Climate Policy with Incomplete Knowledge

Elizabeth Baldwin*

preliminary draft May 2010; this draft May 2014

Abstract

We provide a framework for modelling climate change under uncertainty, that is both more general than the seminal paper of Weitzman (2009a) and, hopefully, more amenable to policy analysis. We first justify certain assumptions on the importance of 'multi-priors' and 'multi-utilities' in evaluating damages from climate change, which give rise to a structure of incomplete preferences. Climate impacts then naturally separate into three sorts: those that are understood well; a final 'catastrophe' at the other extreme; and poorly understood damages, falling short of this final level, in between. By understanding these distinctions, we can better frame the question of how to set the level of climate policy. And the latter two terms have opposing effects on the question of whether, in the long term, price or quantity considerations should guide legislation.

JEL: D04, D81, Q54

Keywords: Climate Change, Catastrophes, Knightian Uncertainty, Incomplete Preferences, Policy Choice

1 Introduction

When plausible outcomes are well beyond our range of experience, and may have profound and lasting effects on society as a whole, standard models for dealing with uncertainty may not suffice.

This paper is written in the context of climate change. Extreme uncertainties in the chain of causality from emissions to welfare impacts mean that very high impact events

*Grantham Research Institute, London School of Economics and Smith School for Enterprise and the Environment, Oxford University. E-mail: e.c.baldwin@lse.ac.uk. Acknowledgements: I am grateful for advice and support from Paul Klemperer, Robert Hahn and Nathaniel Keohane. I also thank Benjamin Bronselaer, David Budescu, Duncan Coutts, David Frame, Alexander Golub, Miles Gould, Britt Groosman, Michael Hanemann, Ian Jewitt, Edi Karni, Randall Kroszner, Oleg Lugovoy, Margaret Meyer, Antony Millner, Peter Neary, Alexander Otto, Ben Parker, Jack Pezzey, Kevin Roberts, Sang Seok Lee, Neil Shephard, Gernot Wagner and James Wang for all their comments and help. I completed this research while funded by the Simon Linnett scholarship at Balliol College, Oxford; I am also very grateful for the funding and inspiration I received at the Environmental Defense Fund.

have a sufficiently high probability that they cannot be disregarded in the analysis. The potentially dramatic impacts of climate change on water availability, ecosystems, food, coasts and health are laid out by, for example, the IPCC (2007, Figure SPM.2).

This paper provides explicit assumptions giving rise to a model of *incomplete preferences and beliefs*, suitable to such circumstances, and discusses the policy relevance of this model, and in particular the implications of the model structure for the question of whether long-term policy targets should be price- or quantity-based.

The assumptions which underlie the model are firstly that a good approximation of damages from climate change cannot be made by only considering those impacts at the low level we understand well. And secondly, we assume that, beyond some threshold level, our estimations of damages from climate change are not objective, but are instead influenced heavily by the *priors* we bring to the analysis, and by how we believe outcomes should be *valued*.[1]

These assumptions are justified by a study of the literature on climate change impacts. Concerns over the extent of uncertainty in these impacts were first raised by Weitzman (2007, 2009a) and similar points have been made recently by other writers.[2] The fundamental points are that we do not yet know how likely various 'bad outcomes' might be – even if we were to know, for example, what temperature changes we should expect – and that we cannot robustly ascribe values to future scenarios very different from our collective experience. (Even if we hypothetically could make such valuations, we must recognise that economists have not yet attempted directly to do so, instead calibrating their 'damage functions' at relatively low levels of warming.) In this context, it is misleading to give a single figure for the social cost of carbon. Any such estimate is heavily dependent on the presumptions we bring to our analysis.

We address these concerns by using a model of incomplete preferences and beliefs (see Bewley, 1986, Galaabaatar and Karni, 2013). No single valuation function and probability density function can capture our knowledge and preferences; we have instead a *set* of such functions, indexed by our priors and tastes. One policy is only robustly preferred to another if it is preferred under all possible such priors and tastes. Frequently this will leave two possible policies incomparable in these terms. Instead of imposing an arbitrary comparability, we should be aware of this issue and re-examine in this light the plausibility of our priors, and what further we can understand about our tastes.

However, we do not assume that preferences are equally 'incomplete' for any possible scenario. On the contrary, we assume that one *can* make objective judgements on risks, up to some threshold level: the more familiar scenarios are more complete.[3] Thus we break down damages into three ranges.

First there are the 'conventional' damages, for which we can sensibly make predictions and valuations: here, the 'priors' we take have little effect. Next, there are 'indeterminate damages' corresponding to climatic changes whose impacts and probabilities we cannot claim to understand well, but which are unlikely to be 'so bad that they cannot get any worse'. These are our 'indeterminate' damages.

On the other hand, there is a limit to loss that climate change might cause. We will

[1]Throughout, 'prior' is used as in the recent literature on decision-making under uncertainty, to refer to subjectively assumed probability distributions; see Gilboa and Schmeidler (1989), Klibanoff et al. (2005), Galaabaatar and Karni (2013), etc.

[2]See e.g. Millner et al. (2013), Pindyck (2013a), Stern (2013), Heal and Millner (2014).

[3]See Karni (2014).

refer to this limiting damage level as ‘endgame’ damages. Computing this limit requires both understanding how bad things can objectively get, and evaluating the subjective level of damages one should assign to such a scenario. Of course, there is scope for much debate – both scientific and philosophical – on the the right way to understand these issues.

Once these assumptions and interpretations are established, a ‘dismal theorem’ is that we cannot make good estimates of damages from climate change without understanding the latter two terms – and that the latter two terms depend on the priors that we take for the problem. Moreover, detailed analyses of the ‘conventional’ impacts we can expect and understand become less and less relevant as the quantity of GHGs increases, raising with it the probability of bad outcomes.

We thus remove Weitzman’s ‘infinity’ from the analysis, but in fact (as we show) this theorem is a generalisation of his. It is also more widely applicable, to more situations of ‘deep’ or ‘Knightian’ uncertainty, and in particular it does not require assumptions about functional forms of damages and uncertainty at the very high end, but weaker assumptions on the importance of priors in estimating the welfare consequences of situations very different from current experience.

It is not *ex ante* clear which of our three terms is dominant, and indeed the answer will depend on our priors and tastes, on the final atmospheric concentration of greenhouse gases, as well as on the explicit limit on catastrophe built into the model. So our model shows explicitly that a response to this uncertainty is *not* necessarily to view the problem as one of insurance against ‘rare disaster’ (as argued by, e.g., Pindyck, 2013a,b). Events that are less bad, but also less unlikely, may be the relevant factor – for example, the possibility of serious welfare damage from 4°C of warming (see also Newbold and Marten, 2014).

Moreover, the relative sizes of these terms has a real policy implication, as is shown by applying the model to the ‘prices versus quantities’ work of Weitzman (1974) to this situation of incomplete preferences. We ask whether, in setting the long-term strategy of our response to climate change, we should let the ‘carbon price’ or the quantity of greenhouse gases accrued in the the atmosphere guide our decision-making: what should determine the targets? We focus on the single price versus quantity comparison as it provides clear counterpoints, but it also is relevant for our analysis which focuses on the large scale and long term: a body such as the UNFCCC seems unlikely to set a nuanced final target in the spirit of Roberts and Spence (1976), or to change their policy frequently.

Whatever their relative sizes, large ‘non-conventional’ damage terms have the effect of increasing total damages. However, on the question of ‘prices versus quantities’, the two terms have opposite effects. A quantity-based mechanism is preferred when ‘indeterminate damages’ are the most significant; in such cases, the risk of additional pollutants is sufficiently unpalatable that one is prepared to countenance their mitigation being more expensive than expected.

However, additional emissions make no difference once the ‘endgame’ is reached. So when this term is significant in the analysis, and when an increase in quantities makes a significant difference to this threshold being passed, there is a relatively smaller downside to overshooting our emissions target. Hence we do not necessarily need the certainty that the quantity tool provides; we may be better off setting a carbon price and avoiding damaging the economy if emission abatement is more expensive than expected. *In extremis*, it is better to choose the price instrument even if we ignore the latter avoided risk: the possible upside of a price instrument giving rise to greater-than-expected abatement might be more significant than the downside of intended targets being missed.

Note that the effect described occurs only when increased quantities do indeed alter the probability of passing the catastrophic threshold. If we are entering the territory of significantly different climate systems, to which our geophysical models are simply not calibrated, we may have to admit that we do not really know what will happen, and that possible outcomes are spread thinly over a wide range. In such circumstances, the effect of the 'endgame' damage term ceases to dominate, and 'indeterminate' damages become relatively more significant, driving us back towards a quantity-based approach.

Moreover, it does not seem palatable to choose a price-based policy because "the risk from 'endgame' is significant". If this term is very significant in the estimation of expected damages, a natural response is to ask if the quantity of emissions allowed should have been reduced. And a lower quantity of emissions, in reducing the dominance of 'endgame' damages in the analysis, shifts favour back in the direction of the quantity mechanism.

This work originated in a desire to apply Weitzman's (2009a) 'dismal theorem' to the question of 'prices versus quantities' (Baldwin, 2010). The model has developed and moved further from Weitzman's original formulation (see also Baldwin, 2013); many other authors have simultaneously addressed the same issues.

The multiple scientific and socio-economic uncertainties in the problem of climate change are discussed in depth and brought up to date by Heal and Millner (2014). A thorough discussion of climate sensitivity (the long-term response of the environment to a doubling in greenhouse gas concentrations) is also given by Millner et al. (2013).

Millner (2013) summarises the debate in the literature on whether the formulation and 'infinity' of Weitzman (2009a) is appropriate (see Nordhaus, 2009, Horowitz and Lange, 2009, Weitzman, 2009b, Costello et al., 2010, and others). By abstracting from the question of functional forms and 'infinity', this paper attempts to regain the spirit of Weitzman's original insight, leaving these debates aside. Weitzman's own subsequent work (Weitzman, 2011) also attempts to move on from some of these technicalities. His work on damages (Weitzman, 2009c, 2010) suggests new models, differing from the present work's emphasis on our ignorance of their correct form. His focus remains the tails Weitzman (2012, 2013a,b), whereas the model of this paper emphasises the importance of the badly-understood mid-range.

Pindyck (2013a)'s damning indictment of integrated assessment models, on the grounds that "IAM damage functions are completely made up", is in close agreement with the premise of this paper. However, his prescription, developed in more detail in Pindyck (2013b), that we must focus on scenarios of rare catastrophes to build a "case for a stringent GHG abatement policy" appears to take those damage functions as a "most likely" starting point. By rejecting such functions, certainly beyond the range for which they are calibrated, the model here illuminates that bad outcomes from plausible environmental scenarios, such as 4 degrees of warming, may neither be unlikely nor represent a 'worst-case' catastrophe, and yet may dominate the analysis. Thus, closest to the conclusions of the first parts of this paper, is the (concurrent) essay of Stern (2013).

Previous authors have applied the model of Weitzman (1974) to climate change, mostly in distinguishing between 'cap and trade' and 'carbon tax' style policy approaches: Newell and Pizer (2003), Hoel and Karp (2001, 2002) provide multi-period analyses, arguing that in the short term, a price-based mechanism is more appropriate when the effects of the pollutant are cumulative, as with climate change. Hepburn (2006) gives a detailed summary of the arguments, and Goulder and Schein (2014) provide an update. Parsons and Taschini (2013) consider the importance of understanding whether shocks are temporary or permanent for choosing between short-term price and quantity instruments. Kornek and Marschinski (2013) consider the strategic consequences of combining this model with the question of coalition formation.

Regarding longer-term strategy as we do, the 'price versus quantity' net present value analysis of Keohane (2009) showed preference for quantity-based targets. Dietz and Fankhauser (2010) also argue that deep uncertainty implies we should set quantity-based targets. We believe this is the first extension of the model to incorporate a model of deep, 'Knightian' uncertainty as well as 'indecisiveness' in tastes.

The paper is organised as follows. In Section 2 we provide background on the uncertainties which make climate change such a hard problem. Section 3 introduces a standard model (Section 3.1), as well as a simplified version of Weitzman (2009a) in Section 3.2 and the theories of ambiguity aversion and incomplete preferences (Gilboa and Schmeidler, 1989, Klibanoff et al., 2005, Galaabaatar and Karni, 2013) in Section 3.3. In Section 4 the new model for this paper is laid out: the key assumptions are given and discussed in Section 4.1 and the generalised 'dismal theorem' is presented in Section 4.2, followed by a discussion of the relative significance of the new, 'unconventional' damage terms (Section 4.3) and, in Section 4.4, an illustrative calibration made using the Matlab model of Baldwin (2014). Section 5 provides the application of this model to the question of 'prices versus quantities'. Section 6 concludes.

2 Sources of uncertainty in the damages from CO_2 emissions

There is considerable uncertainty in what might be the future damages from climate change. A wide range of aspects are covered by Heal and Millner (2014); here we summarise some key points.

One reason that climate change is such a difficult problem to address is that the path of causality from emissions to damages is relatively long. There is, roughly speaking, a five-stage chain of processes, with considerable uncertainty in each stage of the chain. The stages (outlined, for example, by Stern, 2010) are as follows:

1. human activities emit greenhouse gases (GHGs);
2. emissions increase atmospheric concentrations of GHGs;
3. atmospheric concentrations of GHGs change the global temperature;[4]
4. changes in the global temperature cause local climatic and environmental change;

[4]Here one may differentiate between short-term effects – the 'transient climate response' (see, for example, Andrews and Allen, 2008) – and 'climate sensitivity', which is the equilibrium climate response to a doubling of GHG concentrations from pre-industrial levels; the latter is much less well constrained than the former, as reaching this equilibrium may be a very slow process.

5. local changes have impacts on human lives.

In each of these stages, there is uncertainty – both in the level of the damages, and the time-scale over which they take place. The majority of economic models (including that of Weitzman, 2009a) focus on the uncertainty in Stage 3; although this interests us, we emphasise that it is not the deepest source of uncertainty.

What one might call the 'dismal' aspect is that very bad outcomes cannot be ruled out. Here, we distinguish two different questions: firstly the probability that an outcome in the 'very bad' range occurs; and secondly how sharply probabilities decline with increasing temperature, and whether the tail of the pdf of climate sensitivity should be considered formally 'fat'.[5] The latter is a question of functional form and limiting behaviour (see Section 3.2), rather than to the total probability accorded to the tail. The arguments of Weitzman (2009a) focus on this question; we consider the former as well in this paper.

Consider our identified uncertainties in these contexts. The total volume of emissions is imperfectly measured: discrepancies in data from China alone may have amounted to 5% of global emissions in 2010 (Guan et al., 2012).

Currently around 60% of GHG emissions are re-absorbed by the oceans and terrestrial ecosystems. This fraction may decrease with rising temperatures (see for example Friedlingstein et al., 2006, Knorr, 2009) and there are concerns about more severe positive feedbacks, for example from melting permafrost (see, for example Shakhova et al., 2010, Schaefer et al., 2012). Such risks are not quantified as yet.

'Climate sensitivity', the parameter focused on by Weitzman (2009a) is the equilibrium temperature change resulting from a doubling of atmospheric concentrations of carbon dioxide.[6] Regarding the probability that climate sensitivity exceeds some 'bad' level, Solomon et al. (2007, Box 10.2, Figure 2) present a range of studies giving a probability of between 15% and 35% that climate sensitivity exceeds 4.5°C, but note that 'there is no well-established formal way of estimating a single PDF from the individual results'.[7] The 'fat tail' of climate sensitivity, and its importance for policy, is the subject of much debate.[8]

One way to reduce the number of uncertainties to consider is to combine stages 2 and 3, and view warming as a function simply of *cumulative emissions*. Strikingly, it appears that "the peak warming caused by a given cumulative carbon dioxide emission is better

[5]The technical meaning of 'fat' is either that not all moments $E_s[s^n]$ exist, or that all the moments exist, but the moment generating function $E_s[e^s]$ does not. An example of the former situation is the log-logistic distribution; climate sensitivity is modelled in this way by, for example, Dietz (2011). An example of the latter situation is the lognormal distribution; climate sensitivity is modelled in this way by, for example, Golub et al. (2009).

[6]More generally, it drives the logarithmic relationship between stable CO_2 concentrations and long-term temperature change.

[7]One of the tightest constraints on high end probability is provided by Annan and Hargreaves (2006), who estimate that $P(s > 4.5) \approx 5\%$ by using Bayesian methods and a combination of recent and paleo-climate data. However, others argue that these methods are not relevant in this context (see Henriksson et al., 2009).

[8]Roe and Baker (2007) argue that $\mathrm{var}_S(s)$ is infinite (and so that the 'fatness' of the tail is almost as great as could be possible). However, not all are convinced of these arguments: see, for example, Hannart et al., 2009 and Urban and Keller, 2009. That the fat tail in climate sensitivity gives rise to 'dismal' climate outcomes is the heart of Weitzman (2009a)'s argument, but it is argued by Otto et al. (2013), Frame et al. (2013), Millner (2013) that this is not necessarily the case.

constrained than the warming response to a stabilisation scenario" (Allen et al. 2009, see also Stocker et al. 2013). Moreover, this peak warming is found to be "remarkably insensitive" to the emission pathway, and so cumulative emissions are an attractive parameter for economic theorists.

Note that climate sensitivity is not the only uncertain parameter governing the relationship between emissions and temperature pathways. The 'simple coupled carbon cycle and climate' model of Allen et al. (2009, see also Baldwin 2014) additionally allows uncertainty in the diffusion into the ocean of both carbon and surface temperature anomalies, as well as the positive feedback of additional emissions from soils, vegetation and the upper ocean. Incorporating such a suite of uncertainties gives wide spread of possible temperature changes over time.

The local and environmental consequences of a given level of global warming are also hard to predict. The damages themselves are likely to be experienced not simply as a result in a change in mean temperature; changes in the frequency and severity of extreme events, and particularly changes in patterns of precipitation, are more pertinent, but harder to predict. Attempts have been made to quantify plausible impacts on biodiversity, and show great uncertainty: Thomas et al. (2004) produce estimates of biodiversity loss ranging from 9% to 52% of species being 'committed to extinction' by 2050. The passing of 'tipping points', at which a large qualitative change takes place (such as the loss of the Indian summer monsoon the Greenland ice sheet) are intrinsically hard to model, but an attempt to rank the uncertainties of such events is made by Lenton et al. (2008).

Finally, we require an estimate of the socio-economic damages associated to a given level of global warming. Tol (2013, Table 1) lists the main studies in this area; our principal observation is that with two exceptions these are concerned with the damages from at most 2.5 degrees of warming. Modellers then 'fit a curve to a point', so that modelled damages at higher temperatures are purely extrapolations – with little justification of their functional form.[9] Damages from 5 degrees of warming are typically modelled in the range of 1 to 7% of global consumption (see Stern, 2006, Figure 6.2). Stern himself (Stern, 2010) describes as 'ludicrously small'; what we wish to emphasise is that such numbers are based on *almost no actual calculations*, and so the uncertainty surrounding them must be vast.[10]

Although some models incorporate uncertainty in the damage function (see for example Hope, 2011), there have been few attempts to formally estimate a probability distribution of damages.[11] Nordhaus' (1994) survey of experts (the only study to di-

[9] A convention has developed in the literature that damages have the from γT^n where n lies (usually) between 1 and 3 and the function is calibrated using γ at a point estimate. But in fact very little is know about the actual shape of the damage function.

Nordhaus (1993) sets γ on the basis of calibration at 2.5 degrees, and n to be 2 because 'there is evidence that the impact increases non-linearly as the temperature increases, and we assume that the relationship is quadratic'. This choice of a quadratic relationship was *ex post* justified because it matches well the median estimate for 6 degrees of warming obtained by Nordhaus (1994, a survey of a small number of experts which does look at higher levels of warming). Other authors follow Nordhaus in taking $n = 2$ or n uncertain but centred around 2.

[10] See also Weitzman (2011, 2012), Pindyck (2013a), Stern (2013).

[11] Hope (2011) calibrates the probability of 'catastrophe' against Lenton et al. (2008), but takes the *level* of such a catastrophe as 25% from Nordhaus (1994), in which it featured as a question, not survey response (see Footnote 12).

rectly consider higher levels of warming) highlights the 'vast disparities' in estimates obtained, but only the median in only one set of responses was used by Nordhaus and Boyer (2003) to calibrate their damage function.[12]

Further criticisms have been made, both of the calibration provided for the damage function at 2.5 degrees[13] and of the functional form typically associated with damages.[14] Climate change may eventually be associated with mass migration and war; such impacts are difficult to attribute precisely or estimate, but this does not mean they should be ignored in economic models.[15]

In conclusion, the absence of good data means that there is vast uncertainty in the damages that may accrue from climate change, especially at higher levels. And this uncertainty is typically neither estimated nor modelled.

3 Models of Damages from Climate Change

A typical model of climate change considers a social welfare function

$$W(\mathbf{X}, \mathbf{q}, \theta, \phi) = \sum_{t=1}^{\infty} \beta^t U(X_t, \mathbf{q}_{s \le t}, \theta, \phi)$$

where $\mathbf{X}$ is a global consumption pathway over time, $\mathbf{q}$ is the pathway of emissions of GHGs over time, with $\mathbf{q}_{s \le t}$ being those components of $\mathbf{q}$ up to time t, while β is the discount factor, U is a utility function assumed to take negative values (such as greater than unity constant relative risk aversion) and θ and ϕ are realisations of random variables respectively governing uncertainty in the cost of emission reductions and damages from climate change.[16]

[12] Three scenarios were considered, one incorporating 6 degrees of warming by the end of the century. 19 experts were asked to give the probability of a 'high-consequence' (25% consumption loss) outcome from each scenario. The median of the estimates is a 5.0% probability; this is the figure that Nordhaus and Boyer (2003) use (in an adjusted form). However, as is made clear in Figure 3 of Nordhaus (1994), the data has a strong right skew; the mean of the probability estimates is 17.5% and the range is 0.3-95%.

[13] See, for example, Freeman and Guzman (2009) and Ackerman et al. (2008, 2009), who note that it fails to account for catastrophic events, non-market costs and cross-sectoral impacts, and that there is great uncertainty in future growth and productivity, which affect estimates of damages which will principally be in the future.

[14] See in particular Pindyck (2013a). Most modellers of climate change work with a per-period damage function d_t, acting multiplicatively on utility; so if consumption is X_t and temperature change is T_t in period t, then

$$U_t(X_t, T_t) = (1 + d_t(T_t))U_t(X_t)$$

where U_t is the utility function for time t. Weitzman (2010, 2009c) disputes this assumption, on the grounds that it implies damages from climate change would be greater to a rich society than to a poor one. Sterner and Persson (2008) take this view, and consider potential impacts of climate change on relative prices. Pindyck (2012) argues that climate change should be modelled via its effect on the growth rate, rather than consumption level.

[15] US Military, in their recent review (United States of America Department of Defense, 2010), state that 'climate change could have significant geopolitical impacts around the world [...] While climate change alone does not cause conflict, it may act as an accelerant of instability or conflict'.

[16] One might also take into account, for example, population and regional effects. These are not emphasised here for notational simplicity.

3.1 The basic model

To focus on the parts of this welfare of interest to this paper, we first assume that for any fixed cumulative emissions $q := \sum_{t\geq 1} q_t$, the precise pathway $\mathbf{q}$ of emissions is chosen to be welfare optimising.[17] The welfare function W^* subject to this constraint may thus be treated as simply dependent on cumulative emissions q. Then we may focus on welfare damages $d(\mathbf{X}, q, \phi)$ and benefits $b(\mathbf{X}, q, \theta)$ (that is, the welfare gain from undertaking polluting activity, or equivalently the cost of emission reductions) such that

$$W^*(\mathbf{X}, q, \theta, \phi) = [1 + d(\mathbf{X}, q, \phi) - b(\mathbf{X}, q, \theta)]W^*(\mathbf{X}, q, 0, 0).$$

where $\theta, \phi = 0$ respectively denote that benefits from emissions, or abatement costs, are zero.[18]

We will suppress the dependence of b and d on the consumption pathway $\mathbf{X}$ for notational convenience alone; they should be understood to depend on this and, for example, on the discount rate, in the standard way. Thus we consider the The principal 'choice variable' is the quantity of greenhouse gases. We typically think of cumulative emissions; as emphasised by Allen et al. (2009), modelled temperature pathways are very much more sensitive to the total amount of CO_2 emitted, than to the temporal pathway of emissions, and so this net present value presentation makes scientific sense. Similarly, statements on climate policy and the cost of emission reductions tend to be framed around stabilisation concentrations of CO_2, rather than the route to such a stabilisation.

We make a few additional assumptions on these functions. Assume that $b_1(q, \theta)$ and $b_{11}(q, \theta)$ exist and satisfy $b_1(q, \theta) > 0$ and $b_{11}(q, \theta) \leq 0$, for all (q, θ) with $\theta > 0$.[19] Assume similarly that $d_1(q, \phi)$ and $d_{11}(q, \phi)$ exist. The random variable Θ, parametrising uncertainty in the cost of emission reductions, is assumed to take values in $\mathbb{R}$, whereas the random variable Φ, parametrising uncertainty in damages, may take values in $\mathbb{R}^n$ for any fixed $n \geq 1$ (that is, Φ is a vector of n random variables). That is, uncertainty in damages is allowed to be multi-dimensional, reflecting the many issues summarised in Section 2.

Finally, expected (net present value) damages from a stabilisation concentration q are denoted $\mathrm{TD}(q)$, i.e. $\mathrm{TD}(q) := E_\Phi[d(q, \phi)]$.

---

[17] Damages from climate change may in any case be rather insensitive to this choice; Allen et al. (2009, Figure S3 of Supplementary Information) show almost indistinguishable temperature responses for a range of emission pathways with the same cumulative total.

[18] Note that damages are positive, and benefits negative, in this expression because we use a negative-valued utility function to define W. To obtain b and d in this form, one may write $d(\mathbf{X}, q, \phi) = \frac{W^*(\mathbf{X},q,0,0) - W^*(\mathbf{X},q,0,\phi)}{-W^*(\mathbf{X},q,0,0)}$ for relative damages and $b(\mathbf{X}, q, \theta, \phi) = \frac{W^*(\mathbf{X},q,\theta,\phi) - W^*(\mathbf{X},q,0,\phi)}{-W^*(\mathbf{X},q,0,0)}$ and make the additional assumption that benefits b are independent of uncertainty ϕ in damages; this seems reasonable as the greater part of any economic shift away from fossil fuels may take place in the first half of the 21st century, whereas the more serious damages will manifest later than this.

[19] To distinguish derivatives from subscripts which index a function, we use the notation b_1 to denote the partial derivative with respect to the first argument, etc.

3.2 The Dismal Theorem of Weitzman, 2009a

In proposing a ‘dismal’ theorem, Weitzman (2007, 2009a) brought the importance of fat-tailed uncertainty in climate sensitivity to the attention of climate change economics.[20] We describe a version of the ‘Dismal Theorem’ applied to expected damages and their derivatives.[21]

A typical damage function $d(q, \phi)$ satisfies $d_1 > 0, d_{11} \geq 0$, so damages $d(q, \phi)$ tend to infinity with q. However, there must be some finite (though admittedly very large) value to the sum of everything in this finite world, and so some bound λ on damages – a “virtual statistical life” of civilisation, or of life on Earth. This anomaly does not matter in standard problems, because the probability of very high damages is usually so low; whenever this insignificance is the dominating factor, then

$$E_\Phi[d(q,\phi) \,|\, d(q,\phi) \geq \lambda] P(d(q,\phi) \geq \lambda) \to 0 \text{ as } \lambda \to \infty \tag{1}$$

and the expectation $\mathrm{TD}(q) = E_\Phi[d(q, \phi)]$ is finite.

However, if Φ has infinite support, there is no *ex ante* reason to believe this to hold. But we may explicitly impose some bound λ:

$$d_\lambda(q,\phi) := \begin{cases} d(q,\phi) & d(q,\phi) < \lambda \\ \lambda & d(q,\phi) \geq \lambda. \end{cases} \tag{2}$$

Now write $\mathrm{TD}_\lambda(q) := E_\Phi[d_\lambda(q, \phi)]$. If (1) fails to hold, it follows that without truncation at λ, expected damages $\mathrm{TD}(q)$ are infinite and that:

$$\lim_{\lambda\to\infty} \mathrm{TD}_\lambda(q) = \infty. \tag{3}$$

In words, Equation (3) states that *expected damages depend fundamentally on the bounds you place on maximal possible losses* - and on the probability of such a loss.

The “Dismal Theorem” of Weitzman (2009a) is that this is indeed the case for climate change. By considering that our modelling must be based on a finite number of observations, he concluded that the appropriate probability density function as damages become large (or, him, as consumption approaches zero) was such that Equation (3) must hold.

However, the point of Equation (3) is not that losses are unbounded, but that their expectation cannot be approximated by concentrating on the more moderate outcomes.

This presentation of the problem of climate change damages is highly dependent on the functional form of both damages and uncertainty at the very high end (as consumption approaches zero). There has been much discussion of valid forms to use (see

[20]The idea that CRRA utility gives rise to ‘negatively infinite’ expected utility in certain cases goes back to the ‘St. Petersberg paradox’ and was previously explored by Geweke (2001).

[21]Weitzman (2009a) works with the expectation of the ‘stochastic discount factor’ – that is, the marginal willingness to pay for an additional infinitesimal transfer from ‘now’ to ‘the future’, given that only this infinitesimal transfer takes place. As emphasised by Horowitz and Lange (2009) and Karp (2009), a ‘dismal’ result in the stochastic discount factor can be consistent with the optimal transfer from now to the future being a fraction less than one of our entire present wealth, although, as emphasised by Millner (2013), it is also consistent with the optimal transfer *being* our entire wealth. It is useful to understand this subtlety, but we do not wish to dwell on it, preferring to focus more directly on how to understand expectations of damages in these contexts.

Millner, 2013, for a summary) but ultimately one may question whether this is the most important point to emphasise. That expected damages are highly dependent on what we view as the 'worst possible' outcome can remain true even if that outcome is strictly bounded away from zero consumption. And the sensitivity of our expectation to highly uncertain, very bad events need not be limited to those events being the limit of what is possible.

3.3 Ambiguity Aversion and Incomplete Preferences

Before introducing incomplete preferences to the model, we give a brief survey of recent work dealing with 'Knightian uncertainty'.

"A paradigm of rational belief should allow a distinction between assessments that are well-founded and those that are arbitrary." (Gilboa et al., 2009). Thus recent models of decision theory distinguish between risk, in the form of lotteries, and subjective uncertainty over the states of the world, as well as 'indecisiveness' as to our own tastes. The typical framework encompasses a set S of 'states', and the set $\Delta(X)$ of 'lotteries' over 'outcomes' X. States are characteristics of the world which we cannot measure, whereas a lottery comes with a prescribed probability for each outcome. An 'act' is a function $h : S \to \Delta(X)$. That is, our actions will give rise to lotteries over outcomes, but which lottery we play depends on the state of the world.

The original model is Anscombe and Aumann (1963)'s model of subjective expected utility. Here, a preference relation $\succcurlyeq$ satisfies the axioms given for subjective expected utility,[22] if and only if there exists a function $u : \Delta(X) \to \mathbb{R}$, linear across probabilities and unique up to affine transformation, and an unique probability $p \in \Delta(S)$, such that $f \succcurlyeq g$ if and only if $\sum_{s\in S} p(s)u(f(s)) \geq \sum_{s\in S} p(s)u(g(s))$. (The agent is risk-averse if u is concave in outcomes). Thus, an agent must behave as if they are able to ascribe a fixed probability to each of the unknown states. It is this model which Gilboa et al. (2009) criticise.

By weakening the axioms, Gilboa and Schmeidler (1989) provide the model of maxmin expected utility.[23] A preference relation $\succcurlyeq$ satisfies this model if and only if there exists a linear utility $u : \Delta(X) \to \mathbb{R}$, as before, and a unique, non-empty closed convex set $\mathcal{P} \subset \Delta(S)$ such that the acts f are ordered according to the magnitude of $\min_{p\in\mathcal{P}} \sum_s p(s)u(f(s))$. Thus, in this model, the agent only takes into account the worst possible interpretation of the distribution of states of the world – although what is worst may vary from situation to situation.

But focusing only on the very worst state and throwing away all information about others may seem too extreme, and so Klibanoff et al. (2005) propose an alternative model. For simplicity, we describe it here as similar the model of Gilboa and Schmeidler (1989), but with a finite relevant subset $\mathcal{P} \subset \Delta(S)$, an unique probability distribution μ on $\mathcal{P}$ and an unique strictly increasing function ϕ such that acts f are ordered according to the magnitude of $\sum_{p\in\mathcal{P}} \mu(p)\phi\left(\sum_{s\in S} p(s)u(f(s))\right)$.[24] We typically expect a decision-maker to be both risk and ambiguity aversion, in which case both u and ϕ are concave.

[22] These axioms are weak order, independence, continuity, monotonicity and non-degeneracy.

[23] The axiom of independence is rejected, in favour of 'weak constant independence' and 'uncertainty aversion'.

[24] The model of Klibanoff et al. (2005) give the model in the continuous case, but we present a discrete version here for simplicity.

This is the model of 'smooth ambiguity aversion'.

Millner et al. (2013) apply this model to the 'climate sensitivity' parameter, but discuss the difficulty in applying it across all the uncertainties of climate change: it requires a range of possible probability density functions of damages, and some sense of the relative weights one should place on these. Such information is available in the case of the climate sensitivity parameter, but it seems harder to proceed in this way when we consider the welfare effects of higher levels of warming.

Thus, we use an alternative way to handle Knightian uncertainty: by generalising Anscombe and Aumann (1963) in a different way, abandoning the completeness axiom which requires comparability of any two acts. Bewley (1986) first introduced this idea, but as we see incompleteness of information both in ascribing probabilities to states of the world, and in valuing these outcomes, we follow the most general model, that of Galaabaatar and Karni (2013). A (strict) preference relation $\succ$ satisfies this model if there exists a nonempty *set* Π of pairs (p, u) such that

$$f \succ g \iff \sum_s p(s)u(f(s)) > \sum_s p(s)u(g(s)) \text{ for all } (p,u) \in \Pi.$$

If neither the strict inequality, nor its converse, holds for all $(p, u) \in \Pi$, then the decision-maker cannot robustly provide a preference over f and g. Each p is referred to as a belief 'prior'; each u is a taste or 'utility'. Thus this is a multi-prior multi-utility model of incomplete beliefs and preferences.

The contention of this paper is that, absent good information on the potential damages from higher levels of warming, we are unable to provide fully robust arguments for any one mitigation level. Instead of attempting to smooth over this by imposing ad-hoc assumptions, we should instead be explicit about this incompleteness of knowledge, and hence of preferences.

4 A New Model for Climate Damages

We seek here to model our failure to accurately ascribe values to outcomes, beyond a certain level. In particular, recall that the 'global damage functions' used in integrated assessment models of climate change are almost exclusively calibrated at 2.5°C of warming, or less. Those evaluations which we do have must be highly subjective, as outlined in Section 2, but for warming beyond this point we essentially have nothing to work with.

4.1 The Model Extended

We treat unknowns as 'states' rather than lotteries, following the distinction of Anscombe and Aumann (1963). We use a framework of incomplete preferences and beliefs, as in Galaabaatar and Karni (2013, see Section 3.3 above).[25]

[25]This model is not precisely an implementation of Galaabaatar and Karni (2013) as that model has so far only been axiomatised with a discrete state and outcome space (see also Riella (2013)). The model here is a natural extension, and it seems unimportant for our purposes that it has not yet been shown to be equivalent to a set of Savage-style set of axioms. The principle remains that one may recognise a range of plausible valuations and subjective probabilities on states of the world, without retaining any further way to distinguish between them.

To avoid an excess of notation, we take a set $\Pi \subset \mathbb{R}^n$ and allow $\pi \in \Pi$ to represent a prior-taste combination (in the language of Galaabaatar and Karni 2013) so that each π induces a valuation $d_\pi(q, \phi)$ of the outcome under each state ϕ, and a probability density $f_{\pi,\Phi}(\phi)$ on those states. The different components of π distinguishes the multiple distinct deep uncertainties, and the multiple possible choices of taste – for example governing the choice of utility function and discount rate.

We use Λ to denote the set of possible upper bounds for damages (as in Section 3.2). Thus, the final damage function used is the truncated $d_{\pi,\lambda}(q, \phi)$ for some $\pi \in \Pi$ and some $\lambda \in \Lambda$. Although $\Pi \times \Lambda$ should be understood to incorporate both priors and tastes, we shall henceforth refer to it as our set of 'priors', for brevity. We refer to (π, λ) as 'plausible' if $(\pi, \lambda) \in \Pi \times \Lambda$.

Thus we obtain a set of expected valuation of a choice of emission level q, one for each prior (π, λ). We can only 'robustly' recommend one policy above another if it is preferred for *every* choice of (π, λ).

In many cases this will not be possible. But rather than imposing a preference in some case that it itself derived from a prior (for example, a prior over how to aggregate over or pick our priors) it could be more helpful to give the full information. It also focuses attention on the importance of understanding and narrowing our set of priors.

Our λ represents a worst-case, as it is used by Weitzman (2009a), but it need not represent exactly a civilisation-wide 'virtual statistical life': perhaps even the most extreme climate change will not end civilisation; perhaps, even after civilisation ends, there continue to be changes which we currently value as 'worse'. Instead, λ should be found from considering the following questions: 'what is the objective limit to the damages that could be accrued from climate change?'; 'what is the disutility we place on such an eventuality?'; and 'is this the same disutility as we would place now on a less extreme outcome?'. The answers to these questions will depend on value judgements such as inequality aversion and the discount rate.

Our use of explicit priors in this way follows from the situation outlined in Section 2: we cannot make objective estimates of the welfare costs or risks of climate damages when we consider high level damages. However, we allow that it may be possible to estimate climate damages with greater credibility when they are at a lower level: 'familiar' scenarios are more likely to correspond to complete preferences (as with Karni, 2014). Accordingly we introduce a new constant, the 'limit to conventional damages', as well as clarifying the model in the new context:

The limit to conventional damages is l, a known constant; assume $l \leq \lambda$ for all $\lambda \in \Lambda$.

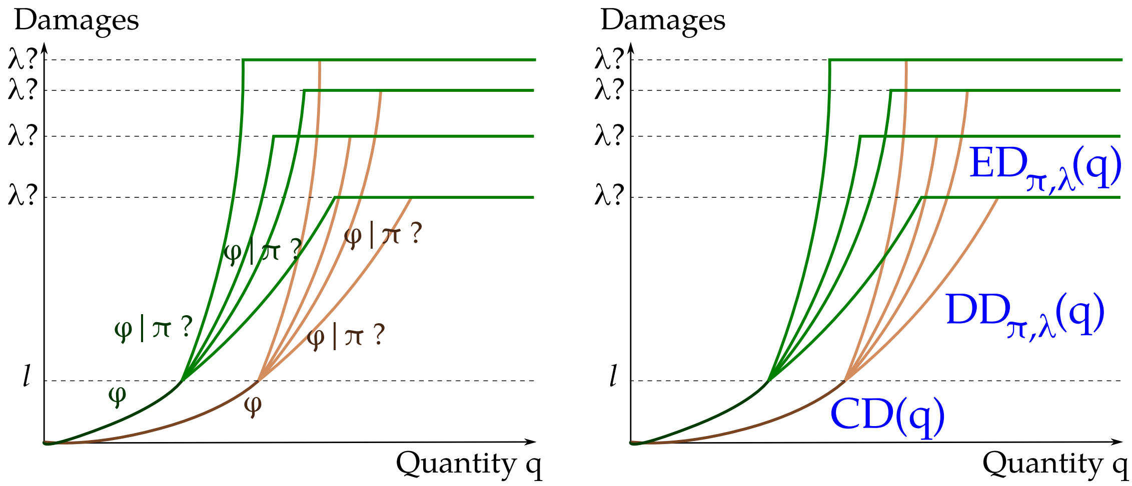
The state of the world is denoted by the previously used 'random variable' Φ.

Damages from climate change are now denoted $d_{\pi,\lambda}(q, \phi)$. The functional form of $d_{\pi,\lambda}$ depends on π, and it is truncated at λ as in Section 3.2.

A pdf on the state space induced by a prior π, is denoted $f_{\pi,\Phi}(\phi)$. Its functional form again depends on π.

Total expected damages will henceforth be written $\mathrm{TD}_{\pi,\lambda}(q) := E_\Phi[d_{\pi,\lambda}(q, \phi) \mid \pi]$. We similarly abbreviate $P(d_{\pi,\lambda}(q, \phi) < a \mid \pi)$ as $P_\pi(d_{\pi,\lambda}(q, \phi) < a)$ for any $a \in \mathbb{R}$.

The relationship between these terms is depicted in Figure 1a.



(a) *Damages $d_{\pi,\lambda}(q, \phi)$ as a function of quantity q, with possible realisations of ϕ given different values of π and λ.*

(b) *The parts of damages assigned to be 'conventional', 'indeterminate' or 'dismal'.*

Figure 1: *Damages $d_{\pi,\lambda}(q, \phi)$ under the new model of Section 4.1.*

Models of decision theory (see Section 3.3) describe how agents should rank 'acts', which map states into lotteries over outcomes. Here, an 'act' is a choice of q. For any plausible priors (π, λ) an action of q maps any state ϕ to the 'lottery' that pays off $d_{\pi,\lambda}(q, \phi)$ with certainty.[26]

We wish to make assumptions as to the relevance of the non-uniqueness of priors for the model. First define what will be meant by a 'reasonable' approximation:

Definition 4.1. For any $A, B \in \mathbb{R}_+$, and any $K > 0$, say that A *is a K-good approximation of B* if $|B - A| \leq KA$.

We typically assume that $A < B$ here.

Now,

Assumption 4.2. For our fixed l, and for some $K > 1$, assume:

1. there exist plausible π, λ such that $\mathrm{TD}_{\pi,l}(q)$ is not a K-good approximation for $\mathrm{TD}_{\pi,\lambda}(q)$;

2. there exist plausible π, λ and π', λ' such that $d_{\pi,\lambda}(q, \phi)$ is not a K-good approximation for $d_{\pi',\lambda'}(q, \phi)$ whenever $d_{\pi,\lambda}(q, \phi) > l$.

These assumptions can be jointly summarised as saying that the priors (which include our tastes as well as beliefs) *matter* for the analysis. The first says that expected damages are not well estimated by considering only those up to the level l which we understand well; the second says that, beyond the limit to our good understanding, priors are critical to estimating the damage outcome of a state.

[26] We could incorporate lotteries in addition by splitting the m components of ϕ into a 'state' part ψ and a 'lottery' part χ, so that an act q on state ψ gives rise to a lottery of payoffs $d_{\pi,\lambda}(q, \psi, \chi)$ such that the distribution of these payoffs is independent of ψ. Doing so simply imposes extra structure on $f_\pi(\phi)$ and adds complexity which seems unnecessary in this context.

Assumption 4.2 is closely connected with Weitzman's 'dismal theorem' in the form of Equation (3). An alternative way to frame that equation is: there exists no level of damages l such that $\mathrm{TD}_\lambda(q)$ is approximated well by $\mathrm{TD}_l(q)$ for all $\lambda > l$.[27] The full relationship is developed in Proposition 4.3 below.

The question, of whether or not Assumption 4.2 holds, can, and in general will, depend on the concentration q in question. Low levels of emissions will result in impacts which we *can* understand. This paper does not pretend to give an upper bound for such a level. However, Section 2 provides evidence that it does hold for *some* level.

Expected damages $\mathrm{TD}_{\pi,\lambda}(q)$ naturally break down as the sum of damages up to l, damages between l and λ, and damages beyond λ. It is useful to have a nomenclature to distinguish between these, which we set as follows:

Conventional expected damages are expected damages up to l:

$$\mathrm{CD}_\pi(q) := E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda}(q,\phi) < l,\, \pi] P_\pi(d_{\pi,\lambda}(q,\phi) < l).$$

Indeterminate expected damages are the non-conventional damages before the ultimate limit λ of damages is met:

$$\mathrm{DD}_{\pi,\lambda}(q) := E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, l \le d_{\pi,\lambda}(q,\phi) < \lambda,\, \pi] P_\pi(l \le d_{\pi,\lambda}(q,\phi) < \lambda).$$

Endgame expected damages are expected damages from catastrophe.

$$\mathrm{ED}_{\pi,\lambda}(q) := \lambda P_\pi(d_{\pi,\lambda}(q,\phi) \ge \lambda).$$

These terms simply represent the natural way to split expectations over the parts of Figure 1a; see Figure 1b. Thus

$$\mathrm{TD}_{\pi,\lambda}(q) \quad = \quad \mathrm{CD}_\pi(q) + \mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q) \tag{4}$$

Consider how restrictive the assumptions are, compared with Weitzman's dismal theorem. Suppose that Assumption 4.2.1 fails: there exists $l > 0$ such that $\mathrm{TD}_{\pi,l}(q)$ *does* form an K-good approximation for $\mathrm{TD}_{\pi,\lambda}(q)$ for all plausible λ. If the set Λ of plausible maximal damages λ is not bounded above, we conclude that Equation (3) must fail – the limiting value of $\mathrm{TD}_{\pi,\lambda}(q)$ is close to $\mathrm{TD}_{\pi,l}(q)$.

However, if the set Λ *is* bounded above, then Equation (3) must also fail. For if $\lambda \le \lambda_0$, say, for all $\lambda \in \Lambda$, then $d_{\pi,\lambda}(q,\phi) \le d_{\lambda_0}(q,\phi)$ for any $\lambda > \lambda_0$ and so $\lim_{\lambda\to\infty} \mathrm{TD}_\lambda(q) \le \mathrm{TD}_{\lambda_0}(q)$. And very similarly, a failure of Assumption 4.2.2 must also imply a failure of Equation (3).

So Equation (3) fails if Λ is bounded above. But such a bound does by no means imply failure of Assumption 4.2; one may easily construct examples incorporating a wide but bounded range Λ of plausible upper bounds. Note also that Equation (3) says little explicitly about our failure to understand damages at high levels below their absolute limit. So:

Proposition 4.3. *'Weitzman's Dismal Theorem' in the form of Equation (3) is sufficient, but not necessary, for Assumption 4.2 to hold.* □

[27] More formally, this equivalence is seen using Cauchy convergence: the statement '$\mathrm{TD}_\lambda(q) \to \infty$ as $\lambda \to \infty$' is equivalent to 'for all $l > 0$ we know $(\mathrm{TD}_\lambda(q) - \mathrm{TD}_l(q)) \to \infty$ as $\lambda \to \infty$'.

Thus Weitzman (2009a) also implicitly shares the view that our damage estimates depend on priors. Otherwise, the 'catastrophic' level λ would be objectively known *ex ante* and the limit as $\lambda \to \infty$ would not be relevant. But it seems implausible that our failure of understanding arises precisely as damage valuations reach their maximum. Thus, we have introduced the limit l to conventional damages, and explicitly model that a prior π determines judgements we make beyond this level.

4.2 A Generalised 'Dismal Theorem'

Our 'dismal theorem', which follows very directly (see Appendix A.A for details) is now:

Theorem 4.4 (The 'Dismal Theorem'). *If Assumption 4.2.1 holds then we cannot make a K-good approximation of $\mathrm{TD}_{\pi,\lambda}(q)$ without including at least one of the expected damage terms $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$.*

If Assumption 4.2.2 also holds then there exist π', λ' such that $\mathrm{TD}_{\pi',\lambda'}(q)$ is not an $(K-1)$-good approximation of $\mathrm{TD}_{\pi,\lambda}(q)$.

That is, our assumptions imply that the priors π and λ, and the terms $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ must be understood in order to calculate expected damages. Though π and λ have a smaller significance when we consider very small limits K on a 'good' approximation, we need not take K as small. Indeed the arguments of Section 2 imply that it is reasonable to assume that $K > 1$ in Assumption 4.2 and hence that both the priors, and the indeterminate and endgame expected damage terms, are of considerable importance. We shall do so for the remainder of this paper.

To calculate expected damages, then, we must consider all three of 'conventional expected damages', 'indeterminate expected damages' and 'endgame expected damages'. The relative importance of the three terms will depend on the values of l and λ we consider legitimate, and will also depend on the quantity q in question. Rather than attempting to answer this question, we develop the implications of its answer for policy choice.

By Proposition 4.3, it follows that 'Weitzman's Dismal Theorem' implies Theorem 4.4, but moreover, that Theorem 4.4 holds in a much wider range of circumstances. It is not affected by scientists being able to say with complete certainty that equilibrium climate change resulting from a doubling of GHG concentrations is bounded by (say) 40°C – if we admit that our knowledge about damages runs out at a much lower level.[28] However, such a distribution is formally not fat tailed at all.

4.3 Relative sizes of $\mathrm{DD}_{\pi,\lambda}$ and $\mathrm{ED}_{\pi,\lambda}$

Which of the terms $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ is more important? The answer is that we do not know – even in the limit as $\lambda \to \infty$, and even when we *do* know, for example, that the more restrictive Dismal Theorem of Weitzman (2009a, our Equation 3) holds.

Proposition 4.5. *Assume expected damages $\mathrm{TD}_{\pi,\lambda}(q)$ satisfy Weitzman's Dismal Theorem (Equation 3).*

[28] See Costello et al. (2010).

1. *It is possible that* $\frac{\mathrm{ED}_{\pi,\lambda}(q)}{\mathrm{DD}_{\pi,\lambda}(q)} \to 0$ *as* $\lambda \to \infty$.

2. *It is possible that* $\frac{\mathrm{ED}_{\pi,\lambda}(q)}{\mathrm{DD}_{\pi,\lambda}(q)} \to \infty$ *as* $\lambda \to \infty$.

3. *Indeterminate damages* $\mathrm{DD}_{\pi,\lambda}(q) \to \infty$ *as* $\lambda \to \infty$.

Proof. See Appendix A.A. □

So it could be the case that cost-benefit analysis is best understood as insurance against a very rare catastrophe – but it could also be the case that the badly understood, but still plausible, damages, drive our valuations. Indeed, the latter 'indeterminate' damages will always go to infinity with λ, whereas endgame damages $\mathrm{ED}_{\pi,\lambda}(q)$ need not necessarily do so.

Note moreover that we need not be in the limiting case; even if $\mathrm{ED}_{\pi,\lambda}(q)$ is dominating for absolutely vast values of λ, it is perfectly consistent that $\mathrm{CD}_{\pi}(q)$ or $\mathrm{DD}_{\pi,\lambda}(q)$ dominates for the value of λ that we consider legitimate.

In particular, the fact that our Dismal Theorem holds does *not* now imply that cost-benefit analysis is best characterised as buying insurance against very rare absolute catastrophe, as is the general interpretation of Weitzman (2009a) and also argued by Pindyck (2013a,b). It may well be the case that the term $\mathrm{DD}_{\pi,\lambda}(q)$ that dominates, and that our analysis is driven by the risk of very large damages from a temperature change of around, say, 4°C; essentially no economic analysis has directly considered such a large temperature change (see Section 2), but for atmospheric concentrations that are twice pre-industrial levels, this temperature lies within the 'most likely range' characterised by Solomon et al. (2007) and so cannot be characterised as 'rare'.

We now consider how changes in policy level, q, and in our subjective understanding of catastrophe, influence the relative importance of these terms.

First, suppose that the quantity q of GHGs increases: how does this affect the relative sizes of the three terms? While q remains small, this increases all the damage terms, though all are bounded above, by l or λ as appropriate.

However, suppose that, in the high-volume limit, the overwhelming probability is that the threshold λ has been passed. At such quantities, then, the terms $\mathrm{CD}_{\pi}(q)$ and $\mathrm{DD}_{\pi,\lambda}(q)$ are in fact *decreasing* with q, while the 'endgame' term $\mathrm{ED}_{\pi,\lambda}(q)$ dominates the analysis.

The evolution of $\mathrm{ED}_{\pi,\lambda}(q)$ with q is more straightforward: it is bounded above by λ by definition, and tends to λ itself if the probability of an 'endgame' damage situation is high enough. Figure 2 provides a schematic representation of these relationships. We can generate examples using the Matlab model of Baldwin (2014); see Section 4.4.

Write $\pi' \geq \pi$ if the cumulative density functions of the beliefs incorporated in π, π' satisfy $F_{\pi',\Phi} \leq F_{\pi,\Phi}$ (first order stochastic dominance) and tastes incorporated in π satisfy $d_{\pi',\lambda}(q) \geq d_{\pi,\lambda}(q)$ for all q.

Proposition 4.6. *Assume expected damages* $\mathrm{TD}_{\pi,\lambda}(q)$ *satisfy Assumption 4.2. Then:*

1. $\mathrm{CD}_{\pi}(q)$ *and* $\mathrm{DD}_{\pi,\lambda}(q)$ *are increasing in* q *for sufficiently small* q.

2. $\mathrm{ED}_{\pi,\lambda}(q)$ *is increasing in* q *and* π *for all* q *and* π.

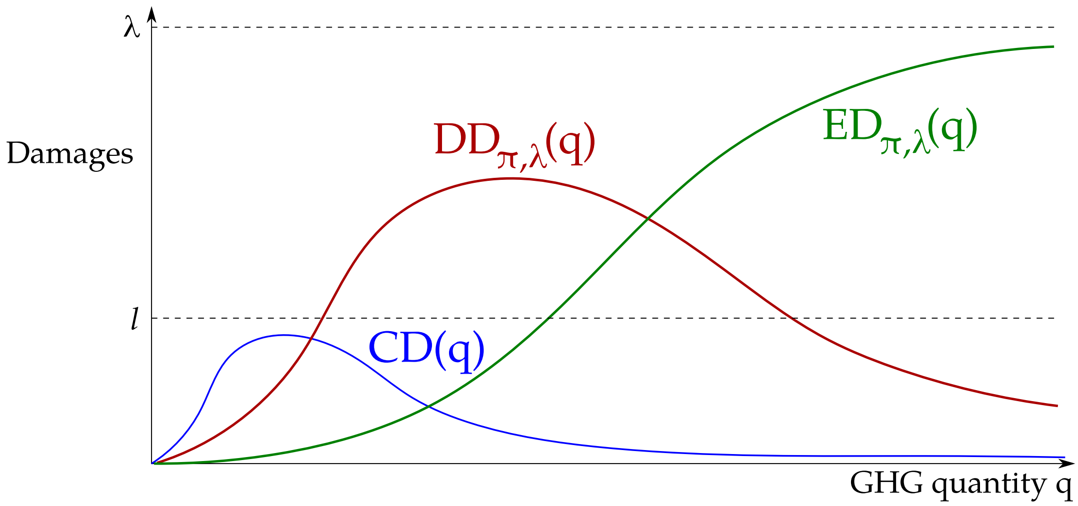


Figure 2: *A schematic representation of how* $\mathrm{CD}_\pi(q)$*,* $\mathrm{DD}_{\pi,\lambda}(q)$ *and* $\mathrm{ED}_{\pi,\lambda}(q)$ *vary with* q*.*

3. $\mathrm{CD}_\pi(q)$ *is bounded above by* l *and* $\mathrm{DD}_{\pi,\lambda}(q), \mathrm{ED}_{\pi,\lambda}(q)$ *are bounded above by* λ*.*

Suppose additionally that there exists a pathway $(q(t), \pi(t))_{t\geq 0}$ *of (weakly) increasing values of* q *and* π *such that* $P_{\pi(t)}(d_{\pi(t),\lambda}(q(t), \phi) \geq \lambda) \to 1$ *as* $t \to \infty$*. Then:*

4. $\mathrm{CD}_{\pi(t)}(q(t)), \mathrm{DD}_{\pi(t),\lambda}(q(t)) \to 0$ *as* $t \to \infty$*.*

5. $\mathrm{ED}_{\pi(t),\lambda}(q(t)) \to \lambda$ *as* $t \to \infty$*.*

Proof. See Appendix A.A. □

$\mathrm{CD}_\pi(q)$ will in general start to decline for a smaller value of q than is the case for $\mathrm{DD}_{\pi,\lambda}(q)$: it takes a lower quantity before the declining probability of staying below the conventional threshold undermines increases in expected damage given the threshold not having been passed.

The policy implication is that detailed analyses of the impacts we *can* understand become less and less relevant as the quantity of GHGs and probability of bad outcomes increase. This observation is particularly pertinent when we consider again that damage functions for climate change are calibrated at either 1 or 2.5°C of warming (with damages at higher levels being simply extrapolations). Given recent warnings that present commitments are unlikely to constrain the global temperature increase to less than 2°C, and may leave us in 2020 unable to prevent warming that exceeds this level,[29] it seems likely that $\mathrm{CD}_\pi(q)$ is indeed declining in the relevant range.

4.4 Sample Calibration

The Matlab model of Baldwin (2014) provides numbers that illustrate these points. We note some of the 'priors' in using this calibration: that the model of Allen et al. (2009) simulates the climate well; that we may convert likelihoods to probability densities using a uniform prior distribution; that the economic model and damages described there are applicable. As there, we let the percentage output damages at 4°C be an explicitly varied

[29] See e.g. Rogelj et al. (2010a,b)

prior, denoted α. Additionally, we set the low threshold l corresponds to a damage level $d_{\pi,\lambda}(q,\phi) = 0.01$ and that the 'endgame' catastrophic threshold λ corresponds to a damage function $d_{\pi,\lambda}(q,\phi) = 1$.

This allows us to illustrate Proposition 4.6, and break down damages into the terms $\mathrm{CD}_\pi(q)$, $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ for a range of possible priors on damages at 4 degrees: see Figure 3: this shows how Figure 2 may look in practice. We see $\mathrm{CD}_\pi(q)$ initially increasing before declining with q; depending on the level of assumed damages, it may dominate (Figure 3a) or be barely visible. Meanwhile $\mathrm{DD}_{\pi,\lambda}(q)$ increases, on a scale determined by our prior on damages, but we only see its eventual decline when both damages and q are extremely large. $\mathrm{ED}_{\pi,\lambda}(q)$ is negligible unless we assume that damages at 4°C are over 30–40% output, but it is only when our prior is for exceptionally high damages from 4°C, that we really see $\mathrm{ED}_{\pi,\lambda}(q)$ take over and $\mathrm{DD}_{\pi,\lambda}(q)$ decline. Depending on one's assumptions and the cumulative emissions relevant, any one of the terms may dominate.

5 Policy level and policy choice when the 'dismal theorem' holds

What does this model say about choosing our response to climate change? As ever, it would be optimal to set marginal benefits to equal marginal costs – if we knew what they were. Given the multiplicity of plausible priors, we do not expect this question to have a single answer. Certainly if one policy prescription gives a higher payoff than another under *all* plausible priors, it is robust to state that this is a better policy. More generally, considering the range of plausible marginal benefits from cutting emissions is a useful guide to policy: it facilitates discussion. By clearly providing the policy implications of different plausible priors, it focuses attention more clearly on which aspects of the problem can in fact be better understood.

Next, the question of whether one should use the long-term carbon price, or a quantity target on emissions, may be addressed using the 'prices versus quantities' model of Weitzman (1974, see Section 5.1). This employs the second derivative of $\mathrm{TD}_{\pi,\lambda}(q)$. Again, multiple priors will mean multiple values for this. In this section we will thus consider the first and second derivatives of $\mathrm{TD}_{\pi,\lambda}(q)$.

Much previous literature treats the social cost of carbon as essentially independent of the final atmospheric concentration, so $\mathrm{TD}''_{\pi,\lambda}(q) \approx 0$. As Hope (2006), van den Bijgaart et al. (2013) show, if one assumes that damages are approximately quadratic in temperature change, then (due to a certain concavity of temperature response in q) then this linearity follows. However, as discussed in Section 2, the routine assumption of a quadratic damage function lacks any empirical foundation. The sensitivity of the social cost of carbon to cumulative emissions, absent this assumption, is a key subject of discussion in Baldwin (2014); see especially Section 3.1.

5.1 The 'prices versus quantities' model (Weitzman, 1974)

We deviate from the original approach of the 'prices versus quantities' model of Weitzman (1974); incorporating uncertainty in benefits *ex ante*, and so working with far

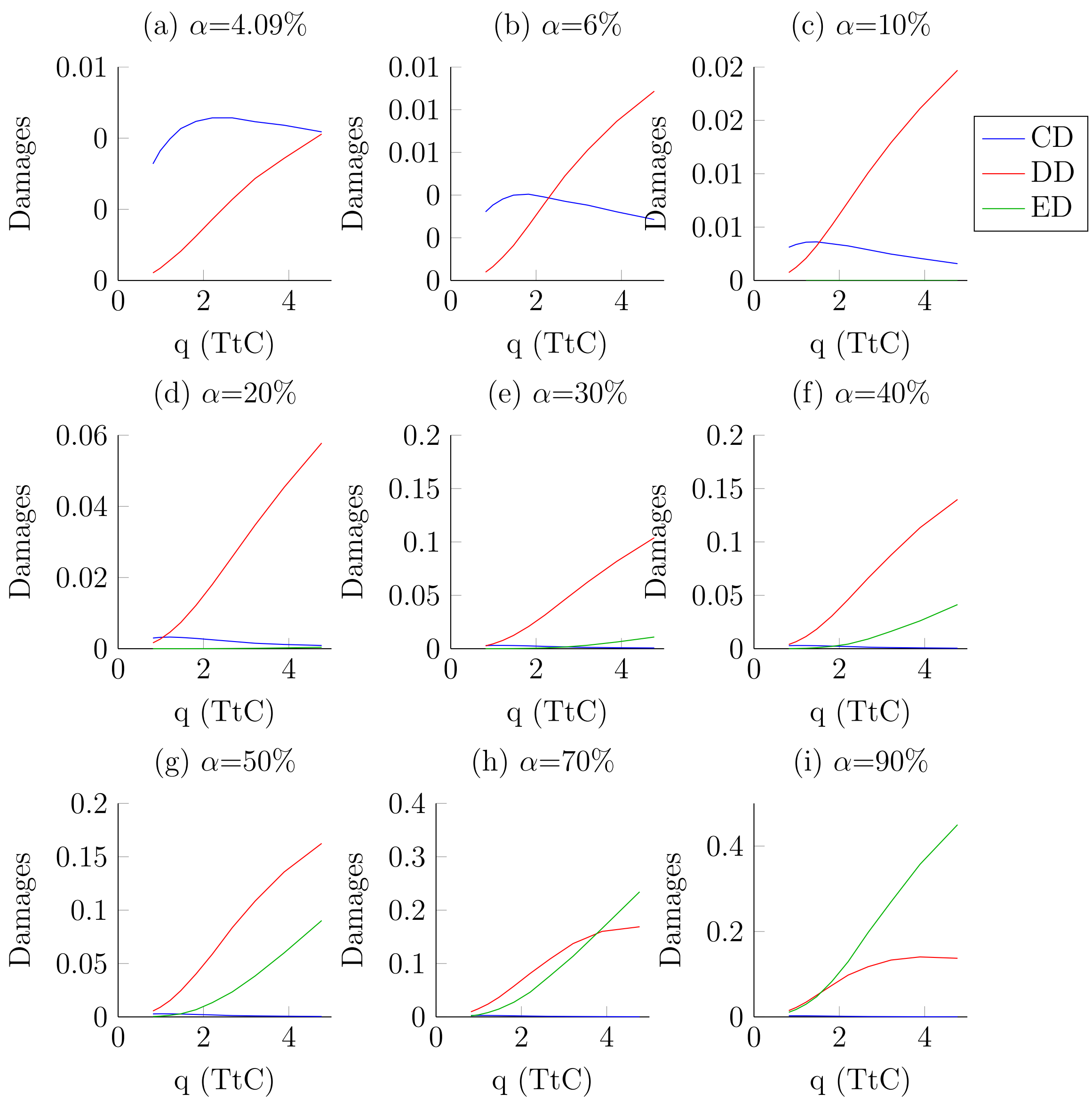


Figure 3: *The break down of damages into 'conventional', 'indeterminate' and 'endgame' terms, for various priors α on the damages from 4°C warming.*

more general forms of uncertainty in damages.[30] We also do not assume that society has chosen q to maximise benefits minus costs from emissions of GHGs, but simply that *some* quantity $\hat{q}$ has been chosen, from political and other considerations. If the regulator is approximately risk-neutral then the corresponding price to choose is $\hat{p} = E_\theta[b_1(\hat{q}, \theta)]$.[31]

We assume that $b_{11}(q, \theta)$ is independent of uncertainty θ in private benefits, and write it as $b''(q)$. Then:[32]

Theorem 5.1 (Weitzman 1974)**.** *Suppose that benefits $b(q, \theta)$ and damages $\mathrm{TD}_{\pi,\lambda}(q)$ may be approximated by their second degree Taylor expansions in a neighbourhood $Q \subset \mathbb{R}$ of $\hat{q}$. Suppose that the quantity outcome of price policy, $\tilde{q}(\theta)$, lies in Q for all realisations of θ. Then the comparative advantage of prices over quantities is equal to*

$$-\frac{\mathrm{var}[b_1(\hat{q}, \theta)]}{2b''(\hat{q})}(b''(\hat{q}) + \mathrm{TD}''(\hat{q})). \tag{5}$$

A derivation of the result in this form is given in Appendix A.B.

Thus, the question of whether prices or quantities form a better long-term policy tool depends on understanding whether damages $\mathrm{TD}_{\pi,\lambda}(q)$ are strongly convex, weakly convex, or even concave, in quantities.

An immediate consequence of working with expected damages in Theorem 5.1 is that the variance of Φ affects both marginal damages and the slope of marginal damages, effects that are lost under Weitzman's assumptions: see Example A.4.

We now apply this theory to the model of Section 4.1. Such analysis is not concerned directly with the question of whether one should, for example, use 'cap and trade' or a carbon tax, but with long-term strategy and intended emission pathways. If a government were to decide on a cumulative emission budget over a decade, say, and implement this using a carbon tax which is regularly updated as its effects are seen, then in the context of the current paper we would regard that as a quantity-based approach.

Of course decisions may be revised over the time-scales in question. But the UNFCCC process has not proved itself fast-moving, so any decision, once taken, will not be readily revised. Moreover, if large-scale investment decisions are to be made in the short term then there must be some assurance that the policy level is being fixed for the medium to long term (see e.g. Brunner et al. 2012). So, targets applicable to the long term are what we consider here – providing an analytical framework for the discussions of Dietz and Fankhauser (2010), for example.

5.2 Prices versus quantities when the 'dismal theorem' holds

Now apply the model of Section 4.1 to the 'prices versus quantities' question. Make Assumption 4.2 for some $K > 1$, so that Theorem 4.4 (the 'dismal theorem') holds.

[30]This approach is valid because the realisation of Φ plays no part in determining the outcome of either policy variable; it would not work for prices.

[31]For details on this point, see Lemma A.2 and Corollary A.3.

[32]We also assume also that we work in a sufficiently small neighbourhood of $\hat{q}$ that we may use quadratic approximations for costs and for expected benefits – note that this is a much weaker assumption than that $d_{\pi,\lambda}(q, \phi)$ is globally quadratic).

Thus, as discussed in Section 4.1, we may split up expected damages into 'conventional expected damages' $\mathrm{CD}_\pi(q)$, 'indeterminate damages' $\mathrm{DD}_{\pi,\lambda}(q)$ and 'endgame damages' $\mathrm{ED}_{\pi,\lambda}(q)$; the terms $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ are significant in this sum.

To emphasise the pattern with the results that follow, we re-state:

$$\begin{aligned}\mathrm{TD}_{\pi,\lambda}(q) &= E_\Phi\big[d_{\pi,\lambda}(q,\phi) \mid d_{\pi,\lambda}(q,\phi) < \lambda,\, \pi\big] P_\pi(d_{\pi,\lambda}(q,\phi) < \lambda) \\ &\quad + \lambda P_\pi(d_{\pi,\lambda}(q,\phi) \geq \lambda) \\ &= \mathrm{CD}_\pi(q) + \mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q)\end{aligned} \tag{6}$$

The most general form of the derivatives of $\mathrm{TD}_{\pi,\lambda}$ are given as follows:

Theorem 5.2. *Suppose that Assumption 4.2 holds, along with the other assumptions of Section 3.1. Suppose additionally that $d_{\pi,\lambda}$ is strictly increasing and differentiable with respect to one coordinate of ϕ, which we label ϕ_n. Then:*

$$\mathrm{TD}'_{\pi,\lambda}(q) = E_\Phi\left[d_{\pi,\lambda,1}(q,\phi) \mid d_{\pi,\lambda}(q,\phi) \leq \lambda,\, \pi\right] P_\pi(d_{\pi,\lambda}(q,\phi) \leq \lambda) \tag{7}$$

$$\begin{aligned}\mathrm{TD}''_{\pi,\lambda}(q) &= E_\Phi\left[d_{\pi,\lambda,11}(q,\phi) \mid d_{\pi,\lambda}(q,\phi) \leq \lambda,\, \pi\right] P_\pi(d_{\pi,\lambda}(q,\phi) \leq \lambda) \\ &\quad - E_{\Phi_{-n}}\left[d_{\pi,\lambda,1}(q,\phi)\frac{\partial}{\partial q}P_{\pi,\tilde{\Phi}_n}(d_{\pi,\lambda}(q,\phi_{-n},\tilde{\phi}_n) \geq \lambda) \,\middle|\, d_{\pi,\lambda}(q,\phi) = \lambda,\, \pi\right]\end{aligned} \tag{8}$$

Proof. See Appendix A.C. □

To understand better, it helps to work with a certain special case: all random variables ϕ acting multiplicatively on damages $d_{\pi,\lambda}$, and the probability $P_\pi(d_{\pi,\lambda}(q,\phi) \geq \lambda)$ declining as some power rule. Then the familiar terms $\mathrm{CD}_\pi(q)$, $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ re-emerge:

Corollary 5.3. *Suppose that both the relative response of damages to GHG quantity $\frac{d_{\pi,\lambda,1}(q,\phi)}{d_{\pi,\lambda}(q,\phi)}$ and that the relative change in probability of reaching 'endgame catastrophe' $\frac{\frac{\partial}{\partial q}P_\pi(d_{\pi,\lambda}(q,\phi)\geq\lambda)}{P_\pi(d_{\pi,\lambda}(p,\phi)\geq\lambda)}$ are independent of ϕ. Then (suppressing arguments for clarity):*

$$\mathrm{TD}'_{\pi,\lambda}(q) = \frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}[\mathrm{CD}_\pi + \mathrm{DD}_{\pi,\lambda}] \tag{9}$$

$$\mathrm{TD}''_{\pi,\lambda}(q) = \frac{d_{\pi,\lambda,11}}{d_{\pi,\lambda}}[\mathrm{CD}_\pi + \mathrm{DD}_{\pi,\lambda}] - \frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}\frac{\frac{\partial}{\partial q}P_\pi(d_{\pi,\lambda} \geq \lambda)}{P_\pi(d_{\pi,\lambda} \geq \lambda)}\mathrm{ED}_{\pi,\lambda}. \tag{10}$$

We discuss Theorem 5.2 and Corollary 5.3 together.

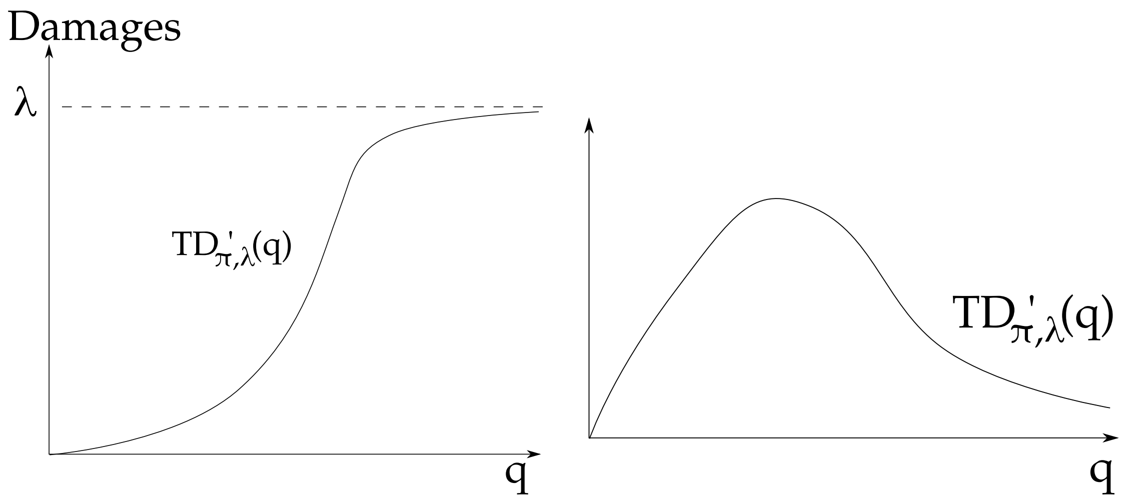
As also observed by Dietz (2011), marginal expected damages depend only on the behaviour of the damage function *up to* the 'endgame' level. Beyond that point there simply are no marginal damages. Thus we obtain (7) and (9). These terms could be split further to present expected damages in the 'conventional' range and those in the 'indeterminate' range.

The special case of (9) shows $\mathrm{TD}'_{\pi,\lambda}(q)$ as some multiple of the sum of conventional and indeterminate damages. So if the latter term is highly significant, due to our assumptions π, λ, then marginal expected damages will also satisfy Assumption 4.2 and hence our 'dismal theorem'. On the other hand, 'endgame' damages are not relevant because marginal damages have ceased from the point at which they reach λ.

The multiple acting on $\mathrm{CD}_\pi(q)$ and $\mathrm{DD}_{\pi,\lambda}(q)$ is the relative response $\frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}$ of damages to total emissions. This highlights the relatively obvious point that, if damages are more sensitive to emissions, then marginal expected damages will be greater in comparison to total damages.

Expression (9) also shows conventional and indeterminate damages having the same relative importance in marginal expected damages as they do in expected damages. This result is an artefact of the assumption that $\frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}$ is independent of ϕ. Were we to assume that values of ϕ representing worse outcomes also gave rise to a greater relative responsiveness of damages to emissions, then the relative weight placed on $\mathrm{DD}_{\pi,\lambda}(q)$ would be greater.

We turn to the second derivative. Our first observation is that 'indeterminate' and 'endgame' damages are opposing in the slope of marginal damages, and that $\mathrm{TD}_{\pi,\lambda}$ may be concave if $\mathrm{ED}_{\pi,\lambda}$ is the dominating term. The role of 'endgame' damages has changed dramatically as we pass from (6) to (8). This is best understood by first thinking about the shapes of the functions. See Figure 4a, a schematic image of how $\mathrm{TD}_{\pi,\lambda}(q)$ varies with q in the case that large enough quantities lead to an overwhelming probability that we do indeed reach the endgame λ.[33] In this case $\mathrm{TD}_{\pi,\lambda}(q) \to \lambda$ as $q \to \infty$ (it is in general bounded above by λ).



(a) *Expected damages* $\mathrm{TD}_{\pi,\lambda}$. **(b)** *Marginal damages* $\mathrm{TD}_{\pi,\lambda}{}'(q)$

Figure 4: *Total and marginal expected damages, given as a function of* q*.*

The upper bound necessarily implies that the function must be concave for large enough q. And if, as shown, the function tends to λ in the large q limit, then marginal damages $d_{\lambda,1}(q,\phi)$ are zero if q is sufficiently large. Thus, expected marginal damages $\mathrm{TD}'_{\pi,\lambda}(q)$ must decrease with q once q is large enough, and so $\mathrm{TD}''_{\pi,\lambda}(q) < 0$ for sufficiently large q. See Figure 4b.

Since the upper bound λ is driving this effect, it corresponds to the situation in which $\mathrm{ED}_{\pi,\lambda}(q)$ is significant. We turn to Equation (10): a large $\mathrm{ED}_{\pi,\lambda}(q)$ term (mitigated by a multiple) does indeed bring down the curvature of total damages, whereas convexity in damages and a large $\mathrm{DD}_{\pi,\lambda}(q)$ brings this curvature up.

[33]For this to hold, it is sufficient to assume, as is usual, that $d_{\pi,\lambda}(q,\phi)$ is convex in the control parameter q.

To understand the these two terms, recall that a small increase in q has two potential effects on marginal damages. It increases marginal damages in the case that the catastrophic limit has not been passed – by $d_{\pi,\lambda,11}$. So a larger relative curvature in this region means a greater significance of effects of this sort. On the other hand, the opposite is possible if $d_{\pi,\lambda,11}$ small or negative – perhaps because of diminishing sensitivity of the environment to greater accumulations of CO_2.

However, in the case that the catastrophic limit *has* been passed, there are no additional damages from a small increase in q, and so this increase brings marginal damages down. Thus an increase in the probability that we pass this threshold *decreases* marginal damages proportionately by the just-pre-catastrophe marginal damage level. And we measure this effect by looking at the effect of q on the *change in probability of passing the catastrophe point.* If increases in q make little difference to this probability then the slope of marginal damages is dominated by the positive term – and this can be the case either if the probability of passing λ is negligible, or if passing λ is a significant risk but quantities are at such a level that small changes in quantity make little difference to this risk.

The latter situation would correspond to being well in the 'tail' of the pdf of Φ – denoting a situation in which, once we are outside the realm of the well-understood, pretty much anything could happen. This may be the case for some assumptions π, and not for others.

Section 4.3 discussed at length the effect of our choice of priors on the relative sizes of $\mathrm{CD}_\pi(q)$ and $\mathrm{DD}_{\pi,\lambda}(q)$. Theorem 5.2 and Corollary 5.3 now show that the question of which term dominates is highly relevant for policy choice. We know that $\mathrm{ED}_{\pi,\lambda}(q)$ will always dominate for large enough q, and so $\mathrm{TD}_{\pi,\lambda}(q)$ will eventually be concave for large enough q, but the question of which dominates in the relevant region will have to be decided by a use of judgement on the plausible priors.

Priors also affect the additional terms which moderate our parts of expected damages in Corollary 5.3. In particular, we observe that the term $\frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}$ is increasing in q if $d_{\pi,\lambda}$ is log convex, and decreasing in q if $d_{\pi,\lambda}$ is log concave. And if we re-write $\frac{d_{\pi,\lambda,11}}{d_{\pi,\lambda}} = \frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}\frac{d_{\pi,\lambda,11}}{d_{\pi,\lambda,1}}$ we can also observe that $\frac{d_{\pi,\lambda,11}}{d_{\pi,\lambda,1}} > \frac{d_{\pi,\lambda,1}}{d_{\pi,\lambda}}$ if $d_{\pi,\lambda}$ is log-convex in q. So, if $d_{\pi,\lambda}$ is log-convex, the weight we place on the $\mathrm{CD}_\pi(q)$ and $\mathrm{DD}_{\pi,\lambda}(q)$ in calculating $\mathrm{TD}'_{\pi,\lambda}(q)$ increases with q, relative to its importance in $\mathrm{TD}_{\pi,\lambda}(q)$ and increases again in relevance as we move to $\mathrm{TD}''_{\pi,\lambda}(q)$. The opposite is true if $d_{\pi,\lambda}(q)$ is log-concave.

Finally, recall why a negative term increases our preference for prices. Using a price-based tool means that we may under- or overshoot our emissions target. If expected damages are in fact concave, then the upside to undershooting the target is greater than the downside to overshooting the target; the usual argument of Weitzman (1974) is reversed.

Concavities in the expected damage function $\mathrm{TD}_{\pi,\lambda}(q)$ arise as a result of any concavity in the damage response to quantities, and as a result of the truncation of damages at the 'endgame' damage level λ. We would wish to use a price-based tool in one of three situations: the 'indeterminate' and 'endgame' damages are not especially large, and the economic downside to a quantity-based approach is significant; damages are relatively concave in quantities; or when the effect of increasing the eventual concentration is to considerably increase an already-important risk of 'endgame' damages.

So there are two fundamental reasons why we may strongly prefer prices: because damages are relatively mild, or because 'the risk of reaching final catastrophe' is high. The latter scenario, which also corresponds generally to larger values of q, seems to suggest the chosen policy target is highly questionable.

6 Conclusion

When there is very great uncertainty, both in what might happen, and in how society should value these outcomes, our economic models must recognise this. And the significance of poorly-understood damages may drive not only the optimal level of policy response, but also the best choice of policy tool: whether to focus on getting the right price, or the right quantity.

To summarise the argument: there is very great uncertainty in the final socio-economic damages that will result from our emissions of greenhouse gases. It follows that we do not know how to estimate their costs, and we should acknowledge this explicitly in our models. We formally model the fact that our 'priors' on the evolution of damages will determine their modelled levels, especially when damages might be large. Our tastes and beliefs are both incomplete, and hence we follow the model of Galaabaatar and Karni (2013).

We separate out damages before some 'conventional' limit, which bounds our good quality estimations, 'indeterminate' damages, whose level depends on the priors we take, and finally damages accruing at some 'endgame' damages level. In particular we are forced to consider and quantify badly-understood but plausible outcomes, and what a putative worst-case might be.

If high-level damages are sufficiently plausible then these terms will be a determining factor in the level of expected damages; one cannot assess damages or start to decide on suitable trade-offs without first considering what the effect of the 'indeterminate' or 'endgame' terms might be. And they are key not only in assessing the levels of damages, but in marginal damages and their slope.

If 'endgame' damages form the crucial term, then marginal damages may still be low. Such outcomes are 'so bad they cannot get worse' and so additional emissions do not add to the problem. However, 'indeterminate' damages do indeed add weight to marginal damages, and the greater the convexity of the damage function, the more important this effect.

We should calculate these effects across a range of plausible priors. If marginal damages could 'plausibly' be very high, society's choice of response will realistically involve and attempt to calibrate that level of 'plausibility'. As there is a limit to the quality of knowledge we can attain, so some final estimation of the right level of action might invoke the Klibanoff et al. (2005) model of smooth ambiguity aversion. However, we argue that keeping track of the range of plausible marginal values provides better information and understanding in this context than does taking such a reduced form in the first place.

Weitzman (1974) showed that 'prices' being better corresponds to concavities, or to convexities less extreme than the concavity of the private benefit function. Expected damages are convex in the volume of gases when damages increase steeply, when 'endgame' damages are not too significant in expectation, and when the probability of 'endgame' is only very slightly affected by a decrease the quantities emitted. It is concave if expected catastrophic damages dominate conventional ones, if quantities are very high, or if the probability of catastrophe is substantially reduced by a small decrease in emissions.

A preference for a price-based target thus corresponds to high levels of emissions, the significant risk of catastrophic damage, and the potential to do a reasonable amount about this by reducing the quantity. Such a scenario really begs the question of whether the concentration one is aiming for is indeed 'optimal'. On the other hand, if one aims for a low concentration, convexities are more assured, and so a quantity-based tool more likely to be relevant.

The use of a one-period net present value model means that we cannot incorporate 'learning' – as uncertainties are resolved, we cannot update our policy choices. A multi-period version of this model would be an interesting extension. However, in the context of climate change, given the speed of the international negotiating process and so its likely ability to update decisions, and the fact that decisions taken in the short term restrict the options available later,[34] the principle insights of this model will continue to hold.

References

F. Ackerman, E. Stanton, C. Hope, S. Alberth, J. Fisher, B. Biewald, and S. Economics. Climate change and the US economy: The costs of inaction. *Medford, MA, Tufts University Global Development and Environment Institute and Stockholm Environment Institute-US Center*, 2008.

F. Ackerman, E. A. Stanton, C. Hope, and S. Alberth. Did the Stern Review underestimate US and global climate damages? *Energy Policy*, 37(7):2717–2721, 2009.

M. Allen, D. Frame, C. Huntingford, C. Jones, J. Lowe, M. Meinshausen, and N. Meinshausen. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458(7242):1163–1166, 2009.

D. G. Andrews and M. R. Allen. Diagnosis of climate models in terms of transient climate response and feedback response time. *Atmospheric Science Letters*, 9(1):7–12, 2008.

J. Annan and J. Hargreaves. Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys. Res. Lett*, 33:L06704, 2006.

F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34(1):199–205, 1963.

E. Baldwin. An essay in economic theory applied to climate change: Prices versus quantities under extreme uncertainty. MPhil thesis, University of Oxford, May 2010.

E. Baldwin. Optimal policy under 'dismal' uncertainty. Mimeo., August 2013.

[34]See, for example, Rogelj et al. (2010a)

E. Baldwin. The social cost of carbon and assumptions on future damages. Mimeo., February 2014.

T. F. Bewley. Knightian decision theory, Part I. Discussion Paper 807, Cowles Foundation, 1986. Published in *Decisions in Economics and Finance*, 25(2):79–110, 2002.

S. Brunner, C. Flachsland, and R. Marschinski. Credible commitment in carbon policy. *Climate Policy*, 12(2):255–271, 2012.

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

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

S. Dietz and S. Fankhauser. Environmental prices, uncertainty, and learning. *Oxford Review of Economic Policy*, 26(2):270–284, 2010.

D. Frame, E. Baldwin, and R. Hahn. Fat-tailed climate change. In preparation, 2013.

J. Freeman and A. Guzman. Sea Walls are Not Enough: Climate Change and US Interests. Public Law Research Paper 1357690, UC Berkeley, 2009.

P. Friedlingstein, P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng. Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate*, 19(14):3337–3353, 2006.

T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

J. Geweke. A note on some limitations of CRRA utility. *Economics Letters*, 71(3): 341–345, 2001.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

I. Gilboa, A. Postlewaite, and D. Schmeidler. Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25:285–296, 2009.

A. Golub, E. Strukova, and J. Wang. Costs and benefits of climate policy under uncertainty. In A. Golub and A. Markandya, editors, *Modelling environment-improving technological innovations under uncertainty*. Routledge, 2009.

L. H. Goulder and A. Schein. Carbon taxes vs. cap and trade: A critical review. *Climate Change Economics*, Forthcoming, 2014.

D. Guan, Z. Liu, Y. Geng, S. Lindner, and K. Hubacek. The gigatonne gap in China's carbon dioxide inventories. *Nature Climate Change*, 2:672–676, 2012.

A. Hannart, J.-L. Dufresne, and P. Naveau. Why climate sensitivity may not be so unpredictable. *Geophysical Research Letters*, 36:L16707, 2009.

G. Heal and A. Millner. Reflections: Uncertainty and decision making in climate change economics. *Review of Environmental Economics and Policy*, 8(1):120–137, 2014.

S. Henriksson, E. Arjas, M. Laine, J. Tamminen, and A. Laaksonen. Comment on "Using multiple observationally-based constraints to estimate climate sensitivity" by J. D. Annan and J. C. Hargreaves, Geophys. Res. Lett., 33, L06704, doi:10.1029/2005GL025259, 2006. *Climate of the Past Discussions*, 5:2343–2349, 2009.

C. Hepburn. Regulation by prices, quantities, or both: A review of instrument choice. *Oxford Review of Economic Policy*, 22(2):226–247, 2006.

M. Hoel and L. Karp. Taxes and quotas for a stock pollutant with multiplicative uncertainty. *Journal of Public Economics*, 82(1):91–114, 2001.

M. Hoel and L. Karp. Taxes versus quotas for a stock pollutant. *Resource and Energy Economics*, 24(4):367–384, 2002.

C. W. Hope. The social cost of carbon: what does it actually depend on? *Climate Policy*, 6(5):565–572, 2006.

C. W. Hope. The social cost of CO_2 from the PAGE09 model. Economics Discussion Paper 2011-39, Kiel Institute for the World Economy, 2011.

J. Horowitz and A. Lange. What's wrong with infinity – a note on Weitzman's dismal theorem. Available from `http://faculty.arec.umd.edu/jhorowitz/weitzman_final.pdf`, 2009.

IPCC. Summary for policymakers. In M. Parry, O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson, editors, *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 7–22. Cambridge University Press, UK, 2007.

E. Karni. Familiarity breeds completeness. *Economic Theory*, 56(1):109–124, 2014.

L. S. Karp. Sacrifice, discounting and climate policy: Five questions. Working Paper Series 2761, CESifo, 2009.

N. Keohane. Cap and trade, rehabilitated: Using tradable permits to control US greenhouse gases. *Review of Environmental Economics and Policy*, 3(1):42–62, 2009.

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

W. Knorr. Is the airborne fraction of anthropogenic CO_2 emissions increasing? *Geophysical Research Letters*, 36(21):L21710, 2009.

U. Kornek and R. Marschinski. Prices vs. quantities for international environmental agreements. Mimeo., 2013.

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

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

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

S. C. Newbold and A. L. Marten. The value of information for integrated assessment models of climate change. *Journal of Environmental Economics and Management*, in press, 2014.

R. Newell and W. Pizer. Regulating stock externalities under uncertainty. *Journal of Environmental Economics and Management*, 45(2S):416–432, 2003.

W. Nordhaus. Rolling the 'DICE': an optimal transition path for controlling greenhouse gases. *Resource and Energy Economics*, 15(1):27–50, March 1993.

W. Nordhaus. Expert opinion on climatic change. *American Scientist*, 82:45–45, 1994.

W. Nordhaus and J. Boyer. *Warming the world: economic models of global warming.* The MIT Press, 2003.

W. D. Nordhaus. An analysis of the dismal theorem. Cowles Foundation Discussion Paper 1686, UCLA Department of Economics, January 2009.

A. Otto, B. J. Todd, N. Bowerman, D. J. Frame, and M. R. Allen. Climate system properties determining the social cost of carbon. *Environmental Research Letters*, 8 (2):024032 (6pp), 2013.

J. E. Parsons and L. Taschini. The role of stocks and shocks concepts in the debate over price versus quantity. *Environmental and Resource Economics*, 55(1):71–86, 2013.

R. S. Pindyck. Uncertain outcomes and climate change policy. *Journal of Environmental Economics and Management*, 63(3):289–303, 2012.

R. S. Pindyck. Climate change policy: What do the models tell us? *Journal of Economic Literature*, 51(3):860–72, 2013a.

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

G. Riella. On the representation of incomplete preferences under uncertainty with indecisiveness in tastes. Mimeo., October 2013.

M. J. Roberts and M. Spence. Effluent charges and licenses under uncertainty. *Journal of Public Economics*, 5(3-4):193–208, 1976.

G. Roe and M. Baker. Why is climate sensitivity so unpredictable? *Science*, 318(5850): 629–632, 2007.

J. Rogelj, C. Chen, J. Nabel, K. Macey, W. Hare, M. Schaeffer, K. Markmann, N. Höhne, K. K. Andersen, and M. Meinshausen. Analysis of the Copenhagen Accord pledges and its global climatic impacts–a snapshot of dissonant ambitions. *Environmental Research Letters*, 5(3):034013 (9pp), 2010a.

J. Rogelj, J. Nabel, C. Chen, W. Hare, K. Markmann, M. Meinshausen, M. Schaeffer, K. Macey, and N. Höhne. Copenhagen Accord pledges are paltry. *Nature*, 464(7292): 1126–1128, 2010b.

K. Schaefer, H. Lantuit, V. Romanovsky, E. Schuur, and I. Gärtner-Roer. Policy implications of warming permafrost. United Nations Environment Programme Special Report, 2012.

N. Shakhova, I. Semiletov, A. Salyuk, V. Yusupov, D. Kosmach, and O. Gustafsson. Extensive Methane Venting to the Atmosphere from Sediments of the East Siberian Arctic Shelf. *Science*, 327(5970):1246–1250, 2010.

S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M. Tignor, H. LeRoy Miller, Jr., and Z. Chen, editors. *Climate Change 2007 – The Physical Science Basis, in Climate Change 2007, IPCC, Fourth Assessment Report (AR4)*. Cambridge University Press, 2007.

N. Stern. *The Economics of Climate Change: The Stern Review*. Cambridge University Press, 2006.

N. Stern. Presidential address imperfections in the economics of public policy, imperfections in markets, and climate change. *Journal of the European Economic Association*, 8(2-3):253–288, 2010.

N. Stern. The structure of economic modeling of the potential impacts of climate change: Grafting gross underestimation of risk onto already narrow science models. *Journal of Economic Literature*, 51(3):838–59, 2013.

T. Sterner and U. Persson. An even sterner review: Introducing relative prices into the discounting debate. *Review of Environmental Economics and Policy*, 2(1):61–76, 2008.

T. Stocker, Q. Dahe, and G.-K. Plattner. Technical summary. In *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 University Press, 2013.

C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips, and S. E. Williams. Extinction risk from climate change. *Nature*, 427(6970):145–148, 2004.

R. S. Tol. Bootstraps for meta-analysis with an application to the impact of climate change. Working Paper Series 6413, Department of Economics, University of Sussex, Sept. 2013.

United States of America Department of Defense. Quadrennial defense review report, February 2010.

N. Urban and K. Keller. Complementary observational constraints on climate sensitivity. *Geophysical Research Letters*, 36(4):L04708, 2009.

I. van den Bijgaart, R. Gerlagh, L. Korsten, and M. Liski. A simple formula for the social cost of carbon. Mimeo., 2013.

M. Weitzman. Prices vs. quantities. *The Review of Economic Studies*, 41(4):477–491, 1974.

M. Weitzman. A review of the Stern Review on the economics of climate change. *Journal of Economic Literature*, 45(3):703–724, 2007.

M. Weitzman. On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1):1–19, 2009a.

M. Weitzman. Reactions to the Nordhaus Critique. Discussion Paper 09-11, Harvard Environmental Economics Program, 2009b.

M. Weitzman. What is the "damages function" for global warming – and what difference might it make? *Climate Change Economics*, 1:57–69, 2010.

M. Weitzman. Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy*, 5:275–292, 2011.

M. Weitzman. GHG targets as insurance against catastrophic climate damages. *Journal of Public Economic Theory*, 14:221–244, 2012.

M. Weitzman. A precautionary tale of uncertain tail fattening. *Environmental and Resource Economics*, 55(2):159–173, 2013a.

M. L. Weitzman. Additive damages, fat-tailed climate dynamics, and uncertain discounting. *Economics: The Open-Access, Open-Assessment E-Journal*, 3(2009-39): 1–22, 2009c.

M. L. Weitzman. Tail-hedge discounting and the social cost of carbon. *Journal of Economic Literature*, 51(3):873–82, 2013b.

A Proofs of results in the text

Appendix A.A Proofs for Section 4

Proof of Theorem 4.4 Assumption 4.2.1 says that there exists $(\pi, \lambda) \in \Pi \times \Lambda$ such that $|\mathrm{TD}_{\pi,\lambda}(q) - \mathrm{TD}_{\pi,l}(q)| \geq K\,\mathrm{TD}_{\pi,l}(q)$. But $\mathrm{TD}_{\pi,\lambda}(q) - \mathrm{TD}_{\pi,l}(q) = \mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q) - lP_\pi(d_{\pi,\lambda}(q,\phi) \geq l)$ so $|\mathrm{TD}_{\pi,\lambda}(q) - \mathrm{CD}_\pi(q)| = \mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q) \geq K\,\mathrm{TD}_{\pi,l}(q) \geq K\,\mathrm{CD}_\pi(q)$, as required.

Next, if Assumption 4.2.2 also holds, then we can choose π, π' and λ, λ' so that $d_{\pi',\lambda'}(q,\phi) > (1+K)d_{\pi,\lambda}(q,\phi)$ whenever $d_{\pi,\lambda}(q,\phi) > l$. Then

$$\begin{aligned} E_\Phi\left[d_{\pi',\lambda'}(q,\phi) - d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda} \geq l,\, \pi\right] P_\pi(d_{\pi,\lambda}(q,\phi) \geq l) & \\ > \;& K E_\phi\left[d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda} \geq l,\, \pi\right] P_\pi(d_{\pi,\lambda}(q,\phi) \geq l) \\ = \;& K\left(\mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q)\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} E_\Phi\left[d_{\pi',\lambda'}(q,\phi) - d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda} \leq l,\, \pi\right] P_\pi(d_{\pi,\lambda}(q,\phi) \leq l) & \\ = \;& \mathrm{CD}_{\pi'}(q) - \mathrm{CD}_\pi(q) > -\mathrm{CD}_\pi(q). \end{aligned}$$

Thus,

$$|\mathrm{TD}_{\pi',\lambda'}(q) - \mathrm{TD}_{\pi,\lambda}(q)| > K\left(\mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q)\right) - \mathrm{CD}_\pi(q)$$

From the first part of the theorem, $\mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q) \geq K\,\mathrm{CD}_\pi(q)$. So

$$|\mathrm{TD}_{\pi',\lambda'}(q) - \mathrm{TD}_{\pi,\lambda}(q)| > (K-1)\left(\mathrm{DD}_{\pi,\lambda}(q) + \mathrm{ED}_{\pi,\lambda}(q)\right) + (K-1)\mathrm{CD}_\pi(1) = (K-1)\mathrm{TD}_{\pi,\lambda}(q)$$

as required. □

Proof of Proposition 4.6.

1. Note that $\mathrm{CD}_\pi(0) = 0$ and that $\mathrm{CD}_\pi(q) \geq 0$ for $q > 0$. Similarly note $\mathrm{DD}_{\pi,\lambda}(0) = 0$ and $\mathrm{DD}_{\pi,\lambda}(q) \geq 0$ for $q > 0$.

2. Since $\mathrm{ED}_{\pi,\lambda}(q) = \lambda P_\pi(d_{\pi,\lambda}(q,\phi) \geq \lambda)$ and since $d_{\pi,\lambda 1} \geq 0$ it follows that $P_\pi(d_{\pi,\lambda}(q,\phi) \geq \lambda)$ is weakly increasing with q and π.

3. Recall that $\mathrm{CD}_\pi(q) = E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda}(q,\phi) < l] P_\pi(d_{\pi,\lambda}(q,\phi) < l)$ and so $\mathrm{CD}_\pi(q) \leq l P_\pi(d_{\pi,\lambda}(q,\phi) < l) \leq l$. Similarly, $\mathrm{DD}_{\pi,\lambda}(q) = E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, l \leq d_{\pi,\lambda}(q,\phi) < \lambda] P_\pi(l \leq d_{\pi,\lambda}(q,\phi) \leq \lambda) \leq \lambda$ and $\mathrm{ED}_{\pi,\lambda}(q) = \lambda P_\pi(d_{\pi,\lambda}(q,\phi) \geq \lambda) \leq \lambda$.

4. Tighter bounds on $\mathrm{CD}_\pi(q)$ and $\mathrm{DD}_{\pi,\lambda}(q)$ are given by $l P_\pi(d_{\pi,\lambda}(q,\phi) < l))$ and $\lambda P_\pi(l \leq d_{\pi,\lambda}(q,\phi) < \lambda)$ respectively. It is clear that if $P_{\pi(t)}(d_{\pi,\lambda}(q(t),\phi) \geq \lambda) \to 1$ then these $\to 0$. 5 is clear. □

Proof of Proposition 4.5.

Note that, by definition

$$\mathrm{DD}_{\pi,\lambda}(q) = E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda}(q,\phi) < \lambda] P_\pi(d_{\pi,\lambda}(q,\phi) < \lambda) - \mathrm{CD}_\pi(q)$$

and that $\mathrm{CD}_\pi(q)$ is independent of λ. But

$$\lim_{\lambda\to\infty} E_\Phi[d_{\pi,\lambda}(q,\phi) \,|\, d_{\pi,\lambda}(q,\phi) < \lambda] P_\pi(d_{\pi,\lambda}(q,\phi) < \lambda) = \lim_{\lambda\to\infty} \mathrm{TD}_{\pi,\lambda}(q).$$

So if $\lim_{\lambda\to\infty} \mathrm{TD}_{\pi,\lambda}(q) = \infty$ then the same is true of $\mathrm{DD}_{\pi,\lambda}(q)$.

We demonstrate that either of the dismal terms may dominate in Example A.1. □

Example A.1. Suppose that Φ takes values in $\mathbb{R}$. Assume that $d_{\pi,\lambda}(q,\phi) = \beta(q\phi)^r$ for some $r > 0$ and some $\beta \in \mathbb{R}_{>0}$. Write $\phi_l := \left(\frac{l}{q^r\beta}\right)^{1/r}$ for the realisation of ϕ giving rise to damages l, and $\phi_\lambda := \left(\frac{\lambda}{q^r\beta}\right)^{1/r}$ for the realisation of ϕ giving rise to damages λ. Our assumption π on the pdf $f_\Phi(\phi)$ is that, for ϕ large enough that $d_{\pi,\lambda}(q,\phi)$ is well above level l, then $\pi(\phi) = \alpha\phi^{-s}$ for some $\alpha \in \mathbb{R}_{>0}$.

Now we distinguish two cases:

1. $s = r + 1$;
2. $s < r + 1$.

1. Suppose $s = r + 1$. Now:

$$\begin{aligned}
\mathrm{DD}_{\pi,\lambda}(q) &= \int_{\phi=\phi_l}^{\phi_\lambda} \alpha(q\phi)^{-r-1}\beta(q\phi)^r d\phi \\
&= \frac{\alpha\beta}{q}\big[\log\phi_\lambda - \log\phi_l\big] \to \infty \text{ as } \phi_\lambda \to \infty
\end{aligned}$$

and moreover note that $\phi_\lambda \to \infty$ as $\lambda \to \infty$. Thus $\mathrm{TD}_{\pi,\lambda}(q) \to \infty$ as $\lambda \to \infty$; the 'Dismal Theorem' of Weitzman's formulation (Equation (3)) holds. However,

$$\mathrm{ED}_{\pi,\lambda}(q) \;=\; \lambda\int_{\phi_\lambda}^{\infty} \alpha(q\phi)^{-r-1} d\phi \;=\; \lambda\frac{\alpha q^{-r-1}}{r}\phi_\lambda^{-r} \;=\; \frac{\alpha\beta}{rq}$$

since by definition, $\beta T_\lambda^r = \lambda$. This, we see, is independent of λ (although $\mathrm{TD}_{\pi,\lambda}(q)$ is not). Thus, in the limit as λ becomes large, $\mathrm{DD}_{\pi,\lambda}(q)$ dominates.

2. Now suppose $s < r + 1$ Now:

$$\mathrm{DD}_{\pi,\lambda}(q) \;=\; \int_{\phi_l}^{\phi_\lambda} \alpha(q\phi)^{-s}\beta(q\phi)^r d\phi \;=\; \frac{\alpha\beta q^{r-s}}{r-s+1}[\phi_\lambda^{r-s+1} - \phi_l^{r-s+1}].$$

However, noting that $\beta(q\phi_\lambda)^r = \lambda$, we see

$$\mathrm{ED}_{\pi,\lambda}(q) \;=\; \lambda\int_{\phi_\lambda}^{\infty} \alpha(q\phi)^{-s} d\phi \;=\; \frac{\lambda\alpha q^{-s}}{s-1}\phi_\lambda^{-s+1} \;=\; \frac{\alpha\beta q^{r-s}}{s-1}\phi_\lambda^{r-s+1}.$$

As ϕ_λ gets big, the term ϕ_l^{r-M+1} becomes negligible and so which out of $\mathrm{DD}_{\pi,\lambda}(q)$ and $\mathrm{ED}_{\pi,\lambda}(q)$ dominates is decided by which is greater out of $\frac{1}{r-s+1}$ and $\frac{1}{s-1}$. Thus, in the limit

$$\begin{aligned}
\frac{\mathrm{DD}_{\pi,\lambda}(q)}{\mathrm{ED}_{\pi,\lambda}(q)} \to 0 \text{ as } \lambda \to \infty &\;\Leftrightarrow\; 2(s-1) < r \\
\frac{\mathrm{DD}_{\pi,\lambda}(q)}{\mathrm{ED}_{\pi,\lambda}(q)} \to \infty \text{ as } \lambda \to \infty &\;\Leftrightarrow\; 2(s-1) > r.
\end{aligned}$$

Appendix A.B Proofs for Section 5.1

We have identified a target quantity $\hat{q}$.

Suppose that the regulator uses a price instrument instead, choosing price $\hat{p}$, and write $\tilde{q}(\theta)$ for the resulting quantity, after θ is realised.

Suppose that, in a neighbourhood $Q \subset \mathbb{R}$ of q, the following second degree Taylor approximations are appropriate:

$$b(q,\theta) \approx b(\hat{q},\theta) + b_1(\hat{q},\theta)(q-\hat{q}) + \frac{1}{2}b''(\hat{q})(q-\hat{q})^2. \tag{11}$$

$$\mathrm{TD}(q) \approx \mathrm{TD}(\hat{q}) + \mathrm{TD}'(\hat{q})(q-\hat{q}) + \frac{1}{2}\mathrm{TD}''(\hat{q})(q-\hat{q})^2. \tag{12}$$

If a price p is fixed, then industry will supply the quantity $q^s(p,\theta)$ which satisfies

$$b_1(q^s(p,\theta),\theta) = p. \tag{13}$$

Now,

Lemma A.2. *Suppose that $b''(\hat{q}) \neq \mathrm{TD}''(\hat{q})$ and $b''(\hat{q}) \neq 0$. Suppose that $q^s(\hat{p},\theta) \in Q$ for all θ, so that approximations (11) and (12) may be used. Then $\hat{p}$, the optimal choice of price instrument, is equal to $\mathrm{TD}'(\hat{q})$.*

Proof. If p satisfies $q^s(p,\theta) \in Q$ for all θ, it follows from differentiating (11) that

$$q^s(p,\theta) = \hat{q} + \frac{p - b_1(\hat{q},\theta)}{b''(\hat{q})}. \tag{14}$$

This is linear in p; the coefficient of p (namely $\frac{1}{b''(\hat{q})}$) is non-zero and independent of θ. We assume that $q^s(\hat{p},\theta) \in Q$ for the optimum price $\hat{p}$, and thus, when we optimise $E_\theta[b(q^s(p,\theta),\theta) - \mathrm{TD}(q^s(p,\theta))]$ with respect to p, the solution $\hat{p}$ satisfies

$$E_\theta[\mathrm{TD}'(q^s(\hat{p},\theta))] = E_\theta[b_1(q^s(\hat{p},\theta),\theta)] = \hat{p}$$

where we have applied the defining equation for $q^s(p,\theta)$. Now, by assumption, $\mathrm{TD}'(q)$ is locally linear in q, so $E_\theta[\mathrm{TD}'(q^s(\hat{p},\theta))] = \mathrm{TD}'(E_\theta[q^s(\hat{p},\theta)])$. We may see from (14) that $E_\theta[q^s(\hat{p},\theta)] = \hat{q} + \frac{\hat{p}-\mathrm{TD}'(\hat{q})}{b''(\hat{q})}$. Applying assumption (12), it follows

$$\hat{p} = \mathrm{TD}'\left(\hat{q} + \frac{\hat{p} - \mathrm{TD}'(\hat{q})}{b''(\hat{q})}\right) = \mathrm{TD}'(\hat{q}) + \mathrm{TD}''(\hat{q})\frac{(\hat{p} - \mathrm{TD}'(\hat{q}))}{b''(\hat{q})},$$

or

$$\hat{p}(b''(\hat{q}) - \mathrm{TD}''(\hat{q})) = \mathrm{TD}'(\hat{q})(b''(\hat{q}) - \mathrm{TD}''(\hat{q})).$$

Thus, $\mathrm{TD}'(\hat{q}) = \hat{p}$ as long as $b''(\hat{q}) \neq \mathrm{TD}''(\hat{q})$, which we assumed. □

Note that $b''(\hat{q}) \neq \mathrm{TD}''(\hat{q})$ whenever the second order condition holds, and *also* whenever its converse holds.

Corollary A.3. *Given the assumptions of Lemma A.2, it follows that:*

1. *the quantity one expects from a price tool is equal to the quantity that one would have set:*

$$E_\theta[q^s(\hat{p},\theta)] = \hat{q};$$

2. *the price that one expects from a quantity tool is equal to the price that one would have set*

$$E_\theta[b_1(\hat{q},\theta)] = \hat{p}.$$

Proof. To prove 2. note the first order condition for $\hat{q}$ implies that $E_\theta[b_1(\hat{q},\theta)] = \mathrm{TD}'(\hat{q})$; the result now follows from Lemma A.2. Now 1. follows from (14) and from $\hat{p} = \mathrm{TD}'(\hat{q}) = E_\theta[b_1(\hat{q},\theta)]$. □

We now prove Theorem 5.1. Recall that we do not assume that $\hat{q}$ is the optimal quantity; it might have been fixed externally by other considerations. If a price policy $\hat{p}$ is used, it satisfies $p_1 = E_\theta[b_1(\hat{q},\theta)]$. By (14), as shown above, it follows that $E_\theta[q^s(\hat{p},\theta)] = \hat{p}$. As in Section 5.1, we write $\tilde{q^s}(\theta)$ for $q^s(\hat{p},\theta)$.

Weitzman (1974) defines the 'comparative advantage of prices over quantities' as:

$$\Delta(\hat{q}) = E_\theta\Big[\big(b(\tilde{q}(\theta),\theta) - \mathrm{TD}_{\pi,\lambda}(\tilde{q}(\theta))\big) - \big(b(\hat{q},\theta) - \mathrm{TD}_{\pi,\lambda}(\hat{q})\big)\Big]. \tag{15}$$

Proof of Theorem 5.1. We follow Weitzman's convention in writing

$$\alpha(\hat{q},\theta) := b_1(\hat{q},\theta) - E_\theta[b_1(\hat{q},\theta)] = b_1(q,\theta) - \hat{p};$$

this implies $E_\theta[\alpha(\hat{q},\theta)] = 0$, and, with (14), that

$$q^s(\hat{p},\theta) - \hat{q} = -\frac{\alpha(\hat{q},\theta)}{b''(\hat{q})}.$$

Also, we may see now that $E_\theta[\alpha(\hat{q},\theta)^2] = \mathrm{var}[b_1(\hat{q},\theta)]$.

Assuming that $q(\theta) \in Q$, we may substitute (11) and (12) into the definition (15) of $\Delta(\hat{q})$ to obtain

$$\Delta(\hat{q}) \quad \approx \quad E_\theta\left[-[b_1(\hat{q},\theta) - \mathrm{TD}'(\hat{q})]\frac{\alpha(\hat{q},\theta)}{b''(\hat{q})(\hat{q})} + \frac{1}{2}[b''(\hat{q}) - \mathrm{TD}''(\hat{q})]\frac{\alpha(\hat{q},\theta)^2}{b''(\hat{q})^2}\right]$$

Since $E_\theta[\alpha(\hat{q},\theta)] = 0$, this is equivalent to

$$\Delta(\hat{q}) \approx E_\theta\left[-(b_1(\hat{q},\theta) - E_\theta[b_1(\hat{q},\theta)])\frac{\alpha(\hat{q},\theta)}{b''(\hat{q})(\hat{q})} + \frac{1}{2}[b''(\hat{q}) - \mathrm{TD}''(\hat{q})]\frac{\alpha(\hat{q},\theta)^2}{b''(\hat{q})^2}\right].$$

We conclude

$$E_{s,\theta}[\Delta(\hat{q})] = -\frac{\mathrm{var}[b_1(\hat{q},\theta)]}{2b''(\hat{q})^2}(\mathrm{TD}''(\hat{q}) + b''(\hat{q}))$$

as required. □

Example A.4. Suppose that $\Phi \subset \mathbb{R}$, We examine the effect simply of incorporating generalised uncertainty in damages $d_{\pi,\lambda}(q,\phi)$. Assume that $d_{\pi,\lambda}(q,\phi)$ is globally quadratic in ϕ. Now, if we write $\bar{\phi} = E_\Phi[\phi]$, then

$$d_{\pi,\lambda}(q,s) = d_{\pi,\lambda}(q,\bar{\phi}) + d_2(q,\bar{\phi})(\phi - \bar{\phi}) + d_{22}(q,\bar{\phi})(\phi - \bar{\phi})^2.$$

It follows that

$$\mathrm{TD}(q) = E_\Phi[d_{\pi,\lambda}(q,\phi)] = d_{\pi,\lambda}(q,\bar{\phi}) + d_{22}(q,\bar{\phi})\mathrm{var}(\phi).$$

Jensen's inequality holds strictly: the non-linearity of d in ϕ means that the expectation over ϕ of $d_{\pi,\lambda}(q,\phi)$ is different from the damages $d_{\pi,\lambda}(q,\bar{\phi})$. Now,

$$\begin{aligned} \mathrm{TD}'(q) &= d_1(q,\bar{\phi}) + d_{221}(q,\bar{\phi})\mathrm{var}(\Phi) \\ \mathrm{TD}''(q) &= d_{11}(q,\bar{\phi}) + d_{2211}(q,\bar{\phi})\mathrm{var}(\Phi). \end{aligned}$$

Both values depend on $\mathrm{var}(\Phi)$ which therefore cannot be ignored in the analysis.[35] The fourth-order derivative $\tilde{d}_{2211}(q,\bar{\phi})$ might seem far-fetched as an important coefficient, but it may be significant even when d is quadratic in ϕ. For example, if $d_{\pi,\lambda}(q,\phi) = \hat{d}(q\phi)$, then $d_{2211}(q,\bar{\phi}) = 2\hat{d}_2$, where $\hat{d}_2$ is the coefficient of the degree 2 term in the quadratic function $\hat{d}$.

Appendix A.C Proofs for Section 5.2

We prove both parts of Theorem 5.2 by developing a simple rule. First we split the domain $[\underline{q},\infty) \times \mathbb{R}^{m+n}$ of our damage function into two closed subsets, which overlap when $d_{\pi,\lambda}(q,\phi) = \lambda$. Note that if $d_{\pi,\lambda}(q,\phi) \neq \lambda$ for any values of (q,ϕ) then one of these sets is empty:

$$\begin{aligned} R_1 &:= \{(q,\phi) \in [\underline{q},\infty) \times \mathbb{R}^{m+n} \mid d_{\pi,\lambda}(q,\phi) \leq \lambda\} \\ R_2 &:= \{(q,\phi) \in [\underline{q},\infty) \times \mathbb{R}^{m+n} \mid d_{\pi,\lambda}(q,\phi) \geq \lambda\}. \end{aligned}$$

For any pair of functions $g : R_1 \to \mathbb{R}$ and $h : R_2 \to \mathbb{R}$, we write $C^{g,h} : [\underline{q},\infty) \times \phi \to \mathbb{R}$ for the function

$$C^{g,h}(q,\phi) := \begin{cases} g(q,\phi) & (q,\phi) \in R_1 \\ h(q,\phi) & (q,\phi) \in R_2 \setminus R_1 \end{cases}.$$

First, we assume that uncertainty is 1-dimensional. Suppose that Φ takes values in $\mathbb{R}$, and that $d_{\pi,\lambda}(q,\phi)$ is strictly increasing and differentiable with respect to both variables. In the case that there exists λ such that $d_{\pi,\lambda}(q,\phi) = \lambda$, let $\phi_\lambda : \mathbb{R} \to \mathbb{R}$ be the function which satisfies $d_{\pi,\lambda}(q,\phi_\lambda(q)) = \lambda$ for all $q \in [0,\infty)$; by the implicit function theorem, ϕ_λ exists (since d is strictly increasing in ϕ) and is differentiable.

Now:

[35] This is especially important when we recall that Roe and Baker (2007) argue that climate sensitivity has infinite variance.

Claim A.5. *Suppose that Φ takes values in $\mathbb{R}$ and that $g : R_1 \to \mathbb{R}$ and $h : R_2 \to \mathbb{R}$ are both continuous and differentiable on their domains. Then*

$$\begin{aligned} \frac{\partial}{\partial q} E_\Phi\left[C^{g,h}(q,\phi)\right] &= E_\Phi\left[C_1^{g,h}(q,\phi)\right] \\ &\quad + \left[h(q,\phi_\lambda(q)) - g(q,\phi_\lambda(q))\right] \frac{\partial}{\partial q} P_\pi(\phi \geq \phi_\lambda(q)) \end{aligned}$$

where we understand the second term on the right hand side to be zero if $d_{\pi,\lambda}(q,\phi) \neq \lambda$ for all q, ϕ.

Note that, although $C^{g,h}$ is not in general differentiable on $R_1 \cap R_2$, this does not matter for its integral over $\mathbb{R}$ as the set has measure 0 (it is a single point).

Proof of Claim A.5. We may elegantly obtain the result in the form stated by re-writing:

$$\begin{aligned} \int_0^\infty C^{g,h}(q,\phi) f_\phi(\phi) d\phi &= \int_{-\infty}^{\phi_\lambda(q)} g(q,\phi) f_\phi(\phi) d\phi \\ &\quad + \int_{\phi_\lambda(q)}^\infty [h(q,\phi) - h(q,\phi_\lambda(q)) + g(q,\phi_\lambda(q))] f_\phi(\phi) d\phi \\ &\quad + [h(q,\phi_\lambda(q)) - g(q,\phi_\lambda(q))]\, P(\phi \geq \phi_\lambda(q)). \end{aligned}$$

Then, applying Leibniz's rule,

$$\begin{aligned} &\frac{\partial}{\partial q} \int_0^\infty C^{g,h}(q,\phi) f_\phi(\phi) d\phi \\ &= \int_{-\infty}^{\phi_\lambda(q)} g_1(q,\phi) f_\phi(\phi) d\phi + \phi'_\lambda(q) g(q,\phi_\lambda(q)) f_\phi(\phi_\lambda(q)) \\ &\quad + \int_{\phi_\lambda(q)}^\infty \left[h_1(q,\phi) - \frac{d}{dq} h(q,\phi_\lambda(q)) + \frac{d}{dq} g(q,\phi_\lambda(q))\right] f_\phi(\phi) d\phi \\ &\quad - \phi'_\lambda(q) \left[h(q,\phi_\lambda(q)) - h(q,\phi_\lambda(q)) + g(q,\phi_\lambda(q))\right] f_\phi(\phi_\lambda(q)) \\ &\quad + \left[\frac{d}{dq} h(q,\phi_\lambda(q)) - \frac{d}{dq} g(q,\phi_\lambda(q))\right] P(\phi \geq \phi_\lambda(q)) \\ &\quad + [h(q,\phi_\lambda(q)) - g(q,\phi_\lambda(q))]\, \frac{\partial}{\partial q} P(\phi \geq \phi_\lambda(q)). \end{aligned}$$

Expanding the second integral shows that almost all terms cancel, leaving the required result. □

We now address the general case, using a distinguished a random variable Φ_n.

Claim A.6. *Let Φ take values in $\mathbb{R}^n$ for some $n \geq 1$. Suppose that $g : R_1 \to \mathbb{R}$ and $h : R_2 \to \mathbb{R}$ are both strictly increasing and differentiable with respect to q and with respect to Φ_n on their domain. Then:*

$$\begin{aligned} &\frac{\partial}{\partial q} E_\Phi\left[C^{g,h}(q,\phi)\right] = E_\Phi\left[C_1^{g,h}(q,\phi)\right] \\ &\quad + E_{\Phi_{-n}}\left[[h(q,\phi) - g(q,\phi)] \frac{\partial}{\partial q} P_{\Phi_n}(d_{\pi,\lambda}(q,\phi_{-n},\tilde{\phi}_n) \geq \lambda | \phi_{-n}) \,\middle|\, d_{\pi,\lambda}(q,\phi) = \lambda\right]. \end{aligned}$$

Proof. First, fix a realisation ϕ_{-n} of the remaining random variables. Apply Claim A.5 to the $C^{g,h}(q,\phi_{-n},\phi_n)$ to obtain

$$\frac{\partial}{\partial q}E_{\phi_n}\left[C^{g,h}(q,\phi_{-n},\phi_n)\right] = E_{\phi_n}\left[C_1^{g,h}(q,\phi_{-n},\phi_n)\right] \\ + \left[h(q,\phi)\big|_{d_{\pi,\lambda}(q,\phi)=\lambda} - g(q,\phi)\big|_{d_{\pi,\lambda}(q,\phi)=\lambda}\right]\frac{\partial}{\partial q}P(d_{\pi,\lambda}(q,\phi_{-n},\phi_n)\geq\lambda|\phi_{-n}), \quad (16)$$

where again we understand the latter term to be zero if there do not exist q, ϕ_n such that $d_{\pi,\lambda}(q,\phi_{-n},\phi_n)=\lambda$.

The additional term in Claim A.5 arises because $C^{g,h}(q,\phi)$ is not differentiable with respect to q for all q (when we fix ϕ). However, $E_{\phi_n}\left[C^{g,h}(q,\phi_{-n},\phi_n)\right]$ *is* differentiable with respect to q for all q; taking the expectation has 'smoothed' the kink and we have its derivative in Equation (16). It follows (using Leibniz's rule) that we may interchange differentiation with respect to q, and integration with respect to the remaining coordinates of ϕ; in other words, when we take expectations over the remaining coordinates of ϕ, we have

$$E_{\Phi_{-n}}\left[\frac{\partial}{\partial q}E_{\phi_n}\left[C^{g,h}(q,\phi_{-n},\phi_n)\right]\right] = \frac{\partial}{\partial q}E_{\Phi}\left[C^{g,h}(q,\phi)\right].$$

Applying $E_{\Phi_{-n}}$ to both sides of (16) then provides the result as stated. □

Now:
Proof of Theorem 5.2. First let $g(q,\phi)$ be $d(q,\phi)$ on the domain R_1 and $h(q,\phi)$ be identically equal to λ on the domain R_2. Then $d_\lambda(q,\phi)=C^{g,h}(q,\phi)$ and so, by Claim A.6, we see that

$$\begin{aligned}\mathrm{TD}'_{\pi,\lambda}(q) &= E_\Phi\left[\frac{\partial}{\partial q}d_\lambda(q,\phi)\right] + (\lambda-\lambda)\frac{\partial}{\partial q}P_\pi(d(q,\phi)\geq\lambda) \\ &= E_\Phi\left[d_1(q,\phi)\mid d(q,\phi)\leq\lambda\right]P_\pi(d(q,\phi)\leq\lambda).\end{aligned}$$

Next we let $g(q,\phi)$ be $d_1(q,\phi)$ on the domain R_1 and we let $h(q,\phi)$ be identically equal to 0 on the domain R_2, apply Claim A.6 again, we obtain:

$$\begin{aligned}\mathrm{TD}''_{\pi,\lambda}(q) &= E_\Phi\left[\frac{\partial^2}{\partial q^2}d_\lambda(q,\phi)\right] \\ &\quad - E_{\Phi_{-n}}\left[d_1(q,\phi)\frac{\partial}{\partial q}P_\pi(d(q,\phi_{-n},\tilde{\phi}_n,\phi)\geq\lambda)\,\middle|\,d(q,\phi)=\lambda\right]\end{aligned}$$

which again provides the required result. □