interactions of economics and climate must make many simplifying assumptions. It is important that such models are robust to these assumptions as conclusions drawn from these models have a real world impact. The scale of these models is often very large (due to the time scale at which climate change occurs), however, and they often have many modelling variables and parameters due to the size and complexity of the real systems they represent. This computational complexity makes full robustness checks nearly infeasible.

In this report, we use the EZ Climate model from Daniel et al. (2018) to compute the SCC. This is a policy optimization model which uses social welfare (i.e. utiltiy) as the optimization objective. In this model, the SCC is computed using the optimal GHG emissions policy as opposed to the current emissions pathway. It is assumed that global consumption is both reduced by damages from climate change, and by climate policies (i.e. a carbon tax may decrease the production of certain goods and lower consumption). While a strong policy will lower consumption, it will also lower climate change damages. A policy/strategy is defined as the set of carbon prices chosen by a representative agent. To model the uncertainty of the future effects climate change will have on the economy, a binomial tree is used. The nodes of the tree are states of the world at different times and in states of nature - these are different scenarios of the impact of climate change. The agent makes a decision at each node. The optimal choice of policy is the one that maximizes the expected lifetime utility of global consumption, modelled using the Epstein-Zin utility function (Epstein and Zin (1989)).

We apply automatic differentiation and gradient-based optimization methods to the EZ Climate model and show that this significantly reduce its computational time. This allows us to increase the size of the tree by adding more decision periods. Additionally, we can check the robustness of the model to assumptions about the cost of emissions mitigation, the effect of technological change on this cost, and the number of decision periods in the model. Following this, the problem is reformulated and solved as a robust optimization problem. Finally, we analyze suboptimal mitigation strategies and determine their monetary costs. This analysis is very important because it is unlikely that in reality the optimal strategy would be able to be implemented world-wide. In particular, we find that the cost of postponing mitigation is equivalent to giving up 5.5 USD - over a sixth of the global economy in 2015.

2 Related Work

In this chapter, a review of related works will be presented. Models combining both climate and economic modules together are called integrated assessment models. These models make large simplifications compared to the climate models used by climate scientists because such simplifications are required to make it computational feasible to evaluate many policy scenarios and compute optimal policies. Three such models are used by the Environmental Protection Agency to estimate the SCC: the DICE model (Dynamic Integrated Model of Climate and the Economy), Nordhaus (1992), Nordhaus (2008), Nordhaus and Sztorc (2013); the PAGE model (Policy Analysis of the Greenhouse Effect), Hope (2008); and the FUND model (Climate Framework for Uncertainty, Negotiation and Distribution), Anthoff and Tol (2012).

DICE refers to a family of models, for example the RICE model (Regional Integrated Climate-Economy) which models particular regions as opposed to the whole world. Like the EZ Climate model, the DICE model aggregates consumption and emissions globally (individual countries are not modelled separately), and mitigation of GHG emissions is assumed to reduce consumption today and increase it in the future (by reducing the effect of climate change). It discretizes time at 5- or 10-year periods. In Cai et al. (2012), a continuous-time version of DICE model was developed. The DICE model is deterministic. In Cai et al. (2015) a stochastic generalization with uncertainty in the economic and climate modules was presented. A constant elasticity utility function is typically used as the objective function in DICE. In Ackerman et al. (2013), an Epstein-Zin utility was used with DICE and the results were found to be insensitive to risk aversion. This highlights one common criticism of the DICE models: catastrophic risks are not well modelled with its polynomial damage function.

The PAGE and FUND model do not attempt to model the economy at the detail of the DICE models. Instead of fully modelling the effect of mitigation on the economy, they use economic scenarios developed elsewhere. FUND simulates 1950 to 3000 in 1-year time steps. In FUND, climate change damages are a function of both the change in global temperature and also the rate of change of the temperature. Damages lower not only consumption but also economic growth. The model PAGE is a stochastic model, 31 of its inputs are random variables. It must be run many times to generate a distribution of returned solutions.

3 Method

In Daniel et al. (2018), carbon emissions pricing is seen as a risk management problem. There is a trade-off between losing wealth or consumption today due to a reduction in emissions, and losing an uncertain and possibly quite large amount of consumption in the future due to the damages from climate change. In this chapter, we will begin by describing how consumption is modelled across time and the problem of finding the optimal trade-off will be formulated. Next, the model of emissions mitigation as a function of the SCC will be given. Additionally, the modelling of uncertain damages will developed. Lastly, we will devise a method of analyzing and assigning a monetary value to suboptimal solutions to this model.

3.1 Discrete Time Binomial Tree

The EZ Climate model runs from 2015 ($t = 0$) to 2400 ($t = 6$) discretized into 7 periods of unequal length. The time gaps between the earlier periods are shorter than the gaps later on. The model not only discretizes time but also states of nature as well. This is because the the effect of climate change on the future economy is uncertain. At time t, the world is in a certain state θ_t that captures the severity of the effect climate change has on the economy (the 'fragility' of the world). At any time, there are only two possible states the agent can enter at the next time step, θ^1_{t+1} and θ^2_{t+1}. This states are equally likely. The world is more fragile to climate change in state θ^1_{t+1} then in state θ^2_{t+1}. Thus, time periods and the states of the world are represented as nodes in a binomial tree, see Figure 3.1.

The model also defines a subinterval length of 5 years. Certain functions are recalculated at intervals equal to this length between periods so that they are sufficently smooth.

The first $N = 6$ time periods are decision periods. A rational representative agent makes decisions about the global climate strategy at these times. In particular, the agent controls the amount of mitigation (as a percentage of the GHG emissions if there were no intervention - the business-as-usual case) at every time step and possible state θ_t, x_{t,θ_t}. Note that the effect of the agent's strategy on the economy is path-dependent and so these levels, x_{t,θ_t},

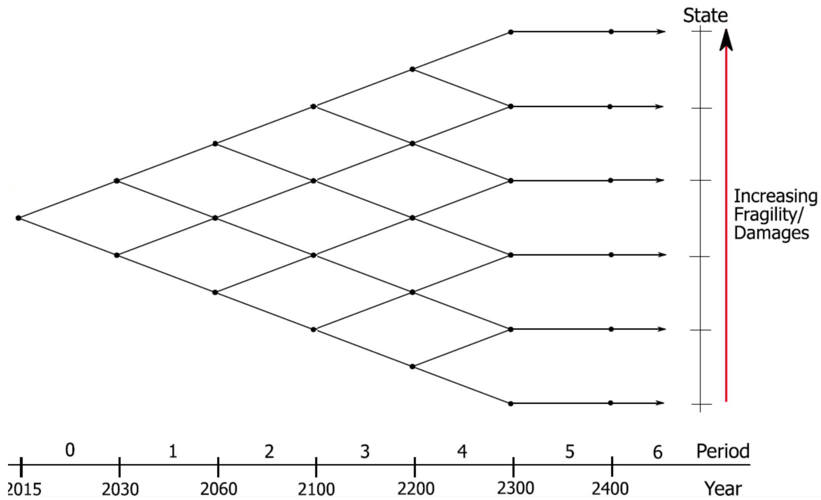


Figure 3.1: *The recombining (path-independent) tree of states of nature over time.*

rescind on a non-recombining binomial tree (see Figure 3.2) not a recombining one, and so the vector of all choices $\boldsymbol{x} = (x_{t,\theta_t})$ has length 63 when $N = 6$. At the final decision period N, it is assumed that all uncertainty about the effect of climate change is resolved.

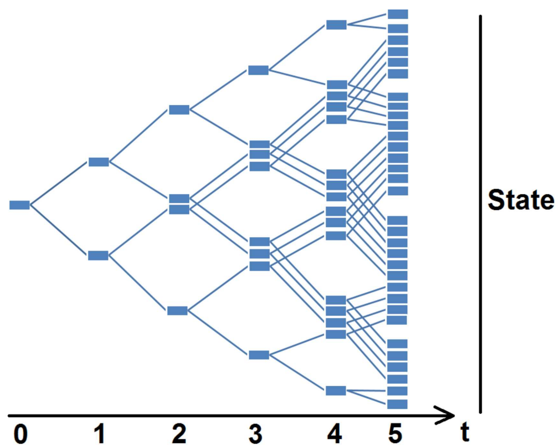


Figure 3.2: *Non-recombining (path-dependent) tree with 63 nodes.*

If $x_{t,\theta_t} = 0$ that means that there is no mitigation and the GHG emissions are what they would be without any climate mitigation policies. While $x_{t,\theta_t} = 0$ means that emissions are mitigated 100% - i.e. zero tons of GHG emissions. Mitigation can be above 100%, $x_{t,\theta_t} > 1$, due to 1) the natural absorption of GHGs from the atmosphere via oceans and trees, and 2) backstop technologies that remove GHGs out of the atmosphere.

3.2 Consumption & Utility

It is assumed in Daniel et al. (2018) that both climate change and the pricing of carbon effect consumption. A part of consumption is lost to damages from climate change (i.e. decreased crop production, loss of land, flooding, extreme weather events, etc.). The price of carbon shapes government climate policies (i.e. a carbon tax or cap-and-trade program). These measures reduce emissions, and thus the amount of consumption lost to damages, while also decreasing production of affected companies, lowering consumption. Thus, a part of consumption can be considered lost to mitigation of climate change.

The consumption at time period t (where $t = 0$ is today and $t = 1$ is one period of time later), c_t, is given as

$$c_0 = \bar{c}_0(1 - \kappa_0), \tag{3.2.1}$$

$$c_t = \bar{c}_t(1 - D_t)(1 - \kappa_t), \quad t \in \{1, 2, ..., N\}, \tag{3.2.2}$$

$$c_{N+1} = \bar{c}_{N+1}(1 - D_{N+1}). \tag{3.2.3}$$

where $\bar{c}_t$ is the consumption at time t if there was no emission pricing or further climate change damages, D_t is the percentage of $\bar{c}_t$ lost due damages from climate change (see Section 3.6), κ_t is the percentage of $\bar{c}_t$ lost due to mitigation policies (see Section 3.4). The term κ_t is deterministic while the damages term D_t is uncertain.

Today, at time $t = 0$, damages do not affect c_0 (Eqn 3.2.1) as any damages today are certain and already included in the current global consumption $\bar{c}_0$. This consumption can only be reduced by the mitigation choice made today by the agent.

The agent must chose the optimal set of mitigation levels, represented by the vector $\boldsymbol{x} = (x_{i,t}) \in \mathbb{R}^{2^N-1}$. The agent would like consumption to be as high as possible, but how does the agent weight the importance of consumption now vs consumption in the future and how concerned is the agent by worst-case scenarios? In economics, a utility function can be used to represent an agent's preferences and allow real numbers to be assigned to a set of options. The option with the maximum expected utility value should be chosen by the agent.

In this model an Epstein-Zin utility function (Epstein and Zin (1989)) is used. This is a recursive utility function; its value at a given time is a function of its value at the next time step. An important feature of Epstein-Zin utility is that it separates two aspects of preferences: risk aversion and intertemporal elasticity of substitution. The Epstein-Zin utility, U_t, of consumption is given by

$$U_0(\boldsymbol{x}) = \left[(1-\beta)c_0^{\rho} + \beta\left(\mathbb{E}_0\left[U_1^{\alpha}(\boldsymbol{x})\right]\right)^{\rho/\alpha}\right]^{1/\rho}, \tag{3.2.4}$$

$$U_t(\boldsymbol{x}) = \left[(1-\beta)c_t^{\rho} + \beta\left(\mathbb{E}_t\left[U_{t+1}^{\alpha}(\boldsymbol{x})\right]\right)^{\rho/\alpha}\right]^{1/\rho},\ t \in \{1, 2, ..., N\}, \tag{3.2.5}$$

$$U_{N+1}(\boldsymbol{x}) = \left[\frac{1-\beta}{1-\beta(1+r)^{\rho}}\right]^{1/\rho} c_{N+1}. \tag{3.2.6}$$

where $0 < \beta < 1$, $\alpha < 1$, $\rho < 1$, and r are parameters. The value $(1-\beta)/\beta$ is called the pure rate of time preference. The higher this value, the more importance the agent places on consumption now vs future consumption. The value $1/(1-\rho)$ is the elasticity of intertemporal substitution. The higher ρ is, the more willing the agent is to substitute consumption across time. The value $(1-\alpha)$ is the coefficient of risk aversion. The lower α is, the more averse the agent is to risk. In this report, we will use the values $\rho = -0.1111$, $\alpha = -9$, and $\beta = (1-0.005)^{\delta t}$ (where δt is the sub-interval length).

The term $\mathbb{E}_t\left[U_{t+1}^{\alpha}\right]$ is called the certainty-equivalent of future lifetime utility; it is the expected value of future utility conditioned on the information the agent has at time t. It is given by

$$\mathbb{E}_t\left[U_{t+1}^{\alpha}(x_{t,\theta_t})\right] = \frac{1}{2}\mathbb{E}_t\left[U_{t+1}^{\alpha}(x_{t,\theta_{t+1}^1})\right] + \frac{1}{2}\mathbb{E}_t\left[U_{t+1}^{\alpha}(x_{t,\theta_{t+1}^2})\right] \tag{3.2.7}$$

At the last time period $N+1$, we assume that all uncertainty on the damages of climate change have been resolved and that consumption simply grows at a constant rate r (we will use $r = 0.9752$).

The optimal $\boldsymbol{x}$ is the choice that maximizes expected utility of lifetime consumption. Thus, we must solve the following optimization problem,

$$\max_{\boldsymbol{x}} \mathbb{E}[U_0] \tag{3.2.8}$$

3.3 Price of Carbon $\tau(x)$

To compute the SCC given the optimal mitigation pathway we first specify a relationship between the fraction of emissions mitigated, x, and the marginal cost (from a societal perspective) of emissions mitigation, τ. In Daniel et al. (2018), this societal cost is used as a stand-in for a carbon taxation rate and, thus, the SCC.

It is assumed that x and τ have a power relation,

$$x(\tau) = c\tau^k. \tag{3.3.1}$$

The parameters c and k are set to fit $x(\tau)$ to a marginal abatement cost curve (MACC). One effort to estimate the MACC is the "Pathways to a Low-Carbon Economy" report from McKinsey & Company (Nauclér and Enkvist (2009)), see Figure 3.3. They used a bottom-up approach, estimating the cost of many individual mitigation technologies and measures. One criticism of the McKinsey report is that they include mitigation opportunities with negative costs (i.e. a positive x from a negative τ). We will use the McKinsey estimates with the modification that $x(0) = 0$ (there is no negative cost abatement) to calibrate $x(\tau)$. The point estimates that this function is fitted to are given in Table 3.1.

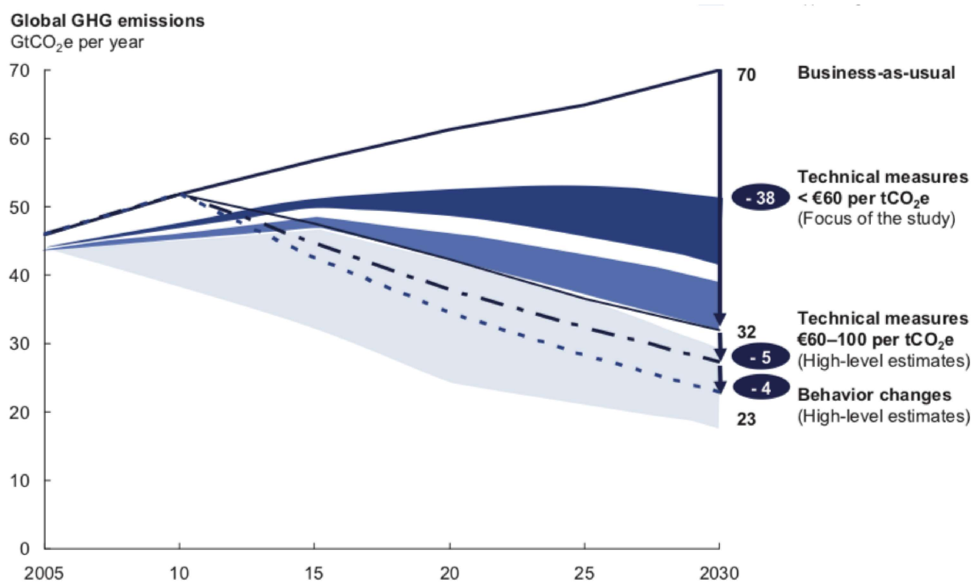


Figure 3.3: *Different GHG pathways and their marginal cost of abatement from Nauclér and Enkvist (2009).*

Table 3.1: *The modified McKinsey point estimates for 2030.*

GHG taxation rate, τ	Fractional GHG reduction, $x(\tau)$
€ 0 per tCO_2e	0
€ 60 per tCO_2e	0.543
€ 100 per tCO_2e	0.671

Thus, we should solve the following system of equations

$$\begin{cases} c \cdot 1.206 \cdot 60^k &= 0.543, \\ c \cdot 1.206 \cdot 100^k &= 0.671, \end{cases} \tag{3.3.2}$$

where 1.206 \$ per e is the exchange rate from 2005 euros to 2015 USD. This yields $c = 0.0923$ and $k = 0.414$. In other words we have

$$x(\tau) = 0.0923\tau^{0.414} \tag{3.3.3}$$

$$\tau(x) = 314.32x^{2.413} \tag{3.3.4}$$

Once we compute the optimal mitigation levels, we use equation 3.3.4 to compute the corresponding carbon prices at all nodes.

It must be stressed that there is a high degree of imprecision in this relationship. Firstly, this model is based on the assumption that mitigation and its cost have a power relationship. Beyond this, the McKinsey estimates themselves have a high degree of uncertainty. In Nauclér and Enkvist (2009), it is stated that key sources of uncertainty are the actual feasibility of implementing the abatement measures they include and the development cost of the abatement technologies. Furthermore, the McKinsey abatement costs are costs from a societal perspective and not a perfect stand-in for the SCC. The implications of the uncertainty of these parameters will be explored later in this report.

3.4 Mitigation cost (κ_t)

We must now define the fraction of consumption lost to mitigation, κ, as a function of the mitigation level x. Increasing the carbon tax rate will decrease consumption at a rate equal to emissions:

$$\frac{dc}{d\tau} = -g(\tau) \tag{3.4.1}$$

where g is the level of GHG emissions.

Solving this, we obtain

$$c(\tau) = \bar{c}_0 - \int_0^\tau g(s)ds \tag{3.4.2}$$

This equation is modified by the assumption that the proceeds from the GHG tax, $\tau g(\tau)$, are entirely refunded and added to the consumption.

$$\begin{aligned} c(\tau) &= \bar{c}_0 - \int_0^\tau g(s)ds + g(\tau)\tau && (3.4.3) \\ &= \bar{c}_0 - g_0\left(\tau x(\tau) - \int_0^\tau x(s)ds\right) && (3.4.4) \end{aligned}$$

where $g_0 = 52$ Gt CO_2e is the baseline level of GHG emissions in 2015. Thus, using equation 3.3.3,

$$\begin{aligned} c(x) &= \bar{c}_0 - g_0 92.08x^{3.413} && (3.4.5) \\ &= \bar{c}_0 - g_0 g x^a && (3.4.6) \end{aligned}$$

where $g = 92.08$, and $a = 3.413$. The cost of mitigation as a fraction of consumption is the change in consumption divided by current consumption:

$$\kappa(x) = \frac{g_0}{c_0} g x^a \tag{3.4.7}$$

where c_0 =\$31 trillion per year is the 2015 global consumption.

This fractional cost of mitigation $\kappa(x)$ is not a function of time and is parameterized only by the McKinsey estimates for current abatement measures. However, in the future one would expect cheaper abatement technologies to be developed, lowering the cost of abatement. In addition, backstop technology to remove GHGs directly from the atmosphere may be developed in the future. We will now discuss three kinds of technological change that will modify κ through time.

The first type of technological change is exogenous technological improvement, these are changes due to external forces (improvements independent of mitigation). All change of this type will be characterized by a single parameter, ϕ_0. The cost κ decreases every year by the constant percent φ_0. The second type of change is endogenous technological improvement, these are changes due to internal forces (improvements dependent on mitigation). It is parameterized by φ_1. The cost κ decreases every year by percent $\varphi_1 X_t$ where X_t is the average mitigation level to date. an dis given by

$$X_t = \frac{\sum_{s=0}^{t} g_s x_s}{\sum_{s=0}^{t} g_s} \tag{3.4.8}$$

where g_s is the GHG emissions in period s if there was no mitigation (in gigatons of CO_2-equivalent emissions per year).

The intuition behind the expression $\varphi_1 X_t$ is that if the average mitigation level, X_t has been high (due to a high taxation rate) then pressure on industries will result in faster development of mitigation technologies and alternative energy sources. This will cause the cost of mitigation to decrease.

The third kind of technological change is the development of backstop technology: technology that would allow GHGs to be directly removed from the atmosphere. The initial marginal cost of removal of one ton of CO2 is given by parameter $\tau_{B,0}$. This is the value of τ at which the marginal cost curve changes from the power function given above to a different shape, see Figure 3.4. Unlimited amounts of CO_2 can be removed at marginal cost $\tau_{B,} \geq \tau_{B,0}$, at this value $x(\tau)$ asymptotes. Note from Figure 3.4 that backstop technology can 'kick in' at $x^B > 1$. This is because GHGs are also removed from the atmosphere due to natural processes and, thus, mitigation levels above 100% can be achieved without the use of backstop technology. When the default parameters of $\tau_{B,}$ =\$2500 and $\tau_{B,0}$ = \$2000 are used, backstop technology is not used in the optimal solution.

The equation for the marginal cost of abatement with backstop technology included is

$$\tau(x) = \begin{cases} 314.42 \cdot x^{2.413}, & x \leq x^B, \\ \tau_{B,\infty} - \left(\frac{k}{x}\right)^{1/b}, & x > x^B, \end{cases} \tag{3.4.9}$$

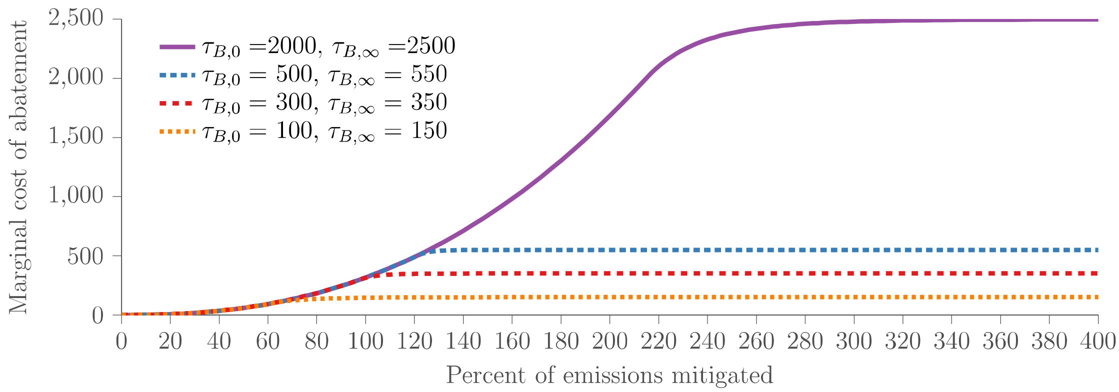


Figure 3.4: *The marginal cost of mitigation at time $t = 0$ for different values of $\tau_{B,\infty}$ and $\tau_{B,0}$. At price $\tau_{B,0}$ (or mitigation level x^B), the rate of marginal cost of abatement increase decreases. It asymptotes to price $\tau_{B,\infty}$.*

All together κ_t is

$$\kappa_t = \underbrace{(1-\varphi_0-\varphi_1 X_t)^t}_{\text{Technological Change Term}} \begin{cases} \kappa(x_t), & x_t \le x^B \\ \kappa(x^B) + \frac{g_0}{c_0} \underbrace{\left[(x_t - x^B)\tau_{B,\infty} - \frac{b}{b-1}\left(x_t\left(\frac{k}{x_t}\right)^{1/b} + x^B\left(\frac{k}{x^B}\right)^{1/b}\right)\right]}_{\text{Backstop Technology Term}}, & x_t > x^B. \end{cases} \tag{3.4.10}$$

where

$$x^B = \left(\frac{\tau_{B,0}}{g \cdot a}\right)^{\frac{1}{2.413}}, \qquad b = \frac{\tau_{B,\infty} - \tau_{B,0}}{2.413 \cdot \tau_{B,0}}, \qquad k = x^B(\tau_{B,\infty} - \tau_{B,0})^b \tag{3.4.11}$$

Note that κ_t depends not just on x_t, the mitigation level at that time, but $(x_s)_{s\le t}$, the path of mitigation levels taken. This is due to the average mitigation, X_t, term.

3.5 Temperature as function of GHG levels ($\Delta\mathbb{T}(G)$)

The global temperature change change, $\Delta\mathbb{T}$, is a random variable whose probability distribution depends on the GHG levels in the atmosphere, G. Wagner and Weitzman (2015) use estimates from the Intergovernmental Panel on Climate Change's (IPCC) Fifth Assessment and the International Energy Agency to assign probabilities of different values of eventual temperature change for different atmospheric concentrations of CO_2. Daniel et al. (2018) extrapolate these values to obtain the probabilities given in Table 3.2. The numbers in the table are $\mathbb{P}(\Delta\mathbb{T}_{100} > \mathbb{T})$, the probabilities that the change in temperature in 100 years ($\mathbb{T}_{100}$) will exceed $T = \{2, 3, 4, 5, 6\}^o$C, for three different scenarios of maximum atmospheric GHG concentrations: 450, 650, and 1,000 parts per million (ppm) of CO_2. These scenarios are correspond to strict, modest, and an ineffective mitigation scenarios.

We will assume that $\mathbb{P}(\Delta\mathbb{T}_{100} > \mathbb{T})$ follows a gamma distribution.

$$f(x;\alpha,\beta,\theta) = \frac{(x+\theta)^{\alpha-1}e^{-\frac{(x+\theta)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, \qquad x \geq -\theta \tag{3.5.1}$$

where $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}ds$ is the Gamma function. Fitting this model to the sets of values in Table 3.2, we obtain calibrations of the gamma distribution for the three mitigation scenarios, see Table 3.3.

Wagner and Weitzman (2015).

Table 3.2: *Point estimates of $\mathbb{P}(\Delta\mathbb{T}_{100} > \mathbb{T})$ for three scenarios of GHG levels.*

	Maximum GHG Level (ppm of CO_2)		
$\mathbb{T}$	**450**	**650**	**1,000**
2^oC	0.396	0.870	0.994
3^oC	0.139	0.566	0.910
4^oC	0.042	0.289	0.696
5^oC	0.011	0.124	0.443
6^oC	0.003	0.047	0.242

Table 3.3: *The gamma distribution parameters fitted to the three scenarios.*

	Gamma distribution parameters		
	450	**650**	**1,000**
$\boldsymbol{\alpha}$	2.810	4.630	6.100
$\boldsymbol{\beta}$	0.600	0.630	0.670
$\boldsymbol{\theta}$	−0.250	−0.500	−0.900

To obtain the temperature change at other times, $\mathbb{T}_t$, the following relationship is used

$$\Delta\mathbb{T}(G(X_t)) = 2\Delta\mathbb{T}_{100}(G(X_t))\left(1 - 0.5^{\frac{t}{100}}\right) \tag{3.5.2}$$

where $\Delta\mathbb{T}_{100}$ is a gamma random variable with parameters that depend on level of GHGs in the atmosphere (see Table 3.2). As time increases, the temperature change asymptotes to $2\Delta\mathbb{T}_{100}$.

3.6 Climate change damages (D_t).

In this section, we will specify the fraction of consumption lost to climate change damages, D_t. The cost of damages is given by

$$D_t = \underbrace{\left(1 - \bar{c}_t e^{-13.97\gamma \Delta\mathbb{T}(G(X_t))^2}\right)}_{\text{Non-catastrophic component}} \times \underbrace{\left[1 - \mathbb{1}_{TP}(1 - e^{-d_{TP}})\right]}_{\text{Catastrophic component}} \tag{3.6.1}$$

where $G(X_t)$ is the level of GHGs in the atmosphere. Variables $\Delta\mathbb{T}(G(X_t))$, γ, $\mathbb{1}_{TP}$, and d_{TP} are independent random variables.

The non-catastrophic component of damages is from Pindyck (2013). It was fitted using data from the IPCC's Fourth Assessment Report. The problem with using this component alone is that it does not include any risk of catastrophic damages due to climate change. To account for this, we will introduce $\mathbb{1}_{TP}$, a random variable.

$$\mathbb{1}_{TP} = \begin{cases} 1, & \text{if a “tipping point” has been reached and there are catastrophic damages} \\ 0, & \text{if the “tipping point” has not been reached.} \end{cases} \tag{3.6.2}$$

The probability of catastrophic damages (of hitting this “tipping point”) will be modelled as

$$\mathbb{P}(\mathbb{1}_{TP} = 1) = 1 - \left(1 - \left[\frac{\Delta\mathbb{T}(G(X_t))}{\max(\Delta\mathbb{T}(G(X_t)), \text{peakT})}\right]^2\right)^{\frac{n}{30}} \tag{3.6.3}$$

where peakT is a parameter; as the parameter peakT increases, the probability of having a catastrophic change decreases. The severity of these catastrophic damages are given by d_{TP}, a gamma random variable.

The stochastic of the damages D_t is modelled using the binomial tree. Monte-Carlo simulations of the damages function are run for four emissions scenarios: GHG concentrations of 450, 650, and 1,000 parts per million (ppm) of CO_2. For each scenario, 6,000,000 simulations are run. In each simulation, a set of random variables $\Delta\mathbb{T}(G(X_t))$, γ, $\mathbb{1}_{TP}$, and d_{TP} is drawn. The simulations are ordered based on their final damages value, D_N and grouped into $2^{N-1} = 32$ percentiles. The average damages of each group are the value of the N period's states of nature in the binomial tree for those emissions scenarios. Results are interpolated to generalize to other GHG levels.

3.7 Suboptimal Mitigation Strategies

In this section we explain our approach to studying suboptimal mitigation strategies. In particular, we would like to answer the question: “How much will it cost to employ a non-optimal mitigation strategy?”. For this reason we will use the suboptimality analysis which is well known in literature (Das and Uppal (2004), Liu and Pan (2003), Larsen and Munk (2012)).

The utility of any strategy x depends on the parameter of initial global consumption c_0.

$$\mathbb{E}[U(C)] = \mathbb{E}[U(C(x; c_0))].$$

Let x^* be the optimal mitigation strategy, and x^{sub} be any other strategy which will be referred to as a suboptimal strategy. Since the utility function is unit-less it is hard to say how bad the suboptimal strategy is. The only thing that we can say is that the strategy x^{sub} is worse than the optimal strategy because it yields lower expected utility. However, we do not know whether x^{sub} is nearly as good as the optimal strategy in some sense. To determine how poor a given suboptimal strategy is, we somehow have to compare expected utilities in dollar terms. This task turns out to be relatively straightforward. First, we evaluate the expected utility $\mathbb{E}[U(C(x^{sub}; c_0))]$ for a suboptimal strategy (by definition we have $\mathbb{E}[U(C(x^{sub}; c_0))] < \mathbb{E}[U(C(x^*; c_0))]$). Second, we solve for z in the following equation:

$$\mathbb{E}[U(C(x^{sub}; c_0))] = \mathbb{E}[U(C(x^*; c_0 - z))], \quad 0 \leq z \leq c_0 \tag{3.7.1}$$

Conceptually, the equation (3.7.1) says that acting suboptimally is equivalent to giving up z dollars of consumption today and then acting optimally. This gives us the monetary cost z of following a suboptimal strategy. A small z indicates that the suboptimal strategy is nearly as good as the optimal one.

Note that solving for z in (3.7.1) is a root finding problem for a nonlinear equation. It must be solved numerically using an iterative procedure. As a result, the suboptimality analysis described in this section could be quite time-consuming because it implies that the optimal strategy should be evaluated many times until (3.7.1) holds with an acceptable precision level. This further highlights the importance of the improved computational performance of the model developed in this paper.

4 Computational Issues and Implementation

The model described in Daniel et al. (2018) was originally coded in Python and is publicly available at GitHub (Litterman and Wagner (2013)). Our version of the EZ Climate model was coded in Matlab and the following chnages were made

- Vectorization: Computations of variables on the binomial tree were vectorized to take advantage of parallel computing. This reduced the time to evaluate the utility function by a factor of 20.
- Automatic Differentiation: We used the Matlab package ADMAT 2.0 (Coleman and Xu (2016)) to compute gradients of the utility function with automatic differentiation. The improved scaling behavior of reverse mode AD compared to the numerical differentiation used in the Python model meant that models with higher numbers of decision periods could be explored. Using 13 decision periods, our AD derivative computation was more than 30 times faster than the numerical differentiation.
- *Optimization method:* We use a Quasi-Newton method to solve for the optimal mitigation strategy, which is must faster than genetic algorithms used in the original implementation.

The first three improvements combined result in much faster computation of solutions. This allows analysis of the model that would otherwise not be feasible. In particular, models with more periods can be studied and a robust optimization approach can be explored. Next we discuss the last two improvements in more detail.

4.1 Automatic Differentiation (AD)

A first-order numerical differentiation (ND) method requires n function evaluations to compute the gradient of a function with respect to a n dimensional input. Thus, we say numerical differentiation has a relative time complexity of $O(n)$. The control variable $\boldsymbol{x}$ has dimension $2^N - 1$ and so computational time utility gradients with scales with the number of decision periods ($O\left(2^N - 1\right)$).This can be observed from the ND columns in Table 4.1. The time ratio is the time of derivative computation divided by the time of evaluating the utility function.

Table 4.1: *Time (in seconds) to compute gradients using numerical differentiation (ND), forward automatic differentiation (FD) and reverse automatic differentiation (RD). Utility time is the time it takes to evaluate the utility function. The times were measured on a PC with 64-bit Windows 7 OS.*

Decision Periods	Decision Nodes	Utility Time	ND Time	ND Ratio	FD Time	FD Ratio	RD Time	RD Ratio
6	63	0.005	0.31	67	0.33	71	0.61	132
7	127	0.007	0.79	114	0.47	68	0.81	117
8	255	0.011	2.36	255	0.79	75	1.24	118
9	511	0.017	8.07	464	1.98	114	2.08	120
10	1023	0.025	24.3	959	5.82	230	3.75	148
11	2047	0.039	73.5	1884	19.1	490	7.24	186
12	4095	0.066	252	3842	68.9	1049	14.7	224
13	8191	0.118	951	8041	245	2072	28.5	241

AD allows us to compute derivatives in forward and reverse modes. Like ND, FD also has a relative time complexity of $O\left(2^N - 1\right)$. From Table 4.1, FD is faster than ND. This is due to the vectorized derivative computation of ADMAT 2.0. On the other hand, the computation time of gradients with RD scales with the dimension of the function's output. The output of the expected utility of lifetime consumption is a scalar value, and so RD has a constant relative time complexity ($O(1)$). Notice that in Table 4.1 that the time ratio of derivative computation remains flat.

FD is generally the fastest method for gradient computations of models with fewer than 10 decision periods. This is because RD has greater computational overhead than FD. For models with more than 10 decision periods, the improved scaling of RD makes it the fastest. As the number of periods grows, the difference in speed becomes greater. For models with $N = 13$, RD is almost 30 times faster than ND and 8 times faster than FD.

The results of Table 4.1 are also illustrated in Figure 4.1.

4.2 Optimization

Our problem, given in Equation 3.2.8, is to find a mitigation strategy that maximizes the expected lifetime utility, $\mathbb{E}[U_0]$. The objective function is non-linear, continuous, and differentiable. To investigate the landscape of the objective function further, its value along a

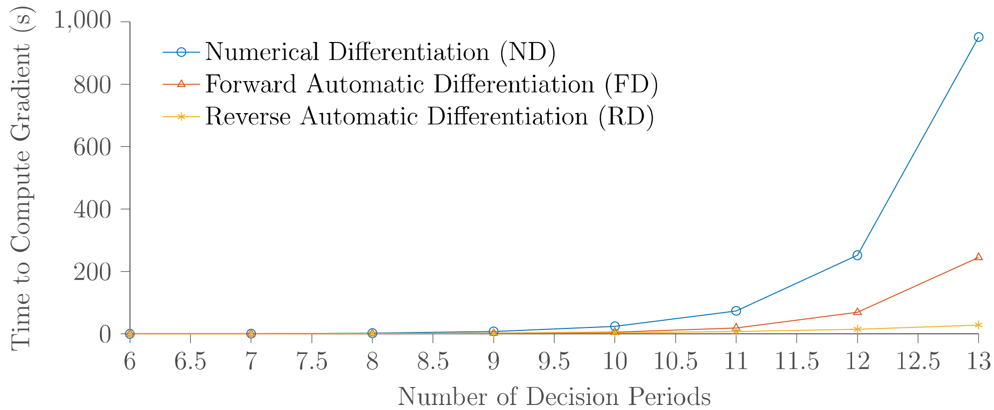


Figure 4.1: *Time of gradient computations using ND, FD, and RD.*

line through the space of mitigation strategies is shown in Figure 4.2. The line is the path of iterations the Quasi-Newton method (discussed more later in this section) took through the space while solving. The starting point of the line is the strategy with 100% mitigation at all nodes in the decision tree. The ending point is the optimal strategy found. Marked on the plot is a point at which the curvature of the objective function along the line appears to be negative.

The eigenvalues of the Hessian of $\mathbb{E}[U_0]$ at these three points are shown in Figure 4.3. At some points the Hessian has negative eigenvalues, confirming the existence of negative curvature. Thus, this is a non-convex optimization problem.

To find a global solution to such a problem, usually one must use a global optimization method. Such methods can be expensive due to the high number of function evaluations required. In Table 4.2, the optimal utility values found using different optimization methods are shown. Both derivative-free global methods and gradient-based local methods are included. All methods converged to solutions with similar utility value with the exception of the genetic algorithm which found a slightly worse solution. Global methods, such as pattern search and the genetic algorithm, required tens of thousands of function evaluations. The local methods, such as Quasi-Newton, found solutions with just as high a utility in run times several orders of magnitude less than the global methods.

A concern when using a local method to solve a non-convex problem is that it will converge to a local maxima instead of the global maxima. The Quasi-Newton method was used to solve the problem for over 200 hundred different initial points with random mitigation levels between 0% and 100%. The results of these runs are given in Table 4.3. The method always converged to a solution with utility value of around 9.791. When initial points with levels well outside this range were used (i.e. levels of -100%, 200%, etc), the method failed to converge. At such points, the utility value attained it's minimum, 0, and the norm of the gradient of the utility is almost zero. The method immediately stops and can make no progress at such points.

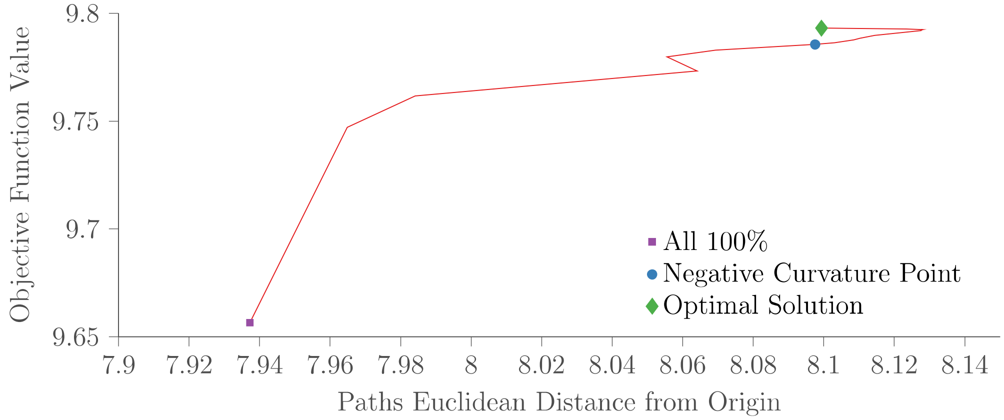


Figure 4.2: *The value of the objective function, $\mathbb{E}[U_0]$, for the 6 period model along the path taken by a Quasi-Newton method through the 63 dimensional space of mitigation strategies. The x axis is the distance of the points along the path from the origin (the business-as-usual case, all 0% mitigation levels). Three points have been marked on the plot: the starting point (all 100% mitigation levels), the final optimal point, and a point where the function's curvature appears to be negative.*

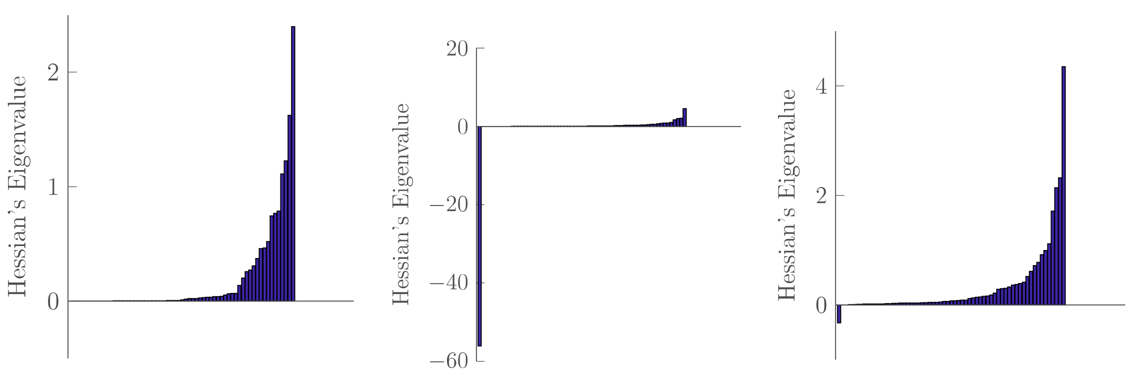


Figure 4.3: *The values of all eigenvalues of the Hessian of the objective function at the three marked points from Figure 4.2. From left to right: the point with all 100% mitigation levels, a point where the function's curvature appeared negative, and the optimal point.*

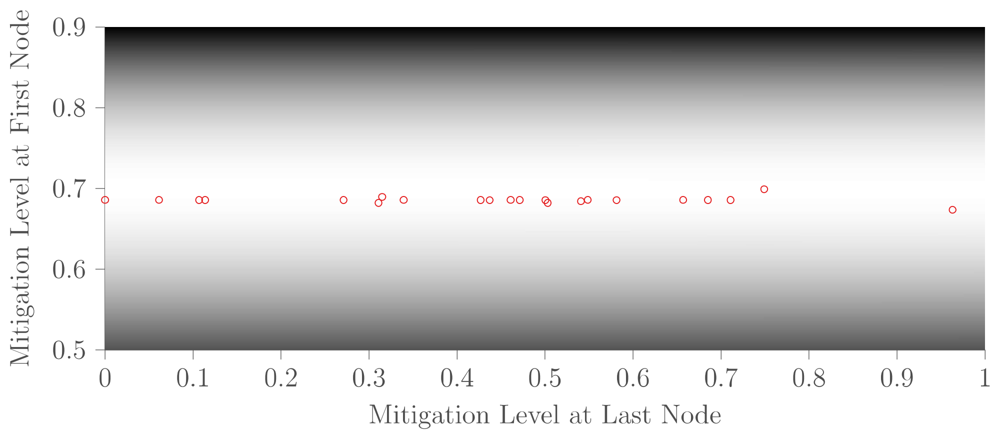


Figure 4.4: *The background colour of this image is the objective function for the optimal solution (found with QN and the all 100% starting point) with varying mitigation levels at just the first and last node – i.e. the decision made at 2015 and one of the possible decisions at 2300.The lighter the background colour, the higher the utility value is. The white strip is the center is where the utility is maximum. The red points are the values solutions found with QN and different random starting points had for these nodes.*

The Quasi-Newton method appears to find solutions with near-optimal utility for all reasonable starting points. But is it finding the same solution each time? In Figure 4.4, these solutions' mitigation levels at the first decision made in 2015 and one of the last decisions made in 2300 are plotted over a contour plot of utility value. This plot shows that the utility stays constant at its maximum value for a large range of mitigation levels assigned at (one of) the last decision and a small range for the first decision. The objective function is flat in a region around the optimum – i.e. the optimal choice of mitigation strategy is insensitive to decisions made at the later periods. This makes sense because while the first decision affects utility for all later periods at all states of nature, the value at one of the nodes in the last period only affects the last period in 3% of damage scenarios since it is one of 32 nodes at that time. This is why the QN method returns different solutions with the same utility value.

Note that global methods will have the same problem of finding different solutions with the same objective value due to the function's landscape. Since this issue is unavoidable, in the interest of computational speed the QN method will be used for the remainder of this report with the knowledge that solution variation in the last 2 periods is expected and not significant.

Table 4.2: *The maximum expected lifetime utility found using different optimization methods. In addition, the number of function evaluations and the run times of the algorithms. Run times are given in seconds and measured on a 2014 Macbook Air running Sierra in MATLAB 2017a (this will be the computing environment in which all run times are measured unless stated otherwise). The first three methods are derivative-free global methods, while the last two are local methods that use automatic differentiation to compute gradients (see section 4.1).*

Method	**Optimal Utility**	**Function Evaluations**	**Time (s)**
Nelder-Mead	9.791	8521	56.830
Genetic Algorithm	9.771	193600	1070.764
Pattern Search	9.792	126000	713.381
Quasi-Newton	9.791	29	11.266
Preconditionined NLCG	9.791	56	18.739

Table 4.3: *The results from the Quasi-Newton method with different initial mitigation strategies. The initials points used were a solution with with levels equal to 0 (the BAU case), all levels equal to 100%, and random points between these extremes. The results given for the random points are the averaged results from 200 different starting points.*

Initial Point	**Utility**	**Gradient Norm**	**Iterations**
All 0%	9.794	0.000	267
All 100%	9.793	8.491×10^{-5}	147
Random between 0% and 100%	9.793	-9.394×10^{-6}	203.6

5 Results

The optimal mitigation strategy computed using the default parameters is shown in Figure 5.1. It has a expected utility value of 9.791 and a SCC in 2015 of 133.13 USD.

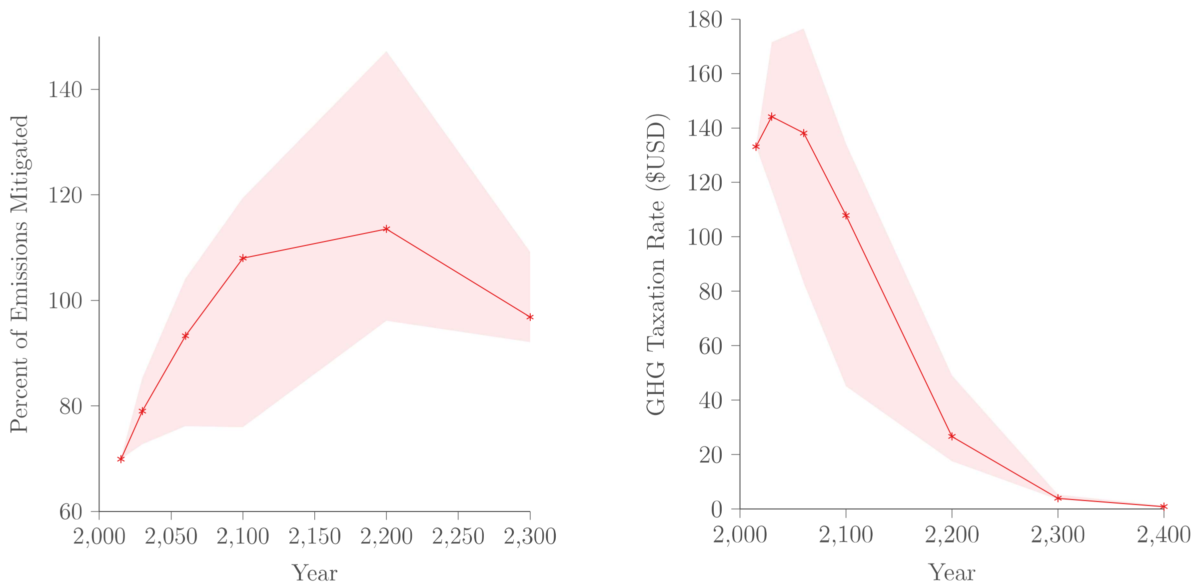


Figure 5.1: *The optimal mitigation strategy for base case parameter values (without jumps). The plot on the left shows the optimal mitigation levels averaged at each decision period, and the plot on the right shows the corresponding averaged taxation rates in 2015 $USD. The large shaded region encloses all of the optimal mitigation levels (for all states of nature represented by nodes on the tree) at a given period.*

It follows from Figure 5.1 that the optimal mitigation level (left panel) and the GHG taxation rate (right panel) can differ quite drastically depending on the state on the tree. For instance, in the year of 2,100, the optimal percent of emissions mitigated could be as low as 75% and as high as 115%. Similarly, the optimal GHG taxation rate for the same year could be as low as 45 USD and as high as 130 USD.

5.1 Model Sensitivity

The EZ Climate model makes many assumptions and simplifications. Complex systems are modelled using relatively simple parameterized equations. Some parameters can be estimated empirical but even they are often uncertain (i.e. the McKinsey abatement estimates). With other parameters it is even more unclear what their value should be (i.e. number of decision periods, utility preferences). It is important that a model for computing the SCC is robust under such parameter uncertainties. In this section we analyze the sensitivity of the model to different parameters.

Unless specified otherwise, we use the following base case specification. We assume that $N = 7$ implying that the first $N - 1 = 6$ periods are decision periods, during which the agent sets a GHG taxation rate. The parameters that determine the technological change in Equation (3.4.10) have values $\varphi_0 = 1.5$, $\varphi_1 = 0$, $\tau_{B,\infty} = \$2500$ and $\tau_{B,0} = \$2000$. To find the optimal mitigation strategy we use a Quasi-Newton method.

5.1.1 Effect of Mitigation Point Estimates

Recall that the McKinsey estimates, $x^{\text{McKinsey}}(€100)$ and $x^{\text{McKinsey}}(€60))$, are the estimated mitigation levels achieved with a taxation of €100 and €60 respectively. These estimates are used to derive parameters for the cost function and are highly uncertain. To investigate the model's sensitivity to these parameters we will find the optimal strategy for different values of these points estimates. To choose sensible alternate values, we assume that they come from independent normal distributions with their mean values given by Table 3.1. To calculate the variance, one assumes that the midpoint between $x^{\text{McKinsey}}(€60)$ and $x^{\text{McKinsey}}(€100)$ is three standard deviations away from either point.

$$\begin{aligned}\sigma &= \frac{1}{3}\left(\frac{x^{\text{McKinsey}}(€100) - x^{\text{McKinsey}}(€60)}{2}\right)\\ &= \frac{0.671 - 0.543}{6} = 0.0213.\end{aligned}$$

This guarantees that $\mathcal{P}(x^{\text{McKinsey}}(€100) > x^{\text{McKinsey}}(€60)) < 10^{-6}$. Let's define X_{McKinsey} to be the set of these parameters:

$$\begin{aligned}X_{\text{McKinsey}} &= [x^{\text{McKinsey}(€60)}, x^{\text{McKinsey}(€100)}] \sim \mathcal{N}(\mu, \Sigma)\\ \mu &= [0.543, 0.671]^{\text{T}}\\ \Sigma &= \begin{bmatrix} 0.0213^2 & 0 \\ 0 & 0.0213^2 \end{bmatrix}\end{aligned}$$

Solutions found using different draws of X_{McKinsey} are shown in Figure 5.2. The base case optimal solution is shown in blue. The optimal mitigation strategy varies drastically for different values of these parameters. The average utility value is 9.78 and ranges from 9.62 to 9.85. The average 2015 carbon price is \$129.43 and ranges from \$116.82 to \$142.67.

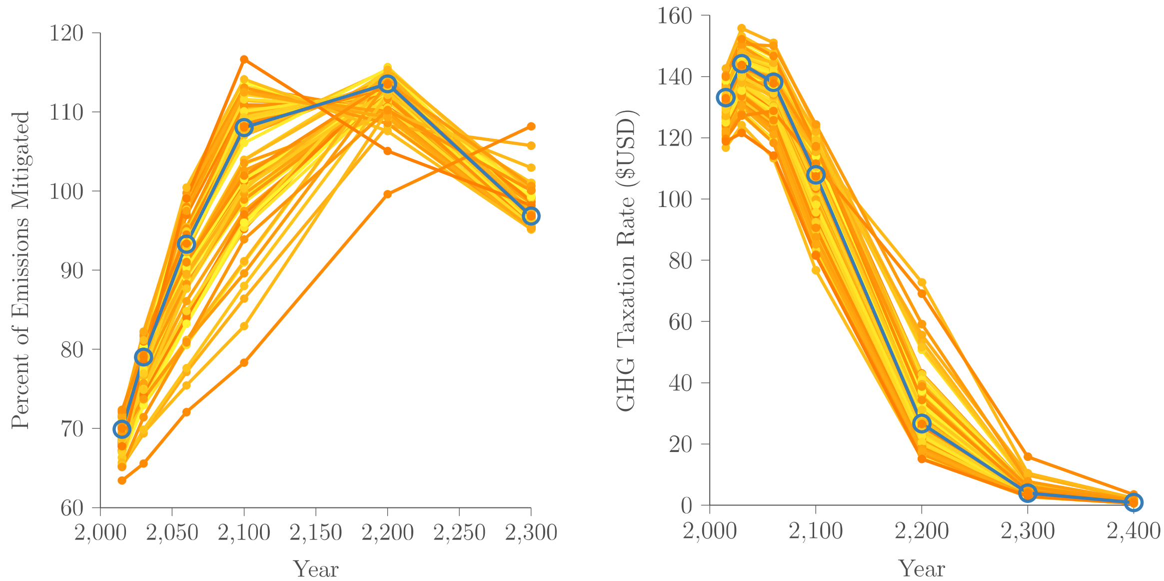


Figure 5.2: *The optimal mitigation levels (left) and GHG taxation rates (right), averaged at each time period, given draws from* $X_{McKinsey} = [x^{McKinsey(€60)}, x^{McKinsey(€100)}] \sim \mathcal{N}(\mu, \Sigma)$. *The blue line is the solution when using the mean values. The other lines are coloured based on our distance from the mean: the darker/more red the line the further from the mean.*

5.1.2 Effect of Technology

As explained in Section 3.4, the cost of mitigation at time t, κ_t, may change in time due to technological change (endogenous and exogenous) and backstop technologies. In this section we explore how the optimal solution depends on our assumptions about the technologies.

Effect of Exogenous Technological Change

The parameter φ_0 in (3.4.10) gives the percent of yearly cost decrease due to exogenous technological improvement, that is, the improvement independent of climate mitigation actions. The optimal solutions for the model with values of φ_0 ranging from 0 to 3% are shown in Figure 5.3.

The following conclusions can be drawn from Figure 5.3. First, lower values of φ_0 resulted in less volatile mitigation levels over time. This is due to the fact that the mitigation cost changes less over time for lower φ_0. Second, higher φ_0 values yield lower carbon taxes at later periods, and higher mitigation levels. This is because the higher φ_0 is, the lower the

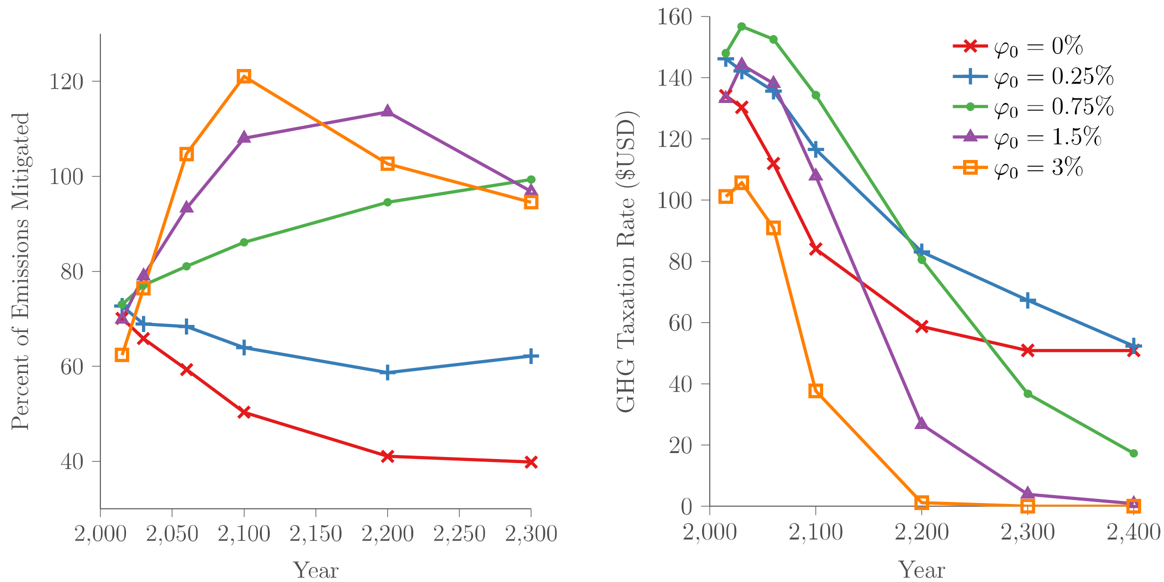


Figure 5.3: *The optimal mitigation levels (left) and GHG taxation rates (right), averaged at each time period for different values of φ_0 – the parameter representing exogenous technological improvement.*

mitigation cost at later times will be. Third, the solution is more sensitive to a change of φ_0 when φ_0 is low.

The initial period GHG taxation rate and the expected lifetime utility values at the optimal solutions for the analyzed values of φ_0 are provided in Table 5.1.

Table 5.1: *The expected utility values at the optimal solution for different values of φ_0*

φ_0	**2015 Carbon Price**	**Expected Utility**
0%	\$ 134.10	9.302
0.25%	\$ 146.19	9.366
0.75%	\$ 147.98	9.548
1.5%	\$ 133.13	9.791
3%	\$ 101.15	10.053

It follows from Table 5.1 that higher values of φ_0 resulted in solutions with a higher expected utility. This is because a higher φ_0 means a lower cost of abatement, and thus a higher fraction of endowed consumption can be retained by the agent.

Effect of Endogenous Technological Change

The parameter φ_1 in (3.4.10) describes yearly cost decrease due to endogenous technological improvement, that is, the improvement that depends on the level of climate mitigation. The optimal solutions for the model with φ_1 varying from 0 to 3% are shown in Figure 5.4.

The following observations are worth noting in Figure 5.4. The higher the value of φ_1, the earlier the mitigation level peaks before decreasing. The early high mitigation levels result in a higher average mitigation later and thus a greater reduction of marginal cost due to endogenous technology. Greater φ_1 values exacerbate this in the sense that for a high φ_1 it is beneficial to start with extremely high mitigation levels to drive down the future price.

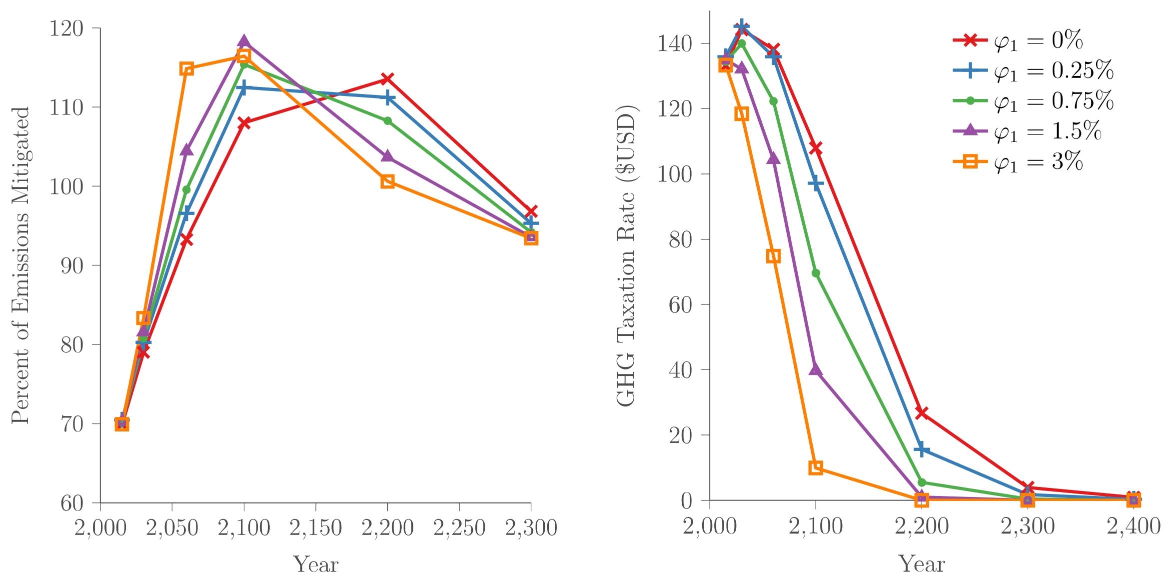


Figure 5.4: *The optimal mitigation levels (left panel) and GHG taxation rates (right panel), averaged at each time period for different values of φ_1 – the parameter representing endogenous technological improvement.*

The initial period GHG taxation rate and the expected lifetime utility values at the optimal solution for different values of φ_1 are shown in Table 5.2.

Table 5.2: *The expected utility values at the optimal solution for different values of φ_1*

φ_1	**2015 Carbon Price**	**Expected Utility**
0%	$ 133.13	9.791
0.25%	$ 135.84	9.843
0.75%	$ 134.46	9.930
1.5%	$ 134.98	10.023
3%	$ 133.28	10.133

It follows from Table 5.2 that higher values of φ_1 resulted in solutions with a higher expected utility. This is because φ_1 results in lower mitigation costs, and thus a higher fraction of endowed consumption can be retained by the agent.

Effect of Backstop Technology

In this section we explore the effect of backstop technology on mitigation cost. The optimal solutions for the model with varying values of $\tau_{B,\infty}$ and $\tau_{B,0}$ are given in Figure 5.5.

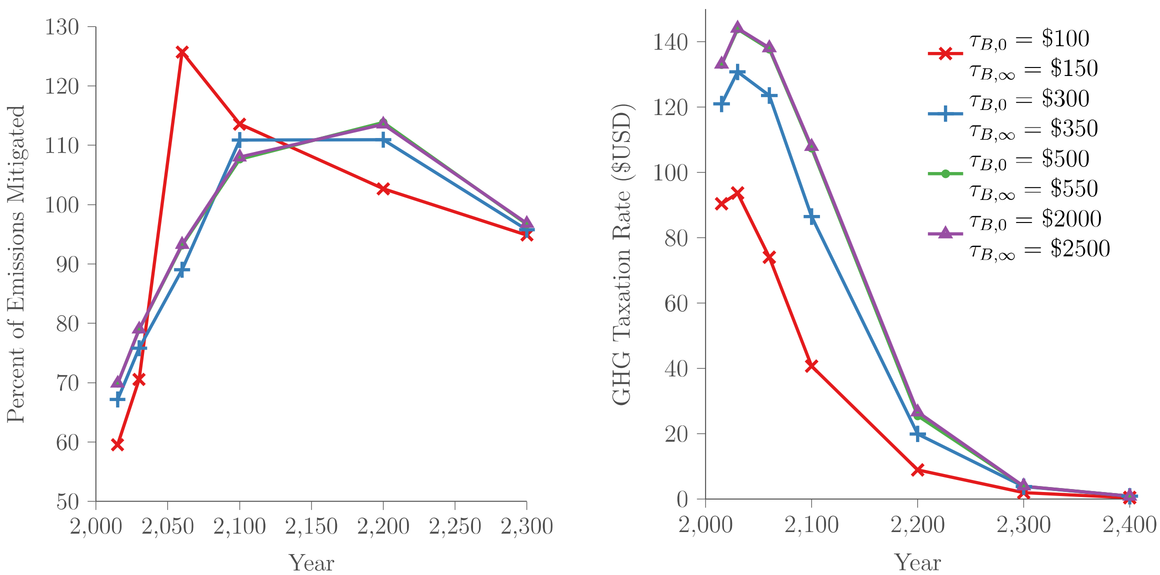


Figure 5.5: *The optimal mitigation levels (left panel) and GHG taxation rates (right panel), averaged at each time period for different values of* $\tau_{B,\infty}$ *and* $\tau_{B,0}$ *– the parameters representing backstop technology. Note that the the solution when* $\tau_{B,0}$ *= $500 (the green line) is exactly underneath the* $\tau_{B,0}$ *= $2000 solution (the purple line). With large enough* $\tau_{B,0}$ *and* $\tau_{B,}$*, backstop technology is not used and solutions stop varying with these parameters.*

We have the following observations:

- The base case scenario is $\tau_{B,\infty} = 2500$ and $\tau_{B,0} = 2000$. For this specification, the mitigation level at which backstop technology is used is $x^B = 2.15$ and it is high enough that the technology is not used if the optimal mitigation strategy is implemented.
- The moderate case is $\tau_{B,\infty} = 350$ and $\tau_{B,0} = 300$, with $x^B = 0.98$. In this case, backstop technology is used to implement the optimal solution. The optimal solution has slightly lower prices and higher mitigation levels in the fourth period compared to the base case.
- The aggressive case corresponds to $\tau_{B,\infty} = 150$ and $\tau_{B,0} = 100$ implying that backstop technology is used for $x \geq x^B = 0.62$. For this specification, high abatement is so cheap that optimal mitigation levels increase quickly to higher values compared with the other cases.

The initial period GHG taxation rate and the expected lifetime utility values for the analyzed parameter values are provided in Table 5.3.

Table 5.3: *The GHG taxation rate and the expected utility values at the optimal solutions found given different values of $\tau_{B,\infty}$ and $\tau_{B,0}$*

$\tau_{\mathbf{B},\infty}$	$\tau_{\mathbf{B},0}$	**2015 Carbon Price**	**Utility Value**
\$ 100	\$ 150	\$ 90.45	9.948
\$ 300	\$ 350	\$ 120.99	9.803
\$ 500	\$ 550	\$ 132.96	9.791
\$ 2000	\$ 2500	\$ 133.13	9.791

From Table 5.3, the aggressive case ($\tau_{B,\infty} = 150$ and $\tau_{B,0} = 100$) implies a very low GHG taxation rate.

5.1.3 Number of Decision Periods and Subinterval Lengths

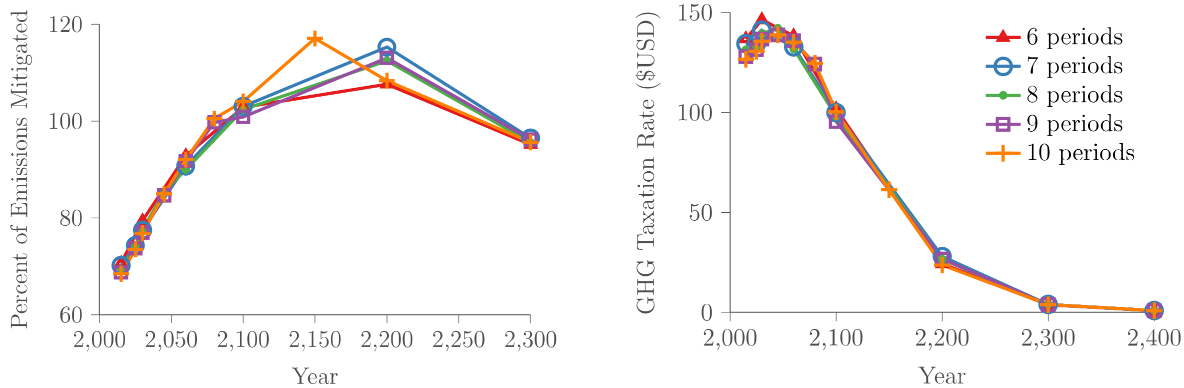
The model assumes that decisons about GHG emissions mitigation can only take place at a discrete set of times. In addition, consumption is recalculated at subintervals between the time periods to smooth the consumption function. The base model uses 6 decision periods and subintervals lengths of 5 years. In this section we explore how decision periods and subinterval lengths impact the optimal mitigation levels.

The optimal mitigation strategies for the model with a range of decision periods and subinterval lengths are shown in Figure 5.6. It should be noticed that there is little qualitative difference in the results and most solutions can be described as the solutions for which mitigation levels start just below 70%, increase until around year 2200, and then decrease. The only outlier is the 10 period model where levels peak in 2150 and then decrease. In addition, the peak mitigation level of this model is higher than the others. This is due to the fact that all of the other models have a large gap in decisions here. If a model with fewer periods had a decision at or around 2150, its solution would look more like the 10 period solution (see Figure 5.7).

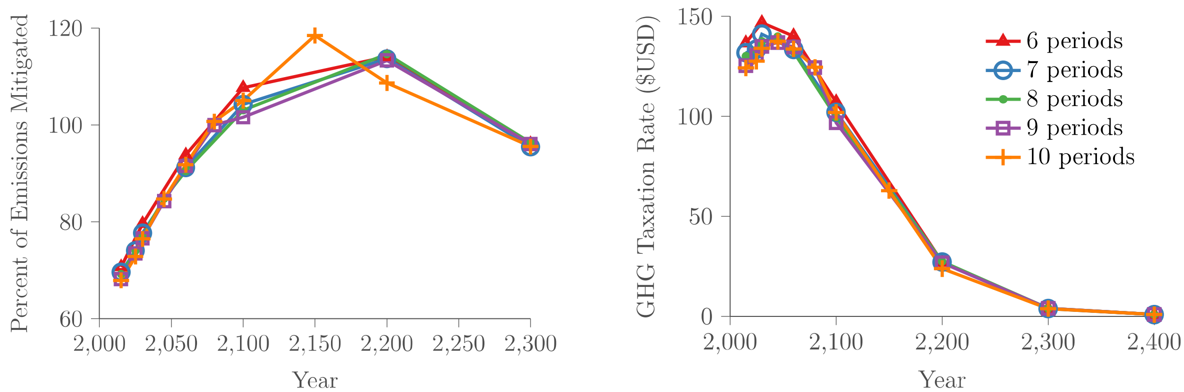
The initial period GHG taxation rate, the expected lifetime utility values, and optimization-specific information for the optimal solutions in Figure 5.6 are provided in Table 5.4.

It follows from Table 5.4 that utility increases with the number of decision periods. Intuitively, more decision periods mean more flexibility in the overall mitigation strategy. In addition, smaller subinterval lengths yield solutions with higher utility value. Indeed, the finer the grid on which values are computed – i.e., the more accurate the utility calculation – the better the optimal solution will be.

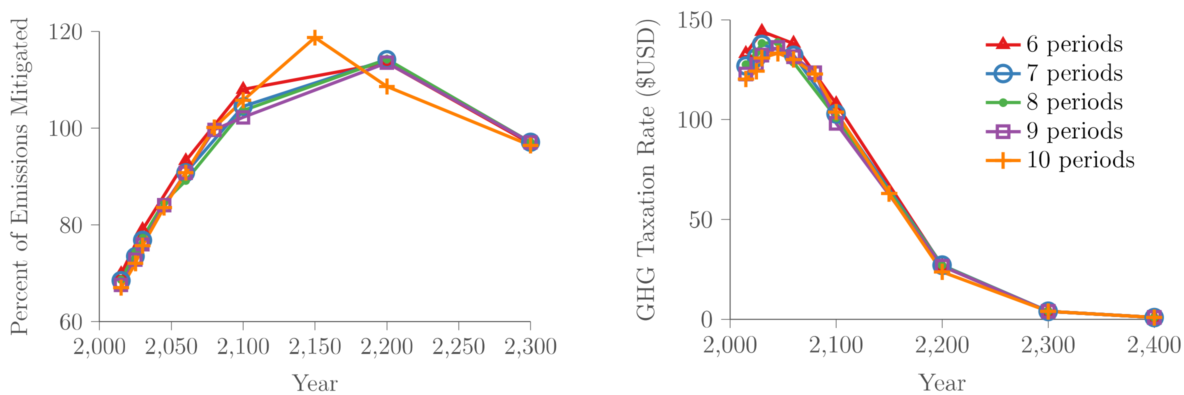
In addition, computation time goes up as the number of decision periods is increased and



(a) *Subinterval length of 1 year.*



(b) *Subinterval length of 2.5 years.*



(c) *Subinterval length of 5 years.*

Figure 5.6: *The optimal mitigation levels and GHG taxation rates, averaged at each time period for models with different numbers of decision periods (number shown in legend).*

Table 5.4: *The initial period GHG taxation rate and the expected utility values at the optimal mitigation levels for different number of decision periods and subinterval lengths. Additionally, the number of Quasi-Newton iterations and the computation times in seconds are given.*

(a) For a subinterval length of 1 year

Decision periods	2015 Carbon Price	Utility Value	Iterations	Time
6	$ 136.74	11.905	55	129.594 s
7	$ 134.37	11.916	35	104.097 s
8	$ 131.43	11.937	68	295.545 s
9	$ 127.89	11.963	77	572.210 s
10	$ 126.61	12.009	90	825.988 s

(b) For a subinterval length of 2.5 years

Decision periods	2015 Carbon Price	Utility Value	Iterations	Time
6	$ 136.05	10.459	30	33.471 s
7	$ 130.95	10.469	40	51.835 s
8	$ 130.12	10.486	43	86.051 s
9	$ 126.62	10.510	54	208.759 s
10	$ 123.11	10.548	80	914.755 s

(c) For a subinterval length of 5 years

Decision periods	2015 Carbon Price	Utility Value	Iterations	Time
6	$ 133.13	9.791	27	17.567 s
7	$ 126.82	9.799	34	28.511 s
8	$ 127.64	9.815	30	51.881 s
9	$ 122.85	9.838	54	220.922 s
10	$ 120.31	9.875	76	491.978 s

the subinterval length is decreased. This is because the function becomes more expensive to evaluate, and the number of iterations required to solve the optimization problem increases. More iterations are required due to the increased number of variables.

It should be noticed that the time at which decisions are made (not just the number of decisions) can also impact the solution. To demonstrate this, let us consider the following three different decision time selections for the 6-period model:

Choice 1: Decisions are made at 2015, 2030, 2060, 2100, 2200, and 2300.

Choice 2: Decisions are made at 2015, 2070, 2130, 2185, 2245, and 2300.

Choice 3: Decisions are made at 2015, 2020, 2025, 2045, 2105, and 2300.

The last (non-decision) period occurs at 2400 for all cases. The first choice of decision times above are the times used in the base case model which has decisions spaced closer early on. The second choice uses approximately evenly spaced decision periods rounded to the nearest multiple of 5 years. The third choice has approximately exponential spacing between the decisions so that decisions are concentrated at earlier times.

The resulting solutions, using a subinterval length of 5 years, are shown in Figure 5.7 and the corresponding utility values are given in Table 5.5.

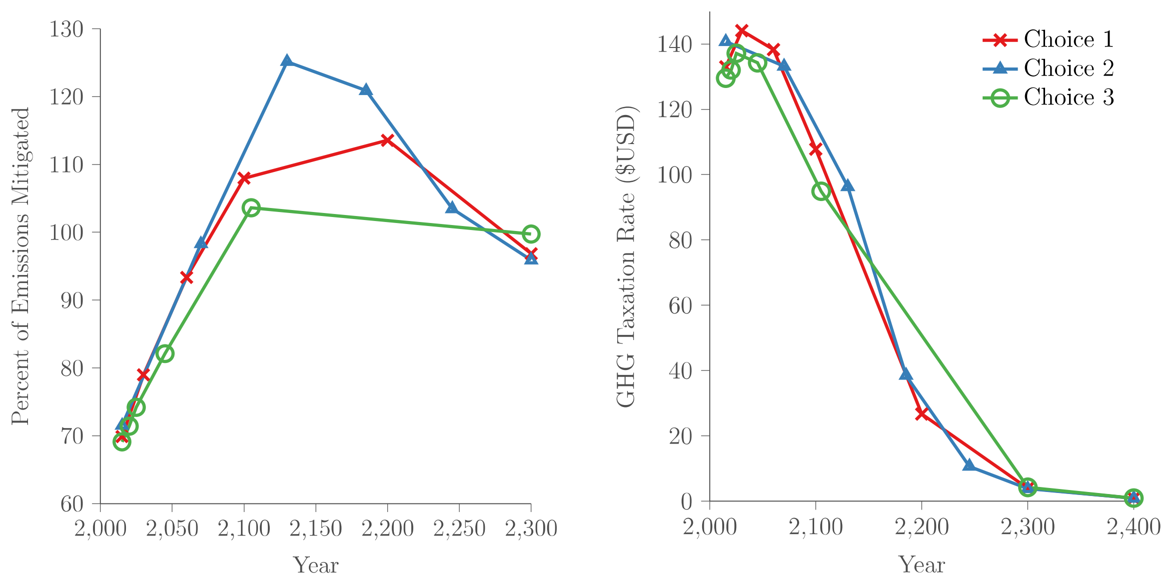


Figure 5.7: *The optimal mitigation levels (left panel) and GHG taxation rates (right panel), averaged at each time period for the three choices of decision times. The legend indicates which of the numbered choices for decision times the solution corresponds to.*

It follows from Table 5.5 and Figure 5.7 that evenly spaced decision times (Choice 2) resulted in a solution with a lower utility value and higher mitigation levels at intermediate periods than for the base case model (Choice 1). This implies that decisions at earlier times have

Table 5.5: *The expected utility values and norm of its gradient at the optimal solutions found for the three choices of decision times.*

Choice of decision times	Utility Value	Norm of Gradient
1	9.790	0.003
2	9.760	0.006
3	9.736	0.004

a greater impact on the utility. Thus, having fewer early decisions on mitigation will mean that higher mitigation levels are needed at later times to counteract the long early period of low mitigation. Exponentially spaced decision times (Choice 3) resulted in a solution with lower utility value and mitigation levels at intermediate periods. While having more early decisions is better, having very large gaps between decisions, even at later times, restricts the maximum possible utility the agent can achieve.

5.2 Robust Mitigation Policies

As seen in Figure 5.2, the model is very sensitive to any deviation of the McKinsey estimates, $p = [x^{\text{McKinsey(€60)}}, x^{\text{McKinsey(€100)}}]$. If the real point estimates are X^1_{McKinsey}, but we compute a mitigation strategy using the values X^2_{McKinsey}, how bad will our solution be? In other words, what is the 'cost' of using the suboptimal solution x^{subop} found with X^2_{McKinsey} compared to using the optimal x^* when X^1_{McKinsey}. To answer this question we can use the suboptimality analysis framework developed earlier.

We say that the 'cost' of using a sub-optimal solution, x^{subop} compared to using the optimum, x^*, is the value z such that $E[U_0(x^{\text{subop}}; c_0; p_1)] = \mathbb{E}[U_0(x^*; c_0 - z; p_1)]$, where c_0 is the global consumption in 2015.

The cost of using the base case solution found using the McKinsey estimates compared to all other solutions shown in Figure 5.2 ranged from 0% of the 2015 consumption to 51.3%. The average was 8.5%. Using this solution is incredibly costly in the worse case scenarios.

We can define a robust solution to SCC problem as the solution which minimizes the expected value of the loss, z, of using it compared to using solutions found with different draws of X_{McKinsey}. Our optimization problem is as follows:

$$\begin{aligned}
\min_{x^{\text{robust}}} \quad & \mathbb{E}[z] \\
\text{such that} \quad & \mathbb{E}[U_0(x^{\text{robust}}; c_0; X_{\text{McKinsey}})] = \mathbb{E}[U_0(x^*; c_0 - z; X_{\text{McKinsey}})], \\
& x^* = \min_x -\mathbb{E}[U_0(x; c_0; X_{\text{McKinsey}})], \\
& X_{\text{McKinsey}} \sim \mathcal{N}(\mu, \Sigma)
\end{aligned} \tag{5.2.1}$$

In practice, the expectation of z is computed as the average over a large set of z_i computed from different draws of $X^i_{\text{McKinsey}} \sim \mathcal{N}(\mu, \Sigma)$.

Instead of minimizing the expected loss of a mitigation strategy under these uncertain parameters, we might want to minimize the conditional value at risk (CVaR) instead:

$$\begin{aligned}
\min_{x^{\text{robust}}} \quad & \text{CVaR}_\alpha(z) \\
\text{such that} \quad & \mathbb{E}[U_0(x^{\text{robust}}; c_0; X_{\text{McKinsey}})] = \mathbb{E}[U_0(x^*; c_0 - z; X_{\text{McKinsey}})], \\
& x^* = \min_x -\mathbb{E}[U_0(x; c_0; X_{\text{McKinsey}})], \\
& X_{\text{McKinsey}} \sim \mathcal{N}(\mu, \Sigma)
\end{aligned} \tag{5.2.2}$$

where $\text{CVaR}_\alpha(z) = \mathbb{E}[z | z > \text{VaR}_\alpha(z)]$ and $\Pr[z > \text{VaR}_\alpha] = \alpha$. Again in practice CVaR_α is approximated as the average of the worst α percent of losses from a large set of z_i. We will use $alpha = 5\%$.

Solving either of these stochastic optimization problems is very costly even with the improved utility evaluation time and optimization run time. This is because computing the value of the objective function of these problems requires many draws of $X_{McKinsey}$ and for each draw a utility maximization problem must be solved and the root of a nonlinear equation must be computed. For the sake of computational feasibility, we will use a fixed set of one hundred $X^i_{McKinsey} \sim \mathcal{N}(\mu, \Sigma)$ and compute the maximal utility solution x^*_i for each. This set will be reused for every evaluation of the robust optimization problem.

The two robust solutions were found and are shown in Figure 5.8. The utility value (computed using the base case McKinsey values) of these solutions and the base case and mean solution along with their average loss and tail loss are given in Table 5.6. The mean of all solutions found with different draws of $X_{McKinsey}$ is not very good: it has a high expected loss and CVaR. The minimized expected loss is around 8% of 2015 global consumption which is quite close to the average loss of the base case solution. The robust solution with minimized expected loss and the base case solution are quite similar and both have a high CVaR.

The robust solution with minimized CVaR differs from the base case solution. The minimized CVaR is 29% of 2015 global consumption – 10% percent lower than the tail loss for the base case solution. This solution's expected loss is almost 16% – double the loss of the min expected loss robust solution.

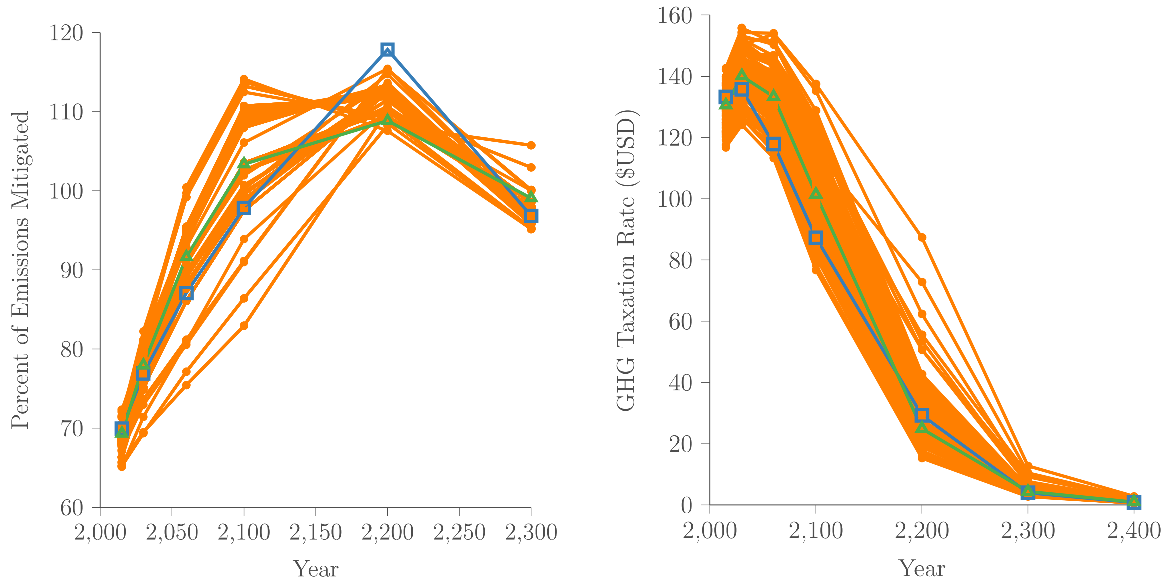


Figure 5.8: *The solutions found using different draws of* $X_{McKinsey} = [x^{McKinsey(€60)}, x^{McKinsey(€100)}] \sim \mathcal{N}(\mu, \Sigma)$ *are in orange. The solution in blue with square points is the robust solution that minimizes expected loss. The solution in green with triangle points is the robust solution that minimizes the CVaR.*

Table 5.6: *A comparison of the different robust solutions. The original solution is the optimal strategy given the McKinsey point estimate values. The mean solution is the average of all one hundred solutions found with different draws of* $X_{McKinsey}$. *The min expected loss solution is the robust solution found when minimizing the expected loss and the min CVaR solution is the one found when minimizing the CVaR. The expected utility values were computed using the mean point estimates. Note that the expectations and CVaR values are approximate averages and are given in terms of percent of initial GDP lost, not a dollar amount.*

Mitigation Strategy	**Utility Value**	$\mathbb{E}[z]$	**CVaR**(z)
Orginal Solution	9.791	0.085	0.380
Mean Solution	9.789	0.116	0.364
Min Expected Loss	9.792	0.080	0.393
Min CVaR	9.780	0.159	0.289

5.3 Suboptimal Mitigation Policies

5.3.1 Cost of Fewer Decisions

To determine the monetary value of using a model with more decision periods, we will compare the optimal solution x^* found for a model with N decision periods to a suboptimal mitigation strategy x^{sub}. The suboptimal strategy used is the optimal solution found for a model with $n \leq N$ decision periods scaled up to fit the N period tree.

An example of how this is done is shown in Figure 5.9. Consider the period at 2100: if the solution computed for the n model was deployed then the mitigation level at 2100 would be equal to the value at 2060 since that was when the last decision was made. This value would depend on the state of nature at 2060. Thus, the strategy shown for the N has the same levels at the same times, given the nature state in the same percentile. Plots of the optimal solutions for models with $N = 7, \ldots, 10$ periods along with suboptimal solutions are shown in Figure 5.10.

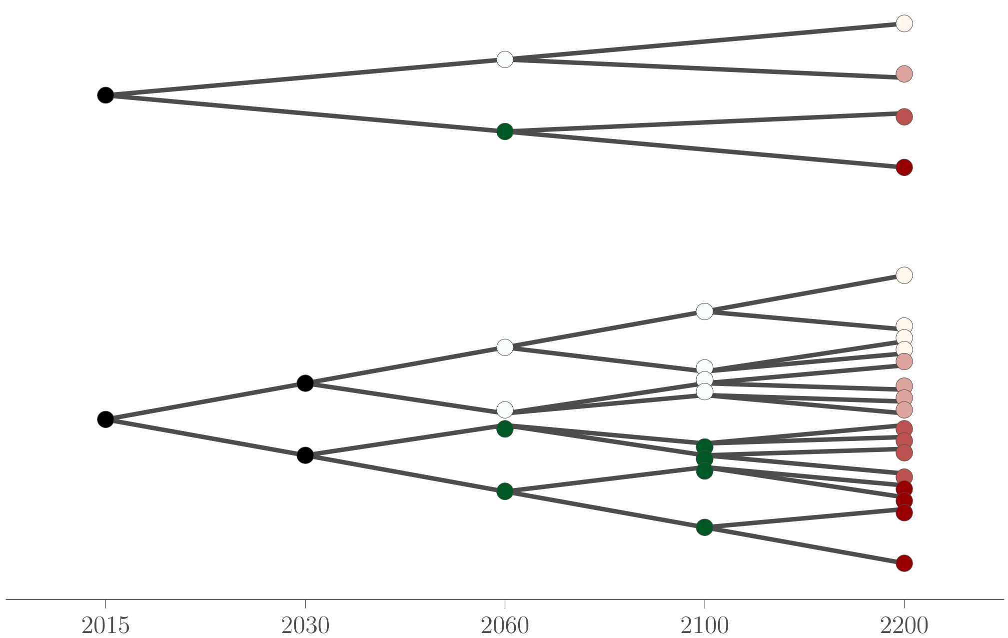


Figure 5.9: *The decision trees for two models. The top with $n = 3$ decision periods and the bottom with $N = 5$. The solution from the n period model is used in the N model. Nodes with the same colour have the same mitigation level.*

The cost of these suboptimal mitigation strategies is shown in Table 5.7. The utility value of the optimal 6 period solution as measured when scaled up to a 10 period tree is equal to

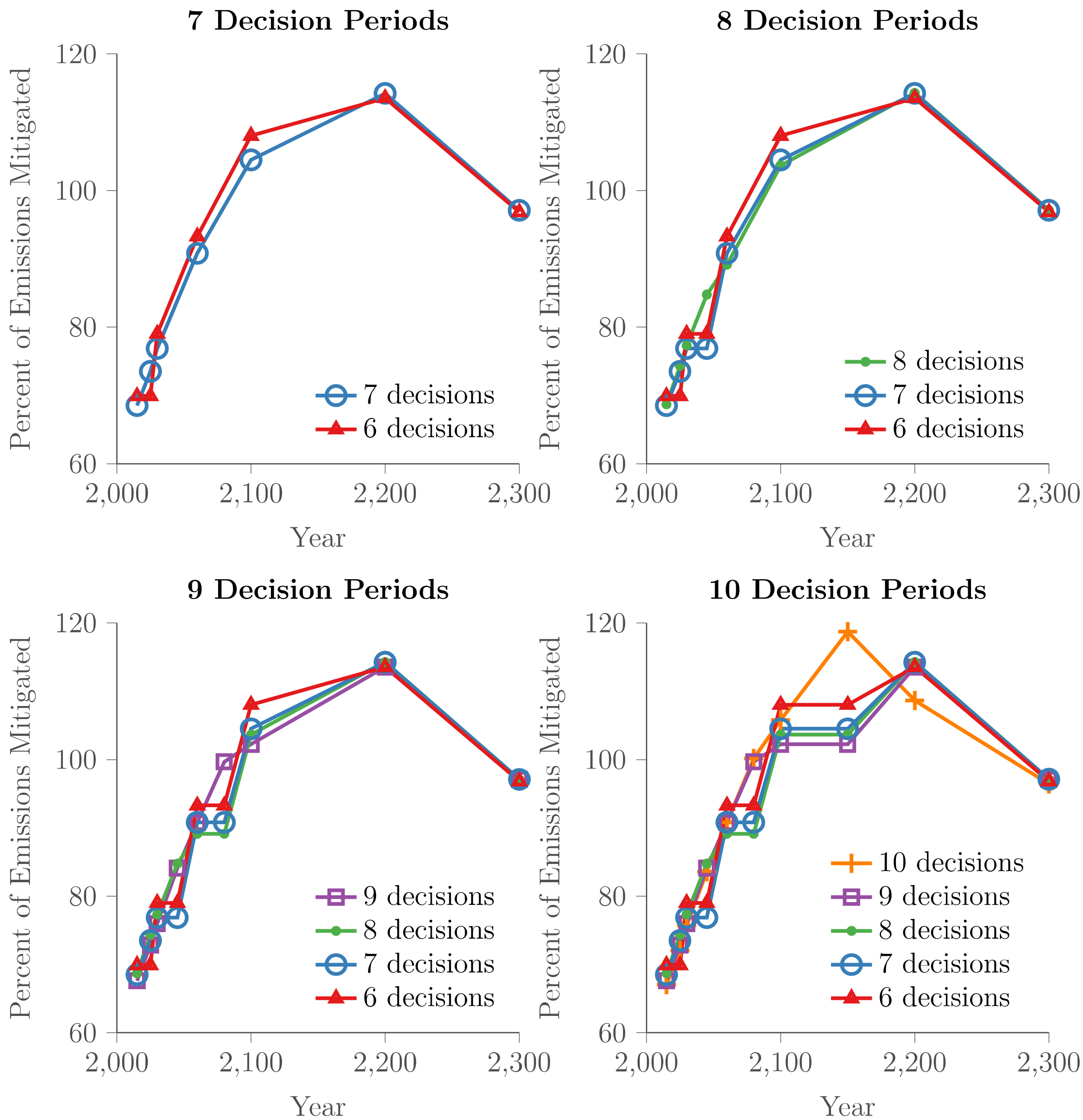


Figure 5.10: *Optimal and suboptimal strategies for models with the number of decision periods, N, given in the title. The suboptimal strategies are the solutions for models with fewer decision times $n \leq N$ (shown in the legend).*

the utility of the optimal 10 period solution when initial global consumption is reduced by 1.9%.

5.3.2 Cost of Postponement

Next we investigate the cost of postponing mitigation to a later date. We only consider models with 6 and 10 decision periods because similar conclusions can be reached for models

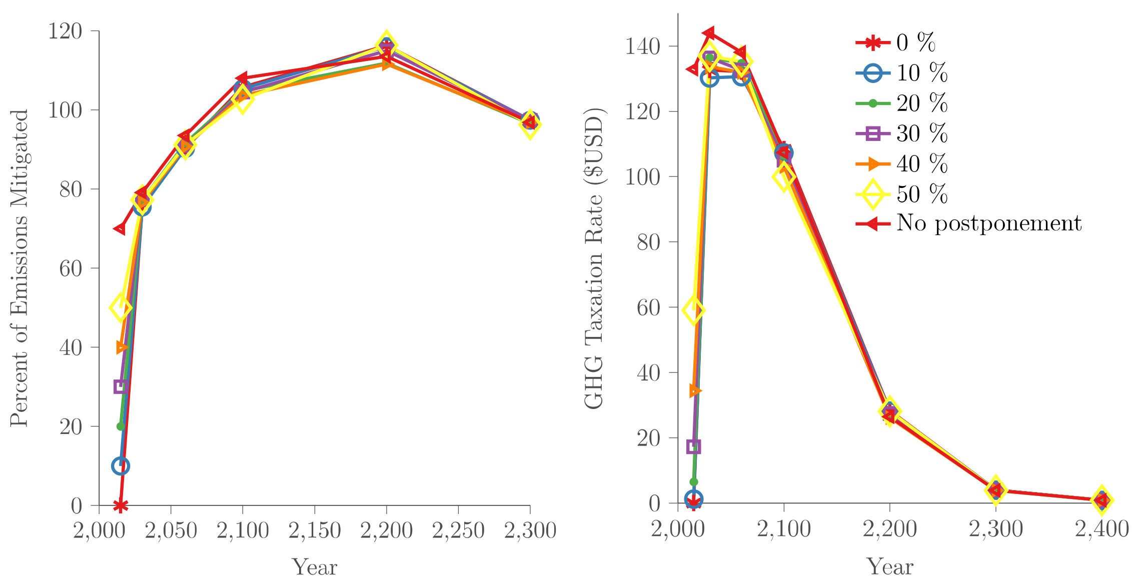
Table 5.7: *The cost of using a model with fewer number of decision periods. The costs are given as a % of the global consumption in 2015 (30.46 trillion $ USD). The number of decisions periods, N, is given along the top row. The suboptimal strategy being considered is the optimal strategy when the agent can make decisions at only $n < N$ periods; n is given in the first column.*

n \ N	6	7	8	9	10
6	0	0.18504	0.73637	1.19	1.9057
7		0	0.1857	0.68881	1.1407
8			0	0.15983	0.65458
9				0	0.16342
10					0

with other numbers of decision periods. For each model, we find the optimal mitigation strategy if the mitigation levels until the year 2030 were fixed to a value (0% to 50%) lower than the optimum value of 70%. These solutions are shown in Figure 5.11.

Figure 5.11: *The optimal mitigation levels (on the left) and GHG taxation rates (on the right), averaged at each time period, for models with different fixed 2015 mitigation levels. This fixed value is given in the legend. The no postponement solution's mitigation level at 2015 is not fixed.*

(a) *A model with 6 decision periods and a subinterval length of 5 years.*



As follows from Figure 5.11, the optimal strategies are very close in all periods except for the very first period. However, the resulting cost of the suboptimal strategies shown in Figure 5.11 can be very high (see Table 5.8).

It is clear from Table 5.8 that the cost of postponing mitigation can very high. For example, not mitigating GHG emissions at all until 2030 costs almost 17% of global consumption –

(b) *A model with 10 decision periods and a subinterval length of 5 years.*

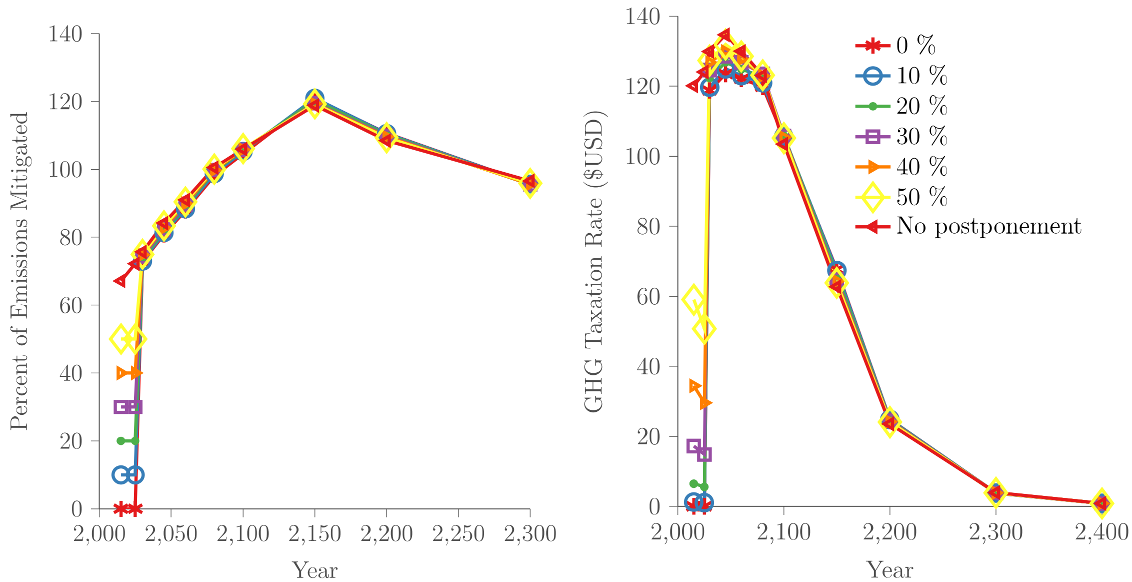


Table 5.8: *The cost of postponing mitigation to 2030, calculated using a 6 decision period model. Mitigation levels for periods after 2030 were optimized while the mitigation level for the 2015 to 2030 period was set to different fixed values, given in the first column. The costs are given as both a % of the global consumption in 2015 (*30.460 *trillion $USD) and in trillion $USD.*

2015 Mitigation Level	Utility Value	Cost as %	Cost in trillion 2015 $USD
0%	9.700	16.587 %	$5.052
10%	9.720	13.590 %	$4.140
20%	9.737	10.764 %	$3.279
30%	9.755	7.570 %	$2.306
40%	9.771	4.567 %	$1.391
50%	9.783	2.053 %	$0.625

over 5 trillion $USD. These costs are even higher when computed using a 10 period model (see Table 5.9). Based on the data shown in Tables 5.8 and 5.9 and overall analysis in this section, it is clear that it is critical to mitigate climate change as soon as possible.

Table 5.9: *The cost of postponing mitigation to 2030, calculated using a 10 decision period model. Mitigation levels for periods after 2025 were optimized while the mitigation levels for the 2015 to 2030 period were set to different fixed values, given in the first column. The costs are given as both a % of the global consumption in 2015,* 30.460 *trillion $USD, and in trillion $USD.*

2015 Mitigation Level	Utility Value	Cost as %	Cost in trillion 2015 $USD
0%	9.786	17.731 %	$5.401
10%	9.804	14.726 %	$4.486
20%	9.822	11.549 %	$3.518
30%	9.839	8.297 %	$2.527
40%	9.854	5.197 %	$1.583
50%	9.866	2.527 %	$0.770

6 Conclusions

In this report, the EZ Climate model from Daniel et al. (2018) was used to compute the social cost of carbon in 2015 assuming an optimal GHG emissions mitigation path. The optimal mitigation strategy was found by maximizing social welfare, modelled using an Epstein-Zin utility function. The SCC was found to be around $130\$/tCO_2$ – within the range of previous estimates but higher the average value. By vectorizing the model and applying automatic differentiation and Quasi-Newton optimization, the model's run time was reduced by several orders of magnitude. This improvement in speed made extending the model to use more decision periods feasible. In addition, robustness checks of the model's assumptions about the number of times climate policy can be changed, the cost of abatement, and effect of technological change could be done. We found that the model was quite sensitive to the parameters used to fit the relationship between the carbon taxation rate and the fraction of emissions mitigated. It was even more sensitive to the parameters modelling technological improvement.

A method to compute a monetary cost to society of acting suboptimality was presented. A robust optimization problem was formulated and used to find solutions that would not be 'costly' to society if certain parameter estimates are inaccurate. This methodology was further used to find the cost of inaction. The cost of postponing the mitigation of GHG emissions to 2030 was found to be 5.4 trillion 2015 USD - more than a sixth of the global economy.

Bibliography

Frank Ackerman, Elizabeth A Stanton, and Ramón Bueno. Epstein–zin utility in dice: Is risk aversion irrelevant to climate policy? *Environmental and Resource Economics*, 56(1): 73–84, 2013.

David Anthoff and Richard SJ Tol. The climate framework for uncertainty, negotiation and distribution (fund): Technical description, version 3.6. *Forschungsstelle für Nachhaltige Entwicklung, Universität Hamburg, Hamburg. Download*, 6:2014, 2012.

Yongyang Cai, Kenneth L Judd, and Thomas S Lontzek. Continuous-time methods for integrated assessment models. Working Paper 18365, National Bureau of Economic Research, September 2012. URL `http://www.nber.org/papers/w18365`.

Yongyang Cai, Kenneth L. Judd, and Thomas S. Lontzek. The Social Cost of Carbon with Economic and Climate Risks. Papers 1504.06909, arXiv.org, April 2015. URL `https://ideas.repec.org/p/arx/papers/1504.06909.html`.

Thomas F Coleman and Wei Xu. Automatic differentiation in MATLAB using ADMAT with applications, 2016.

Kent D Daniel, Robert B Litterman, and Gernot Wagner. Applying asset pricing theory to calibrate the price of climate risk. Technical report, National Bureau of Economic Research, 2018.

Sanjiv Ranjan Das and Raman Uppal. Systemic risk and international portfolio choice. *The Journal of Finance*, 59(6):2809–2834, 2004.

Larry G. Epstein and Stanley E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969, 1989.

Chris W Hope. Optimal carbon emissions and the social cost of carbon over time under uncertainty. *Integrated Assessment*, 8(1), 2008.

Linda Sandris Larsen and Claus Munk. The costs of suboptimal dynamic asset allocation: General results and applications to interest rate risk, stock volatility risk, and growth/value tilts. *Journal of Economic Dynamics and Control*, 36(2):266–293, 2012.

Daniel K. Litterman, R. and G. Wagner. EZClimate. https://github.com/Litterman/EZClimate, 2013.

Jun Liu and Jun Pan. Dynamic derivative strategies. *Journal of Financial Economics*, 69 (3):401–430, 2003.

Tomas Nauclér and Per-Anders Enkvist. Pathways to a low-carbon economy: Version 2 of the global greenhouse gas abatement cost curve. *McKinsey & Company*, 192, 2009.

William Nordhaus. *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press, 2008. ISBN 9780300137484. URL `http://www.jstor.org/stable/j.ctt1npzkh`.

William D. Nordhaus. An optimal transition path for controlling greenhouse gases. *Science*, 258(5086):1315–1319, 1992. ISSN 0036-8075. doi: 10.1126/science.258.5086.1315. URL `http://science.sciencemag.org/content/258/5086/1315`.

William D. Nordhaus and Paul Sztorc. Dice 2013r: Introduction and users manual. *Technical Report*, 2013.

Robert S Pindyck. The climate policy dilemma. *Review of Environmental Economics and Policy*, 7(2):219–237, 2013.

Gernot Wagner and Martin L Weitzman. Climate shock: The economic consequences of a hotter planet, 2015.

Pei Wang, Xiangzheng Deng, Huimin Zhou, and Shangkun Yu. Estimates of the social cost of carbon: A review based on meta-analysis. *Journal of Cleaner Production*, 2018.

Donald J Wuebbles, David W Fahey, and Kathy A Hibbard. Climate science special report: fourth national climate assessment, volume i. 2017.