Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR WIRTSCHAFTS- UND
SOZIALWISSENSCHAFTEN

# Cost Risk Analysis: How Robust Is It in View of Weitzman's Dismal Theorem and Underdetermined Risk Functions?

**Hermann Held**

**WiSo-HH Working Paper Series**
Working Paper No. 55
March 2020

WiSo-HH Working Paper Series
Working Paper No. 55
March 2020

# Cost Risk Analysis: How Robust Is It in View of Weitzman's Dismal Theorem and Underdetermined Risk Functions?

Hermann Held, University of Hamburg

ISSN 2196-8128

Font used: „TheSans UHH" / LucasFonts

Die Working Paper Series bieten Forscherinnen und Forschern, die an Projekten in Federführung oder mit der Beteiligung der Fakultät für Wirtschafts- und Sozialwissenschaften der Universität Hamburg tätig sind, die Möglichkeit zur digitalen Publikation ihrer Forschungsergebnisse. Die Reihe erscheint in unregelmäßiger Reihenfolge.

Jede Nummer erscheint in digitaler Version unter
https://www.wiso.uni-hamburg.de/de/forschung/working-paper-series/

**Kontakt:**
WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

**Cost Risk Analysis: How Robust Is It in View of Weitzman's Dismal Theorem and Underdetermined Risk Functions?**

Hermann Held

Universität Hamburg
Research Unit Sustainability and Global Change
Center for Earth System Research and Sustainability
Grindelberg 5
20144 Hamburg
Germany

Correspondence to: Hermann Held (hermann.held@uni-hamburg.de)

Competing interests: There are not competing interests present.

Acknowledgements. We would like to thank Christian Traeger for helpful discussions, including drawing our attention to Kelly and Tan (2015).

Kelly, D. L. and Tan, Z. (2015) 'Learning and climate feedbacks: Optimal climate insurance and fat tails', *Journal of Environmental Economics and Management,* 72, pp. 98-122.

# Cost Risk Analysis:

# How Robust Is It in View of Weitzman's Dismal Theorem and

# Underdetermined Risk Functions?


## Abstract

Cost risk analysis (CRA) is currently emerging as a noticed decision-analytic framework within the field of climate economics. It combines the expected utility-based structure of cost benefit analysis (CBA) with the target-based approach of cost effectiveness analysis (CEA). As such, it offers a promising candidate for those decision-makers who would like to express their precautionary attitude in view of deeply uncertain global warming impacts through a temperature target, yet who would like to avoid the dynamic inconsistencies of CEA. Here we ask the question whether CRA is potentially subject to the 'dismal theorem' after Weitzman (2009b) like CBA is. We find that in fact, structurally similar issues may arise which, however, can be ameliorated through options specific to CRA. As CRA is not a well-established concept yet, for the convenience of the reader we start by reviewing both its rationale and key results derived from it. This will provide the conceptual basis of justification for CRA-specific solutions of dismal theorem-type effects. The two alternative solutions we offer are as follows. (i) Capping marginal risk at some maximum temperature value beyond which a decision maker caring about the precautionary principle might not show any interest; (ii) sticking to the most conservative risk function of kink-linear type, still in-line with the axiomatic of CRA, thereby avoiding singularities in expected marginal risk.

**Keywords:** dismal theorem; climate targets; fat tails; cost risk analysis; decision under uncertainty; mitigation, precautionary principle.

**Abbreviations:** CBA, CEA, CRA, TSM.

## 1 Introduction

Formulating a rational response to the problem of global warming represents a key challenge for climate economics. Here we argue that the two most prominent schools of thought within climate economics, which we cluster around cost benefit analysis (CBA) and target-based cost effectiveness analysis (CEA), both have their pros and cons in terms of adequately representing society's preferences. We argue that, given the present stock of consolidated knowledge on global warming impacts, a hybrid decision-analytic tool that we call 'cost risk analysis' (CRA) might represent an attractive intermediate alternative that would offer most of the 'pros' of the two methods mentioned above.

In this article we ask whether 'fat tails' in the probability density distribution of climate sensitivity could potentially lead to a 'dismal theorem-type' (Weitzman, 2009b) effect for CRA like it does for CBA. Weitzman (2009b) proposed that fat tails (i.e. slower than exponentially decaying probability density distributions of climate sensitivity) in combination with certain global warming damage and utility functions could lead to infinite expected economic impact in response to greenhouse gas emissions. This would formally ultimately lead to a recommendation of maximum possible mitigation. While the range of applicability of the underlying assumptions was questioned in the aftermath (Horowitz and Lange, 2014, Millner, 2013), we take the position that Weitzman (2009b) in fact pointed to a robust conceptual problem of CBA, given the present situation of deeply uncertain global warming impacts.

As CRA has just been presented as a decision-analytic tool to step-in for CBA as long as this deep uncertainty prevails (Held, 2019), we find it an urgent task to ask whether CRA was sufficiently robust in view of the dismal theorem. As CRA is still not well known to most readers, in the following we review the conceptual and then later the technical basis of CRA, before we tackle its relation to the dismal theorem. Readers who feel sufficiently informed about CRA may want to directly proceed in reading Section 4.

From an environmental economics perspective, cost benefit analysis (CBA) of the climate problem (Nordhaus, 2008, Nordhaus, 2013) is the obvious method of choice for advising society accordingly. When it comes to ‘decision-making under uncertainty’ (in keeping with the latest IPCC report and denoting any lack of information as an ‘uncertainty’), CBA is axiomatically by far the most professional economic tool. It is based on expected utility maximization, for which well-established axioms are available (von Neumann and Morgenstern, 2007, Savage, 1954). However, formally representing the ‘benefit’ of climate policy poses a serious challenge. The complexity of coupled natural and social systems might leave the question whether science had found a good approximation of the totality of global warming impacts open for decades. Furthermore, even if they were known at least approximately, the evaluation of non-market effects is still in its infancy (for an overview of the reliability of CBA see e.g. Stern (2013), Kunreuther et al. (2014)). In view of these structural uncertainties, it is not surprising that virtually the entire spectrum of optimal maximum global mean temperatures has already been derived using CBA. The range from e.g. Nordhaus (2013), who recommends paths close to a no-policy case for the next few decades, to Weitzman (2009a), who essentially suggests maximum mitigation, might be considered illustrative.

This diverse range of recommendations might have left society confused if they had decided to rely on CBA alone as a guide for policy-making. While ongoing research might close the remaining knowledge gaps on global warming impacts and thereby help to establish CBA as the universally accepted tool at the social planner level, this might take too long for present-day decision-making needs. One might argue that this problem is tackled in displaying sensitivity studies regarding the main uncertain inputs of cost benefit analysis (see e.g. Goes et al. (2011) for a successful implementation). Then it might be a matter of personal judgment whether one thinks that decision-makers could really attach to the sometimes rather abstract uncertain inputs cost benefits analysis requires.

In the meanwhile, another main tool has gained in popularity: CEA (cost effectiveness analysis). Over the past few decades the 2-degree target has emerged as a focal point for policy-making (Jaeger and Jaeger, 2011, UNFCCC, 2011). At the heart of that construct lies the assumption that in a context characterized by poor knowledge of global warming impacts, we should not violate the regime of natural fluctuations of global mean temperature that governed the development of humankind (Schellnhuber, 2015). This essentially refers to a Quaternary time scale of at least hundreds of thousands of years, on the basis of which one would prescribe 1.5 degrees as an upper limit for global warming. Adding a bit of faith in the ability of modern civilization to adapt, one arrives at the 2-degree target, a clean-cut number in the decimal system (Schellnhuber, 2015), as a quite possible aggregate target for policy-making (Schellnhuber, 2010). As such, the 2-degree target formalizes the idea of precaution in response to fundamental uncertainties regarding global warming impacts. CEA then seeks to identify the (constrained) welfare-optimal path to complying with such a target, while refraining from formalizing the impacts.

Its non-trivial output is a recommendation on where and when to invest in the energy sector and what welfare losses to expect by compliance with the environmental target (while not economically accounting for the benefits). Roughly 1000 CEA-based scenarios were meta-analyzed in the latest IPCC report (IPCC, 2014). As a key result, it claims the 2-degree target could be complied with by sacrificing 0.06% of annual growth (out of an assumed 1.6%-3% of annual growth (IPCC, 2014)). Assuming society would have found these 0.06% points to be a sufficiently low cost, CEA would have enabled decision-making at the social planner level without the need to wait for academic convergence on the global warming impact function (for a generalized argument along these lines see e.g., Kunreuther et al. (2014), Held and Edenhofer (2008), Patt (1999)).

However, though largely unnoticed by the proponents of CEA, the tool is plagued by a conceptual rather than a lack-of-data issue: Due to an (in any practical terms) unlimited probability density distribution of climate sensitivity (i.e. a distribution of infinite support) to the high end, no temperature target can be met with certainty, and hence CEA should be operated under a probabilistic target (Held et al., 2009). During the preparatory phase of the Conference of the Parties in Copenhagen (2009), the rigid 2-degree target was diluted into a probabilistic target (Allison et al., 2009). Then to our impression that interpretation was accepted in Clarke et al. (2014), IPCC (2014). While most scenarios were formally generated from non-probabilistic CEA, it was apparently assumed that they offered good approximations of some sort of probabilistic CEAs (in fact Held et al. (2009) showed that by properly tuning climate sensitivity in a non-probabilistic CEA, a probabilistic CEA could be mimicked up to an accuracy of a decade for investment purposes). Probabilistic CEA is currently the generally accepted paradigm among this community.

However, if we assume that we might significantly learn about the true value of climate sensitivity within a time horizon that is still relevant for energy system investments, we would need to consider decision-making while reflecting anticipated future learning. Since Blau (1974) at the latest, we know that severe conceptual problems may arise from a probabilistic CEA (which was originally called ‘chance constraint programming’ (Charnes and Cooper, 1959)). In fact Schmidt et al. (2011) demonstrated that the expected value of information (on climate sensitivity (CS)) could become negative for the climate problem. One might even argue, as one of our anonymous reviewers pointed out, that one should not ask for the value of information within CEA at all, as that value were an ill-posed concept within CEA – the benefit side is missing from the welfare functional within CEA.

Furthermore, in view of future learning about rather large values of CS, probabilistic CEA might counter-intuitively recommend abrupt mitigation, somewhat resembling the CBA’s (Weitzman, 2009a). Lastly and even more seriously, learning scenarios on CS are conceivable such that, due to the stock-polluting nature of the climate problem, the probabilistic target might be found infeasible. These conceptual flaws represent the price that was paid for having deviated from the unconstrained expected utility maximization framework when probabilistic CEA was chosen instead of CBA.

Hence, after 20 years of climate economics research, there is still no universally accepted decision-analytic framework that can provide present-day policy advice. In response to this dilemma, Schmidt et al. (2011) proposed a hybrid of CBA and CEA that would do away with the respective shortcomings of both frameworks and combine their pros. They called this hybrid framework 'cost risk analysis' (CRA). In this approach, in order to obtain a welfare measure, the 'risk' of violating a climate target is subtracted from the economic welfare.

This article is organized as follows. Section 2 reviews the current state of research on CRA, both in terms of its foundations as well as findings on its applications. Hereby a complete list of assumptions is detailed, important for the solutions explicated in Section 4. Section 3 identifies a class of applications for which CEA is a good approximation of CRA. This very recent result (Held, 2019) matters as most of the approximately 1000 scenarios as presented in IPCC AR5 WGIII (Clarke et al., 2014) are built on CEA rather than CRA. Hence those 1000 scenarios can be attributed an ex-post interpretation robust under uncertainty and anticipated future learning. Furthermore, Section 3 provides a technical prerequisite for Section 4. Section 4 links CRA to the dismal theorem, identifies a conceptual problem and two possible solutions. Section 5 reviews and rounds out the discussion. Sections 1-3 are to a large extent taken from Held (2019). Section 4 presents unpublished results. Sections 3-4 have never been submitted to an EAERE or a WCERE before.

## 2 Review of cost risk analysis

CRA allows for answering the same questions as CEA plus some additional ones. With CRA, we return to an expected-utility-based framework, though we replace the (difficult to determine) impact function with the 'risk' of violating the 2-degree limit. Here, the construct of a temperature target is preserved. The to-be-optimized welfare functional *W* reads (drawing on Schmidt et al. (2011), Jagannathan (1985) for the static case, and Neubersch et al. (2014) for the dynamic case):

$$\boldsymbol{W}([\boldsymbol{x}(\boldsymbol{t})]) \coloneqq \int_0^\infty \int_s \boldsymbol{p}(\boldsymbol{s})\big(\boldsymbol{U}(\boldsymbol{t},[\boldsymbol{x}(\boldsymbol{t})]) - \beta \boldsymbol{R}(\boldsymbol{t},\boldsymbol{s},[\boldsymbol{x}(\boldsymbol{t})])\big)\, \mathbf{e}^{-\rho t}\, \mathbf{d}\boldsymbol{s}\, \mathbf{d}\boldsymbol{t} \tag{1}$$

Hereby *x* denotes the control variable, [*x*(*t*)] the control path, which can be a scalar or generally multi-dimensional; for the applications of CRA published to date, it represents the collection of investment paths into the energy system. Furthermore, *s* denotes the uncertain 'state of the world,' *p*(.) its probability density measure, *U*(.) consumption-based utility, *t* time, $\beta$ the risk weighting factor, *R*(.) the 'risk' function, and $\rho$ the pure rate of time preference. In general, utility may also depend on the state of the world, which comprises both uncertain properties of the economic or climate subsystem, respectively. For convenience of notation, in the following we focus on climate-based uncertainty so as to illustrate how to formalize decision-making under climate targets.

While for CEA, the climate target represents a constraint on the optimization, for CRA the target is absorbed in *R*(.). Accordingly, for CRA, the maximization of *W* is unconstrained, solving the abovementioned infeasibility problem for a probabilistic target-based CEA. In addition, leaving *p*(.) outside of the brackets ensures the classification of CRA as an expected utility-based framework. Therefore the expected value of future information can no longer be negative (in contrast to a probabilistic target-based CEA; for the formal proof see e.g. Gollier (2004)). Finally, also leaving the exponential discounting outside of the brackets ensures time consistency (Strotz, 1955-56). As such, CRA represents a dynamically consistent decision-analytic framework.

In summary, CRA represents a hybrid out of CBA and CEA in the following sense. From CBA we take the un-boundedness of welfare optimization. This un-boundedness is possible because, like CBA, CRA is a non-lexicographic preference order. Climate risks are traded off against mitigation cost. From CEA we take the concept of an external climate target which liberates us from the need of specifying a damage function.

For a global mean temperature target $T^*$, which shall serve as the upper bounds, we base the risk function on global mean temperature $T(t, s, x)$. To stay as close as possible to the tradition of CEA, no adverse effect on welfare is foreseen for $T<T^*$, hence $R = H(T-T^*)\, r(T)$, with *H*(.) denoting the Heaviside function vanishing for $T<T^*$ and assuming 1 elsewhere.

The risk function is *not* merely a re-named impact function, but rather provides a mathematical instrument for representing preferences in view of informal and fragmented impact knowledge, in conjunction with a precautionary attitude. The user needs to specify both this risk function and a trade-off factor against economic utility, the latter being dependent on the transformation of the energy system.

One might ask why welfare must be a *linear* functional of utility and risk. We found this to be intuitive, as society makes an informal trade-off of costs and its aversion to violating $T^*$. We suggest that for such an informal, intuitive process of judgment, $W$ should be as simple a functional as possible. This supports a linear approach. In fact, this also supports the economist's expectation that a richer society would accept a larger consumption loss for complying with an environmental target, as $U(.)$ is concave.

### 2.1 Specification of the risk weight factor $\beta$

In principle the risk weight factor $\beta$ needs to be specified by eliciting a decision-maker's or stakeholder's willingness to pay for target compliance. However, for the global society the following calibration might be of special interest, as suggested in Neubersch et al. (2014). They propose calibrating $\beta$ such that CRA would optimally mimic the value system encoded in COP2010's proclamation of the 2-degree target (UNFCCC, 2011). They interpreted 'likely' compliance in the IPCC sense of 'calibrated language': as a probability of compliance $P^*$ of at least 66%. This also nicely resonates with the preparatory material for the COP2009 (Allison et al., 2009), which focused on that very compliance level.

Furthermore, Neubersch et al. (2014) proposed that the COP2010's proclamation of the 2-degree target did not have a sophisticated adjustment of anticipated future learning about climate sensitivity in mind, but simply forged ahead with the knowledge structure available at the time. Accordingly the trade-off parameter is set such that in a scenario without anticipated future learning about climate system properties, exactly 66% compliance is achieved. We will stick to this calibration process in the following.

Note that for the limiting case of vanishing climate response uncertainty, i.e. arbitrarily sharp density distributions on climate variables, from this calibration it would follow that $T$=$T$*, regaining the prescription of non-probabilistic CEA.

After this qualitative description we formalize the above. As the calibration as described so far is based on the idea to mimic the target of a probabilistic CEA as much as possible we express the latter first. The CEA-based control problem reads

$$\max_{[x(t)]} W_{\mathrm{CEA}} := \int_0^\infty U(t,[x(t)])\, \mathrm{e}^{-\rho t}\, \mathrm{d}t \qquad \text{subject to} \tag{2}$$

the probabilistic constraint $P_\gamma\left(T(t_{\max}([E(t)];\ \gamma),[E(t)];\ \gamma) \leq T^*\right) = P^*$ .

Hereby $t_{\max}$ (.) denotes an all-time maximum of temperature, [$E$($t$)] the emission path, $\gamma$ the uncertain climate sensitivity, $P_\gamma$ its probability measure, and $P$* the compliance level for $T$* (for a discussion on alternative ways to formulate a probabilistic CEA see Held et al. (2009)). When now calibrating $\beta$, $\beta$ is chosen such that the solution of Eqn. (1) would comply with the probabilistic constraint of Eqn. (2).

(CRA is also defined for the limiting case of certainty. It can be seen as a limit obtained by letting the variance of the distribution going to zero, in analogy of the definition of a delta-distribution.)

**2.2 Specification of the risk function**

How can we specify $r$($T$)? Neubersch et al. (2014) confined the set of admissible functions by means of the 'Axiom of Sacrifice Inhibition'. Here we will review their rather condensed and not particularly formal line of argument (presented in one paragraph only) in a more explicit manner and add a missing (albeit elementary) proof. They refer to a static version of Equation 1, which we have modified into:

$$W(E) = U(E) - \beta R(E) \tag{3}$$

Hereby $E$ denotes the cumulative amount of carbon dioxide emissions over the climate-policy-relevant timeframe. In fact as a key message of IPCC (2013), $E$ is an approximatively good predictor of $T$. Furthermore Lorenz et al. (2012) have analogously shown that $E$ is a good predictor for mitigation costs in the case of the MIND model (Edenhofer et al. (2005), a Ramsey-type economic growth model including an energy sector with induced technological change, as well as a climate module). As such, $E$ might offer a good control variable for studying the link between mitigation costs and climate risks in a static model.

We now digress for a moment, in analogy to Rawls' 'veil of ignorance' (Rawls, 1971), in order to assess some of the necessary structures of $r(T)$. Imagine a world in which the utility losses from mitigation, i.e. $U(E)$, were unknown, as well as $R(E)$. The only things we do know are the signs of the marginals: the more we emit, the smaller the utility losses from transforming the energy system become, as long as $E \leq E_0$, and the larger the climate risk grows.

$$\forall_{E \leq E_0}\ U'(E) \geq 0, \qquad \forall_E\ R'(E) \geq 0. \tag{4}$$

Further, imagine a given concave function $U(E)$. The concavity of that function is highly plausible, as utility must be maximal in the no-policy case $E_0$ (Held, 2015, Roshan et al., 2019), and mitigation costs diverge at a sufficiently low $E$ (see e.g. Luderer et al. (2013)). Suppose $W$ were maximal for $E_1$ ('first best optimum'). Now imagine that, for some policy-related reason, $E_1$ is no longer accessible – e.g. due to delayed participation. As a result, some surplus amount $\Delta E$ has been emitted. In view of the stock-polluting nature of the climate problem and negative emissions not being an option for the time being, then only values $E \geq E_1 + \Delta E$ are accessible for optimization. Now the Axiom of Sacrifice Inhibition dictates that $R(E)$ be such that $W$ is maximal at $E_2 := E_1 + \Delta E$ ('second best maximum') and it is not necessary to assume a value $E_3$ strictly larger than $E_2$ to achieve the second best maximum. In geometric terms this is equivalent to stating that $W(E)$ should not have multiple maxima that are separated by a concave-convex nature of the welfare function.

In plain English: If the first best maximum cannot be achieved as the accessible control space is reduced, the recommended control should then be as close as possible to the original one. We claim that this property codified in the Axiom of Sacrifice Inhibition is highly plausible as a key property for those who advocate temperature targets. If the target is missed, it is not a reason to add further emissions; rather, one should stay as close to the original target as possible.

Hence we conclude that $W(E)$ must not inherit a concave-convex-concave sequence that would induce a 'tipping point-like' behavior of optimal emissions $E_3$ as a function of initial emissions $E_2$.

To make this argument, the following two assumptions are crucial:

- The utility function is concave and hence does not inherit any structure that would cause distinct multiple maxima of the welfare function. This assumption implies that no excessive mitigation cost savings could justify a tipping in optimal emissions.
- In our context, $T^*$ solely represents an expression of the precautionary principle, informed by a historically covered temperature regime, and does *not* refer to any specific knowledge about a tipping point in the climate system (Lenton et al., 2008). Society might want to avoid the latter, but such aspects became part of pro mitigation reasoning much later. However if $T^*$ *did* represent an objective natural threshold, one might argue that after a certain subsystem had been destroyed by violating its tipping point, there would be no case for further mitigation. Quite the contrary: because $T^*$ does not represent an objective threshold, but an expression of precaution, one should always remain as close as possible to $T^*$ once its violation cannot be avoided, as called for by the Axiom of Sacrifice Inhibition.

For those readers who have followed us up to here, we will now go one step further and remind them that the 'veil of ignorance' on the exact functional form of $U(E)$ is still active. Hence, the above restriction on $R(E)$ must hold for *any* concave $U(E)$. This leads us to the following proposition.

**Proposition 1:** *For a static model, let welfare be a linear combination of utility and climate risk. If we stipulate that the welfare functional must comply with the Axiom of Sacrifice Inhibition for any concave utility function, then R(E) must be convex.*

Proof: see Held (2019).

From the above it follows that any risk function that is convex could be a candidate for representing a stakeholder's trade-off between mitigation costs and avoided climate risk. However, can we motivate the choice of a specific risk function out of the set of convex ones? CEA is based on a line of argument establishing $T<T^*$ as a necessary condition for establishing the precautionary principle. Accordingly, $T<T^*$ represents a 'least demanding condition.' To stay consistent with this line of argument we now seek to find the 'least alarmist' risk convex function. To judge what is 'alarmist,' we focus on the high end of emissions and argue that we should consider the least convex risk function. This is the – above $E^*$ leading to $T^*$ – linear one, which we call a 'linear risk function.'

(We would like to note, however, that more convex risk functions are also admissible and could express more risk-prone preferences. While by calibration they would lead to the same $P^*$, in general we have to be prepared for the fact that any quantity diagnosed from CRA would be a function of that convexity. In the following, we focus on the consequences of the 'borderline case' of a linear risk function.)

To now get from the above static model to the dynamic Eqn. (1) we need the following further 'brave' assumptions.

1. Temperature is approximately a linear function of prior cumulative emissions (Allen et al., 2009).
2. The conditions derived for the static model generalize to a dynamic one.
3. In particular a risk function based on $T$-$T^*$
    a. avoids multiple equilibria and
    b. is the least convex risk function based on $T$ which does so.

While assumption number 1 can is backed by the literature, for assumption number 2 here we can only appeal to the intuition that risk is dominated by the temperature maximum which in turn imposes a quasi-static structure. The approximate dynamic model discussed in Section 3 delivers further support of those assumptions. Here we note that above assumptions numerically hold for the MIND model (Neubersch et al., 2014), yet await a completed analytic treatment for the time being. Finally we note that the choice of a kink-linear risk function $R(t)= (T-T^*)\,\mathrm{H}(T-T^*)$ as suggested in item 3 delivers the normatively satisfying property that its marginal is discontinuous at $T^*$. Therefore temperature paths tend to cluster around $T^*$ in the case of learning (Neubersch et al., 2014).

Once a risk function has been chosen, the trade-off parameter needs to be determined, as described in the previous subsection.

### 2.3 Summary of a procedure on how to use CRA

A user of CRA would proceed along the following steps.

1. Choose a function $r(t)$ to complete the risk function. So far by default, a $r(t)$ linear in $T-T^*$ is utilized.
2. For any value of the trade-off parameter $\beta$, Eqn. (1) generates a solution. Choose $\beta$ such that the probabilistic constraint of Eqn. (2) is satisfied.
3. From this solution diagnose the desired control path. It represents an 'elementary solution' in the sense that it represents a case without delay and without future learning. Note that any property beyond $P^*$ is to be diagnosed from the model and is not an input to the model.
4. Generalize Eqn. (1) to exploit CRA on those aspects that CEA might not be able to deliver such as anticipated future learning or delayed climate policy. Hereby keep the value of $\beta$ as determined in step number 2.

### 2.4 Review of CRA-based applications

In a first application of this calibrated CRA, Neubersch et al. (2014) found that the investment paths for the energy sector before learning closely resemble those derived from a probabilistic CEA for the next few decades. That is good news for those who would like to base their decisions on the 1000 CEA-based scenarios gathered in IPCC (2014). They then derived the expected value of perfect climate information, a quantity that by its very nature cannot be derived from CEA. They found that, if perfect climate response information were provided today, on average 1/3 of the mitigation costs could be avoided, and the welfare loss from the climate problem could be reduced by 1/4. In view of their findings, climate research might currently be under-invested in!

CRA also solves the infeasibility problem that climate policy might encounter in connection with delayed participation. Due to the softened target, a policy recommendation is always obtained. While the 2-degree target might have to be sacrificed due to delayed participation, even then, the underlying value system could still be preserved. Roth et al. (2015) found that mitigation costs decrease with delay, while they increase in CEA. In turn, the compliance level $P^*$ is found to be reduced by delay. Climate risks could increase 3 to 4-fold in response to delays until 2050. On a conceptual level Roth et al. (2015) highlight that for the case of delay, CRA needs a third item – in addition to $T^*$ and $P^*$ – to be specified: the point in time $t^*$ to which the calibration of $\beta$ refers. After a delayed mitigation policy still claiming compliance with $P^*$ implies larger mitigation costs than for doing so at the original point in time. Quite the contrary, above reduction in mitigation cost is induced by $P^*$ always referring to the original base year, no matter how much delay would occur. Climate policy may want to become more sensitive to this issue.

Stankoweit et al. (2015) generalized the precautionary principle codified in the 2-degree target to risks emanating from the side-effects of solar radiation management, Roshan et al. (2016) generalized CRA for this, and Roshan et al. (2019) generated a regionalized version. While solar radiation management (SRM) would crowd out mitigation completely due to the lower costs, the picture changes once SRM-induced risks are accounted for. A single risk category (precipitation pattern mismatch) reduces the mitigation cost savings from roughly 100% to about ½ , with all regions weighted equally. Using alternative weightings, or considering further risk categories, might imply a further reduction of the role of SRM in an optimal portfolio.

**3 Cost effectiveness analysis as an emulator for cost risk analysis?**

Above we reported as one of the key findings from cost risks analysis (CRA) the following phenomenon. For the MIND model, in a setting without anticipated future learning and without delay, CRA would produce solutions rather similar to probabilistic cost effectiveness analysis. Here we invent an analytic model that captures key characteristics of the decision problem and that would explain the observed phenomenon.

For this we note that scenarios in-line with the 2°C-target show a time horizon separation of mitigation action and time-maximum of temperature. While mitigation-induced relative consumption losses peak during the first half of the 21st century (Held et al., 2009, Luderer et al., 2012), temperature does so in the second half or later (see e.g. Fig. 6.13 in Clarke et al. (2014)). This suggests that the main decisions are taken in the first half of this century and we can approximately restrict our control space to that time horizon. We call this the 'time horizon separation climate model' (TSM). For convenience, in the context of this article, we lump further assumptions into TSM. The complete list of assumptions shall be the following:

(i) For welfare optimization, variation of control paths only within a first time horizon is relevant.
(ii) For climate risk, only temperature of a second, disjoint time horizon is relevant.
(iii) Temperature only depends on prior time-cumulative emissions.

The justifications for items (ii) and (iii) read as follows. For 2-degree-compatible mitigation scenarios the dominant contribution to the total emission budget stems from the first half of the century (see e.g. Fig 6.8 in IPCC-Ch6). Furthermore the carbon budget defined as the prior time-cumulative emissions is found being a good predictor for temperature (IPCC, 2013).

**Proposition 2:** *Within TSM any solution derived from calibrated cost risk analysis without anticipated future learning is independent of the choice of the risk function R.*

Proof: We note that for CEAs of the climate problem, the emission path [$E(t)$] represents the only variable at the interface between the economy module generally hosting the (generally multi-dimensional) control path [$x(t)$] and the climate module containing the environmental restriction. Hence we can slim down the control problem by disentangling it into a Stackelberg configuration of 'leader' variables [$E(t)$] and 'follower variables' [$x(t)$]:

$$\max_{[x(t)]} W = \max_{[E(t)]} \left( \max_{[x(t)]} W \text{ subject to } [E(t)] \right) \quad (5)$$

This suggests to define a pre-aggregated utility $V$ by

$$V([E(t)]) := \max_{[x(t)]} < U > \text{ subject to } [E(t)] \quad (6)$$

whereby we replace the double integral in Eqn. (1) by the shorthand '<.>'. Utilizing $V$, the control problem then reads in a more compact form

$$\max_{[x(t)]} W = \max_{[E(t)]} (V - \beta < R >) \quad (7)$$

as $R$(.) depends on [$x(t)$] only through [$E(t)$].

Along those lines we assume there is a (for convenience of notation finite) set of independent choices of emissions along the time axes within the 1st half of this century $E_1$, … $E_n$, …, $E_N$ with $n$ indexing the time axes that would determine temperature in the second half of this century $T=T(t, E_1, \dots E_n, \dots, E_N)$ (hereby influences of later emissions are suppressed as we assume there are no significant control changes in the second half of the century). The second half of this century is crucial for diagnosing the climate risk because here temperature might transgress $T^*$.

With these further assumptions the control problem simplifies to

$$\max_{[x(t)]} W = \max_{[E_1,\dots,E_n,\dots,E_N]} (V - \beta < R >) \tag{8}$$

We inspect the first order condition

$$\operatorname{grad} W = 0 \quad \text{with} \quad \operatorname{grad} := \left(\frac{\partial}{\partial E_1}, \dots, \frac{\partial}{\partial E_n}, \dots, \frac{\partial}{\partial E_N}\right) \tag{9}$$

Then Eqn. (9) reads grad $V = \beta$ grad $<R>$. Hence $E_1$, … $E_n$, …, $E_N$ enter $R$ only through $T = T(t, E)$, the emission budget $E$ being defined as $E := E_1 + \dots E_n + \dots + E_N$, hence $R = R(t, E)$. As $\partial E/\partial E_n = 1$,

$$\operatorname{grad} V = \beta \frac{\partial < R >}{\partial E} (1, \dots, 1) \tag{10}$$

This implies that for any $E_1$, … $E_n$, …, $E_N$, grad $V$ is parallel to (1, …, 1). This parallelism induces a 1-dimensional solution manifold within the $N$-dimensional solution space of emission paths. The carbon budget $E$ suggests itself as the variable to parameterize this remaining degree of freedom.

For given $\beta$, $E$ is then fixed by the scalar pre-factor left in Eqn. (10). In the calibration process $\beta$ is chosen such that the $E$ resulting from it is compatible with the temperature target compliance probability $P^*$. The latter formally reads (as being replicated from Eqn. (2))

$$P_\gamma (T(t_{\max}(E;\ \gamma), E;\ \gamma) \leq T^*) = P^* \tag{11}$$

In summary, the fact that temperature approximately depends only on the carbon budget and not on the timing of emissions – a key message of IPCC AR5 – leads to a solution $E_1$, … $E_n$, …, $E_N$ that is solely determined by ($U$(.), $E$), however not by the functional form of $R(t, T)$. (End of proof.)

From the above line of argument we readily can link CRA and probabilistic CEA.

**Proposition 3:** *Within TSM and without anticipated future learning, CRA and probabilistic CEA lead to identical policy recommendations.*

Proof: By definition probabilistic CEA seeks maximizing <*U*> – equivalent to maximizing V – under a constraint identical to Eqn. (11) (and hence Eqn. (2)). While for CRA, that equation is used to fix the calibration, in CEA it is operational to define the solution:

$$\mathrm{grad}\, V \ = \lambda \, \frac{\mathrm{d}P_\gamma}{\mathrm{d}E}\, (1, \dots, 1) \tag{12}$$

Hereby $\lambda$ serves as Lagrange multiplier to enforce the boundary condition of Eqn. (11). One readily identifies that both equations (10) and (12) induce the identical vector field structure on the emission path (i.e. the gradient must parallel (1, …, 1)). As for both also the identical emission budget constraint hold (i.e. given by Eqn. (11)), the CEA- and the CRA-based solutions must be identical. (End of proof.)

We have shown that for TSM, the solution does not depend on the choice of the risk function. How does this relate to the finding of the previous section that certain risk functions must be excluded to satisfy the Axiom of Sacrifice Inhibition? Above reasoning is derived from a local (gradient-based) analysis in control space. Hence the above statements hold only for cases that would not display multiple optima, i.e. within the allowed subset of risk functions.

## 4 Cost risk analysis in view of fat tails

Any decision-analytic framework based on (unconstrained) expected utility maximization is in principle prone to Martin Weitzman's 'dismal theorem' (Weitzman, 2009b). In a cost-benefit framework, a 'fat tail' in the distribution of the very system property, which would describe climate response uncertainty, could lead to infinite expected losses and subsequently infinite marginal expected losses. How does this effect play out for cost risk analysis which might include statistical moments of such a fat tailed distribution as well?

For the sake of analytic accessibility, we draw on the static cost risk model as of Section 2.2

$$\boldsymbol{W}(\boldsymbol{E}) = \boldsymbol{U}(\boldsymbol{E}) - \beta \boldsymbol{R}(\boldsymbol{E}) \quad . \tag{13}$$

For this static model, we link temperature and cumulative emissions *E* in a linear manner (Allen et al., 2009) through (potentially fat-tailed) climate sensitivity $\gamma$

$$\boldsymbol{T}(\boldsymbol{E}) = \gamma \boldsymbol{E} \tag{14}$$

hereby counting cumulative emissions in suitable, dimensionless units. The static counterpart of Eq. (1) reads, assuming no further uncertainties rather than for climate sensitivity

$$\boldsymbol{W}(\boldsymbol{E}) = \int_{\mathbf{0}}^{\infty} \boldsymbol{p}(\gamma)\,(\boldsymbol{U}(\boldsymbol{E}) - \beta \boldsymbol{R}(\gamma \boldsymbol{E}))\,\mathbf{d}\gamma \ . \tag{15}$$

Noting that the economic contribution *U* does not depend on climate sensitivity and assuming the standard linear risk function *r*(*T*) as motivated in Section 2.2, we obtain

$$\boldsymbol{W}(\boldsymbol{E}) = \boldsymbol{U}(\boldsymbol{E}) - \beta \int_{\mathbf{0}}^{\infty} \boldsymbol{p}(\gamma)\,\mathrm{H}(\gamma \boldsymbol{E} - \boldsymbol{T}^{*})\,(\gamma \boldsymbol{E} - \boldsymbol{T}^{*})\,\mathbf{d}\gamma \ . \tag{16}$$

In exploiting the Heaviside function we can simplify this expression to

$$\boldsymbol{W}(\boldsymbol{E}) = \boldsymbol{U}(\boldsymbol{E}) - \beta \int_{\gamma^{*}(\boldsymbol{E})}^{\infty} \boldsymbol{p}(\gamma)\,(\gamma \boldsymbol{E} - \boldsymbol{T}^{*})\,\mathbf{d}\gamma \quad \text{with} \quad \gamma^{*}(\boldsymbol{E}) \coloneqq \frac{\boldsymbol{T}^{*}}{\boldsymbol{E}} \quad . \tag{17}$$

The solution of cost risk analysis is the optimum of the thereby defined welfare. Hence we search for the solution of *W'*(*E*)=0 and accordingly for

$$\boldsymbol{U}'(\boldsymbol{E}) = \beta \int_{\gamma^{*}(\boldsymbol{E})}^{\infty} \boldsymbol{p}(\gamma)\,\gamma\,\mathbf{d}\gamma - \beta \boldsymbol{p}(\gamma^{*})(\gamma^{*}\boldsymbol{E} - \boldsymbol{T}^{*}) \quad . \tag{18}$$

As by definition, the last bracket vanishes, we obtain

$$\boldsymbol{U}'(\boldsymbol{E}) = \beta \int_{\gamma^{*}(\boldsymbol{E})}^{\infty} \boldsymbol{p}(\gamma)\,\gamma\,\mathbf{d}\gamma \tag{19}$$

as the defining equation for the solution of our cost risk analysis (the last two equations are taken from Roshan et al. (to be resubmitted)). Whether we face a singularity problem like for Weitzman's dismal theorem is hence equivalent to a potential divergence of the first moment of climate sensitivity's probability distribution *p*.

Applications of cost risk analysis (Neubersch et al., 2014, Roshan et al., 2019, Roth et al., 2015) have so far relied on a log-normal distribution (Wigley and Raper, 2001) for climate sensitivity, for which the first moment is finite (Johnson et al., 1994). However, Roe and Baker (2007) traced back the distribution of climate sensitivity to a Gaussian in the overall climate feedback, leading to

$$p_{\mathrm{Roe\&Baker}}(\gamma) = \frac{1}{\sigma\sqrt{2\pi}}\,\frac{\gamma_0}{\gamma^2}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{1-f_{\mathrm{m}}-\frac{\gamma_0}{\gamma}}{\sigma}\right)^2} \qquad (20)$$

The 'fat-tailedness' of this distribution was the starting point for Weitzman's discussion. The distribution asymptotically scales like $1/\gamma^2$, hence the first moment grows like $\log(\Gamma)$ if $\Gamma$ denotes the upper bound of the integral determining the marginal expected risk. Therefore we conclude that the distribution according to Roe and Baker would lead to infinite expected marginal risk. Can cost risk analysis deal with this phenomenon?

Technically, it can. We propose generalizing above defining equation into

$$U'(E) = \beta(\Gamma)\int_{\gamma^*(E)}^{\Gamma} p(\gamma)\,\gamma\,\mathrm{d}\gamma \quad . \qquad (21)$$

Hereby $\beta(\Gamma)$ is defined through the request that the calibration procedure (see Section 2.3), hence above equation, should guarantee that $E_{2/3}$ is found as the optimal solution. The latter denotes the emission level leading to a probability of target compliance of 2/3. Therefore the defining equation for $\beta(\Gamma)$ is

$$U'\left(E_{2/3}\right) = \beta(\Gamma)\int_{\gamma^*(E_{2/3})}^{\Gamma} p(\gamma)\,\gamma\,\mathrm{d}\gamma \quad . \qquad (22)$$

Apparently, $\lim_{\Gamma\to\infty} \beta(\Gamma) = 0$. This can be handled by requesting that the CRA problem should be solved for for any $\Gamma$, and finally $\Gamma\to\infty$. Due to Proposition 2 (see Section 3) we know that the solution that is generated from the calibration process, approximately does not depend on the risk functional. Hence also for the time-dependent problem we expect that the control path would lead in the course of $\Gamma\to\infty$ to a finite solution. The findings on CRA as demonstrated on Section 3 should therefore be robust against a choice of a fat tailed distribution as that by Roe and Baker (2007).

How about applications of CRA in which we transfer a once calibrated $\beta(\Gamma)$ to another problem, such as delayed participation or climate engineering (recall Section 2.4)? We leave this for another investigation. Instead, here we note a very problematic aspect in the solution as just proposed in terms of $\lim_{\Gamma\to\infty} \beta(\Gamma)$.

We invite the reader on the following thought experiment. CRA should also hold with fixed, once calibrated trade-off parameter $\beta$ for consecutive Bayesian learning, the very reason why it was invented as an alternative to probabilistic CEA. Bayesian learning may in fact reduce the spread of the density distribution of climate sensitivity over time (Kelly and Tan, 2015). Suppose there were a type of observation which allowed to convert a fat tailed density function into a thin tailed one. Schneider von Deimling et al. (2006b) proposed such a mechanism in Bayesian learning from the glacial/interglacial transition. In essence, the thin-tailedness becomes possible because climate sensitivity is inferred from two alternative equilibrium situations, rather than a process-based approach (Roe and Baker, 2007) or the global warming signal (to mention as prototypical examples Forest et al. (2002) and Knutti et al. (2002)). While the proposal as of Schneider von Deimling et al. (2006b) remained controversial, within the context of this article it is sufficient to imagine that such tail-thinning might happen in the future.

Then the expected marginal risk would not diverge any longer, however, as $\beta$ should be preserved after calibration, and in particular throughout Bayesian learning, still $\lim_{\Gamma\to\infty} \beta(\Gamma) = 0$. Hence, in response to the tail-thinning, the decision maker would abandon mitigation. While formally there is nothing wrong with this finding, it raises the question whether CRA as introduced above is already the most adequate model for a climate target-oriented decision maker. Does Weitzman's dismal theorem also hamper CRA as a decision-analytic tool?

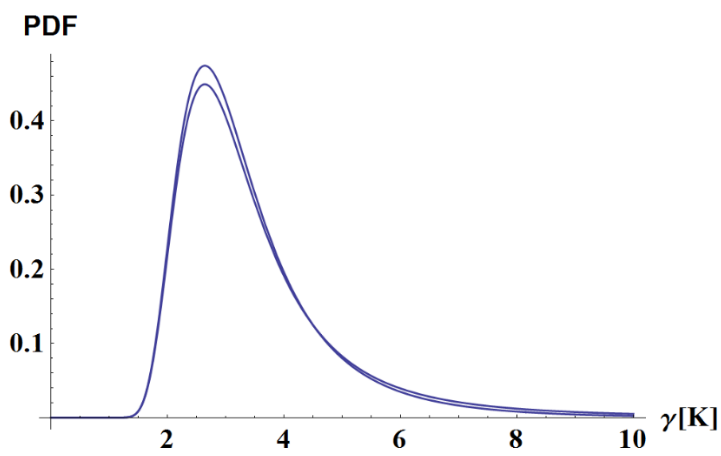


Figure 1: Probability density distribution of climate sensitivity after Roe and Baker (2007) prior and posterior updating with an imagined Gaussian likelihood. The taller distribution denotes the posterior. Updating almost leaves the distribution intact, however transforms the fat tail ($\propto 1/\gamma^2$) into a slim one.

### 4.1 Cost risk analysis: Defining a finite temperature horizon of precaution

Above paradox is triggered by the finding that fat tails could also matter for CRA, at least inversely quadratic tails do in fact matter. The reason is that CRA, as it is introduced above, cares about long tails. Here we propose to further exploit the fact that the risk function is not based on a damage function but rather expresses the decision maker's aversion against transgressing $T^*$. Would a decision maker care if we increased the temperature from Venus' temperature to Venus's temperature plus one unit? Clearly, marginal risk should vanish at Venus' temperature, the latter not allowing for life as we know it. Our suggestion somewhat resembles what had been discussed by Costello et al. (2010) for CBA. However the Axiom of Sacrifice Inhibition (see Section 2.2.) requests a convex risk function and hence positive marginal risk at arbitrarily large temperature values.

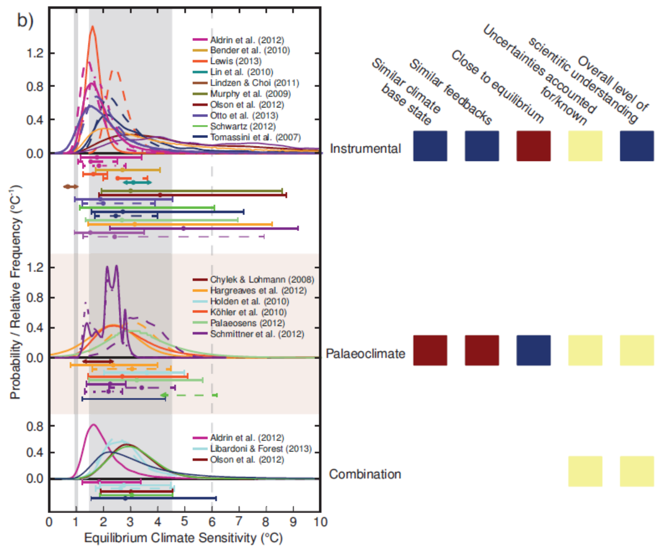


Figure 2: Probability density distributions of climate sensitivity according to IPCC AR5 WG-I (Bindoff et al. (2013), Fig. 10.20). The grey rectangle denotes the overall 'likely' range, denoting 66% of probability.

Here we put up for discussion whether the domain of validity for the Axiom of Sacrifice Inhibition should be restricted to a horizon compatible with the horizon that led to the operationalization of the precautionary principle in terms of the temperature target $T^*$.

We recall that the 2-degree target was defined as a rounded up version of the very regime experienced during the development of humankind. We can now contrast this in defining 'the un-experienced'. This scale could be set by the regime beyond 5°C of cooling which characterized the last ice age (Schneider von Deimling et al., 2006a). While this amplitude was survived by humankind to the cool end, this was definitely not experienced to the hot end ('hothouse') (Zachos et al., 2001). We hence propose that $R(T)$ shall stabilize for $T>T^{**}>T^*$, a regime 'beyond imagination' for a decision maker with a precautionary attitude. We would hence modify the Axiom of Sacrifice Inhibition for a temperature regime below $T^{**}$.

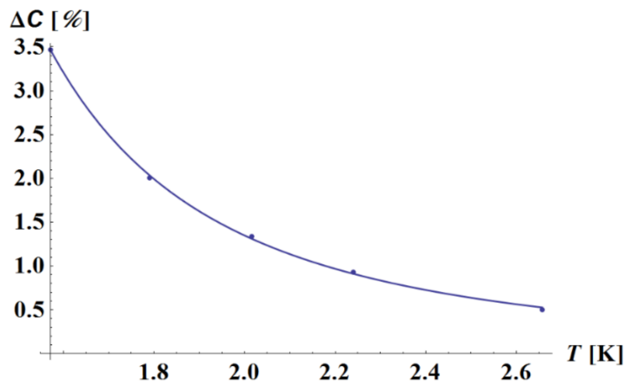


Figure 3: Relative consumption loss vs. temperature target $T^*$. The more restrictive the target, the larger the consumption loss. Dots: Diamonds from Luderer et al. (2013), curve: power-law fit for our analysis.

We numerically illustrate the consequences of this suggestion for $T^*$=2°C, $T^{**}$=5°C, $\sigma$= 0.13, $f_{\mathrm{m}} = 0.62$, $\gamma_0$ = 1.2°C (the latter three coefficients guaranteeing one of the representative distributions according to Roe and Baker (2007), see **Figure 1**; this is within the ensemble of climate sensitivity distributions by the IPCC (Bindoff et al., 2013), see Figure **2**). The goal is to show that – after the invention of $T^{**}$ – a mere change from a 'fat' to a 'slim', but otherwise unchanged, distribution does not change our CRA solution significantly.

For $U(E)$ we choose a %GWP loss curve as a function of the temperature target from Luderer et al. (2013) (based on the complex energy-economy-climate model REMIND (Kriegler et al., 2017)) that happens to be compatible with our $P^*$=2/3. We approximate %-loss in consumption as %-loss in production and identify for our static model $U$=-$\Delta C(E)/C_0$ (note that consumption smoothing via CRRA is already being taken care of in the time-aggregated losses as reported by Luderer et al. (2013)). We find we can fit what we sampled from their diamonds by a power function (see Figure 3 and Eq. (23)) to a degree sufficient for our illustrative static model.

$$U(E) = -\kappa(T - T_0)^{-\alpha} \tag{23}$$

for which we find the last-square fit parameters $\kappa$=2.06, $T_0$=0.78 (in °C), $T$ to be measured in °C, $\alpha$=2.17. Note that this approximation claims validity only for the regime sampled as displayed in Figure 3.

Now we define some Bayesian updating with a Gaussian likelihood function which would almost not change the distribution by Roe and Baker (2007) except for thinning the tale. For this, first we normalize the distribution as given in Eq. (20) such that it integrates to one (some probability weight in the order of a per mill is missing; for details see Section 4.2). Then we update it by a Gaussian which peaks at the peak of the original distribution and which would deliver a posterior with the identical density at a climate sensitivity value of 4.5°C, the high end of IPCC AR5 'likely' range for climate sensitivity (see Figure **2**). We find we need to utilize the Gaussian N(2.64, 5.64) (in terms of mean and standard deviation).

We determine the prior 2/3-quantile of climate sensitivity as 3.70°C. From this we determine the emission level $E_{2/3}$, ensuring that the 2/3-quantile of temperature would match $T^*$. Then we fix (i.e. calibrate) $\beta$. When we search for the new optimum for CRA after updating, the emission level increases due to reduced expected risk pressure. However the 2/3-quantile of temperature decreases from 2°C to 1.95°C because of an overcompensation through a reduced 2/3-quantile of climate sensitivity. A change of 0.05°C is below the standard deviation of interannual variability in global mean temperature and hence might appear tolerable. Hence we have shown that a truncation of marginal risk above $T^{**}$=5°C leads to a robustification of CRA against fat tails. Again, the findings of Section 3 remain valid as no assumptions about the risk function were made in it.

## 4.2 Accepting Weitzman's modification of Roe and Baker's distribution

While Roe and Baker's finding in terms of asymptotically power-law-type distributions might be seen as robust, the quadratic asymptotics might be seen as an artifact. To show this, we recall their key assumption

$$\gamma = \frac{\gamma_0}{1-f} \quad , \qquad (24)$$

$f$ denoting the feedback parameter which is 0 for vanishing feedback and 1 for a climate system at the borderline of a runaway greenhouse effect. We follow Weitzman (2009a) in that values $f$

> 1 should be excluded as we assume a (prior anthropogenic intervention) stable climate. Hence a Gaussian distribution for *f* (as assumed by Roe and Baker) cannot be exact as it comes with infinite support. Rather, the density of *f* must vanish for *f*=0,1. Accordingly Weitzman requested a linear behavior of the density of *f* around *f*=1 (Weitzman, 2009a).

For our illustrative purpose we interpret the Gaussian for *f* as a likelihood function which in the course of Bayesian updating must be multiplied with a prior density which vanishes at the singular points *f*=0,1. We chose 6 *f* (1-*f*) as symmetric prior, of lowest-possible order if interpreted as a Taylor expansion. In-line with Weitzman's results, now the distribution in terms of $\gamma$ displays inverse cubic asymptotics. Hereby the transformation law $p_f(f)\mathrm{d}f = p_\gamma(\gamma)\mathrm{d}\gamma$ was utilized. Accordingly, the first moment of expected risk exists and no invention of *T*** is necessary.

We repeat the updating procedure as of Section 4.1., while we keep the linear risk metric up to infinity. The updating leads to a negligible change in the 2/3-quantile of temperature 2°C $\rightarrow$ 1.996°C. It seems that once we avoid the somewhat unrealistic inverse quadratic asymptotics, and allow for at least an order faster decay of density, CRA reacts robustly with respect to fat tails, even without truncating marginal risk at some *T***.

**4.3 Pros and cons of the two approaches of dealing with the dismal theorem in CRA**

Which of those two solutions to the dismal theorem is to be seen as the preferred one? Cutting marginal risk at some high-end temperature value creates additional need for normative input in terms of *T***. This represents a 'con' for the solution as of Subsection 4.1.

This drawback is absent for the solution as of Subsection 4.2. However, the latter crucially depends on a linear risk function. More convex risk functions cannot be excluded for CRA. The linear risk function is usually implemented to be in-line with the structurally conservative approach of CEA ('What is the minimum mitigation action required?') as representing the least alarmistic representation of climate risk. If CRA is to be applied for stakeholders with a less conservative attitude, a risk function with infinite support might not be applicable.

In that sense, capping marginal risk at some $T^{**}$ has to be seen as the superior solution.

As next steps we need to augment above results by an analytically supported comprehensive sensitivity study.

## 5 Conclusion

Cost risk analysis (CRA) allows for preserving the concept of temperature targets while avoiding the inconsistencies plaguing lexicographic preference orders (Schmidt et al., 2011). Temperature targets rather than cost benefit analysis (CBA) might be preferred if a decision maker

- found the global warming impact function too uncertain for formal analysis, or
- was willing to accept a formal analysis based on a global warming impact function, however would neither trust a probabilistic or Knightian (Millner et al., 2013) measure and even found a subsequent sensitivity study too abstract, or
- simply accepted the prescribed political target (UNFCCC, 2011) as boundary condition for an economic analysis.

For this reason cost risk analysis (CRA) was invented to preserve as much of the original target idea as possible, yet to avoid former inconsistencies. Here we present a completed list of further assumptions that have been necessary to make CRA operational. Furthermore, we invented a semi-static climate model, based on a time-separation idea of causes and effects of a mitigation policy ('TSM'). By means of TSM we could prove that for the case of instantaneous mitigation, the further choice of the risk function is irrelevant for determining the CRA-based solution in the absence of learning. Moreover, the solution is identical to the original solution based on strict probabilistic climate targets. This is of highest importance to the climate community as thereby the 2-degree-adjusted scenarios out of about a thousand as reported in IPCC (2014) which were based on probabilistic targets, plagued by inconsistencies, can retroactively be given an approximate interpretation as dynamically consistent CRA-based solutions.

We then utilized this conceptual and formal background to provide solutions to the fall-out of Weitzman's dismal theorem on CRA. We offer to either cap marginal risk at some high-end temperature value or sticking to linear as the most conservative risk functions. Capping marginal risk is possible as CRA does not describe real impacts but rather a decision maker's attitude towards violating a temperature target.

We conclude that cost risk analysis can be seen as a valid decision analytic tool for those decision makers who do not trust in global warming impact functions for now. In the meanwhile, research on the mechanistic understanding of global warming impacts and adaptation options, and their evaluation will move on and at some point allow for replacing CRA by a then well-informed CBA.

**References:**

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M. and Meinshausen, N. (2009) 'Warming caused by cumulative carbon emissions towards the trillionth tonne', *Nature,* 458(7242), pp. 1163–6.

Allison, I., Bindoff, N. L., Bindschadler, R. A., Cox, P. M., Noblet, N., England, M. H., Francis, J. E., Gruber, N., Am, H., Karoly, D. J. and others (2009) 'The Copenhagen Diagnosis', 60, *http://www.copenhagendiagnosis.org*.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y. and Jain, S. (2013) 'Detection and attribution of climate change: from global to regional', in Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (ed.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Blau, R. A. (1974) 'Stochastic programming and decision analysis', *Management Science,* 21(3), pp. 271-276.

Charnes, A. and Cooper, W. W. (1959) 'Chance-constrained programming', *Management Science,* 6(1), pp. 73-79.

Clarke, L., K. Jiang, K. Akimoto, M. Babiker and G. Blanford, K. F.-V., J.-C. Hourcade, V. Krey, E. Kriegler, A. Löschel, D. McCollum, S. Paltsev, S. Rose, P. R. Shukla, M. Tavoni, B. C. C. van der Zwaan, and D.P. van Vuuren (2014) 'Assessing Transformation Pathways', *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Costello, C. J., Neubert, M. G., Polasky, S. A. and Solow, A. R. (2010) 'Bounded uncertainty and climate change economics', *Proceedings of the National Academy of Sciences,* 107(18), pp. 8108-8110.

Edenhofer, O., Bauer, N. and Kriegler, E. (2005) 'The impact of technological change on climate protection and welfare: Insights from the model MIND', *Ecolog Econ,* 54(2-3), pp. 277–292.

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. and Webster, M. D. (2002) 'Quantifying uncertainties in climate system properties with the use of recent climate observations', *Science,* 295(5552), pp. 113-117.

Goes, M., Tuana, N. and K, K. (2011) 'The economics (or lack thereof) of aerosol geoengineering', *Climatic Change,* 109, pp. 719-744.

Gollier, C. (2004) *The economics of risk and time.* The MIT Press.

Held, H. (2015) 'Lecture on cost risk analysis at University of Hamburg'.

Held, H. (2019) 'Cost risk analysis: dynamically consistent decision-making under climate targets', *Environmental and Resource Economics,* 72(1), pp. 247-261.

Held, H. and Edenhofer, O. (2008) 'Restructuring the problem of Global Warming Mitigation: Climate Protection vs. Economic Growth as a false Trade off', in Hirsch Hadornn, G., Hoffmann-Riem, H., Biber-Klemm, S., Grossenbacher-Mansuv, W., Joye, D., Pohl, C. & al., e. (eds.) *Handbook of Transdisciplinary Research*. Heidelberg: Springer.

Held, H., Kriegler, E., Lessmann, K. and Edenhofer, O. (2009) 'Efficient Climate Policies under Technology and Climate Uncertainty', *Energy Economics* 31, pp. S50–S61

Horowitz, J. and Lange, A. (2014) 'Cost–benefit analysis under uncertainty—A note on Weitzman's dismal theorem', *Energy Economics,* 42, pp. 201-203.

IPCC (2013) 'Summary for Policymakers.', in Press, C.U. (ed.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA.

IPCC (2014) 'Summary for Policymakers', *Climate Change 2014, Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jaeger, C. C. and Jaeger, J. (2011) 'Three views of two degrees', *Regional Environmental Change,* 11(1), pp. 15-26.

Jagannathan, R. (1985) 'Use of sample information in stochastic recourse and chance-onstrained programming models', *Manage Sci,* 31(1), pp. 96–108.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994) 'Lognormal distributions', *Continuous univariate distributions,* 1, pp. 207-227.

Kelly, D. L. and Tan, Z. (2015) 'Learning and climate feedbacks: Optimal climate insurance and fat tails', *Journal of Environmental Economics and Management,* 72, pp. 98-122.

Knutti, R., Stocker, T. F., Joos, F. and Plattner, G.-K. (2002) 'Constraints on radiative forcing and future climate change from observations and climate model ensembles', *Nature,* 416(6882), pp. 719.

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., Baumstark, L., Bodirsky, B. L., Hilaire, J. and Klein, D. (2017) 'Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century', *Global environmental change,* 42, pp. 297-315.

Kunreuther, H., Gupta, S., Bosetti, V., Cooke, R., Dutt, V., Ha-Duong, M., Held, H., Llanes-Regueiro, J., Patt, A., Shittu, E. and others (2014) 'Integrated risk and uncertainty assessment of climate change response policies', *IPCC AR5 WG3 Ch2*.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S. and Schellnhuber, H. J. (2008) 'Tipping elements in the Earth's climate system', *Proceedings of the National Academy of Sciences,* 105(6), pp. 1786-1793.

Lorenz, A., Kriegler, E., Held, H. and Schmidt, M. G. W. (2012) 'How to measure the importance of climate risk for determining optimal global abatement policies? ', *Climate Change Economics,* 3(1), pp. 1250004

Luderer, G., Bosetti, V., Jakob, M., Leimbach, M., Steckel, J. C., Waisman, H. and Edenhofer, O. (2012) 'The economics of decarbonizing the energy system — results and insights from the RECIPE model intercomparison', *Climatic Change,* 114, pp. 9–37.

Luderer, G., R. C. , Pietzcker, Bertram, C., Kriegler, E., Meinshausen, M. and Edenhofer, O. (2013) 'Economic mitigation challenges: how further delay closes the door for achieving climate targets', *Environmental Research Letters,* 8, pp. 0-8.

Millner, A. (2013) 'On welfare frameworks and catastrophic climate risks', *Journal of Environmental Economics and Management,* 65(2), pp. 310-325.

Millner, A., Dietz, S. and Heal, G. (2013) 'Scientific Ambiguity and Climate Policy', *Environmental and Resource Economics,* 55(1), pp. 21-46.

Neubersch, D., Held, H. and Otto, A. (2014) 'Operationalizing climate targets under learning', *Climatic Change,* 126(3-4), pp. 305-318.

Nordhaus, W. D. (2008) *A question of balance.* New Haven: Yale University Press.

Nordhaus, W. D. (2013) *The climate casino.* Yale University Press.

Patt, A. (1999) 'Separating Analysis from Politics: Acid Rain in Europe', *Rev. Policy Res. ,* 16, pp. 104-137.

Rawls, J. (1971) *A Theory of Justice.* Cambridge, M.A.: Oxford University Press.

Roe, G. H. and Baker, M. B. (2007) 'Why is climate sensitivity so unpredictable?', *Science,* 318(5850), pp. 629-632.

Roshan, E., Mohammadi Khabbazan, M. and Held, H. (2016) 'Cost-risk trade-off of solar radiation management and mitigation under probabilistic information on climate sensitivity.', *EAERE*.

Roshan, E., Mohammadi Khabbazan, M. and Held, H. (2019) 'Cost-Risk Trade-Off of Mitigation and Solar Geoengineering: Considering Regional Disparities Under Probabilistic Climate Sensitivity', *Environ Resource Econ,* 72(1), pp. 263-279.

Roshan, E., Mohammadi Khabbazan, M. and Held, H. (to be resubmitted) 'A Scheme for Jointly Trading-off Costs and Risks of Solar Radiation Management and Mitigation under Long-tailed Climate Sensitivity Probability Density Distributions'.

Roth, R., Neubersch, D. and Held, H. (2015) 'Evaluating Delayed Climate Policy by Cost-Risk Analysis', *EAERE*, *http://www.webmeets.com/eaere/2015/prog/viewpaper.asp?pid=609*, Available: EAERE.

Savage, L. J. (1954) *The foundations of statistics.* New York; London: Wiley & Sons Chapman & Hall.

Schellnhuber, H. J. (2010) 'Tragic triumph', *Climatic change,* 100(1), pp. 229-238.

Schellnhuber, H. J. (2015) *Selbstverbrennung.* C. Bertelsmann Verlag.

Schmidt, M. G. W., Lorenz, A., Held, H. and Kriegler, E. (2011) 'Climate targets under uncertainty', *Climatic change,* 104(3), pp. 783-791.

Schneider von Deimling, T., Ganopolski, A., Held, H. and Rahmstorf, S. (2006a) 'How cold was the last glacial maximum?', *Geophysical Research Letters,* 33(14).

Schneider von Deimling, T., Held, H., Ganopolski, A. and Rahmstorf, S. (2006b) 'Climate sensitivity estimated from ensemble simulations of glacial climate', *Climate Dynamics,* 27(2-3), pp. 149-163.

Stankoweit, M., Schmidt, H., Roshan, E., Pieper, P. and Held, H. (2015) 'Integrated mitigation and solar radiation management scenarios under combined climate guardrails', *Geophysical Research Abstracts*. EGU General Assembly 2015: EGU, pp. EGU2015-7152.

Stern, N. (2013) 'The structure of economic modeling of the potential impacts of climate change', *Journal of Economic Literature,* 51(3), pp. 838-859.

Strotz, R. H. (1955-56) 'Myopia and inconsistency in dynamic utility maximization.', *The Review of Economic Studies,* 23(3), pp. 165-180.

UNFCCC 2011. *RE: Report of the Conference of the Parties on its seventeenth session, held in Durban from 28 November to 11 December 2011. Addendum. Part two: Action taken by the Conference* *of the Parties at its seventeenth session. Decision 1/CP.17: Establishment of an Ad Hoc Working Group on the Durban Platform for Enhanced Action FCCC/CP/2011/9/Add 1 .*

von Neumann, J. and Morgenstern, O. (2007) *Theory of games and economic behavior.* Princeton University Press.

Weitzman, M. (2009a) 'Additive damages, fat-tailed climate dynamics, and uncertain discounting', *Economics,* 3, pp. 2009-39.

Weitzman, M. L. (2009b) 'On modeling and interpreting the economics of catastrophic climate change', *The Review of Economics and Statistics,* 91(1), pp. 1-19.

Wigley, T. M. and Raper, S. C. (2001) 'Interpretation of high projections for global-mean warming', *Science,* 293(5529), pp. 451-454.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. (2001) 'Trends, rhythms, and aberrations in global climate 65 Ma to present', *science,* 292(5517), pp. 686-693.