DISCUSSION PAPER SERIES

DP13708
(v. 2)

DISCOUNTING THE FUTURE: ON CLIMATE CHANGE, AMBIGUITY AVERSION AND EPSTEIN-ZIN PREFERENCES

Stan Olijslager and Sweder van Wijnbergen

FINANCIAL ECONOMICS

ISSN 0265-8003

DISCOUNTING THE FUTURE: ON CLIMATE CHANGE, AMBIGUITY AVERSION AND EPSTEIN-ZIN PREFERENCES

Stan Olijslager and Sweder van Wijnbergen

Discussion Paper DP13708
First Published 01 May 2019
This Revision 05 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programmes:

- Financial Economics

DISCOUNTING THE FUTURE: ON CLIMATE CHANGE, AMBIGUITY AVERSION AND EPSTEIN-ZIN PREFERENCES

Abstract

We show that empirically strongly supported deviations from standard expected time separable utility have a major impact on estimates of the willingness to pay to avoid future climate change risk. We propose a relatively standard integrated climate/economy model but add stochastic climate disasters. The model yields closed form solutions up to solving an integral, and therefore does not suffer from the curse of dimensionality of most numerical climate/economy models. Introducing Epstein-Zin preferences with an elasticity of substitution higher than one and ambiguity aversion leads to much larger estimates of the social cost of carbon (SCC) than obtained under power utility. The dominant parameters are the risk aversion coefficient and the elasticity of intertemporal substitution. Ambiguity aversion has more complicated consequences but overall also leads to substantially higher estimates of the SCC.

JEL Classification: Q51, Q54, G12, G13

Keywords: Social cost of carbon, ambiguity aversion, Epstein-Zin Preferences, Stochastic Differential Utility, climate change

Stan Olijslager - s.w.j.olijslagers@uva.nl
University of Amsterdam

Sweder van Wijnbergen - s.j.g.vanwijnbergen@uva.nl
University of Amsterdam and CEPR

Discounting the Future: on Climate Change, Ambiguity Aversion and Epstein-Zin preferences*

Stan Olijslagers† Sweder van Wijnbergen‡

January 4, 2021

Abstract

We show that empirically strongly supported deviations from standard expected time separable utility have a major impact on estimates of the willingness to pay to avoid future climate change risk. We propose a relatively standard integrated climate/economy model but add stochastic climate disasters. The model yields closed form solutions up to solving an integral, and therefore does not suffer from the curse of dimensionality of most numerical climate/economy models. Introducing Epstein-Zin preferences with an elasticity of substitution higher than one and ambiguity aversion leads to much larger estimates of the social cost of carbon (SCC) than obtained under power utility. The dominant parameters are the risk aversion coefficient and the elasticity of intertemporal substitution. Ambiguity aversion has more complicated consequences but overall also leads to substantially higher estimates of the SCC.

JEL codes: Q51, Q54, G12, G13
Keywords: Social Cost of Carbon, Ambiguity Aversion, Epstein-Zin preferences, Stochastic Differential Utility, Climate Change

*We thank Bob Pindyck, Nick Stern and Martin Weitzman (†) for comments and pension administrator and wealth manager MN for financial support.

†University of Amsterdam, s.w.j.olijslagers@uva.nl
‡University of Amsterdam, Tinbergen Institute, CEPR, s.j.g.vanwijnbergen@uva.nl

1 Introduction

Climate change is one of the main risks the economy will face in the upcoming decades or possibly even centuries. However, there is still much uncertainty about climate change. While (almost) all scientists agree on the fact that climate change will have possibly dramatic negative consequences for the environment and economic growth, we are still not able to accurately estimate the extent and timing of future damages induced by climate change. But one thing we do know is that such consequences will take place far in the future, while if they are to be avoided policies need to be implemented today. This has made the issue of discounting future uncertain costs of climate change back towards today arguably the most important element of the climate change debate and that is the subject of this paper. We show that taking into account ambiguity aversion leads to a sizeable increase in the social cost of carbon.

Rather than arguing about specific numerical values for parameters such as time preference, we challenge the structure of preferences commonly assumed to derive the appropriate discounting procedures and discount rates.[1] Specifically, in this paper we model climate damages as disaster risk and assume that there is ambiguity about the arrival rate and size of future climate disasters. We show that implementing these extensions leads to estimates of the social cost of carbon that are substantially larger than have been derived so far using conventional approaches to time and risk discounting.

The impact of climate change on the economy is often modeled using combined economy/climate models called Integrated Assessment Models (IAMs). IAMs integrate the knowledge of different domains into one model. In the case of climate change, IAMs combine an economic model with a climate model. Three well-known IAMs are DICE (W. Nordhaus, 2014), PAGE (Hope, 2006) and FUND (Tol, 2002).[2] These models are, among others, used as policy tools for cost-benefit analyses. They provide a conceptual framework to better understand the complex problem of climate change by combining different fields and allowing for feedback effects between those fields.

But IAMs also have major drawbacks. To quote Pindyck (2017): "*IAM-based analyses of climate policy create a perception of knowledge and precision that is illusory, and can fool policy-makers into thinking that the forecasts the models generate have some kind of scientific legitimacy.*" His critique is that the models are (1) very sensitive to the choices of parameters and functional forms, especially the discount rate. Besides, we know very little about (2) climate sensitivity and (3) damage functions. Lastly, (4) IAMs don't incorporate tail risk. He recommends simplifying the problem by focusing on the catastrophic outcomes of climate change, instead of modeling the underlying causes. In line with that view we focus on disaster risks and the associated ambiguities and risks.

[1]For a very different (and strongly worded) view focusing on the social welfare aspects of the rate of time preference rather than on individual preferences, see Stern (2015) and Chichilnisky, Hammond, and Stern (2018) who look at a positive rate of time preference as discrimination between generations that happen to have been born at different moments in time.

[2]The references do not contain the most recent versions of the IAMs.

The three main IAMs are deterministic, largely because stochastic models with many state variables are more difficult to solve than deterministic models. To nevertheless capture uncertainty, some authors perform a Monte Carlo-like approach by analyzing several deterministic runs with different parameter values and then taking a weighted average of all runs (Dietz, 2011; W. D. Nordhaus, 2014). Such an analysis is useful if we are interested in the sensitivity of the models to different parameter values. However, it is conceptually different from explicitly using stochastic variables, since for each run all uncertainty is resolved at time 0. Crost and Traeger (2013) compare the Monte Carlo approach to a model that actually uses random variables and find that the Monte Carlo approach underestimates the impact of climate damages. And as we will discuss below, particularly under the structure of preferences we are analyzing, the timing of the resolution of uncertainty matters a great deal.

We propose an analytically solvable IAM (Integrated Assessment Model) that addresses both the critiques of Pindyck (2017) and of Crost and Traeger (2013) on the use of deterministic IAMs. Since there is so little known about the damage functions, we investigate the impact of both attitudes towards well defined measurable risks and ambiguity aversion towards unmeasurable uncertainty on the willingness to pay for avoiding climate risk. Furthermore we model climate risk as disaster risk instead of assuming that temperature increases generate a certain amount of damage every year. The model is transparent due to the closed form solutions for the social cost of carbon. Where stochastic numerical IAMs commonly take hours or more to be solved, solving this model only requires numerical integration and is therefore solved within seconds.

The economy is modeled as a pure exchange economy with exogenous but stochastic endowments. We extend the general equilibrium Consumption-based Capital Asset Pricing Model (CCAPM), also known as Lucas-tree model, developed in Lucas Jr (1978) in several directions. In the literature, this model is widely used in conjunction with a lognormal distribution.[3] The diffusion component of the endowment captures fluctuations in consumption. But we take into account that the nature of climate risk is different from 'normal' economic risk as captured by a diffusion term. Climate disasters are events that occur rarely and take place abruptly (Goosse, 2015). To model this feature, we add a jump process to the endowment consumption stream to capture the climate disaster risk.

The intensity of the disasters is temperature-dependent. We model emissions, atmospheric carbon concentration and the temperature anomaly. The arrival rate of climate disasters is increasing in temperature. Furthermore we explicitly take into account that it is hard to estimate the probability that a disaster occurs and its expected impact by assuming that the agent does not know the exact probability distributions of the arrival rate of climate disasters and the size of the disasters: there is so called ambiguity about the characteristics of the jump risk component. And the agent is assumed to be averse to this ambiguity or Knightian uncertainty. Finally we use the continuous time version of Epstein-Zin utility, also called stochastic differential utility (SDU), which allows us to separate the intertemporal elasticity of

[3] Although Lucas Jr (1978) doesn't assume a specific distribution for the endowment stream.

substitution from the degree of risk aversion. In the widely used power utility specification risk aversion and elasticity of intertemporal substitution (EIS) are captured by one parameter, they are equal to each other's inverse. There is strong empirical evidence placing the relative degree of risk aversion in the range of 5 - 10 (Cochrane, 2009). Using such estimates in combination with power utility then results in implied estimates for the EIS much lower than direct empirical estimates of the EIS suggest. But especially for long term problems such as climate change intertemporal choices play an important role and restricting parameters such as the EIS is a severe limitation. SDU preferences make it possible to separate risk aversion and the elasticity of intertemporal substitution. We can therefore disentangle risk aversion effects (known probabilities), ambiguity aversion effects (unknown probabilities) and substitution effects. The Epstein-Zin preferences also allow for the possibility that the agent has a preference for early resolution of risk, clearly of relevance in a discussion on climate risks. We show that the specification of the agent's preferences in combination with stochastic disaster risk has large effects on how much one is willing to pay to reduce climate risk.

We explicitly focus on the valuation of climate risk in the Business As Usual (BAU) scenario, like in W. Nordhaus (2014) and do not analyse optimal abatement policies at this stage yet. The idea is that an analysis of the environmental costs of current policies (not current plans...) is useful in the climate policy debate. A commonly used measure for the cost of carbon emissions is the social cost of carbon (SCC), the long term discounted damage in dollar terms of emitting one ton of carbon today. Note that the social cost of carbon in our model is not equal to the globally optimal Pigouvian carbon tax, since we do not consider abatement policy in this model. The social cost of carbon using a baseline scenario can be interpreted as the monetized welfare loss of emitting one additional unit of carbon today, given the current global carbon abatement policy scenario under the assumption that no measures will be taken in the future either, like in W. Nordhaus (2014). This seems to us an important first step to take for as long as effective international policies are not yet agreed upon and future agreement is not yet certain. The cost of doing nothing surely is an important input in the debate, but we elaborate on the differences between the SCC under the BAU scenario and the SCC assuming optimal abatement policies in Olijslagers (2020) and Olijslagers, van der Ploeg, and van Wijnbergen (2020).

Our base calibration yields a sizable social cost of carbon. Similar to the numerical IAMs, the SCC in our model is very sensitive to the choice of the input parameters. But in addition we can easily explore the implications of ambiguity aversion, preferences for early resolution of uncertainty and (related to that) a higher EIS than implied by commonly accepted values for the degree of risk aversion. In spite of incorporating all these generalizations we can still derive analytic expressions for the SCC, up to an integral, in our core model setup, making it transparent how ambiguity aversion and Epstein-Zin preferences influence the SCC. Our numerical example using best estimates of the various parameters indicates that introducing ambiguity aversion yields a SCC that is between 28% and 36% higher depending on the structure of climate risk. Moreover we highlight that the social cost of carbon is also sensitive to choices about time discounting, either via the pure rate of time preference, risk aversion or the elasticity of intertemporal substitution, and that all these parameters interact with the cost of ambiguity aversion in complicated ways. But the overall conclusion remains: insufficient attention to risk pricing leads to substantial underestimation of the SCC.

2 Related literature

This paper is related to two strands of literature. First, our methodology is related to consumption based asset pricing models with disaster risk and/or non-expected utility. And second, the paper is related to research on the impact of climate change on the economy.

The model we develop is an extension of the Consumption based Capital Asset Pricing Model (CCAPM) by Lucas Jr (1978). Mehra and Prescott (1985) point out that for plausible parameter values, the CCAPM produces a way too low equity premium and correspondingly a too high risk-free rate. These puzzles are called respectively the Equity Premium Puzzle and the Risk-Free Rate Puzzle. Jump risk or disaster risk has been proposed as a possible solution of these puzzles (Barro, 2006; Rietz, 1988). Extensions to the early disaster/jump risk models are the use of Stochastic Differential Utility (SDU) instead of power utility, and the introduction of time-varying disaster probabilities and multi-period (i.e. persistent) disasters (Barro, 2009; Tsai & Wachter, 2015; Wachter, 2013). Climate change induced disasters fit in the rare disaster literature since climate change is increasingly thought to give rise to abrupt destructive changes in the Earth's environment (Goosse, 2015). We define disaster shocks as shocks whose occurrence has a small probability at any given moment of time but with possibly large and persistent negative effects on the economy once they do take place.

Ambiguity aversion, aversion of unmeasurable or Knightian uncertainty, is the second extension of the CCAPM we introduce to our climate model. Liu, Pan, and Wang (2004) consider a general equilibrium model with rare disasters and ambiguity aversion in their analysis of option pricing 'smirks'. Their agent is only concerned about misspecification of the jump process, a logical choice that we follow, since the probability distribution of rare events is by their very nature much harder to estimate than the diffusion component.

Risk aversion and ambiguity aversion are obviously important in a climate change setting, but since abrupt climate change is anticipated to take place far into the future, intertemporal choices play an important role as well. Power utility is therefore an unsatisfactory framework since with that structure of preferences, risk aversion and EIS cannot be varied independently. This is why we adopt the Stochastic Differential Utility framework introduced by Duffie and Epstein (1992b) since with SDU the risk aversion parameter and the elasticity of intertemporal substitution (EIS) are no longer restricted to be each other's inverse. We go beyond the setting of Tsai and Wachter (2015) who also use SDU to analyze the consequences of disaster risk for asset prices by in addition introducing ambiguity aversion. This extension is especially relevant in a climate disaster model since there is no clear history of events on which we can base our estimates of the damages.

The second strand in the literature our paper is obviously related to is the literature on climate change economics and especially to the part of that literature that considers climate disaster risk, non-expected utility and analytic approaches to solve their models. This paper is to the best of our knowledge the first paper to consider ambiguity aversion in a framework with climate disaster risk.

Barro (2015) extends his disaster risk model with environmental disasters and focuses on discount rates and optimal environmental investment. He does not incorporate a climate model but rather assumes that the disaster probability is constant and that it can be reduced by environmental investment. Bansal, Kiku, and Ochoa (2016) propose a climate model based on the Long-Run-Risk (LRR) model of Bansal and Yaron (2004). In the LRR-model, the agent has Epstein-Zin preferences and consumption growth contains persistent shocks. Bansal et al. (2016) model climate disasters as a jump process that affects both consumption itself and the growth rate of consumption. They show that the outcomes of their model are very sensitive to choices of the EIS. Karydas and Xepapadeas (2019) consider a dynamic asset pricing framework with both macroeconomic disasters and climate change related disasters and analyze the implications for portfolio allocation. Our approach differs from these papers by including ambiguity aversion.

Furthermore, our paper is related to the literature on climate change economics that considers risk and non-expected utility. The most well-known integrated assessment model is the DICE model W. D. Nordhaus (2017). This model is deterministic and the representative agent is assumed to have power utility. Several papers have recently studied the impact of risk and more complex preference structures on the social cost of carbon. For instance Cai and Lontzek (2019), Hambel, Kraft, and Schwartz (2019) and Jensen and Traeger (2014) study integrated assessment models with Epstein-Zin preferences and different types of economic and climate risk. Epstein-Zin preferences can have a substantial effect on the discount rate, for obvious reasons a very important parameter in climate models. Traeger (2014) studies the effect of ambiguity aversion on discount rates. Millner, Dietz, and Heal (2013) look at the effect of ambiguity about the climate sensitivity on optimal policy, where Lemoine and Traeger (2016) focus on ambiguity about tipping points that affect the climate dynamics. All three papers use the smooth ambiguity approach proposed by Klibanoff, Marinacci, and Mukerji (2005). In contrast, we consider the multiple priors approach to model ambiguity aversion, in which the worst case within a specified set of priors is chosen following Gilboa and Schmeidler (1989). Lastly, Barnett, Brock, and Hansen (2020) introduce ambiguity aversion into a climate-economy model, also based on the smooth ambiguity approach of Klibanoff et al. (2005). Barnett et al. (2020) consider even three different types of uncertainty: they distinguish between risk, ambiguity and model misspecification. Our key contribution to this literature is that 1) we study ambiguity aversion in a climate disaster risk framework, 2) we use a different ambiguity approach (the multiple priors approach), and 3) we provide a closed form expression for the social cost of carbon (up to solving an integral) which facilitates conceptual understanding of the results particularly where non-linear effects are at play.

The three IAMs that were mentioned in the introduction are all solved using numerical methods. However, since it has become clear that the choice of the input parameters has a large influence on the results, we think it is useful to know how these parameters exactly influence the outcomes and therefore opt for models allowing for analytical solutions. There are a few recent papers that also focused on obtaining analytic solutions. Golosov, Hassler, Krusell, and Tsyvinski (2014) were the first to obtain closed form solutions in an IAM. However, this required quite strict assumptions such as logarithmic utility and full depreciation of capital every decade. Bretscher and Vinogradova (2018) develop a stylized production-based model where the current carbon concentration directly enters the damage function and obtain closed form solutions for the optimal abatement policy. Bremer and van der Ploeg (2018) consider a rich stochastic production-based model with Epstein-Zin preferences, convex damages, uncertainty in state variables, correlated risks and skewed distributions to capture climate feedbacks. Since the model is too complex to obtain exact analytic solutions, they obtain closed form approximate solutions using perturbation methods. Lastly, Traeger (2018) extends the model of Golosov et al. (2014). Where in other analytic other models the atmospheric carbon concentration often directly enters the damage function (Bretscher & Vinogradova, 2018; Golosov et al., 2014), ACE explicitly models the carbon cycle and the temperature anomaly while damages are induced by an increasing temperature. Additionally Traeger (2018) considers the effect of stochastic state variables.

3 The Model

In this section we first outline the setup for the economy, then extend that setup to incorporate a climate model and finally discuss the utility specification.

Since we do not consider mitigation policies in this paper, we opt for assuming a pure exchange economy, where agents are endowed with an exogenous stochastic income stream. Agents can buy risky stocks, which give a claim on the endowment. Consumption goods are perishable, transferring wealth to the future is only possible by buying stocks. The income stream can intuitively be seen as a tree that produces an uncertain amount of fruit every time period. All agents can buy stocks, which are shares in the tree. The fruit is non-storable, so it must be consumed at the period of the endowment. This implies that aggregate endowment equals aggregate consumption at every moment in time. It is assumed that all agents have identical preferences and endowments, so the separate agents can be replaced by one representative agent. We extend the standard pure exchange model by assuming that the stochastic endowment stream is subject to climate disasters, where the probability of a climate disaster depends on the temperature level.

3.1 The economy

The aggregate endowment process follows a geometric Brownian motion with an additional jump component that represents climate disasters. Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a standard Brownian motion Z_t, a Poisson process N_t with arrival rate λ_t and a random variable J_t are defined. The arrival rate of climate disasters is a function of the temperature level. The distribution of the size of disasters is assumed to be the same for any t. The three types of shocks, namely Brownian motions, Poisson arrivals and disaster sizes, are assumed to be independent. Assume there is a filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$. We will use the following notation throughout this paper: $E_t[.] = E[.|\mathcal{F}_t]$. Consider the following process for aggregate endowments:

$$dC_t = \mu C_t dt + \sigma C_t dZ_t + J_t C_{t-} dN_t. \tag{1}$$

The endowment follows the usual geometric Brownian motion dynamics, with an additional jump process. C_{t-} denotes aggregate endowment just before a jump ($C_{t-} = \lim_{h \downarrow 0} C_{t-h}$). In equilibrium aggregate consumption must equal aggregate endowment and therefore the process is also referred to as the aggregate consumption process. The growth rate $\mu \geq 0$ and the volatility $\sigma > 0$ are constant. When a climate disaster arrives at time t, the size of the disaster is controlled by the random variable J_t. We assume that J_t has the following density: $f(x) = \eta(1+x)^{\eta-1}$ where $-1 < x < 0$. J_t can thus be seen as the percentage loss of aggregate consumption after a disaster. The expected disaster size equals $E_t[J_t] = \frac{-1}{\eta+1}$ and the moments $E_t\Big[(1+J_t)^n\Big] = \frac{\eta}{\eta+n}$ can be easily calculated. In line with the subject of climate disasters, jumps can only be negative.

3.2 The climate model

The arrival rate of disasters is assumed to be temperature dependent. We assume that damages are linearly increasing in temperature: $\lambda_t = \lambda_T T_t$. However, our derivations remain valid for non-linear specifications of the arrival rate. We discuss this assumption in the calibration section. We make some simplifying assumptions to allow for analytic solution of the model. The main requirement for that is that the state variables of the climate submodel are deterministic, an assumption we have relaxed elsewhere (Olijslagers, 2020; Olijslagers et al., 2020).

Industrial emissions (from fossil fuel burning) are usually modeled as the product of the carbon intensity of aggregate output and aggregate output (or aggregate consumption) itself. In addition to industrial emissions, land-use change such as deforestation also causes carbon emissions.[4] We simplify the problem by modeling emissions as exogenous, which in the current setting is not all that important because output growth itself is not yet endogenized. Thus we directly model total emissions, which are the sum of industrial emissions and emissions caused by for example land-use change. This simplification is necessary to keep the state variables deterministic, which in turn is necessary for analytical solvability. If we would not make this assumption, emissions are stochastic and this would make it impossible to solve the model analytically. We therefore assume that emissions are growing at a rate $g_{E,t}$.

[4] For an extensive report on the relation between land-use change and emissions we refer to the special IPCC report (Noble, Bolin, Ravindranath, Verardo, & Dokken, 2000).

The growth rate itself moves towards the long-run equilibrium $g_{E,\infty}$ at a rate δ_E. By assuming that $g_{E,\infty} < 0$, this specification allows us to have growing emissions today, but in the long run the growth rate will then become negative and emissions will go to zero. This is a logical assumption since there is a point where the stock of fossil fuels will be depleted. This gives us the following process for emissions:

$$\begin{aligned} dE_t &= g_{E,t} E_t dt, \\ dg_{E,t} &= \delta_E (g_{E,\infty} - g_{E,t}) dt. \end{aligned} \tag{2}$$

We calibrate exogenous emissions to match the baseline scenario in W. D. Nordhaus (2017). In our setup, it is not a great loss to lose the direct connection between the economy and the carbon emissions since we use a Lucas-tree model where the economy already has an exogenous growth rate. We do not analyse optimal policy and therefore the causes of economic growth and emissions are not of first order importance. What is important for the valuation of the risk is that the climate model is in line with reality.

Since we focus on disaster risks which through our climate model depend on cumulative emissions, not incorporating any short term correlation between economic growth and emissions has no major consequences for the answers to the questions addressed in this paper. It does matter once abatement policies are incorporated, even for our narrow focus on discount rates since the correlation of abatement costs with consumption growth clearly is going to have an impact on the risk premium. We consider this issue elsewhere (Olijslagers, 2020). In reality, emissions are low when the economy is in a recession and vice-versa, there is a substantial correlation between economic growth and worldwide carbon emissions. However, due to thermal inertia it takes some time for temperature to react on emissions and the contemporaneous correlation between consumption and temperature will be lower. When climate risk is high in good states, one would be willing to pay less to reduce the risk. So the correlation between temperature and the consumption process does play a role in the valuation of damages. However, since the contemporaneous correlation between aggregate consumption and temperature is smaller compared to the correlation between aggregate consumption and emissions we expect that this does not play a large role given our focus on disaster risks.

We use the climate model (carbon cycle and temperature model) discussed in Mattauch et al. (2018), which they call the IPCC AR5 impulse-response model. This model is in line with recent insights from the climate literature and is also used in IPCC (2013). Specifically, this climate model incorporates the fact that thermal inertia play a smaller role than commonly assumed in the climate modules in economic models. Climate modules commonly used in economic models tend to overstate the time it takes for the earth to warm in response to carbon emissions (cf Dietz, van der Ploeg, Rezai, and Venmans (2020)).

The first step is to model how the carbon concentration evolves over time given a path of carbon emissions. Define by M_t the atmospheric carbon concentration compared to the pre-industrial level M_{pre}. We then assume that the carbon concentration is the sum of four artificial carbon boxes: $M_t = \sum_{i=0}^{3} M_{i,t}$. This specification can capture that the decay of carbon has multiple time scales and that a fraction of emissions will stay in the atmosphere forever. The dynamics of carbon box i are given by:

$$dM_{i,t} = \nu_i \Big(E_t - \delta_{M,i} M_{i,t} \Big) dt. \tag{3}$$

ν_i is the fraction of emissions that ends up in carbon box i, which implies that $\sum_{i=0}^{3} \nu_i = 1$. $\delta_{M,i}$ controls the decay rate of carbon in box i. We assume that all carbon that ends up in box 0 will permanently stay in the atmosphere, such that $\delta_{M,0} = 0$. The other three boxes have a positive decay rate: $\delta_{M,i} > 0, i = \{1, 2, 3\}$.

The next step is to model the impact of carbon concentration on temperature. This requires modeling what is called radiative forcing: radiative forcing is the difference between energy absorbed by the earth from sunlight and the energy that is radiated back to space. A higher atmospheric carbon concentration strengthens the greenhouse effect and therefore leads to higher radiative forcing. We propose a logarithmic relation between atmospheric carbon concentration and radiative forcing:

$$F_{M,t} = \alpha \frac{\upsilon}{\log(2)} \log\Big(\frac{M_t + M_{pre}}{M_{pre}} \Big). \tag{4}$$

α equals the climate sensitivity: the long-run change in temperature due to a doubling of the carbon concentration compared to the pre-industrial level. υ is a parameter that is also part of the temperature module and this parameter will be discussed later. We also include non-carbon related (exogenous) forcing $F_{E,t}$, which follows:

$$dF_{E,t} = \delta_F (F_{E,\infty} - F_{E,t}) dt. \tag{5}$$

Total radiative forcing is the sum of carbon-related radiative forcing and exogenous forcing: $F_t = F_{M,t} + F_{E,t}$.

The final step moves from F_t to the actual surface temperature T_t. T_t is the difference between the actual temperature compared to the pre-industrial temperature level. The change in surface temperature is a delayed response to radiative forcing. Call the heat capacity of the surface and the upper layers of the ocean τ while τ_{oc} equals the heat capacity of the deeper layers of the ocean. The parameter κ captures the speed of temperature transfer between the upper layers and the deep layers of the ocean. The dynamics of temperature are then given by:

$$\begin{aligned} dT_t &= \frac{1}{\tau} \Big(F_t - \upsilon T_t - \kappa (T_t - T_t^{oc}) \Big) dt, \\ dT_t^{oc} &= \frac{\kappa}{\tau_{oc}} (T_t - T_t^{oc}) dt. \end{aligned} \tag{6}$$

From this equation, one can verify that the long run equilibrium temperature for a given level of radiative forcing equals: $T_t^{eq} = \frac{F_t}{\upsilon}$. The parameter υ therefore controls the equilibrium temperature response to a given level of forcing. Note that when $M_t = 2M_{pre}$, we obtain that $F_t = \alpha\upsilon + F_t^E$ and $T_t^{eq} = \alpha + \frac{F_t^E}{\upsilon}$. Therefore the parameter α can indeed be interpreted as the equilibrium temperature response to doubling of the carbon concentration. We can rewrite the first equation to:

$$dT_t = \frac{1}{\tau}\Big(\upsilon(T_t^{eq} - T_t) - \kappa(T_t - T_t^{oc})\Big). \tag{7}$$

This equation is more intuitive, since it captures the fact that the temperature moves towards its equilibrium level at a rate proportional to $T_t^{eq} - T_t$. The second part shows that the oceans are delaying this convergence. It takes time for T_t^{oc} to adjust towards T_t and this will also delay the convergence of T_t towards the equilibrium level T_t^{eq}. As specified earlier, the arrival rate of climate disasters is a linear function of temperature T_t.

3.3 Utility Specification

The representative agent maximizes his utility of consumption over an infinite planning horizon. We consider the continuous time version of Epstein-Zin preferences (Epstein & Zin, 1989), called stochastic differential utility (SDU) by (Duffie & Epstein, 1992b). Epstein and Zin (1989) consider the following class of preferences in discrete time: $V_t = [(1-\beta)C_t^{1-1/\epsilon} + \beta ce_t(V_{t+1})^{1-1/\epsilon}]^{\frac{1}{1-1/\epsilon}}$ where $\epsilon = EIS$, β is the time preference parameter and $ce_t(.)$ is a certainty equivalent function. When considering a deterministic consumption program, V_t is a constant elasticity of substitution (CES) utility function. In the other extreme case where only a static gamble is considered, there are no intertemporal choices and the utility is entirely determined by the certainty equivalent function $ce_t(.)$. The certainty equivalent function (or risk aggregator) that is widely used throughout the literature is $ce_t(V_{t+1}) = E_t(V_{t+1}^{1-\gamma})^{\frac{1}{1-\gamma}}$ where γ is the coefficient of relative risk-aversion, which we assume to be constant. This specification of $ce_t(.)$ yields a special case of the preferences studied by Kreps and Porteus (1978) and is therefore also called Kreps-Porteus utility. Static gambles are evaluated as if the agent has power utility, but in a dynamic stochastic setting EIS and risk aversion are decoupled under SDU: this specification allows to separate risk aversion γ from the elasticity of intertemporal substitution ϵ. An important property of this utility specification is that the agent has preferences for early resolution of uncertainty if $\epsilon > \frac{1}{\gamma}$ and for late resolution if $\epsilon < \frac{1}{\gamma}$.

We consider a special case of SDU, the continuous time equivalent of Kreps-Porteus utility, or rather an ordinally equivalent utility process. Similar to the discrete time case, SDU can be represented by a combination of an aggregator f that determines the degree of intertemporal substitution and a certainty equivalent operator ce. In the case of Kreps-Porteus utility, $f(C,V) = \frac{\beta}{1-1/\epsilon}\frac{C^{1-1/\epsilon} - V^{1-1/\epsilon}}{V^{-1/\epsilon}}$ and $ce(\sim V) = [E(V^{1-\gamma})]^{\frac{1}{1-\gamma}}$. In this case the drift of the value function consists of the aggregator $f(C,V)$ and a variance multiplier A that belongs to ce. Duffie and Epstein (1992b) show that there exists an ordinally equivalent utility process with aggregator f as in (8). In this case $ce(\sim V) = E(V)$ and the variance multiplier A that belongs to c is zero. The agent's utility or value function then becomes:

$$V_t = E_t\Big[\int_t^\infty f(C_s, V_s)ds\Big]$$

where

$$f(C,V) = \frac{\beta}{1-1/\epsilon}\frac{C^{1-1/\epsilon} - \big((1-\gamma)V\big)^{\frac{1}{\zeta}}}{\big((1-\gamma)V\big)^{\frac{1}{\zeta}-1}} \quad \text{for } \epsilon \neq 1 \tag{8}$$

with $\zeta = \dfrac{1-\gamma}{1-1/\epsilon}$.

Throughout this paper, we refer to this utility specification as stochastic differential utility (SDU) although Duffie and Epstein (1992b) actually consider a more general class of utilities under that label. Similar to the discrete time counterpart, γ denotes risk-aversion, ϵ is the elasticity of intertemporal substitution and β equals the time preference parameter. We will focus on the case where $\epsilon \neq 1$ and therefore will derive our results only for this case. For the case $\epsilon = 1$ we can take the limit $\epsilon \to 1$ or follow the same derivation but with $f(C,V) = \beta(1-\gamma)V\Big(\log C - \frac{1}{1-\gamma}\log\big((1-\gamma)V\big)\Big)$. If $\gamma = \frac{1}{\epsilon}$, the utility specification reduces to standard power utility.

3.4 Ambiguity

There is much uncertainty regarding the arrival rate and magnitude of climate disasters. And, as stressed by Pindyck (2017), we know very little about the damage functions. Where consumption growth and volatility can be estimated more accurately from historical data, the estimation of the climate disaster parameters will be much harder since climate disasters do not happen that often. It is fair to state that we simply do not know the exact distribution of climate damages. We consider it therefore desirable to account for the possibility that the 'best estimate' model is not the true model: there is ambiguity. We assume that the representative agent is ambiguity averse.

It is important to stress the difference between risk and ambiguity. When we are talking about risk, an agent knows the probabilities and possible outcomes of all events. When the agent has to deal with ambiguity, the probabilities attached to particular events are unknown. The distinction between risk and ambiguity is already extensively discussed in Knight (1921), which is why ambiguity is often referred to as Knightian uncertainty. Ellsberg (1961) shows using the Ellsberg Paradox that people are ambiguity averse, i.e. they prefer known probabilities over unknown probabilities.

We use the *recursive multiple priors utility* developed in continuous time by Chen and Epstein (2002) to model ambiguity. For an overview of different methods to model ambiguity we refer to appendix A. An advantage of this method compared to other methods is that it preserves the homotheticity of the value function.

To apply approach of Chen and Epstein (2002) to model ambiguity, we begin by defining the 'best estimate' model or reference model as the agent's most reliable model with probability measure $\mathbb{P}$. But the agent also takes into account other, alternative models. The alternative models have measure $\mathbb{Q}^{a,b}$; the jump arrival rate becomes $\lambda_t^{\mathbb{Q}} = a_t\lambda_t$ and the jump size parameter becomes $\eta_t^{\mathbb{Q}} = b_t\eta$. Remember that the expected jump size equals $\frac{-1}{\eta+1}$, which implies that a low b_t leads to a more negative jump size. The agent takes into account that his reference model is not the true model and he therefore specifies a set of models that he considers possible. Given the set of models, he then considers the worst case (Chen & Epstein, 2002; Gilboa & Schmeidler, 1989).

The size of the set of models is assumed to depend on the ambiguity aversion parameter θ. All models with a distance smaller than θ are allowed in the set of admissible models. The distance between the reference model $\mathbb{P}$ and an alternative model $\mathbb{Q}^{a,b}$ is measured using the concept of *relative entropy*, a common metric for the distance between two probability measure (see for example Hansen and Sargent (2008)). The distance or *relative entropy* between the reference and alternative model depends on the parameters a_t and b_t and can therefore be written as $RE(a_t, b_t)$. The relative entropy metric satisfies $RE(a_t, b_t) \geq 0\, \forall(a_t, b_t)$ and $RE(1, 1) = 0$: the distance of the reference model to itself is by definition equal to 0. If θ is large, the agent is very ambiguity averse and thus considers a large set of models. The preferences of the agent then become:

$$\begin{aligned} &V_t = \min_{\mathbb{Q}\in\mathcal{P}^\theta} V_t^{\mathbb{Q}} \\ &\text{where } \; V_t^{\mathbb{Q}} = E_t^{\mathbb{Q}}\Big[\int_t^\infty f(C_s, V_s^{\mathbb{Q}})ds\Big] \\ &\text{and } \; \mathcal{P}^\theta = \{\mathbb{Q}^{a,b} : RE(a_t, b_t) \leq \theta \;\forall t\}. \end{aligned} \tag{9}$$

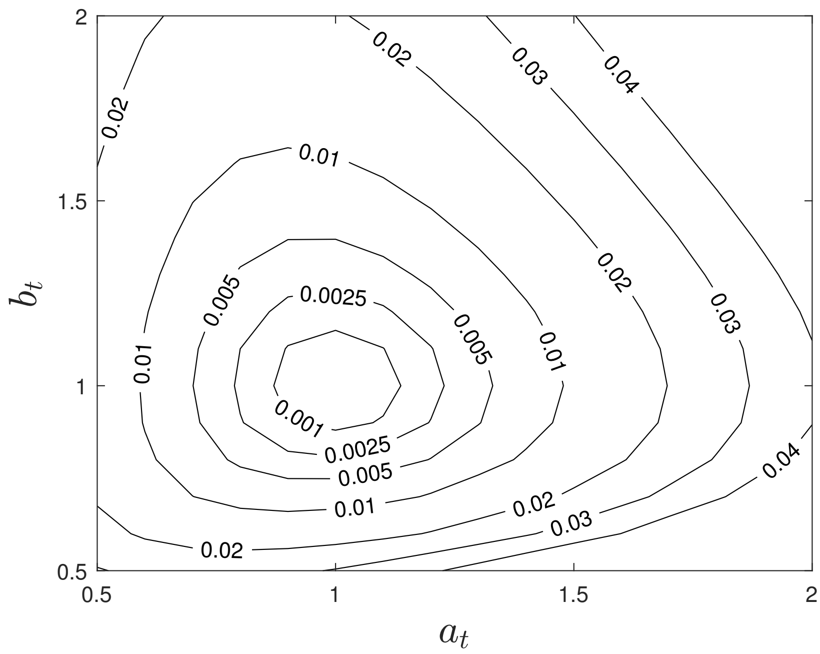
Here $V_t^{\mathbb{Q}}$ is the SDU utility process given the measure $\mathbb{Q}$. $\theta = 0$ implies that $\mathcal{P}^\theta = \{\mathbb{P}\}$ and the agent only considers one measure, namely the reference measure. Thus there is no ambiguity aversion when $\theta = 0$. Where the risk aversion parameter γ can be seen a parameter that is relevant for any risky bet, the parameter θ captures intrinsic ambiguity aversion (one person might be more ambiguity averse than another), but it is also source dependent. If there is a lot of information and data available about a process, the set of admissible priors will be smaller compared to a process about which not much is known.

It is not necessary to have a constant θ, one could for example incorporate learning by assuming that θ_t is a decreasing function over time as the actual stochastic processes unfold. The agent then obtains more information about a process over time and therefore one could argue it is plausible that the set of priors will be shrinking over time. However, there does not (yet) seem to exist a generally accepted framework to determine how the set of priors should shrink over time based on new observations. In particular the multiple prior approach does not lend itself to Bayesian updating since we do not define model probabilities in this approach. Similar to Chen and Epstein (2002) we will therefore focus on the case with a constant θ.

In appendix B we derive that relative entropy equals:

$$RE(a_t, b_t) = (1 - a_t)\lambda_t + a_t\lambda_t\Big(\log(a_t b_t) + \frac{1}{b_t} - 1\Big). \tag{10}$$

Figure 1: Relative entropy for different values of a_t and b_t. Results are given for $\lambda_t = 0.1$.



If we take a look at this expression for the relative entropy, it is clear that for $(a_t, b_t) = (1, 1)$, the relative entropy equals zero. When one or both of the two variables deviate from the reference model, the relative entropy increases. Every contour in figure 1 gives a set of combinations (a_t, b_t) that yields the same relative entropy. If for example $\theta = 0.01$, then all (a_t, b_t) combinations within that contour line are included in the set of admissible priors. The worst case probability measure will be the probability measure for which either a_t is large (high arrival rate) and/or b_t is small, since the expectation of the jump size under the alternative measure is inversely related to b_t: $E_t^{\mathbb{Q}}[J_t] = \frac{-1}{b_t \eta + 1}$.

From the current setup, it is hard to argue what a reasonable value for ambiguity aversion θ would be. In order to give more guidance about reasonable values for θ, we use the concept of *detection error probabilities* introduced by Anderson, Hansen, and Sargent (2003).[5] Consider the following thought experiment. Assume that the representative agent would be able to observe the process of consumption over the next N years, and after observing the process the agent has to choose which of the two models (the reference model or the worst-case model) is most likely. There are two types of errors in this case. The agent could choose the reference model while the process was actually generated by the worst-case model and he could also make the opposite error. The detection error probability is defined as the average of the probability of the two errors. Appendix C describes how the detection error probability is calculated.

[5]See for example Maenhout (2006) for another application of detection error probabilities.

The detection error probability depends on N, since when the agent observes the process for a longer period, the probability of a mistake will be smaller. The detection error probability also depends on the ambiguity aversion parameter. When θ is small, the reference and worst-case model are similar to each other and the probability of a mistake is large. On the other hand, when the agent is extremely ambiguity averse the reference and worst-case models are very different and the detection error probability becomes small. The representative agent wants to make the set of models sufficiently large to make a robust decision, but on the other hand does not want to take into account implausible models. Since the detection error also depends on the other parameters of the model, we come back to the issue of calibrating the ambiguity aversion parameter in the calibration section.

4 Solving the model

We first derive for each alternative probability measure $\mathbb{Q}^{a,b}$ the corresponding Hamilton-Jacobi-Bellman (HJB) equation and find an expression for the value function $V_t^{\mathbb{Q}}$. Then we derive the HJB-equation for $V_t = \min_{\mathbb{Q}\in\mathcal{P}^\theta} V_t^{\mathbb{Q}}$. At the end of the section we discuss our solution method.

4.1 The HJB-equation

The value function is a function of aggregate consumption and all the climate state variables. Let $V_C^{\mathbb{Q}}$ denote the first derivative of the value function with respect to aggregate consumption, similar notation is used for the second derivative. For notational purposes, define the vector of climate state variables:

$$X_t = [g_{E,t}\ E_t\ M_{0,t}\ M_{1,t}\ M_{2,t}\ M_{3,t}\ F_{E,t}\ T_t\ T_t^{oc}]'. \tag{11}$$

The vector of state variables then follows: $dX_t = \mu_X(X_t)dt$. Denote by $V_X^{\mathbb{Q}}$ the row vector of partial derivatives of the value function $V_t^{\mathbb{Q}}$ with respect to the vector of state variables X_t: $V_X^{\mathbb{Q}} = \left[\frac{\partial V^{\mathbb{Q}}(C_t,X_t)}{\partial g_{E,t}} \ \ldots \ \frac{\partial V^{\mathbb{Q}}(C_t,X_t)}{\partial T_t^{oc}}\right]$.

We show in appendix D that under the measure $\mathbb{Q}^{a,b}$, the value function $V_t^{\mathbb{Q}}$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} 0 &= f(C_t, V_t^{\mathbb{Q}}) + V_C^{\mathbb{Q}}\mu C_t dt + \frac{1}{2}V_{CC}^{\mathbb{Q}}\sigma^2 C_t^2 + V_X^{\mathbb{Q}}\mu_X(X_t) \\ &+ \lambda_t^{\mathbb{Q}} E_t^{\mathbb{Q}}\big[V^{\mathbb{Q}}\big((1+J_t)C_{t-}, X_t\big) - V^{\mathbb{Q}}(C_{t-}, X_t)\big]. \end{aligned} \tag{12}$$

The HJB equation is a partial differential equation. We conjecture and verify that the value function under the measure $\mathbb{Q}^{a,b}$ is of the following form:

$$V^{\mathbb{Q}}(C_t) = g^{\mathbb{Q}}(X_t)\frac{C_t^{1-\gamma}}{1-\gamma}, \tag{13}$$

where $g^{\mathbb{Q}}(X_t)$ is some function of X_t. Substituting this form of the value function into the HJB-equation and calculating the expectation gives the following reduced HJB-equation (see appendix E):

$$
\begin{aligned}
0 = &\frac{\beta}{1-1/\epsilon}\Big(g^{\mathbb{Q}}(X_t)^{-\frac{1}{\zeta}}-1\Big)+\mu-\frac{\gamma}{2}\sigma^2+\frac{g_X^{\mathbb{Q}}(X_t)}{g^{\mathbb{Q}}(X_t)(1-\gamma)}\mu_X(X_t) \\
&+a_t\lambda_t\frac{-1}{b_t\eta+1-\gamma}.
\end{aligned} \tag{14}
$$

Given a probability measure $\mathbb{Q}^{a,b}$, we could solve this equation to find $g^{\mathbb{Q}}(X_t)$. Now let us return to the problem with ambiguity. We are not interested in the solution for every single measure $\mathbb{Q}^{a,b}$, but we want to find the solution to $V_t = \min_{\mathbb{Q}\in\mathcal{P}^\theta} V_t^{\mathbb{Q}}$. We can replace the global minimization problem of equation (9) by an instantaneous optimization problem at every time period t, since relative entropy is a function of a_t, b_t and λ_t, which are all three known at time t . The HJB-equation of the problem with ambiguity then becomes:

$$
\begin{aligned}
0 = &\min_{(a_t,b_t)\ s.t.\ RE(a_t,b_t)\le\theta}\Big\{\frac{\beta}{1-1/\epsilon}\Big(g^{\mathbb{Q}}(X_t)^{-\frac{1}{\zeta}}-1\Big)+\mu-\frac{\gamma}{2}\sigma^2 \\
&+\frac{g_X^{\mathbb{Q}}(X_t)}{g^{\mathbb{Q}}(X_t)(1-\gamma)}\mu_X(X_t)+a_t\lambda_t\frac{-1}{b_t\eta+1-\gamma}\Big\}.
\end{aligned} \tag{15}
$$

4.2 Optimal control variables

From the HJB-equation we can then calculate the optimal control variables a_t^* and b_t^*. This is a constrained optimization problem with Lagrangian:

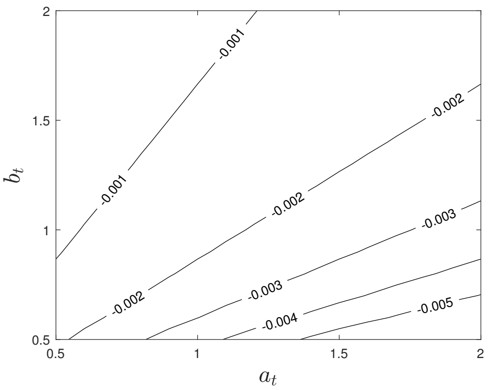
$$
L(a_t,b_t,l_t) = a_t\lambda_t\frac{-1}{b_t\eta+1-\gamma} - l_t\Big(RE(a_t,b_t)-\theta\Big). \tag{16}
$$

Here l_t is the Lagrange multiplier. a_t^* and b_t^* and the Lagrange-multiplier l_t are the solutions to the following first order conditions:

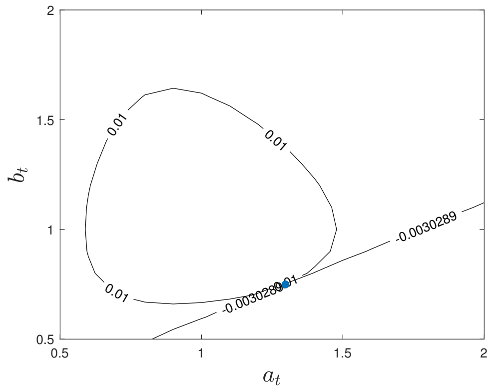
$$
\begin{aligned}
&\frac{\partial}{\partial a_t}L(a_t,b_t,l_t) = \lambda_t\frac{-1}{b_t\eta+1-\gamma} - l_t\lambda_t\Big(\log(a_tb_t)+\frac{1}{b_t}-1\Big) = 0, \\
&\frac{\partial}{\partial b_t}L(a_t,b_t,l_t) = a_t\lambda_t\frac{\eta}{(b_t\eta+1-\gamma)^2} - l_ta_t\lambda_t\frac{b-1}{b^2} = 0, \\
&\frac{\partial}{\partial l_t}L(a_t,b_t,l_t) = \theta-(1-a_t)\lambda_t - a_t\lambda_t\Big(\log(a_tb_t)+\frac{1}{b_t}-1\Big) = 0.
\end{aligned} \tag{17}
$$

Figure 2 illustrates the optimization problem. Given an entropy budget θ and the arrival rate λ_t, one can determine the feasible set of (a_t,b_t). Figure 1 shows the feasible sets for several budgets. A contour plot of the objective function for several (a_t,b_t) combinations is given in subfigure 2a. Clearly combinations in the bottom right corner (high a_t, low b_t) give the lowest objective function. The goal is to minimize this function, given the relative entropy constraint. Subfigure 2b shows how the optimal combination (a_t^*,b_t^*) is determined. The point where objective function touches the feasible region is the optimal solution. From now on we use the following notation for the optimal arrival rate and jump size: $\lambda_t^* = a_t^*\lambda_t$ and $\eta_t^* = b_t^*\eta$. Since a_t^* and b_t^* are a function of λ_t, they are implicitly a function of temperature T_t as well. Furthermore

Figure 2: Selection of the optimal a and b.



(a) Contour plot of the objective function of the constrained minimization problem for different values of a_t and b_t.



(b) Illustration of selection of optimal (a_t, b_t). The oval area shows all admissible values for a_t and b_t that are within the relative entropy budget of 0.01. The straight line is the objective function.

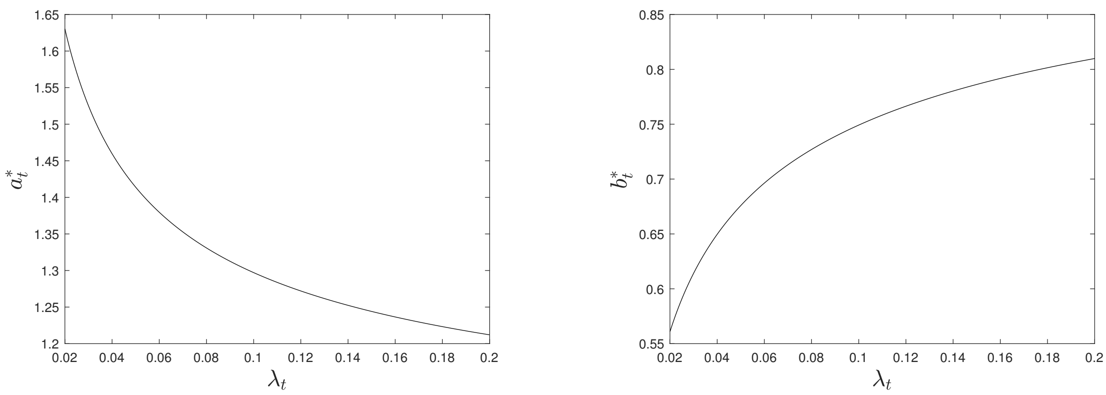
we define by $g(X_t)$ the function that solves that HJB-equation with parameters a_t^* and b_t^*.

Figure 3 shows the optimal a_t^* and b_t^* as a function of λ_t. For each λ_t one finds the corresponding a_t^* and b_t^* by solving the first order conditions. A constant relative entropy budget implies that a_t is decreasing in λ_t and b_t is increasing in λ_t. The idea behind the time-varying parameters is illustrated in the following example. Assume $\theta = 0.01$. At time t, the arrival rate equals 0.05 and at time t' the arrival rate equals 0.1. At every time point the following equality must hold at the optimum: $RE(a_t, b_t, \lambda_t) = \theta$. For $\lambda_t = 0.05$ the optimal parameters are $(a_t^*, b_t^*) = (1.41, 0.68)$ and $RE(1.41, 0.68, 0.05) = 0.01$. Now consider time t' with arrival rate 0.1. If we would use the same optimal parameters as at time t, the relative entropy exceeds the budget: $RE(1.41, 0.68, 0.1) > 0.01$. The distance or relative entropy between the reference model and the worst-case model is increasing in the arrival rate λ_t. For a larger arrival rate, an x% increase in the arrival rate generates a larger 'distance' between the two measures. Intuitively, when the arrival rate is larger, more disasters are observed. With the same a_t^* and b_t^*, detecting which probability distribution is the true distribution is easier when disasters occur frequently. Therefore the optimal a_t and b_t must adjust to make sure that the relative entropy remains within the constant budget. At time t', the optimal parameters become: $(a_t^*, b_t^*) = (1.30, 0.75)$.

4.3 Solution method

It is typically not possible to solve the partial differential equation of the problem with climate state variables (except when the highly restrictive assumption assumption of a unit EIS is made). However we are able to obtain exact solutions for the value function and the consumption-to-wealth ratio without making restrictive assumptions

Figure 3: Optimal parameters a_t and b_t with time-varying arrival rate λ_t and constant ambiguity aversion parameter θ. Input parameters: $\theta = 0.01$, $\eta = 62$, $\gamma = 5$.



(a) Optimal a_t as function of λ_t. (b) Optimal b_t as function of λ_t.

like $EIS = 1$, and the consumption-to-wealth ratio is what we need for assessing the SCC. We will now sketch our approach.

Duffie and Epstein (1992a) derive that the pricing kernel (or stochastic discount factor) with stochastic differential utility equals $\pi_t = \exp\left\{ \int_0^t f_V(C_s, V_s)ds \right\} f_C(C_t, V_t)$. However, the pricing kernel has to be adjusted for the ambiguity aversion preferences. Chen and Epstein (2002) show that the pricing kernel in the ambiguity setting should be multiplied by the Radon-Nikodym derivative $\xi_t^{a^*,b^*}$ of the measure corresponding to the optimal a^* and b^*. $\xi_t^{a,b}$ is defined in (36). When calculating the pricing kernel, we obtain an expression that depends on the unknown function $g(X_t)$. But by substituting the HJB-equation into the pricing kernel we obtain an expression that only depends on known parameters.

As an intermediate step it is helpful to introduce the concept of consumption strips. A consumption strip is an asset that pays a unit of aggregate consumption C_s at time time $s > t$. Call its value at time t: $H_t(C_t, X_t, u)$, where u denotes the time to maturity; $u = s - t$. The price of a consumption strip paying out at time $s > t$ equals:

$$\begin{aligned} H_t &= H(C_t, X_t, u) \\ &= E_t\left[\frac{\pi_s}{\pi_t} C_s\right] = \exp\left\{ -\int_t^{t+u} CDR_s ds \right\} C_t. \end{aligned} \tag{18}$$

We will refer to CDR_t as the consumption discount rate. We can use the fact that every asset multiplied by the pricing kernel must be a martingale to calculate the value of such an asset.

Furthermore, we can define a stock S_t that gives a claim to the Lucas-tree and therefore it pays a continuous stream of dividends C_t. The value of such a stock then obviously becomes:

$$S_t = \int_0^\infty H(C_t, X_t, u)du. \tag{19}$$

In equilibrium aggregate wealth must be equal to the value of the stock. The state-dependent consumption-wealth ratio therefore equals:

$$k(X_t) = \frac{C_t}{S_t} = \frac{C_t}{\int_0^\infty H(C_t, X_t, u)du} = \left(\int_0^\infty \exp\Big\{-\int_t^{t+u} CDR_s ds\Big\}du\right)^{-1}. \quad (20)$$

Using the expression for the consumption-wealth ratio, we can calculate the value function. At the optimum (see for example Munk (2015), Ch. 17), we have the envelope condition that $f_C = V_S$. Furthermore, we derived that $V(C_t, X_t) = \frac{g(X_t)C_t^{1-\gamma}}{1-\gamma}$. Using the chain rule we get:

$$V_S = V_C \frac{\partial C}{\partial S} = V_C k(X_t) = g(X_t) C_t^{-\gamma} k(X_t). \quad (21)$$

Also we have for the intertemporal aggregator:

$$f_C = \beta g(X_t)^{\frac{1/\epsilon - \gamma}{1-\gamma}} C_t^{-\gamma}. \quad (22)$$

Together this gives us:

$$g(X_t) = \Big(\frac{k(X_t)}{\beta}\Big)^{-\frac{1-\gamma}{1-1/\epsilon}}. \quad (23)$$

We can now derive an expression for the pricing kernel in terms of known parameters and using this pricing kernel, we can calculate the price of a consumption strip. We will analyse the consumption strips in detail in subsection 5.2. Integrating over the maturities of consumption strips with different maturities gives us the value of the stock, which in turn enables us to calculate the consumption-wealth ratio. Lastly, we can link $g(X_t)$ to the consumption-wealth ratio, which then allows us to derive an expression for the value function.

5 Asset prices and discounting

5.1 Asset market

Before going to the main part of this paper, the analysis of the Social Cost of Carbon, we first calculate the risk-free rate and the risk premium as an input in the analysis of the SCC. Assume that the representative agent has the possibility to invest in two assets, namely a risk-free asset and a risky stock. The risk-free asset with price B_t pays a continuously compounded interest rate r_t. The stock pays continuous dividends at a rate C_t and has ex-dividend price S_t. We denote the cum-dividend stock price by S_t^d. Using equation (20) we can write $S_t = \frac{C_t}{k(X_t)}$. The assets have the following distribution:

$$dB_t = r_t dt, \quad (24)$$

$$
\begin{aligned}
dS_t^d &= dS_t + C_t dt = \frac{1}{k(X_t)} dC_t - \frac{C_t}{k(X_t)^2} dk(X_t) + k(X_t) S_t dt \\
&= \Big(\mu - \frac{k_X(X_t)}{k(X_t)} \mu_X(X_t) + k(X_t)\Big) S_t dt + \sigma S_t dZ_t + J_t S_{t-} dN_t.
\end{aligned} \tag{25}
$$

Chen and Epstein (2002) show that the pricing kernel with ambiguity and stochastic differential utility equals $\pi_t = \xi_t^{a^*,b^*} \exp\Big\{ \int_0^t f_V(C_s, V_s) ds \Big\} f_C(C_t, V_t)$. This allows us to derive an explicit stochastic differential equation for the pricing kernel. Using this pricing kernel, we can calculate the endogenous risk-free rate and the endogenous risk premium of the stock. [6] The interest rate r_t equals:

$$
\begin{aligned}
r_t = \beta + \frac{\mu}{\epsilon} - \Big(1 + \frac{1}{\epsilon}\Big)\frac{\gamma}{2}\sigma^2 - \Big(\gamma - \frac{1}{\epsilon}\Big) a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1 - \gamma} \\
- a_t^* \lambda_t \Big(\frac{b_t^* \eta}{b_t^* \eta - \gamma} - 1 \Big).
\end{aligned} \tag{26}
$$

The risk premium of the dividend paying stock (the expected excess return of that asset compared to investing in the risk-free asset) then equals:

$$
rp_t = \gamma \sigma^2 + a_t^* \lambda_t \Big(\frac{-1}{b_t^* \eta + 1} - \frac{b_t^* \eta}{b_t^* \eta + 1 - \gamma} + \frac{b_t^* \eta}{b_t^* \eta - \gamma} \Big). \tag{27}
$$

5.2 The consumption discount rate

As shown in appendix appendix F, the price of a consumption strip at time t that pays aggregate consumption C_s and has time to maturity $u = s - t$ equals:

$$
\begin{aligned}
H_t &= \exp\Big\{ - \int_t^{t+u} CDR_s ds \Big\} C_t, \quad \text{where} \\
CDR_t &= \underbrace{r_t}_{A} + \underbrace{rp_t}_{B} - \underbrace{\Big(\mu + a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1}\Big)}_{C} \\
&= \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2 + a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1 - \gamma}\Big).
\end{aligned} \tag{28}
$$

We discuss the consumption discount rate CDR_t in detail, since it is also an essential component of the social cost of carbon. The effective discount rate on a consumption strip consists of three terms, labeled A, B and C. Part A is the risk-free rate, which is used to discount a risk-free cashflow. The consumption strip is a risky asset and therefore the risk-free rate is increased with a risk-premium, part B. Lastly, the discount rate should be corrected for the growth of the aggregate consumption process. On average, consumption grows at a rate $\mu + a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1}$. Note that the average growth rate is smaller than μ since climate disasters have a negative impact on consumption.

[6] The derivations are given in appendix F.

Consider first the most simple case without climate disasters and risk, i.e. the case with $(\sigma, \lambda_T) = (0, 0)$; then consumption strips are not risky anymore so the risk premium is zero. The interest rate reduces to the well-known Ramsey rule for the interest rate (Ramsey, 1928):

$$r_t = \beta + \frac{\mu}{\epsilon}, \tag{28a}$$

which implies a growth corrected discount rate $r_{n,t}$ for the case of $(\sigma, \lambda_T) = (0, 0)$ equal to:

$$r_{n,t} = \beta + (1/\epsilon - 1)\mu. \tag{28b}$$

Now add just diffusion risk: $\sigma > 0, \lambda_T = 0$. In a general equilibrium setting this will both affect the interest rate (due to a flight to safety) and the risk premium, in this case $\gamma\sigma^2$:

$$r_t = \beta + \frac{\mu}{\epsilon} - (1 + 1/\epsilon)\frac{\gamma}{2}\sigma^2, \tag{28c}$$

$$rp_t = \gamma\sigma^2. \tag{28d}$$

Adding the risk premium to the risk-free rate but correcting for the growth rate once again gives the growth-adjusted discount rate, now for $\sigma > 0, \lambda_T = 0$:

$$r_{n,t} = \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2\Big). \tag{28e}$$

One would intuitively expect that adding risk to the consumption stream and the associated risk premium $\gamma\sigma^2$ to the interest rate would lead to a higher risk-adjusted discount rate. However, due to the flight to safety effect the risk-free rate decreases which in itself lowers the discount rate. Which of the two effects dominates depends on the elasticity of intertemporal substitution.

- When $\epsilon = 1$, both the interest rate and risk premium effect cancel out and the discount rate simply becomes β.
- When $\epsilon < 1$, the discount rate in the presence of risk ($\sigma > 0$) is actually smaller than the discount rate in the absence of risk ($\sigma = 0$). This implies that for $\epsilon < 1$ adding risk to the consumption stream increases the value of the consumption strip.
- When $\epsilon > 1$ we get the more intuitive outcome. In that case the risk premium effect dominates and the discount rate in the presence of risk is larger than the discount rate in the absence of risk.

But for empirically plausible parameter values for the growth rate and consumption variance[7], we also need to take the impact of the growth rate into account. Taking the impact of growth and risk both into account, and reasonably assuming

[7] $\mu - \frac{\gamma}{2}\sigma^2 > 0$.

$\mu - \frac{\gamma}{2}\sigma^2 > 0$, it becomes clear from 28 that the overall effect of a higher value for ϵ implies a lower growth adjusted discount rate.

Where ϵ determines the relative importance of the interest rate and risk premium effects, risk aversion γ determines the magnitude. A large risk aversion amplifies the effect of risk on the discount rate. When the agent is risk neutral ($\gamma = 0$), risk has no effect on the discount rate. The preceding discussion makes abundantly clear that using a General Equilibrium framework endogenizing the risk-free rate is essential in this context.

Assume now that in addition to adding risk, climate disasters also play a role: $\sigma > 0, \lambda_T > 0$; the discount rate of the consumption strip then becomes state dependent.

First assume that there is no ambiguity aversion ($\theta = 0$). Once again, adding climate disaster risk has an effect on both the interest rate and the risk premium. And similarly to the changes in the σ case, when $\epsilon < 1$ the interest rate effect dominates so that in that case adding disasters leads to a lower discount rate. But when $\epsilon > 1$, the risk premium effect dominates and adding climate disasters actually leads to higher discount rates.

Equation (28) indicates that the climate-risk related term in the discount rate for the reference case $a_t^* = 1, b_t^* = 1$ equals $\lambda_t \frac{-1}{\eta+1-\gamma}$. The term scales with the arrival rate λ_t: more frequent disasters have a larger effect on discount rates. The term $\frac{-1}{\eta+1-\gamma}$ can be interpreted as the certainty equivalent of the climate shock. When $\gamma = 0$, the certainty equivalent is equal to $E_t[J_t] = \frac{-1}{\eta+1}$.

Including ambiguity aversion leads to a larger worst case arrival rate: $a_t^* > 1 \Rightarrow a_t^* \lambda_t > \lambda_t$ and a more negative certainty equivalent term since $b_t^* < b_t$. Therefore we can unambiguously conclude that ambiguity aversion amplifies the effect of climate risk on discounting. And once again assuming a reasonable parameterization[8], increasing ϵ still leads to a lower discount rate.

6 The social cost of carbon

Given the value function, we can calculate the Social Cost of Carbon (SCC), which we define as the marginal cost (in terms of reduced welfare) of increasing carbon emissions by one ton carbon scaled by the marginal welfare effect of one additional unit of consumption to obtain the social cost of carbon in terms of the price of time t consumption units terms (to which we refer as 'in dollar terms', for brevity's sake). With a single carbon box, the marginal cost of increasing carbon emissions by one unit is the derivative of the value function with respect to the carbon concentration M_t: $\frac{\partial V_t}{\partial M_t}$. However, with multiple carbon boxes, emitting one unit of carbon leads to an increase of ν_i units in box i, $i = 0, 1, 2, 3$. We slightly abuse notation and define $\frac{\partial}{\partial M_t} \equiv \nu_0 \frac{\partial}{\partial M_{0,t}} + \nu_1 \frac{\partial}{\partial M_{1,t}} + \nu_2 \frac{\partial}{\partial M_{2,t}} + \nu_3 \frac{\partial}{\partial M_{3,t}}$. Differentiation of the value function gives:

[8]Specifically, we assume throughout the rest of the analysis that $\mu - \frac{\gamma}{2}\sigma^2 + a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1 - \gamma} > 0 \ \ \forall t$.

$$
\begin{aligned}
SCC_t &= -\frac{\partial V_t/\partial M_t}{f_C(C_t,V_t)} = -\frac{\frac{\partial}{\partial M_t} g(X_t)}{(1-\gamma)g(X_t)k(X_t)}C_t = -\frac{\frac{\partial}{\partial M_t}\big(\frac{k(X_t)}{\beta}\big)^{-\frac{1-\gamma}{1-1/\epsilon}}}{(1-\gamma)\big(\frac{k(X_t)}{\beta}\big)^{-\frac{1-\gamma}{1-1/\epsilon}}k(X_t)}C_t \\
&= -\frac{C_t}{1/\epsilon-1}\frac{\frac{\partial}{\partial M_t}k(X_t)}{k(X_t)^2} = \frac{C_t}{1/\epsilon-1}\frac{\partial}{\partial M_t}\int_0^\infty \exp\Big\{-\int_t^{t+u} CDR_s ds\Big\}du \\
&= C_t\int_0^\infty \underbrace{\exp\Big\{-\int_t^{t+u} CDR_s ds\Big\}}_{\text{A}}\int_t^{t+u}\frac{\partial}{\partial M_t}\Big(\underbrace{a_s^*\lambda_s}_{B}\underbrace{\frac{1}{b_s^*\eta+1-\gamma}}_{C}\Big)dsdu
\end{aligned}
\tag{29}
$$

We will first discuss the general formula and then the implications of different preferences.

Equation (29) shows that the social cost of carbon is proportional to C_t, the aggregate consumption level: when the current aggregate consumption level doubles, the SCC doubles as well. For a given consumption level, the SCC depends on three terms, labeled A, B and C respectively. The social cost of carbon measures the marginal welfare loss of emitting an additional unit of carbon today. It is, in discrete time terms, the discounted sum of all future damages done by emitting one ton of carbon today. The outer integral indicates that all future marginal damages are included in the SCC. Future damages are discounted with the consumption discount rate (term A). The integral over the terms B and C captures the marginal damage for a given maturity u. What matters is the cumulative effect of a unit of carbon emissions at time t on the terms B and C over the time period t to $t+u$. Not only the impact of M_t on T_{t+u} plays a role, but the whole path of the temperature between t and $t+u$, since any climate damages that occur within this period have an effect on consumption at time $t+u$.

Without ambiguity aversion ($\theta = 0$) the marginal effect of M_t on terms B and C has a simple expression. Term C is independent of M_t and $\frac{\partial}{\partial M_t}\lambda_s = \lambda_T \frac{\partial}{\partial M_t}T_s$. If we now consider the marginal damages, λ_T captures the increase in the arrival rate when temperature increases by one degree. $\frac{\partial}{\partial M_t}T_s$ gives the marginal increase of temperature at time s due to an increase of atmospheric carbon concentration at time t. In the case without ambiguity, term C equals $\frac{1}{\eta+1-\gamma}$. Without risk aversion, this is equal to the expected value of a climate disaster.

Consider first the impact of γ and ϵ. When the agent is risk averse, term C can be interpreted as the certainty equivalent of the loss after a disaster. The certainty equivalent is clearly increasing in risk aversion. But risk aversion also has an effect on the discount rate CDR_t. As discussed before, increasing risk aversion increases the discount rate when $\epsilon > 1$. In this case the discounting effect works in opposite direction of the effect on the certainty equivalent: for $\epsilon > 1$ the impact of γ on the SCC is therefore ambiguous. But for $\epsilon < 1$ the two effects reinforce each other and the SCC is then an increasing function of γ. Consider next the impact of ϵ. The elasticity of intertemporal substitution ϵ only plays a role in the discount rate. When ϵ increases, the willingness to substitute over time increases which leads to lower discount rates. So a higher ϵ unambiguously leads to a higher SCC.

When ambiguity aversion is present, i.e. $\theta > 0$, we obtain that $a_s^* > 1$ (higher worst-case arrival rate) and $b_s^* < 1$ (higher worst-case jump size). The ambiguity aversion parameter θ does not directly show up in the formula, but its effect works via a_s^* and b_s^*. With $\theta > 0$ both the arrival rate of disasters in the expression is higher (so term B is larger) and the certainty equivalent, which with ambiguity aversion becomes $\frac{1}{b_s^* \eta + 1 - \gamma}$, is also higher. Through these two channels ambiguity aversion leads to a higher social cost of carbon. But similar to risk aversion, ambiguity aversion also affects discount rates and the sign again depends on the elasticity of intertemporal subtitution ϵ. When $\epsilon < 1$, ambiguity aversion additionally leads to a lower discount rate and thus an even higher SCC. When $\epsilon = 1$, the discount rate is simply β and ambiguity has no effect on the discount rate. Lastly, when $\epsilon > 1$, increasing θ leads to higher discount rates. Therefore increasing ambiguity aversion has two offsetting effects in this case. We will focus in the numerical section on the empirically supported case where $\epsilon > 1$.

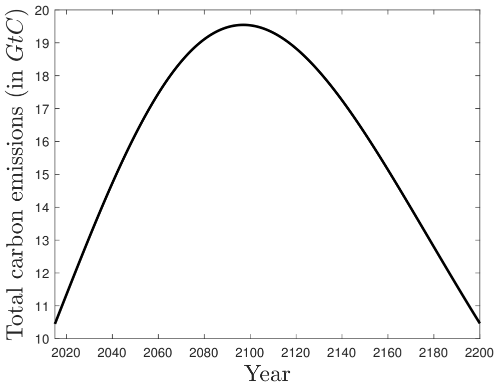
Summarizing, when considering the effect of ambiguity aversion on the social cost of carbon we can identify two effects. First, including ambiguity aversion leads to a higher arrival rate and a larger certainty equivalent, which pushes the social cost of carbon up. We call this effect the direct effect of ambiguity aversion. Second, there is a more indirect general equilibrium effect through the impact of ambiguity aversion on discount rates. We call this the discounting effect. The discount rate that should be used to discount future climate disasters is the consumption discount rate, and when the elasticity of substitution is larger than 1, ambiguity aversion leads to a higher consumption discount rate. This is an intuitive result: if the representative agent is very ambiguity averse about climate disasters, he would rather like to consume today than to postpone consumption since the future consumption level is uncertain. Ambiguity aversion therefore increases the consumption discount rate and decreases the price of a future consumption strip. Thus for $\epsilon > 1$ it is ultimately a numerical issue which of the two effects will dominate. We will highlight both effects separately in the numerical section and show that for our calibration, the first effect dominates. In our numerical analysis the net impact is positive: more ambiguity aversion leads to a higher SCC. We turn to that numerical analysis in the next section.

7 Climate change and the social cost of carbon: numerical results

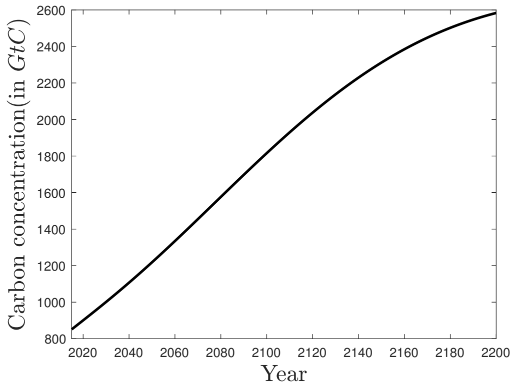
7.1 Calibration

Appendix G gives the full details of the calibration of the climate model. Parameters for the growth rate of emissions and the initial level are chosen to match the baseline scenario of the DICE-2016 calibration (W. D. Nordhaus, 2017). The parameters of the carbon cycle and temperature model are taken from Mattauch et al. (2018). In addition, and different from Mattauch et al. (2018), we also include a base level of non-carbon related radiative forcing and calibrate it to match exogenous forcing in DICE-2016. Figure 4 shows the future path of the climate state variables using our

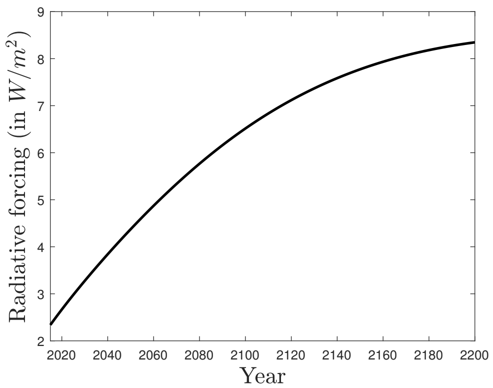
Figure 4: Future path of climate variables.



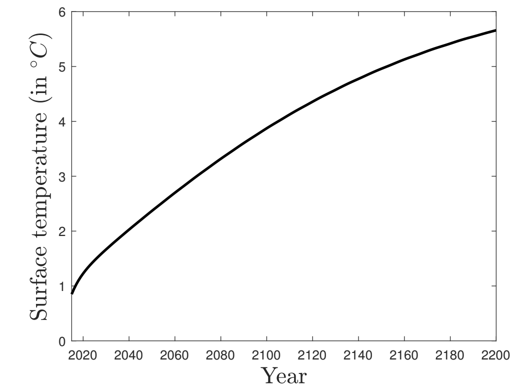
(a) Total emissions



(b) Carbon concentration



(c) Radiative forcing



(d) Surface temperature

emissions path and climate model. Under the Business-As-Usual scenario, emissions are projected to peak at the end of the century, and decline from then on. The surface temperature will then rise by almost 4 degrees in 2100.

The calibration of the economic parameters is given in table 1. Since we consider an exogenous endowment economy, output and consumption are the same thing in our model. That leaves the question open whether we should calibrate the endowment to output or to consumption data. The focus of the paper is on the social cost of carbon. What ultimately matters for the social cost of carbon is consumption, since utility depends on consumption and not on output. To make our results more comparable to other models, we therefore calibrate endowment to consumption data. The next choice to be made is whether one should aggregate output or consumption data using market exchange rates or using purchasing power parities (PPP). In line with the DICE-2016 model we use purchasing power parity exchange rates. Consumption data is not directly available in PPP. To obtain a proxy for world consumption in PPP we first obtain output data in PPP. Then we determine the world consumption ratio using market exchange rates. Our proxy for world consumption in PPP is then output in PPP multiplied by the world consumption ratio. Real world GDP (PPP)

Table 1: Parameters for the economic model

Par.	Description	Value
C_t	Initial consumption level (PPP, in trillion 2015$)	83.07
λ_T	Arrival rate parameter	0.02 / 0.04
η	Disaster size parameter	30.25 / 62.5
$E[J]$	Expected disaster size	-0.016
γ	Risk aversion	5
θ	Ambiguity aversion	0.01
ϵ	Elasticity of substitution	1.5
CDR_0	Consumption discount rate	1.5%

in 2015 equals 114.137 trillion 2015$ (IMF World Economic Outlook October 2016). World consumption in 2015 using market exchange rates equals 55.167 (in trillion 2010 $), while world GDP using market exchange rates equals 75.803 (in trillion 2010 $) (Worldbank Database). This yields a consumption-output ratio of 72.78%. Applying this ratio to World GDP (PPP) then gives 83.065 (in trillion 2015 $) for aggregate consumption in PPP terms.

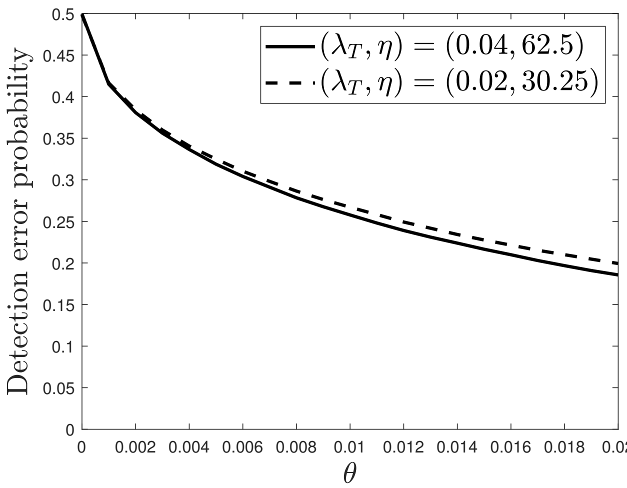
The next step is to calibrate the climate disaster distribution, and in particular the parameters λ_T and η. Our setup does allow for an arrival rate that is convex in temperature, but we do not consider this extension since it would give another free parameter to calibrate. Karydas and Xepapadeas (2019) also consider climate disasters and assume, based on natural disaster data, that for every degree warming the arrival rate increases by 6%. The disaster size is calibrated to 1.6%. This implies that the expected growth loss due to climate change would be $6\% \times 1.6\% = 0.096\%$ per degree global warming. W. D. Nordhaus (2017) models the economic impact of climate change as the percentage loss of output as a function of temperature (level impact). Hambel et al. (2019) consider both a level and a growth impact of climate damages. They find that a loss of 0.026% per degree warming gives the same GDP loss in the year 2100 as the level impact of W. D. Nordhaus (2017). Setting the disaster size to 1.6% and calibrating λ_T such that on average climate disasters lead to a loss of 0.026% gives $\lambda_T = 1.63\%$, much lower than the arrival rate assumed in Karydas and Xepapadeas (2019). The latter obviously get much higher expected damages than the calibration of W. Nordhaus (2014) yields.

We decide to choose $\lambda_T = 4\%$, which is in between these two calibrations and set $\eta = 62.5$ which yields $E_t[J_t] = -1.6\%$, in line with Karydas and Xepapadeas (2019). Additionally, we consider a variant with less frequent but on average larger disasters: $\lambda_T = 2\%$, and a disaster size parameter $\eta = 30.25$ which gives $E_t[J_t] = -3.2\%$. While both calibrations have on average the same impact, their impact on risk premia is very different.

We now turn to the calibration of risk aversion and ambiguity aversion. We set risk aversion equal to 5. This level of risk aversion can be seen as conservative if we compare it to common values in the asset pricing literature.[9]

[9] A coefficient of relative risk aversion between 5 and 10 is common in the asset pricing literature

Figure 5: Detection error probabilities as a function of θ.

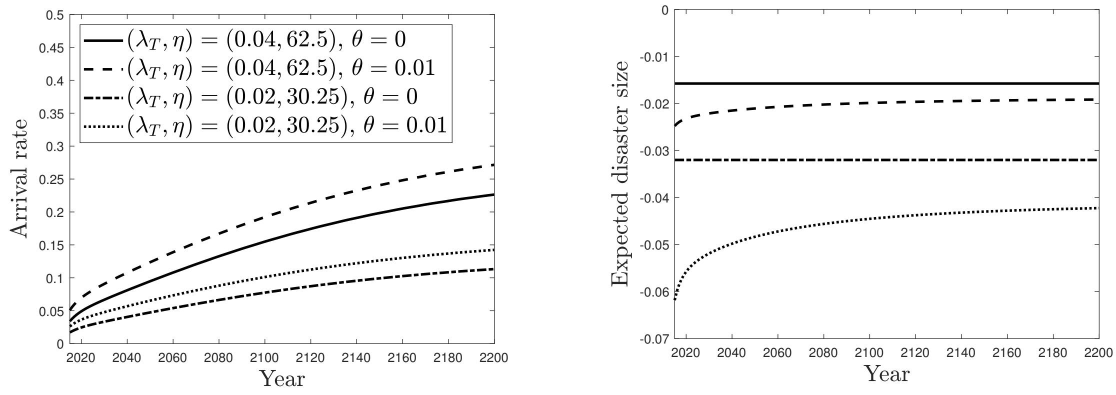


The level of ambiguity aversion is harder to calibrate. To get a feeling for reasonable values of ambiguity aversion, we use the concept of detection error probabilities. The ambiguity aversion parameter θ pins down the arrival rate and the expected jump size in the worst-case scenario. A higher θ leads to a higher worst-case arrival rate and a more negative worst-case expected jump size. The detection error probability is the probability of choosing the wrong model (so choosing the reference model $\mathbb{P}$ when the worst-case $\mathbb{Q}$ is true and vice-versa). When θ is higher, the two models are more different and the probability of making a mistake is therefore lower. When the detection error probability is close to 50%, the two models are very similar. This is an indication of a low ambiguity aversion parameter. On the other hand, when the detection error probability is close to 0, it is easy to distinguish the worst-case model from the reference model. This indicates that the worst-case model is extreme and the ambiguity aversion parameter very high.

We calculate the detection error probability assuming that the consumption process can be observed over a period of 100 years. The ambiguity aversion parameter θ is varied between 0 and 0.02. The results are given in figure 5. Detection error probabilities are decreasing in θ and are higher for a lower λ_T. This is intuitive, since a lower λ_T implies that there are less disasters over the observed time period and the probability of choosing the wrong model is therefore larger. We choose to set $\theta = 0.01$ in the base calibration, which gives a detection error probability of 26.7% for $(\lambda_T, \eta) = (0.02, 30.25)$ and 25.8% for $(\lambda_T, \eta) = (0.04, 62.5)$ (cf figure 5). This level of ambiguity aversion balances the trade-off between wanting to make a robust decision, but not taking into account too extreme models. The detection error probabilities for $\theta = 0.01$ are sufficiently far away from 50%, which implies the two models are not too close to each other. On the other hand, the detection error probabilities are also not close to 0, which would indicate an extreme amount of ambiguity aversion. However, since this parameter remains hard to calibrate we do vary θ in robustness checks.

according to (Cochrane, 2009).

Figure 6: Arrival rate and expected disaster size over time with and without ambiguity aversion.



(a) Arrival rate λ_t^*

(b) Expected disaster size $\frac{-1}{b_t^* \eta + 1}$

Figure 6 shows the resulting arrival rate and expected disaster size with ambiguity aversion ($\theta = 0.01$) and without ambiguity aversion ($\theta = 0$) in both cases.

The parameters that still have to be calibrated affect the social cost of carbon only indirectly, via the discount rate. Equation (28) shows that one can separate the expression for the Consumption Discount Rate (the relevant discount rate for the social cost of carbon) CDR_t in a time-independent part CDR_0 and a part that does depend on time as:

$$\begin{aligned} CDR_t &= CDR_0 + (1/\epsilon - 1) a_t^* \lambda_t \frac{-1}{b_t^* \eta + 1 - \gamma} \\ CDR_0 &= \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2\Big). \end{aligned} \tag{30}$$

CDR_0 is the consumption discount rate in the absence of climate disasters. First, the value of the elasticity of intertemporal substitution ϵ determines whether additional risk increases or decreases the discount rate. Generally, there is strong empirical evidence of an EIS larger than one (Van Binsbergen, Fernández-Villaverde, Koijen, & Rubio-Ramírez, 2012; Vissing-Jørgensen & Attanasio, 2003). When $\epsilon > 1$, we are in the realistic situation that additional risk decreases asset prices. We choose $\epsilon = 1.5$, which is a common value in the literature on Epstein-Zin preferences. The growth rate μ, the volatility σ and the pure rate of time preference β only affect the social cost of carbon via CDR_0. The calibration of β has been widely discussed in the climate change literature. Additionally, we could calibrate σ from observed consumption volatility. However, as Mehra and Prescott (1985) point out, the model in that case would generate a way too low risk premium (the equity premium puzzle). A way to circumvent this is to calibrate σ to the volatility of stock prices, but this solution is also not very satisfactory. There have been several (partial) solutions proposed to the equity premium puzzle, for example including economic disaster risk. Solving the equity premium puzzle goes beyond the scope of this paper. Since both β and σ only affect the SCC via CDR_0, we choose to directly calibrate the consumption discount rate in the absence of climate risk. In our base calibration, we choose $CDR_0 = 1.5\%$, but we show our results for values of CDR_0 between 0.5% and 2.5%. The parameter combinations $(\beta, \mu, \sigma) = (2.25\%, 2.5\%, 3\%)$ and $(\beta, \mu, \sigma) = (1.5\%, 2.5\%, 10\%)$ for example yield a consumption discount rate $CDR_0 = 1.5\%$. Note that the actual consumption discount rate CDR_t is higher because of the impact of climate disasters on discounting.

7.2 Social cost of carbon

Our base calibration yields a social cost of carbon of \$461 per ton of carbon with $(\lambda_T, \eta) = (0.04, 62.5)$ and \$533 per ton carbon with $(\lambda_T, \eta) = (0.02, 30.25)$.[10] Comparing the two cases shows that it matters whether the disasters are frequent but small (large η) or more infrequent but larger (smaller η). The two sets of assumptions yield the same expected disaster shock, but in the low frequency/large-shock case risk aversion and ambiguity aversion play a larger role and the social cost of carbon is correspondingly higher.

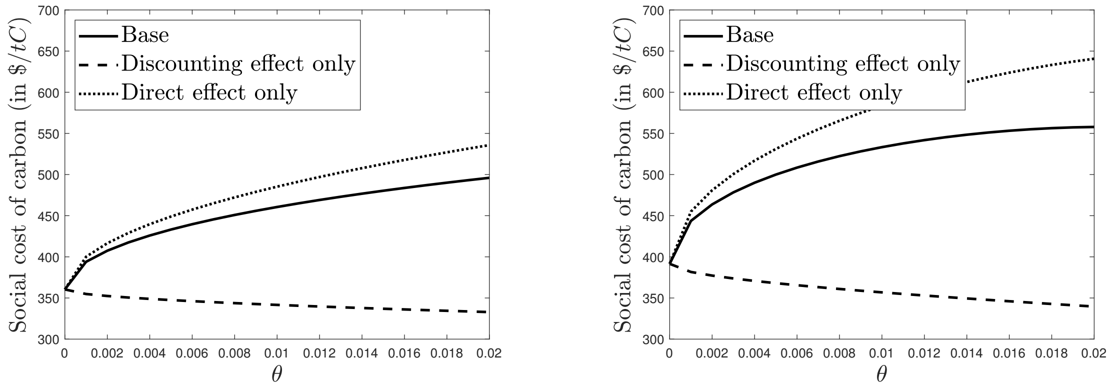
Ambiguity aversion and the SCC

Figure 7 shows for each of the two sets of assumptions on the disaster risk parameters the social cost of carbon for different values of θ. Ambiguity aversion clearly leads to a substantially higher social cost of carbon. For the $(\lambda_T, \eta) = (0.04, 62.5)$ case, the SCC is 28% higher with $\theta = 0.01$ compared no the case without ambiguity aversion. The relative increase is even larger when we consider the $(\lambda_T, \eta) = (0.02, 30.25)$ case: the SCC is then 36% higher with ambiguity aversion. The intuition behind this difference is that the relative entropy between the reference model and the worst-case model is increasing in λ_t. When the arrival rate is smaller, disasters happen less frequently and the two probability distributions are harder to distinguish. With the same ambiguity aversion level θ, this then implies that a_t^* is larger for $\lambda_T = 0.02$ compared to the a_t^* for $\lambda_T = 0.04$ and also that b_t^* is lower for $\lambda_T = 0.02$ compared to $\lambda_T = 0.04$. Therefore the relative increase in the SCC due to ambiguity aversion is larger with $\lambda_T = 0.02$ than it is with $\lambda_T = 0.04$.

From equation (30) it is clear that ambiguity aversion does not only affect the arrival rate and certainty equivalent of climate disasters, but also the discount rate: in our calibration with $\epsilon = 1.5$, more ambiguity aversion leads to a higher discount rate. We disentangle the two effects on the SCC by first considering the *discounting only* effect, in which we assume ambiguity aversion only affects the discount rate CDR_t, but we leave the arrival rate and the certainty equivalent in the SCC formula unaffected by ambiguity aversion (the line with label *discounting only* in figure 7). And second we consider the opposite case, where we leave the consumption discount rate CDR_t unchanged, but we do take into account the direct effect of ambiguity aversion on the arrival rate and certainty equivalent of the climate disasters, with the label *direct only* in figure 7. The two effects are combined in the base case where

[10] We express the social cost of carbon in dollars per ton carbon. To convert in dollars per ton CO_2, divide by 3.67.

Figure 7: Social cost of carbon as a function of θ.



(a) $(\lambda_T, \eta) = (0.04, 62.5)$ (b) $(\lambda_T, \eta) = (0.02, 30.25)$

This figure shows the social cost of carbon as a function of the ambiguity aversion parameter θ. The total effect of ambiguity aversion on the SCC is given by the solid line *(base)*. We additionally distinguish two special cases. In the *discounting effect only* case (dashed line) we assume that increasing θ does lead to an increase in the discount rate, but does not change the arrival rate and the certainty equivalent in the SCC formula. In the *direct effect only* case (dotted line) we look at the opposite case, where increasing θ is assumed have an effect on the arrival rate and the certainty equivalent, but not on the consumption discount rate CDR_t.

both the direct effect and the impact via the discount rate are incorporated (labeled *Base* in figure 7. Figure 7 clearly indicates that ambiguity aversion increases the discount rate and therefore the SCC is decreasing in θ when only the discounting effect of ambiguity aversion is considered. This also implies that when we look at the direct only effect, the SCC is above the base case SCC since the negative impact of discounting is left out. Overall we can conclude that ambiguity aversion does indeed lead to a higher discount rate, but that the direct effect on the SCC dominates and that ambiguity aversion therefore leads to a higher social cost of carbon, and in our calibration substantially so.

The elasticity of intertemporal substitution ϵ and the SCC

The sign of the discounting effect depends on the choice of ϵ. When $\epsilon < 1$, additional risk, more risk aversion or more ambiguity aversion would lower discount rates and both the discounting effect and the direct effect of ambiguity aversion would have the same sign. However, this would lead to counter-intuitive effects. For example $\epsilon < 1$ implies that the price of a consumption strip increases when the volatility of consumption increases. For $\epsilon = 1$, the consumption discount rate CDR_t simply equals β and risk, risk aversion and ambiguity aversion do not affect discount rates.

Risk aversion, ambiguity aversion and the SCC

In table 2 we compare the effect of risk aversion and of ambiguity aversion. By definition, the SCC is the same for both calibrations when risk aversion γ and ambiguity aversion θ are both 0. In that case the expected value of both calibrations is the same and since risk is then not priced, the SCC is the same for both calibrations. Introducing risk aversion has a negligible effect on the SCC for the frequent disasters

Table 2: Social cost of carbon as function of risk aversion and ambiguity aversion

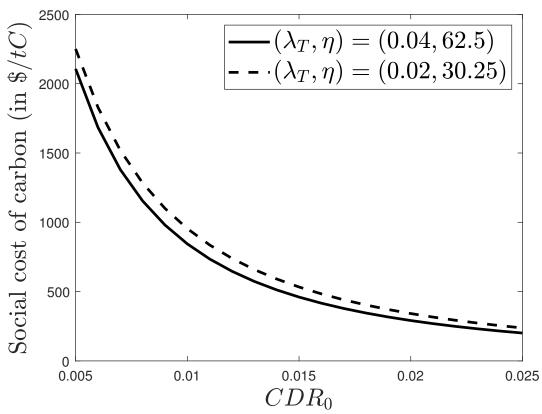
Social Cost of Carbon	$(\lambda_T, \eta) = (0.04, 62.5)$	$(\lambda_T, \eta) = (0.02, 30.25)$
$\gamma = 0$, $\theta = 0$	363	363
$\gamma = 5$, $\theta = 0$	360	392
$\gamma = 5$, $\theta = 0.01$	461	533

with low disaster size: for $(\lambda_T, \eta) = (0.04, 62.5)$ the direct impact of risk aversion on the certainty equivalent is small and is canceled out by the *discounting only* effect: for this configuration the SCC is even slightly lower than what it is without risk aversion. This changes when damages are more infrequent but larger. In the alternative calibration with $(\lambda_T, \eta) = (0.02, 30.25)$, risk aversion does increase the social cost of carbon from \$363 to \$392. This increase is still modest compared to the effect of ambiguity aversion. In both cases, introducing ambiguity aversion leads to a significantly higher value of the social cost of carbon. The table shows the very different implications that risk aversion and ambiguity aversion have for the valuation of climate risk.

Discount rates and the SCC

Figure 8 shows the dependence of the SCC on the time-independent part of the consumption discount rate CDR_0, the core discount rate. Note that the actual discount rate that is used to discount future damages (CDR_t) is higher than CDR_0 due to the effect of climate disasters itself on discounting. When core discount rates are close to zero, the social cost of carbon becomes very high. With $CDR_0 = 0.5\%$, the SCC is even above \$2000 in both cases, around four times higher than in the base calibration. On the other hand, setting $CDR_0 = 2.5\%$ gives a social cost of carbon that is less than half the value in the base calibration. This figure highlights the importance of the discount rate when analyzing climate change and in particular its impact on the social cost of carbon.

Figure 8: Social cost of carbon as a function of CDR_0.



8 Conclusions

Climate change will beyond reasonable doubt have a large impact on economic growth in the future. However, because of the complex nature of the problem and the lack of data, it is not possible to accurately estimate the timing and extent of its impact. But one thing we do know is that potentially large and irreversible consequences are likely to take place unless mitigating policies are implemented. But these changes will happen far into the future, while mitigating policies are (or should be) under consideration right now. That discrepancy puts the discussion on discounting at the center of the debate about the social cost of carbon and what we should do about climate change: to compare uncertain future damages with costs today, those future damages need to be discounted back towards today. The debate in the literature has largely zeroed in on the rate of time preference; the problem there is that to be consistent with capital market data, discount rates must be relatively high which in turn does not leave much once climate change consequences a century out are discounted back towards today (cf Weitzman (2007) for a very lucid overview of this debate). In this paper we squarely focus on the discounting question, but we take a different approach. Rather than discussing numerical values of certain parameters, we explore alternative specifications of preferences, and show the implications for the social cost of carbon.

We focus on the effect of Epstein-Zin recursive preferences on outcomes of the model, on the impact of unmeasurable risk (ambiguity) and the interaction between those two. Both breaking the link between γ and the EIS (by introducing stochastic differential utility, the continuous time implementation of Epstein-Zin preferences) and introducing ambiguity aversion are conceptually relevant in the climate change setting. The first extension is relevant because climate change problems have a very long horizon and therefore the elasticity of intertemporal substitution (EIS) unavoidably plays an important role. Arbitrarily restricting its value to $1/\gamma$ is then surely unsatisfactory. Second, conceptually ambiguity aversion is a logical extension, since we have no accurate estimation of climate damages nor in particular of their probability density function in the future. The assumption of unmeasurable risk ("Knightian uncertainty") then is a natural framework to use. Finally we highlight the sometimes complciated interactions between ambiguity aversion and intertemporal substitution elasticities for the value of the Social Cost of Carbon.

To do all this we set up an analytic IAM by extending a disaster risk model with a climate change model and a temperature dependent arrival rate. Furthermore we model climate risk as tail risk instead of assuming that temperature increases generate a certain amount of damage every year. The model is transparent because we manage to derive closed form solutions for the social cost of carbon. Where stochastic numerical IAMs can take hours to be solved, solving our model only requires numerical integration and is therefore solved within seconds.

Our base calibration generates a substantial social cost of carbon which is between \$461 and \$533 per ton of carbon. This is both much higher than for example the estimate of \$114 that is obtained using the DICE-2016R model (W. D. Nordhaus, 2017), and also much higher than current market prices in for example the European

Emissions Trading System.[11] Moreover we use our model to highlight how ambiguity aversion changes the social cost of carbon.

Analysing the effect of ambiguity aversion on the SCC is a complicated exercise since multiple potentially offsetting effects play a role: ambiguity aversion has both an effect on the arrival rate and certainty equivalent of disasters for given discount rates (more ambiguity aversion leads to a higher certainty equivalent) and on the discounting component. The effect of ambiguity aversion on discounting depends on the EIS. When $EIS < 1$, increasing ambiguity aversion leads to a smaller effective discount rate on climate damages. For the more interesting (because empirically supported) case $EIS > 1$, the opposite is true, in which case increasing ambiguity aversion has two offsetting effects on the SCC. However we show that the direct effect dominates and therefore that the presence of ambiguity aversion leads to a (substantially) higher social cost of carbon.

Lastly, we also show the importance of the consumption discount rate on the social cost of carbon. It is of course well known that the social cost of carbon is very sensitive to changes in the discount rate. However we stress that analyzing the discount rate impact of climate change involves more than a discussion of the pure rate of time preference on the discount rate; a low discount rate can also be caused by a high elasticity of intertemporal substitution, and additionally the discount rate depends in elaborate ways on the growth rate of the economy, volatility, risk aversion, climate disaster risk and ambiguity aversion. Disentangling these various effects and their interactions is the key contribution of this paper. One major theme emerges: proper risk pricing and incorporating ambiguity aversion leads to much higher estimates of the Social Cost of Carbon, literally by orders of magnitude. These findings are surely of more than just academic interest.

References

Anderson, E. W., Hansen, L. P., & Sargent, T. J. (2003). A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association*, *1*(1), 68–123.

Bansal, R., Kiku, D., & Ochoa, M. (2016). *Price of long-run temperature shifts in capital markets* (Tech. Rep.). National Bureau of Economic Research.

Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, *59*(4), 1481–1509.

Barnett, M., Brock, W., & Hansen, L. P. (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies*, *33*(3), 1024–1066.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, *121*(3), 823–866.

Barro, R. J. (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review*, *99*(1), 243–64.

[11] These prices are usually quoted per ton of carbon dioxide, which involves a conversion factor of 3.67.

Barro, R. J. (2015). Environmental protection, rare disasters and discount rates. *Economica*, *82*(325), 1–23.

Bremer, T., & van der Ploeg, F. (2018). The risk-adjusted carbon price. *OxCarre Research Paper*, *203*.

Bretscher, L., & Vinogradova, A. (2018). Escaping damocles' sword: Endogenous climate shocks in a growing economy. *Economics Working Paper Series*, *18*.

Cai, Y., & Lontzek, T. S. (2019). The social cost of carbon with economic and climate risks. *Journal of Political Economy*, *127*(6), 2684–2734.

Chen, Z., & Epstein, L. (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica*, *70*(4), 1403–1443.

Chichilnisky, G., Hammond, P. J., & Stern, N. (2018). Should we discount the welfare of future generations? ramsey and suppes versus koopmans and arrow. *Warwick economics research papers series (WERPS)(1174), http://wrap. warwick. ac. uk/107726*.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Crost, B., & Traeger, C. P. (2013). Optimal climate policy: uncertainty versus monte carlo. *Economics Letters*, *120*(3), 552–558.

Dietz, S. (2011). High impact, low probability? an empirical analysis of risk in the economics of climate change. *Climatic Change*, *108*(3), 519–541.

Dietz, S., van der Ploeg, R., Rezai, A., & Venmans, F. (2020). Are economists getting climate dynamics right and does it matter?

Duffie, D., & Epstein, L. G. (1992a). Asset pricing with stochastic differential utility. *The Review of Financial Studies*, *5*(3), 411–436.

Duffie, D., & Epstein, L. G. (1992b). Stochastic differential utility. *Econometrica: Journal of the Econometric Society*, 353–394.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, 643–669.

Epstein, L. G., & Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, *113*(1), 1–31.

Epstein, L. G., & Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica: Journal of the Econometric Society*, 937–969.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141 - 153. Retrieved from `http://www.sciencedirect.com/science/article/pii/0304406889900189` doi: https://doi.org/10.1016/0304-4068(89)90018-9

Golosov, M., Hassler, J., Krusell, P., & Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, *82*(1), 41–88.

Goosse, H. (2015). *Climate system dynamics and modelling*. Cambridge University Press. Retrieved from `https://books.google.at/books?id=2-NBCgAAQBAJ`

Hambel, C., Kraft, H., & Schwartz, E. (2019). *Optimal carbon abatement in a stochastic equilibrium model with climate change* (Tech. Rep.). National Bureau of Economic Research.

Hansen, L. P., & Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, *91*(2), 60–66.

Hansen, L. P., & Sargent, T. J. (2008). *Robustness*. Princeton university press.

Hansen, L. P., Sargent, T. J., Turmuhambetova, G., & Williams, N. (2001). Robustness and uncertainty aversion. *Manuscript, University of Chicago*.

Hope, C. (2006). The marginal impact of co2 from page2002: An integrated assessment model incorporating the ipcc's five reasons for concern. *Integrated assessment*, *6*(1).

IPCC. (2013). *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* [Book]. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Retrieved from `www.climatechange2013.org` doi: 10.1017/CBO9781107415324

Jensen, S., & Traeger, C. P. (2014). Optimal climate change mitigation under long-term growth uncertainty: Stochastic integrated assessment and analytic findings. *European Economic Review*, *69*, 104–125.

Karydas, C., & Xepapadeas, A. (2019). *Climate change risks: pricing and portfolio allocation* (Tech. Rep.). CER-ETH-Center of Economic Research (CER-ETH) at ETH Zurich.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, *73*(6), 1849–1892.

Knight, F. H. (1921). Risk, uncertainty and profit. *New York: Hart, Schaffner and Marx*.

Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, 185–200.

Lemoine, D., & Traeger, C. P. (2016). Ambiguous tipping points. *Journal of Economic Behavior & Organization*, *132*, 5–18.

Liu, J., Pan, J., & Wang, T. (2004). An equilibrium model of rare-event premia and its implication for option smirks. *The Review of Financial Studies*, *18*(1), 131–164.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, 1429–1445.

Maenhout, P. J. (2004). Robust portfolio rules and asset pricing. *Review of financial studies*, *17*(4), 951–983.

Maenhout, P. J. (2006). Robust portfolio rules and detection-error probabilities for a mean-reverting risk premium. *Journal of Economic Theory*, *128*(1), 136–163.

Mattauch, L., Millar, R., van der Ploeg, R., Rezai, A., Schultes, A., Venmans, F., ... Teytelboym, A. (2018). Steering the climate system: an extended comment.

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of monetary Economics*, *15*(2), 145–161.

Millner, A., Dietz, S., & Heal, G. (2013). Scientific ambiguity and climate policy. *Environmental and Resource Economics*, *55*(1), 21–46.

Munk, C. (2015). *Dynamic asset allocation*.

Noble, I., Bolin, B., Ravindranath, N., Verardo, D., & Dokken, D. (2000). *Land use, land use change, and forestry*. Cambridge University Press.

Nordhaus, W. (2014). Estimates of the social cost of carbon: concepts and results from the dice-2013r model and alternative approaches. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 273–312.

Nordhaus, W. D. (2014). *A question of balance: Weighing the options on global warming policies*. Yale University Press.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 201609244.

Olijslagers, S. (2020). The social cost of carbon: Optimal policy versus business-as-usual.

Olijslagers, S., van der Ploeg, R., & van Wijnbergen, S. (2020). Commit to a credible path of rising co2 prices.

Pathak, P. (2002). Notes on robust portfolio choice. *unpublished paper, Harvard University*, *3*, 14.

Pindyck, R. S. (2017). The use and misuse of models for climate policy. *Review of Environmental Economics and Policy*, *11*(1), 100–114.

Ramsey, F. P. (1928). A mathematical theory of saving. *The economic journal*, *38*(152), 543–559.

Rietz, T. A. (1988). The equity risk premium a solution. *Journal of monetary Economics*, *22*(1), 117–131.

Stern, N. (2015). *Why are we waiting?: The logic, urgency, and promise of tackling climate change*. Mit Press.

Strzalecki, T. (2011). Axiomatic foundations of multiplier preferences. *Econometrica*, *79*(1), 47–73.

Tol, R. S. (2002). Estimates of the damage costs of climate change. part 1: Benchmark estimates. *Environmental and resource Economics*, *21*(1), 47–73.

Traeger, C. P. (2014). Why uncertainty matters: discounting under intertemporal risk aversion and ambiguity. *Economic Theory*, *56*(3), 627–664.

Traeger, C. P. (2018). Ace–analytic climate economy (with temperature and uncertainty).

Tsai, J., & Wachter, J. A. (2015). Disaster risk and its implications for asset pricing. *Annual Review of Financial Economics*, *7*, 219–252.

Van Binsbergen, J. H., Fernández-Villaverde, J., Koijen, R. S., & Rubio-Ramírez, J. (2012). The term structure of interest rates in a dsge model with recursive preferences. *Journal of Monetary Economics*, *59*(7), 634–648.

Vissing-Jørgensen, A., & Attanasio, O. P. (2003). Stock-market participation, intertemporal substitution, and risk-aversion. *American Economic Review*, *93*(2), 383–391.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, *68*(3), 987–1035.

Weitzman, M. L. (2007). A review of the stern review on the economics of climate change. *Journal of economic literature*, *45*(3), 703–724.

A Overview of methods to model ambiguity

There are several different approaches that are commonly used in the literature to model ambiguity about parameters. A widely used approach in the static setting is the maxmin approach of Gilboa and Schmeidler (1989). Assume the agent does not know the distribution of a random variable. The idea is to first specify a set of reasonable probability measures $\mathbb{Q}$. The agent is ambiguity averse and given this set of measures he considers the worst case measure. Utility is then of the form $V_t = \min_{\mathbb{Q}\in\mathcal{P}} E_{\mathbb{Q}}[U(C_t)]$ for some utility specification $U(.)$.

It is not straightforward to extend the Gilboa-Schmeidler maxmin preferences to a dynamic setting. We will discuss two approaches that have been proposed by Chen and Epstein (2002) and by Hansen, Sargent, Turmuhambetova, and Williams (2001) in the setting of our model.

Consider the agent's problem. In the setting without ambiguity, the value function is given by:

$$V_t = E_t\Big[\int_t^\infty f(C_s, V_s)ds\Big]. \tag{31}$$

However, in the model with ambiguity the agent takes into account the fact that he is not certain about the true value of the arrival rate λ_t and the jump size parameter η.

Hansen et al. (2001) propose two approaches to model ambiguity: the *multiplier* approach and *constraint* approach. We first consider the *multiplier* approach. The 'best estimate' model or reference model is the agent's most reliable model with measure $\mathbb{P}$. But he also takes into account other, alternative models. The alternative models have measure $\mathbb{Q}^{a,b}$, the jump arrival rate becomes $\lambda_t^{\mathbb{Q}} = a_t\lambda_t$ and the jump size parameter becomes $\eta_t^{\mathbb{Q}} = b_t\eta$. Deviating from the reference model is penalized since the agent does not choose the 'best estimate' model. The size of the penalty is proportional to $d(a_t, b_t)$, which represents the distance between the alternative model and the reference model. An alternative model that has a large distance to the reference model is considered less likely to be true and therefore using it receives a larger penalty. The distance function should satisfy $d(a_t, b_t) \geq 0\ \forall(a_t, b_t)$ and $d(1, 1) = 0$. Therefore using the reference model carries a zero penalty. The penalty is scaled by θ, which is the ambiguity aversion parameter. This parameter controls the importance of the penalty term. Then the agent solves the following problem:

$$V_t = \min_{\{a_s,b_s\}_{s\geq t}} E_t^{\mathbb{Q}}\Big[\int_t^\infty \Big(f(C_s, V_s) + e^{-\beta(s-t)}\theta d(a_s, b_s)\Big)ds\Big]. \tag{32}$$

$E_t^{\mathbb{Q}}$ denotes the expectation under the alternative model with parameters $\lambda_t^{\mathbb{Q}}$ and $\eta_t^{\mathbb{Q}}$. The expected utility of consumption is lower for high a_t and low b_t. We see that the agent faces a trade-off between how likely a combination of (a_s, b_s) is in terms of distance to the reference model and how bad it is in terms of expected utility. This trade-off results in optimal values of a_s and b_s.

The *constraint* approach is closely related to the *multiplier* approach. Instead of adding a penalty function, the agent can put a constraint on the distance function

$d(a_t, b_t)$. Then the problem becomes:

$$
\begin{aligned}
V_t = &\min_{\{a_s, b_s\}_{s \geq t}} E_t^{\mathbb{Q}} \Big[\int_t^\infty f(C_s, V_s) ds \Big], \\
&\text{s.t.} \int_t^\infty e^{-\beta(s-t)} d(a_s, b_s) ds \leq \phi.
\end{aligned}
\tag{33}
$$

So in this approach, ϕ to controls the size of the set of alternative models that seem reasonable to him. ϕ can again be seen as an ambiguity aversion parameter. A high ϕ implies a large set of priors and therefore corresponds to high ambiguity aversion. Given the constraint the worst-case model is chosen. This approach is very similar to the penalty approach and the two problems are related via Lagrangian optimization where θ can be seen as the Lagrange multiplier.

Hansen and Sargent (2001) consider how the penalty and constraint approaches are related. They show that when the consumption process follows a pure geometric Brownian motion (i.e. no jumps), there exists a ϕ for the constraint approach and a θ for the multiplier approach such that both problems yield the same optimal outcome. The constraint approach is directly motivated from the Gilboa-Schmeidler maxmin utility. Since the multiplier approach is weakly related to the constraint approach, these approaches are indirectly also motivated by the static maxmin utility. Furthermore, the multiplier utility is axiomatized by Strzalecki (2011).

A disadvantage of both these approaches is that utility is not homothetic. Maenhout (2004) proposes to use a state-dependent Lagrange-multiplier $\theta(V_t)$ in the framework of the multiplier approach to obtain homothetic utility. This approach is also adopted by Liu et al. (2004). However, by assuming that the ambiguity aversion parameter θ can be state dependent, the relation with the constraint preferences is lost. Pathak (2002) extensively discusses this issue. He argues that the main motivation of the multiplier approach by Hansen et al. (2001) is through the constraint approach. But with the state-dependent ambiguity aversion parameter this new utility specification cannot be seen anymore as a dynamic extension of the Gilboa-Schmeidler utility. Furthermore, the axiomatic foundation is not valid anymore. Pathak (2002) proposes an alternative method to model ambiguity: *recursive multiple priors utility* developed in continuous time by Chen and Epstein (2002).

We follow the advise of Pathak (2002) and in contrast to Liu et al. (2004) we choose to use the approach of Chen and Epstein (2002). The approach is closely related to the constraint approach of Hansen et al. (2001), but does preserve the homotheticity of the preferences. Consider the following problem:

$$
\begin{aligned}
V_t = &\min_{\{a_s, b_s\}_{s \geq t}} E_t^{\mathbb{Q}} \Big[\int_t^\infty f(C_s, V_s) ds \Big], \\
&\text{s.t. } d(a_t, b_t) \leq \theta \quad \forall t.
\end{aligned}
\tag{34}
$$

The main difference with the constraint approach of Hansen et al. (2001) is that not the lifetime distance between the reference measure and the alternative measure is bounded, but at every time period t the distance between the measures is bounded.

This approach leads to more tractable solutions. The recursive multiple priors utility is axiomatized by Epstein and Schneider (2003).

Lastly we will briefly discuss the *smooth ambiguity* model, since it is often used in the literature as well. Assume that the agent does not know the true values of λ and η. In this approach the agent first constructs a probability distribution that reflects his beliefs on λ and η. Define $p(x, y) = P(\lambda = x, \eta = y)$. To incorporate ambiguity aversion, he then transforms this distribution to put more weight on the events that give him low utility and less weight on the events that give high utility. This results in the following problem:

$$V_t = \int_0^\infty \int_0^\infty \Big(p(x,y)\phi\Big(E_t\Big[\int_t^\infty f(C_t, V_t)ds \Big| \lambda = x, \eta = y \Big]\Big)\Big) dxdy. \tag{35}$$

Here the function ϕ controls the ambiguity aversion of the agent. When ϕ is a concave function, the agent is ambiguity averse. This approach was introduced by Klibanoff et al. (2005). This may be a matter of taste, but we think that the assumption of probabilities attached to the different priors is in fact at variance with the basic assumption that ambiguity is about unmeasurable processes, i.e. we cannot map events to probability densities, or in this case priors to model probabilities. And since the recursive multiple priors utility is intuitive and leads to tractable results, we chose not to move in the direction of the smooth ambiguity model.

B Derivation of Relative Entropy

For each $a = (a_t)_{t\geq 0}$ and $b = (b_t)_{t\geq 0}$ we define the measure $\mathbb{Q}^{a,b}$ which is equivalent to $\mathbb{P}$ and has Radon-Nikodym derivative $\frac{d\mathbb{Q}^{a,b}}{d\mathbb{P}}|_{\mathcal{F}_t} = \xi_t^{a,b}$ where $\xi_t^{a,b}$ follows:

$$d\xi_t^{a,b} = (\lambda_t - \lambda_t^{\mathbb{Q}})\xi_t^{a,b}dt + \Big(\frac{\lambda_t^{\mathbb{Q}} f^{\mathbb{Q}}(J_t)}{\lambda_t f(J_t)} - 1 \Big)\xi_{t-}^{a,b} dN_t. \tag{36}$$

and $\xi_0^{a,b} = 1$. $\xi_t^{a,b}$ is chosen such that the jump distribution under $\mathbb{Q}^{a,b}$ has arrival rate $\lambda_t^{\mathbb{Q}}$ and such that the pdf of the jump distribution equals $f^{\mathbb{Q}}(x)$. Specifically we assume that $\lambda_t^{\mathbb{Q}} = a_t\lambda_t$ and $\eta_t^{\mathbb{Q}} = b_t\eta$. We can calculate in this case the fraction of the two probability distributions: $\frac{f^{\mathbb{Q}}(x)}{f(x)} = \frac{\eta_t^{\mathbb{Q}}(1+x)^{\eta_t^{\mathbb{Q}}-1}}{\eta(1+x)^{\eta-1}} = b_t(1+x)^{(b_t-1)\eta}$. Substituting this into 36 gives:

$$d\xi_t^{a,b} = (1-a_t)\lambda_t\xi_t^{a,b}dt + \Big(a_t b_t (1+J_t)^{(b_t-1)\eta} - 1 \Big)\xi_{t-}^{a,b} dN_t. \tag{37}$$

The Radon-Nikodym derivative $\xi_t^{a,b}$ that we have specified is the ratio between the alternative measure $\mathbb{Q}^{a,b}$ and the reference measure $\mathbb{P}$. We can use it to determine the relative entropy between the two measures. The relative entropy between $\mathbb{Q}^{a,b}$ and $\mathbb{P}$ over time unit Δ is defined as $E_t^{\mathbb{Q}}\Big[\log(\frac{\xi_{t+\Delta}^{a,b}}{\xi_t^{a,b}}) \Big]$. Here $E_t^{\mathbb{Q}}$ denotes the expectation with respect to the alternative measure $\mathbb{Q}^{a,b}$. Then divide by Δ and let $\Delta \to 0$ to obtain the instantaneous relative entropy: $RE(a_t, b_t) = \lim_{\Delta\to 0} \frac{1}{\Delta} E_t^{\mathbb{Q}}\Big[\log\Big(\frac{\xi_{t+\Delta}^{a,b}}{\xi_t^{a,b}} \Big) \Big]$.

Applying Itô's lemma for jump processes to $\xi_t^{a,b}$, we obtain the following dynamics for $\log(\xi_t^{a,b})$:

$$d\log(\xi_t^{a,b}) = (1-a_t)\lambda_t dt + \Big(\log(a_t b_t) + (b_t - 1)\eta\log(1+J_t)\Big)dN_t. \tag{38}$$

Using integration by parts we can calculate that $E_t^{\mathbb{Q}}[\log(1+J_t)] = -\frac{1}{\eta_t^{\mathbb{Q}}}$. Therefore the (instantaneous) relative entropy at time t equals:

$$\begin{aligned} RE(a_t, b_t) = \lim_{\Delta\to 0}\frac{1}{\Delta}E_t^{\mathbb{Q}}\Big[\log\Big(\frac{\xi_{t+\Delta}^{a,b}}{\xi_t^{a,b}}\Big)\Big] = \\ (1-a_t)\lambda_t + a_t\lambda_t\Big(\log(a_t b_t) + \frac{1}{b_t} - 1\Big). \end{aligned} \tag{39}$$

C Calculating the detection error probability

After observing the process of consumption over a period N years, what is the probability of choosing the wrong model? Let us start with the case that the reference model $\mathbb{P}$ is the true model and the agent considers the alternative model $\mathbb{Q}^{a,b}$. Note that the Radon-Nikodym derivative informs us about the likelihood ratio of both models. When this derivative is larger than one after N years, the worst-case model $\mathbb{Q}^{a,b}$ is the most likely and the agent will choose the wrong model. The probability of making this error is equal to (see for example Maenhout (2006)):

$$Pr\Big(\xi_N^{a,b} > 1|\mathbb{P}\Big) = Pr\Big(\log(\xi_N^{a,b}) > 0|\mathbb{P}\Big). \tag{40}$$

We calculate this probability by simulating the process of $\log(\xi_t^{a,b})$ forward. Simulation is done via a standard Euler method. Similarly, we can define the opposite mistake where the alternative model is actually true and the agent chooses the reference model. We now define the inverse Radon-Nikodym derivative: $\frac{d\mathbb{P}}{d\mathbb{Q}^{a,b}}|_{\mathcal{F}_t} = \tilde{\xi}_t^{a,b}$ where $\tilde{\xi}_t^{a,b}$ follows:

$$d\tilde{\xi}_t^{a,b} = (a_t - 1)\lambda_t\tilde{\xi}_t^{a,b}dt + \Big(\frac{1}{a_t b_t}(1+J)^{(1-b_t)\eta} - 1\Big)\tilde{\xi}_{t-}^{a,b}dN_t. \tag{41}$$

Applying Itô's lemma gives:

$$d\log(\tilde{\xi}_t^{a,b}) = (a_t - 1)\lambda_t dt + \Big(-\log(a_t b_t) + (1-b_t)\eta\log(1+J_t)\Big)dN_t. \tag{42}$$

The probability of choosing the wrong model when actually the alternative model $\mathbb{Q}^{a,b}$ is true equals:

$$Pr\Big(\tilde{\xi}_N^{a,b} > 1|\mathbb{Q}\Big) = Pr\Big(\log(\tilde{\xi}_N^{a,b}) > 0|\mathbb{Q}\Big). \tag{43}$$

Again this probability can be calculated by simulating the process $\log(\tilde{\xi}_t)$ forward. The detection error probability is then defined as:

$$\frac{1}{2}Pr\Big(\log(\xi_N^{a,b}) > 0|\mathbb{P}\Big) + \frac{1}{2}Pr\Big(\log(\tilde{\xi}_N^{a,b}) > 0|\mathbb{Q}\Big). \tag{44}$$

D Hamilton-Jacobi-Bellman equation

We will first derive the Hamilton-Jacobi-Bellman equation for every measure $\mathbb{Q}^{a,b}$. Duffie and Epstein (1992b) show that the HJB-equation for stochastic differential utility equals:

$$0 = f(C_t, V_t^{\mathbb{Q}}) + \mathcal{D}\mathcal{V}^{\mathbb{Q}}. \tag{45}$$

Here $\mathcal{D}\mathcal{V}^{\mathbb{Q}}$ is the drift of the value function. In order to calculate the drift of the value function, we will apply Itô's lemma. By Itô's lemma for jump processes we have:

$$\begin{aligned} dV_t^{\mathbb{Q}} &= V_C^{\mathbb{Q}}\Big(\mu C_t dt + \sigma C_t dZ_t^{\mathbb{Q}}\Big) + V_X^{\mathbb{Q}}\mu_X(X_t)dt + \frac{1}{2}V_{CC}^{\mathbb{Q}}\sigma^2 C_t^2 dt \\ &\quad + \Big(V^{\mathbb{Q}}\big((1+J_t)C_{t-}, X_t\big) - V^{\mathbb{Q}}(C_{t-}, X_t)\Big)dN_t \end{aligned} \tag{46}$$

Then the drift under $\mathbb{Q}^{a,b}$ equals:

$$\begin{aligned} \mathcal{D}\mathcal{V}^{\mathbb{Q}} &= V_C^{\mathbb{Q}}\mu C_t + V_X^{\mathbb{Q}}\mu_X(X_t) + \frac{1}{2}V_{CC}^{\mathbb{Q}}\sigma^2 C_t^2 \\ &\quad + \lambda_t^{\mathbb{Q}} E_t^{\mathbb{Q}}\big[V^{\mathbb{Q}}\big((1+J_t)C_{t-}, X_t\big) - V^{\mathbb{Q}}(C_{t-}, X_t)\big] \end{aligned} \tag{47}$$

This gives the following Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} 0 &= f(C_t, V_t^{\mathbb{Q}}) + V_C^{\mathbb{Q}}\mu C_t + V_X^{\mathbb{Q}}\mu_X(X_t) + \frac{1}{2}V_{CC}^{\mathbb{Q}}\sigma^2 C_t^2 \\ &\quad + \lambda_t^{\mathbb{Q}} E_t^{\mathbb{Q}}\big[V^{\mathbb{Q}}\big((1+J_t)C_{t-}, X_t\big) - V^{\mathbb{Q}}(C_{t-}, X_t)\big] \end{aligned} \tag{48}$$

E Reduced HJB-equation

Substituting our conjecture $V^{\mathbb{Q}}(C_t, X_t) = \frac{g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}}{1-\gamma}$ into $f(C_t, V_t)$ gives:

$$\begin{aligned} f(C_t, V^{\mathbb{Q}}(C_t, X_t)) &= \frac{\beta}{1-1/\epsilon}\frac{C_t^{1-1/\epsilon} - \Big(g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}\Big)^{\frac{1}{\zeta}}}{\Big(g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}\Big)^{\frac{1}{\zeta}-1}} \\ &= \frac{\beta}{1-1/\epsilon}\Big(g^{\mathbb{Q}}(X_t)^{1-\frac{1}{\zeta}}C_t^{1-\gamma} - g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}\Big) \\ &= \beta\zeta\Big(g^{\mathbb{Q}}(X_t)^{-\frac{1}{\zeta}} - 1\Big)V^{\mathbb{Q}}(C_t, X_t). \end{aligned} \tag{49}$$

The partial derivatives of V are given by:

$$\begin{aligned} V_C^{\mathbb{Q}} &= g^{\mathbb{Q}}(X_t)C_t^{-\gamma}, \qquad V_{CC}^{\mathbb{Q}} = -\gamma g^{\mathbb{Q}}(X_t)C_t^{-\gamma-1}, \\ V_X^{\mathbb{Q}} &= \frac{g_X^{\mathbb{Q}}(X_t)C_t^{1-\gamma}}{1-\gamma}. \end{aligned} \tag{50}$$

Here $g_X^{\mathbb{Q}}$ denotes the row vector with partial derivatives to each of the state variables, similar to $V_X^{\mathbb{Q}}$. Additionally we can calculate the expectation:

$$\begin{aligned} E_t^{\mathbb{Q}}\big[V^{\mathbb{Q}}\big((1+J_t)C_{t-},X_t\big)-V^{\mathbb{Q}}(C_{t-},X_t)\big] &= \frac{E_t^{\mathbb{Q}}\big[(1+J_t)^{1-\gamma}\big]-1}{1-\gamma}g^{\mathbb{Q}}(X_t)C_t^{1-\gamma} \\ = \frac{\frac{b_t\eta}{b_t\eta+1-\gamma}-1}{1-\gamma}g^{\mathbb{Q}}(X_t)C_t^{1-\gamma} &= \frac{-1}{b_t\eta+1-\gamma}g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}. \end{aligned} \tag{51}$$

Substituting $f(C_t,V^{\mathbb{Q}}(C_t,X_t))$ together with the partial derivatives of $V_t^{\mathbb{Q}}$ and the expectation into (12) yields the following equation:

$$\begin{aligned} 0 = &\frac{\beta}{1-1/\epsilon}\Big(g^{\mathbb{Q}}(X_t)^{-\frac{1}{\zeta}}-1\Big)g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}+\mu g^{\mathbb{Q}}(X_t)C_t^{1-\gamma} \\ &-\frac{\gamma}{2}\sigma^2 g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}+\frac{g_X^{\mathbb{Q}}(X_t)C_t^{1-\gamma}}{1-\gamma}\mu_X(X_t)+a_t\lambda_t\frac{-1}{b_t\eta+1-\gamma}g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}. \end{aligned} \tag{52}$$

Dividing by $g^{\mathbb{Q}}(X_t)C_t^{1-\gamma}$ gives:

$$\begin{aligned} 0 = &\frac{\beta}{1-1/\epsilon}\Big(g^{\mathbb{Q}}(X_t)^{-\frac{1}{\zeta}}-1\Big)+\mu-\frac{\gamma}{2}\sigma^2+\frac{g_X^{\mathbb{Q}}(X_t)}{g^{\mathbb{Q}}(X_t)(1-\gamma)}\mu_X(X_t) \\ &+a_t\lambda_t\frac{-1}{b_t\eta+1-\gamma}. \end{aligned} \tag{53}$$

F Asset prices

F.1 The Pricing Kernel

Duffie and Epstein (1992a) derive that the pricing kernel with stochastic differential utility equals $\pi_t = \exp\Big\{\int_0^t f_V(C_s,V_s)ds\Big\}f_C(C_t,V_t)$. However, the pricing kernel has to be adjusted for the ambiguity aversion preferences. Chen and Epstein (2002) show that the pricing kernel in the ambiguity setting should be multiplied by the Radon-Nikodym derivative $\xi_t^{a^*,b^*}$ of the measure corresponding to the optimal a^* and b^*. $\xi_t^{a,b}$ is defined in (36).

We will start with deriving the explicit stochastic differential equation of the pricing kernel. First we calculate the derivatives of $f(C_t,V_t)$ with respect to C_t and V_t:

$$\begin{aligned} f_C(C,V) &= \frac{\beta C^{-1/\epsilon}}{\big((1-\gamma)V\big)^{\frac{1}{\zeta}-1}}, \\ f_V(C,V) &= \beta\zeta\Big\{\Big(1-\frac{1}{\zeta}\Big)\Big((1-\gamma)V\Big)^{-\frac{1}{\zeta}}C^{1-1/\epsilon}-1\Big\}. \end{aligned} \tag{54}$$

Substituting $V_t = g(X_t)\frac{C_t^{1-\gamma}}{1-\gamma}$ into $f_C(C_t,V_t)$ and $f_V(C_t,V_t)$ we obtain:

$$\begin{aligned} f_C(C_t,V_t) &= \beta g(X_t)^{1-\frac{1}{\zeta}}C_t^{-\gamma}, \\ f_V(C_t,V_t) &= \beta\zeta\Big\{g(X_t)^{-\frac{1}{\zeta}}\Big(1-\frac{1}{\zeta}\Big)-1\Big\}. \end{aligned} \tag{55}$$

This gives:

$$\pi_t = \xi_t^{a^*,b^*} \exp\left(\int_0^t \beta\zeta \Big(g(X_s)^{-\frac{1}{\zeta}} (1 - \frac{1}{\zeta}) - 1 \Big) ds \right) \beta g(X_t)^{1-\frac{1}{\zeta}} C_t^{-\gamma}. \tag{56}$$

Take the logarithm and write as a differential equation:

$$\begin{aligned} d\log(\pi_t) &= \beta\zeta \Big(g(X_t)^{-\frac{1}{\zeta}} (1 - \frac{1}{\zeta}) - 1 \Big) dt - \gamma d\log(C_t) + d\log(\xi_t^{a^*,b^*}) \\ &\quad + (1 - \frac{1}{\zeta}) d\log\big(g(X_t)\big). \end{aligned} \tag{57}$$

Apply Ito's lemma to $\log(C_t)$, $\log(\xi_t^{a^*,b^*})$ and $\log\big(g(X_t)\big)$ and substitute the results; this leads to the following differential equation:

$$\begin{aligned} d\log(\pi_t) =& \Big\{ \beta\zeta \Big(g(X_t)^{-\frac{1}{\zeta}} (1 - \frac{1}{\zeta}) - 1 \Big) - \gamma\big(\mu - \frac{\sigma^2}{2}\big) + \lambda_t(1 - a_t^*) \\ &+ (1/\epsilon - \gamma) \frac{g_X(X_t)}{g(X_t)(1-\gamma)} \mu_X(X_t) \Big\} dt \\ &- \gamma\sigma dZ_t + \Big(\log(a_t^* b_t^*) + \big((b_t^* - 1)\eta - \gamma\big) \log(1 + J_t) \Big) dN_t. \end{aligned} \tag{58}$$

After applying Ito's lemma to $\log(\pi_t)$ we find:

$$\begin{aligned} d\pi_t =& \Big\{ \beta\zeta \Big(g(X_t)^{-\frac{1}{\zeta}} (1 - \frac{1}{\zeta}) - 1 \Big) - \gamma\big(\mu - (\gamma+1)\frac{\sigma^2}{2}\big) + \lambda_t(1 - a_t^*) \\ &+ (1/\epsilon - \gamma) \frac{g_X(X_t)}{g(X_t)(1-\gamma)} \mu_X(X_t) \Big\} \pi_t dt + -\gamma\sigma\pi_t dZ_t \\ &+ \Big(a_t^* b_t^* (1 + J_t)^{(b_t^*-1)\eta - \gamma} - 1 \Big) \pi_{t-} dN_t. \end{aligned} \tag{59}$$

We can now substitute the HJB equation (53) into the pricing kernel. Several terms cancel out and we are left with:

$$\begin{aligned} d\pi_t =& \Big\{ -\beta - \frac{\mu}{\epsilon} + \Big(1 + \frac{1}{\epsilon}\Big) \frac{\gamma}{2}\sigma^2 + \Big(\gamma - \frac{1}{\epsilon}\Big) \lambda_t^* \frac{-1}{b_t^*\eta + 1 - \gamma} \\ &+ \lambda_t(1 - a_t^*) \Big\} \pi_t dt - \gamma\sigma\pi_t dZ_t \\ &+ \Big(a_t^* b_t^* (1 + J_t)^{(b_t^*-1)\eta - \gamma} - 1 \Big) \pi_{t-} dN_t. \end{aligned} \tag{60}$$

F.2 The interest rate

By the no-arbitrage argument, r_t should be such that $\pi_t B_t$ is a martingale, where B_t is the price of the risk-free asset. Now write $d\pi_t = \mu_{\pi,t}\pi_t dt + \sigma_\pi \pi_t dZ_t + J_{\pi,t}\pi_{t-} dN_t$. The product with B_t then follows:

$$d\pi_t B_t = (r_t + \mu_{\pi,t})\pi_t B_t dt + \sigma_\pi \pi_t B_t dZ_t + J_{\pi,t}\pi_{t-} B_t dN_t. \tag{61}$$

This is a martingale if $r_t + \mu_\pi + \lambda_t E_t[J_{\pi,t}] = r_t + \mu_\pi + \lambda_t\Big(a_t^* \frac{b_t^*\eta}{b_t^*\eta-\gamma} - 1\Big) = 0$. Therefore the interest rate equals:

$$
\begin{aligned}
r_t =& -\mu_\pi - \lambda_t\Big(a_t^* \frac{b_t^*\eta}{b_t^*\eta - \gamma} - 1\Big) \\
=& \beta + \frac{\mu}{\epsilon} - \Big(1 + \frac{1}{\epsilon}\Big)\frac{\gamma}{2}\sigma^2 - \Big(\gamma - \frac{1}{\epsilon}\Big)a_t^*\lambda_t \frac{-1}{b_t^*\eta + 1 - \gamma} \\
& - a_t^*\lambda_t\Big(\frac{b_t^*\eta}{b_t^*\eta - \gamma} - 1\Big).
\end{aligned} \tag{62}
$$

Substituting r_t into the pricing kernel gives:

$$
\begin{aligned}
d\pi_t =& \Big\{ - r_t - \lambda_t\Big(a_t^* \frac{b_t^*\eta}{b_t^*\eta - \gamma} - 1\Big)\Big\}\pi_t dt - \gamma\sigma\pi_t dZ_t \\
& + \Big(a_t^* b_t^*(1 + J_t)^{(b_t^*-1)\eta-\gamma} - 1\Big)\pi_{t-}dN_t.
\end{aligned} \tag{63}
$$

F.3 The equity premium

Using equation (25), we know that the drift of the stock equals $\mu_{S,t} = \mu - \frac{k_X(X_t)}{k(X_t)}\mu_X(X_t) + k(X_t)$. From (23) we have: $k(X_t) = \beta g(X_t)^{-\frac{1-1/\epsilon}{1-\gamma}}$. This gives: $\frac{k_X(X_t)}{k(X_t)} = -\frac{1-1/\epsilon}{1-\gamma}\frac{g_X(X_t)}{g(X_t)}$. Rewriting the HJB equation (53) gives:

$$
\begin{aligned}
\frac{1-1/\epsilon}{1-\gamma}\frac{g_X(X_t)}{g(X_t)}\mu_X(X_t) + k(X_t) =& \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2 \\
& + a_t^*\lambda_t \frac{-1}{b_t^*\eta + 1 - \gamma}\Big).
\end{aligned} \tag{64}
$$

Substituting this into $\mu_{S,t}$ gives:

$$
\begin{aligned}
\mu_{S,t} =& \mu - \frac{k_X(X_t)}{k(X_t)}\mu_X(X_t) + k(X_t) \\
=& \mu + \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2 + a_t^*\lambda_t \frac{-1}{b_t^*\eta + 1 - \gamma}\Big).
\end{aligned} \tag{65}
$$

The risk premium is then equal to the excess return of the stock over the interest rate:

$$
\begin{aligned}
rp_t =& \mu_{S,t} + a_t^*\lambda_t \frac{-1}{b_t^*\eta + 1} - r_t \\
=& \gamma\sigma^2 + a_t^*\lambda_t\Big(\frac{-1}{b_t^*\eta + 1} - \frac{b_t^*\eta}{b_t^*\eta + 1 - \gamma} + \frac{b_t^*\eta}{b_t^*\eta - \gamma}\Big)
\end{aligned} \tag{66}
$$

F.4 Consumption strips

Let $H_t = H(C_t, X_t, s - t) = E_t\Big[\frac{\pi_s}{\pi_t}C_s\Big]$ be the price of an asset that pays out the aggregate consumption at time s. H_t is also called a consumption strip. Conjecture that $H(C_t, X_t, u) = \exp\Big\{-\int_t^{t+u} CDR_s ds\Big\}C_t$. u denotes the time to maturity of the consumption strip. Clearly, $H(C_t, X_t, 0) = C_t$. Applying Ito's lemma to H_t gives:

$$
\begin{aligned}
dH_t = H_C dC_t + H_X dX_t - \frac{\partial H_t}{\partial u}dt = \frac{1}{C_t}H_t dC_t \\
- \frac{\partial}{\partial X_t}\Big(\int_t^{t+u} CDR_s ds\Big)\mu_X(X_t)H_t dt \\
+ \frac{\partial}{\partial u}\Big(\int_t^{t+u} CDR_s ds\Big)H_t dt.
\end{aligned} \tag{67}
$$

We can calculate both derivatives:

$$
\begin{aligned}
\frac{\partial}{\partial X_t}\Big(\int_t^{t+u} CDR_s ds\Big)\mu_X(X_t) = \frac{\partial}{\partial t}\Big(\int_t^{t+u} CDR_s ds\Big)\frac{\partial t}{\partial X_t}\mu_X(X_t) \\
= \frac{\partial}{\partial t}\Big(\int_t^{t+u} CDR_s ds\Big) = CDR_{t+u} - CDR_t,
\end{aligned} \tag{68}
$$

$$
\frac{\partial}{\partial u}\Big(\int_t^{t+u} CDR_s ds\Big) = CDR_{t+u}. \tag{69}
$$

Therefore dH_t becomes:

$$
dH_t = \Big(\mu + CDR_t\Big)H_t dt + \sigma H_t dZ_t + J_t H_{t-} dN_t. \tag{70}
$$

Now define $dH_t = \mu_{H,t}H_t dt + \sigma H_t dZ_t + J_t H_{t-} dN_t$. By the no arbitrage condition, $\pi_t H_t$ must be a martingale:

$$
\begin{aligned}
d\pi_t H_t = (\mu_{\pi,t} + \mu_H + \sigma\sigma_\pi)\pi_t H_t dt + (\sigma + \sigma_\pi)\pi_t H_t dZ_t \\
+ \Big((1 + J_t)(1 + J_{\pi,t}) - 1\Big)\pi_{t-}H_{t-}dN_t.
\end{aligned} \tag{71}
$$

We can calculate the expectation of the jump term:

$$
\begin{aligned}
E_t[(1 + J_t)(1 + J_{\pi,t}) - 1] = E_t[a_t^* b_t^*(1 + J_t)^{(b_t^*-1)\eta+1-\gamma} - 1] \\
= a_t^* \frac{b_t^*\eta}{b_t^*\eta + 1 - \gamma} - 1.
\end{aligned} \tag{72}
$$

Therefore $\pi_t H_t$ is a martingale if:

$$
0 = \mu_\pi + \mu_H + \sigma\sigma_\pi + \lambda_t\Big(a_t^*\frac{b_t^*\eta}{b_t^*\eta + 1 - \gamma} - 1\Big). \tag{73}
$$

Substituting μ_π, μ_H and $\sigma\sigma_\pi = -\gamma\sigma^2$ gives:

$$
\begin{aligned}
0 = \mu + CDR_t - r_t - \lambda_t\Big(a_t^*\frac{b_t^*\eta}{b_t^*\eta - \gamma} - 1\Big) - \gamma\sigma^2 \\
+ \lambda_t\Big(a_t^*\frac{b_t^*\eta}{b_t^*\eta + 1 - \gamma} - 1\Big).
\end{aligned} \tag{74}
$$

Note that this implies that: $CDR_t = r_t + rp_t - (\mu + a_t^*\lambda_t\frac{-1}{b_t^*\eta+1})$. Lastly, we can substitute r_t and rp_t, which yields:

$$
CDR_t = \beta + (1/\epsilon - 1)\Big(\mu - \frac{\gamma}{2}\sigma^2 + a_t^*\lambda_t\frac{-1}{b_t^*\eta + 1 - \gamma}\Big). \tag{75}
$$

G Calibration of Climate model

Table 3: Parameters for the Climate model

Par.	Description	Value
E_t	Initial level of total emissions (in GtC, 2015)	10.45
g_t^E	Initial growth rate of emissions (2015)	0.017
g_∞^E	Long-run growth rate of emissions	-0.02
δ_{g^E}	Speed of convergence of growth rate of emissions	0.0075
M_t	Initial carbon concentration compared to pre-industrial (in GtC, 2015)	263
M_{pre}	Pre-industrial atmospheric carbon concentration (in GtC)	588
$M_{0,t}$	Initial carbon concentration box 0 (in GtC, 2015)	139
$M_{1,t}$	Initial carbon concentration box 1 (in GtC, 2015)	90
$M_{2,t}$	Initial carbon concentration box 2 (in GtC, 2015)	29
$M_{3,t}$	Initial carbon concentration box 3 (in GtC, 2015)	4
$\delta_{M,0}$	Decay rate of carbon box 0	0
$\delta_{M,1}$	Decay rate of carbon box 1	0.0025
$\delta_{M,2}$	Decay rate of carbon box 2	0.027
$\delta_{M,3}$	Decay rate of carbon box 3	0.23
ν_0	Fraction of emissions carbon box 0	0.217
ν_1	Fraction of emissions carbon box 1	0.224
ν_2	Fraction of emissions carbon box 2	0.282
ν_3	Fraction of emissions carbon box 3	0.276
F_0^E	Initial level of exogenous forcing (in W/m^2, 2015)	0.5
F_∞^E	Long-run level of exogenous forcing (in W/m^2)	1
δ_F	Speed of convergence exogenous forcing	0.02
T_0	Initial surface temperature compared to pre-industrial (in $°C$, 2015)	0.85
T_0^{oc}	Initial ocean temperature compared to pre-industrial (in $°C$, 2015)	0.0068
κ	Speed of temperature transfer between upper and deep ocean	0.73
υ	Equilibrium temperature response to radiative forcing	1.13
α	Equilibrium temperature impact of CO_2 doubling (in $°C$)	3.05
τ	Heat capacity of the surface	7.34
τ_{oc}	Heat capacity of the oceans	105.5