An essay in economic theory applied to climate change: Prices versus quantities under extreme uncertainty[1]

Elizabeth Baldwin[2]

May 15, 2011

[1]A thesis submitted in partial fulfilment of the requirements for the degree of Master of Philosophy in Economics. Candidate number: 190635. Word count: 26702.

[2]Acknowledgements: I am extremely grateful for all the advice, support and encouragement I have received for the past two years from my supervisor Paul Klemperer. I started this project at the Environmental Defense Fund, New York, and am very grateful for the support I received there from Nathaniel Keohane. I have also benefited greatly from conversations with Robert Hahn. I would also like to thank Samantha Burn, David Frame, Alexander Golub, Miles Gould, Britt Groosman, Anthony Harris, Oleg Lugovoy, Margaret Meyer, Peter Neary, Kevin Roberts, Gernot Wagner and James Wang for all their comments and help. My master's course has been generously funded by a scholarship from Nuffield College, Oxford; I am also very grateful for the funding, inspiration and support I received at the Environmental Defense Fund. Moreover I would like to thank all those who encouraged and supported me as I took up economics: Antonia Custance Baker, Cath Howdle, Frances Kirwan, Nazmiye Ozkan, and especially my family and Duncan Coutts.

Abstract

In this thesis I ask whether it is better to set a long-term price or quantity target for the regulation of greenhouse gases.

The prime example of the price approach is a carbon tax, which charges polluters some centrally-fixed amount per unit of pollution. The alternative is to fix the quantity of emissions allowed, and allow the market to set the price of tradable permits: the principle behind cap and trade. I focus on the implications of scientific uncertainty for long-term strategy, in a more sophisticated way than has been done to date.

Weitzman's (1974) classical analysis of 'prices versus quantities' balances the risk that emissions are higher than was intended against the risk that cuts are more expensive than expected. However, the uncertainty in this case is extreme. Weitzman (2009d) argues that a 'Dismal Theorem' applies: catastrophic outcomes are sufficiently likely that they dominate any computation of expected costs. My contributions are first to re-frame the Dismal Theorem, and second to apply it to the question of whether one should regulate emissions by price or by quantity.

To clarify Weitzman's 'dismal' effect, I separate expected damages from climate change into three types. First come 'conventional' expected losses, namely those up to the limit of what we can easily predict and understand. When the Dismal Theorem holds we cannot estimate damages in this way alone; we must include some 'dismal' terms. Next, one must recognise that there is a limit to the damages we face, which I call 'endgame catastrophe'. It is partly objective: there must be some worst-case scenario. It is also partly subjective, limited by how we value loss. The Dismal Theorem says that this 'endgame catastrophe' level is important to our analysis. I thus define two separate dismal expected damage terms: 'intermediate dismal expected damages', corresponding to loss beyond the 'conventional' limit but before endgame catastrophe; and 'endgame catastrophe dismal expected damages'. I then show how the relative sizes of these terms informs policy choice.

The results are as follows. If expected damages are 'dismal', then 'dismal' terms are significant in policy choice. Large 'intermediate dismal expected damages' argue for a quantity-based mechanism; an overshoot in emission levels is very risky. However, 'endgame catastrophe dismal expected damages' argue against a quantity mechanism, as extra emissions make no difference once the threshold is passed; the influence of this term is moderated by the effect of concentration levels on the probability of passing the threshold. However, if the 'endgame catastrophe' term is very significant in the estimation of expected damages, then this might indicate that the concentration being aimed for is not optimal. And a lower emission target tends to favour the quantity mechanism.

Contents

1 Introduction

How should we go about cutting greenhouse gas emissions? Economic theory favours pricing these externalities, and then letting the market determine how the cuts are made. There are two complementary mechanisms for doing so: one can set the price centrally, or one can set the quantity of emissions allowed, and leave the market to set the price.

This thesis considers instrument choice over the very long run, and so is concerned more with strategic guiding principles than precise policy choice. However, it is helpful to illustrate the two approaches with the principal associated mechanism: a carbon tax, or cap and trade. There has been much debate in recent years as to which mechanism would be better.[1] Following Weitzman (1974), we compare a price and a quantity strategy by looking at the associated risks. With a price tool, we run the risk of excess emissions; with a quantity tool, we run the risk of severe economic loss. Policy choice, then, requires a consideration and balancing of these risks.[2]

These issues have been considered by economists before. However, as recently pointed out by Weitzman (2007, 2009d), the scale of the risk run by releasing greenhouse gases has not typically been modelled well by economists. They typically focus too heavily on a 'best guess' assumption of the way in which the climate system responds to such emissions. Scientists are not in fact certain about the level of this response; they are not able to rule out the possibility that the climate system is highly sensitive to greenhouse gases. Additionally, economists tend to model future damages from climate change with an undue level of certainty; having calculated the damages they expect from a moderate amount of warming, they extrapolate the expected damages corresponding to far more severe warming without direct evidence and without modelling the risk that their 'best guess' numbers are completely wrong.

Weitzman (2009d) argues that a 'Dismal Theorem' applies to expected losses from climate change: catastrophic outcomes are sufficiently likely that they may not be neglected, and they may in fact dominate any computation of expected costs.[3] In this thesis, we reformulate the 'Dismal Theorem' to present an intuitively simpler and more convenient form. We then develop the implications of this for whether one should regulate greenhouse gases via a price or via specifying the quantity of emissions we will allow, over the long run.[4] The re-framing of the Dismal Theorem separates the range of possible temperature changes into three intervals. First there is the range within which we can sensibly make predictions. We call loss due to climate change of this scale 'conventional'. The Dismal Theorem broadly says that expected damages cannot be well estimated over this range alone; we must include some 'dismal' terms.

On the other hand, there is a limit to loss that climate change might cause. We will refer to this limiting damage level as 'endgame catastrophe'. Computing

[1] It is argued by economists such as Nordhaus (2008) that taxes are superior, although others disagree for political reasons; see Hahn and Stavins (2010). Cap and trade is used in the EU Emissions Trading Scheme and is under serious consideration in the US, Japan, Australia and New Zealand. Carbon taxes are applied in Sweden, Finland and British Columbia.

[2] There are other factors that influence the choice between prices and quantities – fluctuations in prices, lobbying incentives, political feasibility – but we will not consider them in the model of this thesis.

[3] Precisely what the analysis says and how it may be interpreted is developed at length in Section 4.

[4] The model is thus a one-period net present value model. We do not take into account the problem of governments making a credible commitment to a precise emission limit over the very long run. This problem is very real in the current context, but would be mitigated in a series of shorted commitment periods; future work will examine such cases.

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. The Dismal Theorem tells us that we must know this level in order to estimate expected damages. And finally, there are 'intermediate expected damages' corresponding to climatic changes whose impacts we cannot claim to understand well, but which are unlikely to be 'so bad that they cannot get any worse'. This thesis shows how the two dismal terms inform policy choice.

The results are laid out in Section 5. If expected damages are 'dismal', then 'dismal' terms are also significant for policy choice. However, the effect that they have may go either way. If 'intermediate dismal damages' are the most significant, then a quantity-based mechanism is strongly preferred; an overshoot in emission levels is too risky and must be avoided. However, additional emissions make no difference once 'endgame catastrophe' has taken place. Thus, if the most significant term is expected loss from 'endgame catastrophe', then a price mechanism is preferred.

Also important is the way in which the probability of endgame catastrophe is affected by the precise quantity of gases emitted. If a small increase in emission levels makes little difference to the probability of this threshold being passed, then 'intermediate dismal' damages are relatively more important. For the reasons outlined above, this situation thus gives preference to the quantity-based mechanism. On the other hand, if a small increase in emission levels greatly increases the probability of endgame catastrophe, then this emission level increase may make no difference to the damage level; under such circumstances, a price policy is preferred.

However, it is not reassuring to say that a price policy is preferred because 'the risk from endgame catastrophe 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 tends to favour the quantity mechanism.

The originality of this thesis lies in its restructuring of the Dismal Theorem, and in its application to the 'prices versus quantities' framework of Weitzman (1974). Unlike some previous applications of the latter model to the climate change situation, this is a one-period model; net present values are used for both damages from climate change, and the economic benefit derived from polluting activities.[5] We assume that one can set the quantity of emissions allowed as a single unit, allowing businesses 'banking and borrowing' of permits between trading periods. This is of course an over-simplification, and also assumes away issues of credibility and commitment. However, it enables us to focus entirely on the effect on climate policy of the inclusion of 'dismal' damages terms, and comment on whether long-term government strategy should be price or quantity-based. It paves the way for a multi-period analysis, discussed at the end of

[5] Previous, multi-period applications include Newell and Pizer (2003), Hoel and Karp (2001, 2002). The one-period net present value approach we follow is taken by the brief calculations of Keohane (2009).

Section 6.

We proceed as follows. Section 2 provides background information, primarily on uncertainty in climate change science and economics. Section 3 outlines the 'prices versus quantities' model. In Section 4 we develop the 'Dismal Theorem' aspects of the model, first providing a review of Weitzman's (2009c) model and then reformulating it as described above. In Section 5 we apply the 'prices versus quantities' model to the dismal situation. Section 6 concludes.

2 Background material: The uncertain path from emissions to damages

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. At each stage, there is uncertainty in the level of the effect. The stages (outlined, for example, in Stern, 2009) are as follows:

1. human activities emit greenhouse gases (GHGs);
2. emissions of GHGs lead to increase atmospheric concentrations of GHGs;
3. atmospheric concentrations of GHGs lead to a change in the global mean temperature;
4. changes in the global temperature cause local climatic and environmental change;
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 timescale over which they take place. Our model, however, simplifies this sequence.

We model uncertainty in stage 3, explicitly, with uncertainty in climate sensitivity, as outlined in Section 2.1 (although we consider only the long-run temperature change and not the timescale over which this takes place). We amalgamate stages 4 and 5 by modelling damages as a function of temperature change. We will not formally model uncertainty here, although doing so would be an obvious extension.

One could argue that uncertainty in stages 1 and 2 is incorporated by modelling uncertainty in the cost of reducing emissions; this will be higher against a counter-factual of steeply increasing emissions, and will be higher again if the global carbon cycle stops absorbing such a large fraction of human emissions.[6] However, as outlined in Section 3.2 and Appendix A.3, our treatment of uncertainty here is somewhat cursory. Future work could examine these issues more thoroughly.

[6]Currently around 60% of GHG emissions are re-absorbed by the oceans and terrestrial ecosystems. This fraction does not appear to be decreasing at present, but there are concerns that it might; see for example Knorr (2009).

2.1 Climate sensitivity: scientific uncertainty, and economic models

Climate sensitivity underlies the relationship between atmospheric concentrations of carbon dioxide and equilibrium temperature change of the surface of the Earth.[7] Here, we will outline scientific understanding of uncertainty in climate sensitivity, and economic modelling of the parameter. As noted by Weitzman (2009d), the majority of economic analyses of climate change have incorporated inaccurate models of uncertainty in climate sensitivity. The importance of this is not just that quantitatively different outcomes result (although this is true, see Appendix A.2 and especially Tables 5 and 6); the point is that incorrect models are qualitatively different.

Climate sensitivity s is the equilibrium temperature change resulting from a doubling of atmospheric concentrations of carbon dioxide.[8] In their summary for policymakers, Solomon et al. (2007) state that climate sensitivity 'is likely to be in the range 2°C to 4.5°C with a best estimate of about 3°C, and is very unlikely to be less than 1.5°C.' However, they also note that the possibility that $s > 4.5$ 'cannot be excluded'. With this in mind, an economist should look at the actual probability density functions and cumulative density functions of climate sensitivity estimated by the science, as pictured, for example, by Solomon et al. (2007, Box 10.2, Figures 1 and 2.)

There are two questions for economic models incorporating climate sensitivity. The first is the probability weight represented by the 'tail' of the probability density function of climate sensitivity: what is the probability that it exceeds, for example, 4.5°C? The second is, whether this tail should be considered 'fat': how sharply probabilities decline with increasing temperature.

Considering the weight in the tail, one may read off the approximate probability that a variety of studies accord to climate sensitivity exceeding 4.5°C from Solomon et al. (2007, Box 10.2, Figure 2). The range of probabilities is approximately 5% to 50%, with the majority of studies giving a probability between 15% and 35%. It is not clear, however, how these estimates should be aggregated.[9] 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.,

[7] Carbon dioxide is not the only greenhouse gas, but it is the most important. Moreover, the total effect of all non-CO_2 forcings is small, because aerosols have a negative forcing effect, approximately cancelling out the positive forcing caused by methane, nitrous oxides, etc (see Hansen et al., 2008). Additionally, although the climate response to carbon dioxide is logarithmic, that is not the case with other greenhouse gases. Therefore, in this model, we refer solely to carbon dioxide.

[8] The relationship between more general CO_2 concentrations and temperature change is logarithmic; see equation (1).

[9] There is a discussion between Weitzman (2009d,a) and Nordhaus (2009) over whether one should use a Bayesian approach or 'classical-frequentist' inspired approach for aggregation. However, Solomon et al. (2007), who have more familiarity with the data in question, note that 'there is no well-established formal way of estimating a single PDF from the individual results'.

2009).

The question of whether a tail is 'fat' refers to its functional form and limiting behaviour, rather than to the total probability accorded to the tail. Precisely what is meant by a 'fat tail' varies in the literature; it can be taken to mean 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.[10] One key paper on the uncertainty in climate sensitivity is by Roe and Baker (2007). They argues that $\text{var}_S(s)$ is infinite; the tail of climate sensitivity is almost as fat as it could possibly be, with moments ceasing to exist at the second moment.[11] An intuitive interpretation of the result is given by Allen and Frame (2007):

> The fundamental problem is that the properties of the climate system that we can observe now do not distinguish between a climate sensitivity, s, of 4°C and $s > 6$°C. In a sense, this should be obvious: once the world has warmed by 4°C, conditions will be so different from anything we can observe today (and still more different from the last ice age) that it is inherently hard to say when the warming will stop.

If it is really impossible to say what will happen once 4 degrees of warming have been exceeded, then our model should provide as diffuse a probability distribution as possible for values beyond this bound. However, not all are convinced by the arguments of Roe and Baker (see, for example, Hannart et al., 2009 and Urban and Keller, 2009).

It would appear, then, that the PDF of climate sensitivity has a significant tail, but it is not quite clear how much weight it carries, how fat it is, or how precisely it should best be modelled in economic analyses. It does seem plausible to argue that not all moments of the probability density should exist. Differing models make a qualitative difference to the analysis, as discussed by Weitzman (2009d) and further developed here.

Many economists have used a 'best guess' or 'triangular distribution' to model climate sensitivity.[12] However, as scientists are not able to rule out values of s above 4.5, nor may they be ignored in economic models. Doing so gives rise to estimates of damages which are wildly different from those which take the 'fat tail' into account, as we illustrate in Tables 5 and 6 in Appendix A.2. The importance of this error was pointed out by Weitzman (2007, 2009d). He argues that in fact the nature of the uncertainty in climate sensitivity gives rise to a 'Dismal Theorem' (which is outlined and developed in Section 4). The point is

[10] An example of the former situation is the log-logistic distribution; climate sensitivity is modelled in this way by, for example, Dietz (2009). An example of the latter situation is the lognormal distribution; climate sensitivity is modelled in this way by, for example, Golub et al. (2009).

[11] The heart of the argument, as outlined by Weitzman (2009b), is that climate sensitivity varies inversely as the sum of observables; if each of these has Gaussian uncertainty, then climate sensitivity itself has an inverse Gaussian distribution and so a fat tail.

[12] Nordhaus typically uses a value of 3 (see,for example, Nordhaus, 2008). Stern (2006) uses a triangular distribution, with minimum 1.5, maximum 4.5, and mode 3.

not simply to compare the precise quantitative results of slightly different models of this uncertainty. The point is that differing modelling possibilities make a considerable qualitative difference to the analysis.

2.2 Uncertainty in socio-economic damages from climate change

In order to compute the expected economic damages from climate change corresponding to a given concentration of carbon dioxide, one is required to know both the probability distribution of temperature changes, and the economic impacts associated to each. We thus need to know the damages $b(T)$ associated to a temperature change T, for a range of possible values of T.[13]

As is outlined by Tol (2009), various methods have been used to calculate the total economic impact of the damages from a given level of climate change. The main studies are listed in his Table 1; with one exception these are concerned with the damages from 2.5 degrees of warming, or less. Modelled damages at higher temperatures are thus principally extrapolations. The end product of these processes is that damages from 5 degrees of warming are typically modelled in the range of 1 to 7% of GWP (see Stern, 2006, Figure 6.2). Stern (2009) describes such damage functions as 'ludicrously small'; our point, however, is to emphasise that such numbers are based on *almost no actual calculations*, but purely on extrapolations, and that the choice of functional form defining them lacks justification. In fact very little is known about the shape of the damage function.

Damage functions in use are usually of the form

$$b(T) = \gamma T^n$$

where γ and n are parameters to be found. (It is implicitly assumed that $b(0) = 0$). 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, the exceptional study mentioned by Tol (2009) which does look at higher levels of warming). Other authors follow Nordhaus in taking $n = 2$ or n uncertain but centred somewhere around 2. However, models are generally based on only *one* calibration point, in addition to $b(0) = 0$.

Although some models incorporate uncertainty in the damage function (see for example Hope, 2006), there have been few attempts to formally estimate a probability distribution of damages. We briefly return to Nordhaus' (1994) survey of experts, which does consider warming of 6 degrees. The abstract highlights the 'vast disparities' in estimates obtained, but this uncertainty is

[13] It is of course a serious simplification to have damages just as a function of equilibrium temperature change, ignoring timescales, etc. Understanding the relationship between damages and temperature change is just the first part of the process of modelling the socio-economic effects of climate change.

not reflected when the study's figures are used for economic models. Nordhaus and Boyer (2003) simply take the median response from one of the questions concerning 6 degrees of warming to calibrate their damage function.[14]

The calibration point provided for the damage function at 2.5 degrees is not above criticism either. Freeman and Guzman (2009) and Ackerman et al. (2008, 2009) 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. Climate change may eventually be associated with mass migration and war; such impacts are difficult to attribute precisely or estimate, but that does not mean that they should be ignored. Indeed the US Military, in their recent review (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'.

Another criticism levelled at economic modelling of damages from climate change concerns the way in which the damage function interacts with GWP. 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.[15] Weitzman (2009c,b) argues that this assumption is highly suspect; why should damages from climate change 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 (2010) argues that climate change should be modelled via its effect on the growth rate, rather than GWP level.

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 The model 1: Framework and prices versus quantities

We use a one-period model, working with net present value cost and benefit functions. The 'quantities' of carbon dioxide we refer to are long-term equilib-

[14]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% GWP 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%. This range of uncertainty does not appear to have influenced any models.

[15]If U_t is CRRA with coefficient of relative risk aversion greater than 1, then $U_t(X_t) < 0$ and so *positive* multiplicative damages decrease utility; if one were to use a utility function which takes positive values then of course the '+' sign should be a '−'.

rium concentrations, and the temperature changes are the eventual equilibrium temperature change resulting from these concentrations.

This approach is for convenience; by working with a single period model we can isolate the effects of the fat tail without too many other complicating factors. However, it is also logical. As emphasised by Allen et al. (2009), modelled temperature pathways are highly sensitive to the total amount of CO_2 emitted, but reasonably insensitive to the temporal pathway of emissions. Thus it is not a great simplification to neglect the emission pathway in the model of damages. Moreover, statements on climate policy and the cost of emission reductions tend to be framed around stabilisation concentrations of CO_2. Future work, as outlined in Section 6, will consider a multi-period model.

3.1 Notation and global assumptions

We now define the notation to be used throughout the thesis, and give the assumptions that will always be held.

Quantities, denoted q, are the final atmospheric concentration of CO_2, relative to pre-industrial levels. Assume throughout that $q \geq 1$.

This may naturally be regarded as the quantity of a 'bad'. However, to avoid dealing throughout with negative functions, **we will in fact regard q as a 'good'**: the 'benefits' in our cost-benefit analysis will be the 'private benefits' accrued by industries and individuals indulging in polluting activities, over and above that which they would attain from clean economic activities (in other words, the avoided cost of emission cuts not undertaken); the 'costs' will be 'public costs' or 'damages', resulting from climate change. Our nomenclature is thus the opposite of that of Weitzman (1974).

Let θ be a random variable parametrising uncertainty in the cost of abating emissions of CO_2. Let the domain of θ be Θ.

Private benefits of emissions are denoted $b(q, \theta)$.[16] Assume that the function $b : \mathbb{R}_{\geq 1} \times \Theta \to \mathbb{R}$ is twice differentiable with respect to the first variable, and that $b_1(q, \theta) \geq 0$ and $b_{11}(q, \theta) \leq 0$.

In other words, there are positive marginal benefits from polluting economic activity and these are smaller for a higher eventual concentration (or equivalently, marginal cuts in emissions are more costly if one aims for a lower eventual concentration).

Regarding the 'costs' of a concentration q, we make the following definitions.

Climate sensitivity, introduced in Section 2.1, is denoted s.

[16]Formally, if $\{X_t\}$ is the GWP pathway over time t and U is global utility, then utility in the presence of emission cuts $U(\{X_t\}, q, \theta)$ is given by $(1 - b(q, \theta))U(\{X_t\})$. Benefits are subtracted rather than being added because we will typically use CRRA utility with coefficient of risk aversion greater than 1; thus utility is negative and so multiplying by a number less than 1 increases utility. However, these 'benefits' are in general negative.

Equilibrium temperature change from pre-industrial levels is denoted by T. Assume throughout that $T \geq 0$; note $T = 0$ precisely when $q = 1$.

Throughout, these parameters and q are linked by the relationship[17]

$$T(q, s) = \frac{1}{\log 2} s \log q. \tag{1}$$

Damages accrue as a result of temperature change, not directly from emissions.

Public costs or **damages** from climate change are denoted $d(T)$.[18] Assume that the function $d : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is twice differentiable, and that $d(0) = 0$, that $d'(T) \geq 0$ and that $d''(T) \geq 0$ – as long as damages have not reached 'catastrophic' levels. (Damages will later be bounded; see Section 4).

These assumptions correspond to the belief that climate change is costly, the more so the more extreme it is.

Expected damages from a stabilisation concentration q are denoted $D(q)$, so that

$$D(q) := E_s[d(T(q, s))]. \tag{2}$$

Expression (2) defines a function $D : \mathbb{R}_{\geq 1} \to \mathbb{R}$ (recall we assume throughout that $q \geq 1$). Importantly, one must note that it is not necessarily the case that $D''(q) \geq 0$; see Section 5.4. It will sometimes be convenient to neglect s, and consider T to be a random variable, given q. When we write

$$E_{T|q}[d(T) \mid q]$$

we formally mean precisely the expression of (2).

Benefits and damages from greenhouse gas emissions are assumed to accrue as follows: if $\{X_t\}$ is the GWP pathway in the absence of climate change, and U is the NPV utility function, if the long-term temperature change is T degrees Celsius, if q is the stabilisation level and the state of the world regarding the cost of emission cuts is θ, then net present value utility incorporating the effects of climate change and mitigation attempts is[19]

$$U(\{X_t\}, q, T, \theta, s) = [1 - b(q, \theta) + d(T(q, s))]U(\{X_t\}). \tag{3}$$

[17]The 'log 2' term in this formula enables one to characterise s as the equilibrium temperature change resulting from a doubling of pre-industrial concentrations of CO_2. It is slightly cumbersome to include this scaling factor, but it enables us to use precisely the same uncertain parameter s as climate scientists do.

[18]Formally, if $\{X_t\}$ is the GWP pathway over time t and U is global utility, then utility in the presence of temperature change $U(\{X_t\}, T)$ is given by $(1 + d(T))U(\{X_t\})$; the sign is plus rather than minus because utility is typically negative, as discussed in Footnote 16.

[19]For discussion of the signs here, see Footnote 16. This is consistent with Footnotes 16 and 18 if one assumes that the term $b(q, \theta)d(T)$ is small enough to be ignored, and that any interactions between climate damages and mitigation costs are insignificant. Note also that, although this model may appear multiplicative and so open to the criticisms of such models (see Weitzman, 2009b,c), the expressions in Footnotes 16 and 18 need not be taking to define $b(q, \theta)$ and $d(T)$ as being independent of $\{X_t\}$; we simply assume $\{X_t\}$ constant throughout this model.

3.2 Calibrations

Although this model is principally theoretical, it is useful to provide some calibrations of the functions in use and some sample calculations, to understand the implications in the context of climate change. Here we outline and justify the numbers and functions that will be used; details are given in Appendix A.

Values of q: We perform calculations for a range of possible stabilisation concentrations of CO_2, given below. To convert these numbers to 'q', we divide by the pre-industrial concentration of CO_2.[20] The numbers we use are:

- 350 ppm ($q = 1.24$): the focus of environmental campaign groups;[21]
- 450 ppm ($q = 1.59$): commonly referred to as the policy choice to limit warming to 2 degrees;[22]
- 550 ppm ($q = 1.94$): a target suggested by Stern (2006);[23]
- 650 ppm ($q = 2.30$): a possible result of lighter policy.[24]

Sokolov et al. (2009) gives 866 as the median figure for CO_2 concentrations in 2095 without emission reduction policy. As this thesis is about comparing policy options, we do not consider the effects of a total absence of policy.

Calibrations of the NPV damage function: Per-period damages are typically used in the literature. As we work with a net present value damage function, we have attempted a conversion and estimated a range of NPV damage functions using the temperature pathway and economic growth model of DICE2007. This is outlined in Appendix A.1; see Table 4 there. For analytic convenience, and to derive stylised facts elegantly, we often use power functions for our damage functions; we do not argue that this is accurate.

[20] Pre-industrial concentrations of CO_2 were 283 parts per million by volume (ppm) (see Indermühle et al., 1999).

[21] 350ppm is the stabilisation concentration recommended by Hansen et al. (2008); making this a global policy objective is the main focus of '350.org', and is supported by other campaigners. According to the website http://www.350.org, this target is supported by 87 countries and has been personally endorsed by Al Gore, Rajendra Pachauri (the chair of the Intergovernmental Panel on Climate Change) and Nicholas Stern.

[22] 450ppm leads to 2°C warming if $s = 3.0$, which is roughly the median estimate of s; see Solomon et al. (2007) Box 10.2, Figures 2. Thus there is a 50% chance that stabilisation at 450ppm indeed limits warming to at most 2°C.

[23] Strictly speaking, Stern (2006) refers to CO_2-equivalent concentrations, in which they include other greenhouse gases such as methane but do not include particulates (which cause negative forcing). As an illustration of the difference this makes, Stern's estimates of 2005 concentrations were 380 ppm CO_2 and 430 ppm CO_2-equivalent. Thus the policy figure of 550 ppm CO_2-equivalent should correspond to a lower one in terms of CO_2 concentrations alone. In any case Stern himself has since apparently given his support to a long-term target of 350ppm; see Footnote 21.

[24] This number is included to fit the pattern of the others, as a high end benchmark. It is the outcome of the optimistic 'Blueprints' scenario in Shell (2008); see Prinn et al. (2008). It is also approximately the concentration in 2200 under the 'optimal' policy of Nordhaus (2008); note that, in that model, the absence of policy gives rise to concentrations of 685ppm in 2100 and so Nordhaus' projections of emission levels are more optimistic than Sokolov et al. (2009).

The probability density function of climate sensitivity: In Appendix A.1, we use a lognormal distribution to we provide finite estimates from a probability density function of roughly the right shape.[25] In our formal analysis we will sometimes assume the tail of $f_s(s)$ is locally declining as a power function, for analytic convenience.

Estimates of $D(q)$ and its derivatives: The results of our calculations are laid out in Appendix A.2, especially in Tables 5, 6, 7 and 8.

The cost of mitigating greenhouse gas emissions: This thesis is mostly concerned with the shape of damages from climate change. However, we have obtained rough estimates of the cost of stabilising concentrations of greenhouse gases (modelled here as the private benefits associated with emissions), as laid out in Appendix A.3.

3.3 Prices versus quantities

The heart of this thesis is an application of Weitzman's model of 'prices versus quantities' (Weitzman, 1974). In this section we lay out the key principles; further details are given in Appendix B.

Note again that our treatment of pollution as a 'good' with associated private benefits (economic production) and public costs (climate change) is the opposite way round from Weitzman's; the exposition here is consistent with the rest of this thesis. We also deviate from Weitzman's convention by incorporating uncertainty in benefits *ex ante*. This is a valid approach because the realisation of s plays no part in determining the outcome of either policy variable; it also allows us to work with far more general forms of uncertainty in damages. Such an approach would not be valid for uncertainty θ in prices: once a price is set, θ determines the quantity set, and *vice versa*.

Society should wish to choose q to maximise

$$E_{\theta,s}[b(q,\theta) - d(T(q,s))] = E_\theta[b(q,\theta)] - D(q).$$

The first-order solution is found at $\hat{q}$ satisfying

$$D'(\hat{q}) = E_\theta[b_1(\hat{q},\theta)], \tag{4}$$

only if

$$E_\theta[b_{11}(\hat{q},\theta)] - D''(\hat{q}) < 0. \tag{5}$$

(We will see that this does not always hold.) We can then mandate the quantity $\hat{q}$.

Alternatively, one can use a price instrument p. The eventual quantity $\tilde{q}(p,\theta)$ will be set by the market and determined by the realisation of θ, satisfying

$$b_1(\tilde{q}(p,\theta),\theta) = p.$$

[25] A lognormal $f_s(s)$ will give rise to 'dismal' damages when damages increase exponentially with temperature change; although this is essentially the case considered by Weitzman (2009d), we will be interested in the more straightforward cases of polynomially increasing damages.

If quadratic approximations of benefits and damages are appropriate (as below), then marginal benefits and marginal damages are linear, i.e. the regulator is risk-neutral;[26] the optimal price $\hat{p}$ gives rise to quantity $\hat{q}$ in expectation. Thus $\hat{p} = D'(\hat{q})$. Conversely, under the optimal quantity $\hat{q}$, the expectation of the market price $b_1(\hat{q}, \theta)$ is the optimal price $\hat{p}$. For details, see Lemma B.1 and Corollary B.2.

We need not in general assume that we are aiming for the true optimal quantity $\hat{q}$; the following results hold more generally. We assume instead that someone (the government) has fixed upon some $\hat{q}$ to aim for – either through use of (4) or from other considerations. They have now turned to the economists, to ask which the best policy to attain $\hat{q}$ might be. Assume that the risk-neutrality discussed above still applies: under a price policy, we will choose $\hat{p} = E_\theta[b_1(\hat{q}, \theta)]$ so that the expected quantity outcome is still $\hat{q}$.[27] We contract our notation by writing

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

Weitzman defines the 'comparative advantage of prices over quantities' as the expected difference between net economic benefits from the price tool, and from the quantity tool:

$$\Delta(\hat{q}) = E_\theta\Big[\big(b(\tilde{\tilde{q}}(\theta), \theta) - D(\tilde{\tilde{q}}(\theta))\big) - \big(b(\hat{q}, \theta) - D(\hat{q})\big)\Big]. \tag{6}$$

He then assumes that uncertainty θ in private benefits affects abatement costs such that $b_{11}(q, \theta)$ is independent of θ; write $b''(q)$ for this. He assumes also that we work in a sufficiently small neighbourhood of $\hat{q}$ that we may use quadratic approximations for costs and benefits. The result is:

Theorem 3.1 (Weitzman (1974))**.** *Suppose that benefits $b(q, \theta)$ and damages $D(q)$ 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, $\tilde{\tilde{q}}(\theta)$, lies in Q for all realisations of θ. Then*

$$\Delta(\hat{q}) = -\frac{\text{var}[b_1(\hat{q}, \theta)]}{2b''(\hat{q})}(b''(\hat{q}) + D''(\hat{q})). \tag{7}$$

A derivation of the result in this form is given in Appendix B.[28] Note that the assumption that a local quadratic approximation is valid for $D(q)$ is much weaker than an assumption that $d(T)$ is globally quadratic; see equations (13), (14) and (15) in Section 5.1.

Thus, if we are able to show that damages $D(q)$ are strongly convex in quantities, we may argue that long-term policy should be based on quantity. On the other hand, if damages $D(q)$ are only weakly convex in quantities, or even are concave, then it is more important to fix the 'carbon price' long-term.[29]

[26]For the risk-averse central planners, see Meyer (1984).

[27]Conversely, if the government were to first identify their ideal $\hat{p}$, then the corresponding quantity policy would be to set $\hat{q}$ such that $E_\theta[b_1(\hat{q}, \theta)] = \hat{p}$.

[28]In particular, the minus sign may be explained by the fact that we have reversed the roles of 'costs' and 'benefits' (for convenience elsewhere in the model).

[29]We focus here on the damage function because understanding this function is the aim of subsequent sections. Some concrete calculations estimating $b''(\hat{q})$ and $D''(\hat{q})$ for various plausible benefit and damage functions are given in Appendix A.2.

4 The model 2: Reformulation of the Dismal Theorem

The concept of a 'Dismal Theorem' is not unique to climate change or Weitzman's work. As pointed out by Nordhaus (2009), the idea that CRRA utility gives rise to 'negatively infinite' expected utility in certain cases was previously explored by Geweke (2001). However, Weitzman (2007, 2009d) brought the importance of uncertainty in climate sensitivity to the attention of climate change economics.

4.1 The Dismal Theorem due to Weitzman

The framework of Weitzman (2009d) applies a 'Dismal Theorem' to the expectation of the 'stochastic discount factor'.[30] We have found it more straightforward to work in terms of expected damages and their derivatives.

A modeller will usually allow the damage function $d(T)$ to tend to infinity with T. This is despite the fact that we do not really recognise unbounded damages; 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. This anomaly does not matter in standard problems, because the probability of very high damages is usually so low. If the expectation $D(q) = E_{T|q}[d(T) \,|\, q]$ is finite, then the expectation of high level damages must be small; in such cases

$$E_{T|q}[d(T) \,|\, d(T) \geq \lambda, q] \to 0 \text{ as } \lambda \to \infty \tag{8}$$

Thus, the fact that we may have included larger damages than we really meant is unimportant.

However, the unbounded nature of the damage function need not be irrelevant in this way; property (8) need not hold. If the support of the PDF $f_{T|q}(T)$ of temperature change T is infinite,[31] then

$$E_{T|q}[d(T) \,|\, q] = \lim_{h \to \infty} \int_{T=0}^{h} d(T) f_{T|q}(T) dT, \tag{9}$$

a limit which we have no *ex ante* reason to believe exists as a finite number. One may characterise the question of a finite limit's existence as a race: as $T \to \infty$, the question is whether $f_{T|q}(T) \to 0$ faster than $d(T) \to \infty$. In fact:

- if there exist constants $k > 0$ and $a > 1$ such that $d(T) f_{T|q}(T) \leq k T^{-a}$ for all sufficiently large T, then $f_{T|q}(T)$ 'wins'; a finite limit exists;
- if there exists a constant $k > 0$ such that $d(T) f_{T|q}(T) \geq k T^{-1}$ for all sufficiently large T, then $d(T)$ 'wins'; a finite limit does not exist.

[30] That is, the amount of consumption an agent would be willing to give up this period, to be sure of attaining one unit of extra consumption next period; Weitzman divides time into 'now' and 'the distant future'.

[31] It is equivalent to say that the support of the PDF $f_s(s)$ of climate sensitivity is infinite; in this section we work with T for conceptual and notational simplicity.

Since we model $d(T) \geq 0$ for all T, the integral $\int_0^h d(T) f_T(T) dT$ is monotone increasing with h. Thus, if a finite limit does not exist, we may conclude that the limit of (9) is infinity.[32] This situation corresponds to the 'dismal' case.

The question, then, is how to deal with this situation. In Weitzman (2009d), the approach adopted is to speak of a 'virtual statistical life'-like parameter λ which represents the disutility of the end of civilisation. We will follow this approach and bound damages at a constant λ; interpretations and implications of this parameter are discussed at the end of Section 4.2.

We thus define the truncated damage function

$$d_\lambda(T) := \begin{cases} d(T) & d(T) < \lambda \\ \lambda & d(T) \geq \lambda \end{cases}$$

Expected damages, dependent on λ, are now

$$D_\lambda(q) := E_{T|q}[d_\lambda(T) \mid q]. \tag{10}$$

Weitzman argues that, for plausible choices of the probability density function $f_s : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and damage function $d : \mathbb{R}_{\geq 0} \to \mathbb{R}$, the following holds:

Theorem 4.1 (Weitzman's Dismal Theorem)**.** *The expectation of damages truncated at λ satisfies*

$$\lim_{\lambda \to \infty} D_\lambda(q) = \infty.$$

□

This holds whenever the limit in (9) does not exist.

Some have read this as saying 'the expected disutility of anything we do is infinite; cost-benefit analysis means nothing; we must throw up our hands in despair'. However, this is not the right interpretation, and fails to understand the importance of the truncation of damages at catastrophe. In fact, the Dismal Theorem tells us that *any attempt to calculate* the expected damages from climate change depends fundamentally on the bounds you place on maximal possible losses - and on the probability of such a loss. Of course, once one realises that there is always some limit to damages, one realises that this truncation is always the correct way to view expected damages - but only in the case of the Dismal Theorem holding, and (8) failing to hold, does that matter.

4.2 Reformulation and DT_λ terms

Here, we extract what seems to be the key feature of the Dismal Theorem, and re-formulate it.[33] The outcome is that the 'dismal' situation is much more widely applicable than appears in Weitzman (2009d). It also becomes easier to

[32] Formally, for all $R \in \mathbb{R}$ there exists $h_R > 0$ such that $\int_{s=0}^h d(T) f_{T|q}(T) dT > R$ for all $h > h_R$. This is distinct from, for example, cases in which the function in question oscillates, and no limit of any kind can be said to exist.

[33] For the work in this section I have benefited enormously from conversations with Robert Hahn.

start asking questions about implications for marginal damages, and the slope of marginal damages.

Weitzman's Dismal Theorem depends upon a precise relationship between the probability density function of uncertainty in damages on the one hand, and the rate of increase in damages on the other. From the mathematics of his presentation, the important factor is the behaviour of these functions at the very top end. At a temperature change of 19°C, 'the human race would very likely be extinct' according to Stern (2009), but the precise analysis of Weitzman's Dismal Theorem depends on the behaviour of functions at temperature changes of beyond even 190°C. Seen in this context, the use of infinity within the Dismal Theorem seems somewhat a distraction. More important may be the necessity of using some 'catastrophic' damage level λ in order to close the model.

Using Cauchy convergence, one may reformulate '$D_\lambda(q) \to \infty$ as $\lambda \to \infty$' as 'for all $l > 0$ it is the case that $(D_\lambda(q) - D_l(q)) \to \infty$ as $\lambda \to \infty$'. This implies that *there exists no level of damages l such that $D_\lambda(q)$ is well approximated by $D_l(q)$ for all $\lambda > l$*; we cannot truncate damages without thinking carefully about the level of truncation. In particular, in the dismal situation, we cannot simply truncate at some level l which represents the limit of our good understanding of damages.

We thus may re-formulate the Dismal Theorem as follows:

Definition 4.2. *We say that expected damages $D(q)$* satisfy the Dismal Theorem *if, for any level l of damages such that we can understand and well model $d_l(T)$, it is not the case that $D_\lambda(q)$ is well approximated by $D_l(q)$ for all $\lambda > l$.*

We introduce the parameter l as the limit to our good understanding of damages. In particular, $l \leq \lambda$ for any possible 'catastrophic' bound λ. In the dismal situation such a limit must exist. For if there is no such limit, we can state λ precisely; then $d(T) = d_\lambda(T)$ and we may precisely calculate $D(q) = D_\lambda(q)$. Such cases then are, by definition, not dismal. On the other hand, the parameter λ is the conceptual limit to damages – which is very likely to lie well above that which we can reasonably quantify. Thus in general $l \neq \lambda$.

We assume that l is known, whereas λ is unknown. We do not formally model uncertainty in λ, as we have no starting point for how to go about forming a probability density function for it. Instead we emphasise which expressions depend on it and so on a consideration of the level of 'catastrophe'. We also do not formally model uncertainty in $d(T)$ beyond l, although doing so would be an obvious extension.

The existence of this bound $l < \lambda$ does not necessarily imply that we are in the dismal situation. Expected damages still do not satisfy the Dismal Theorem if we believe $D_\lambda(q) \approx D_l(q)$ for all relevant $\lambda > l$. However, if we have insufficient confidence in our ability to bound damages, and if there is sufficient probability that higher level damages will occur, then expected damages do satisfy the Dismal Theorem.

It is useful to have a nomenclature to distinguish between the various damages in play. We define and denote them as follows.

Conventional expected damages are expected damages up to l:

$$\mathrm{CT}(q) := E_{T|q}[d_\lambda(T)\,|\,d_\lambda(T) \leq l, q]P(d_\lambda(T) < l\,|\,q).$$

Total dismal expected damages are expected damages beyond l:

$$\mathrm{DT}_\lambda(q) := E_{T|q}[d_\lambda(T)\,|\,d_\lambda(T) \geq l, q]P(d_\lambda(T) \geq l\,|\,q).$$

Now

$$D_\lambda(q) = \mathrm{CT}(q) + \mathrm{DT}_\lambda(q), \tag{11}$$

i.e. expected damages break down as conventional and dismal expected damages. Moreover, note that

$$\begin{aligned}\mathrm{DT}_\lambda(q) &= E_{T|q}[d(T)\,|\,l \leq d(T) < \lambda, q]P(l \leq d(T) < \lambda\,|\,q)\\ &\quad + \lambda P(d(T) \geq \lambda\,|\,q).\end{aligned}$$

We may thus subdivide dismal expected damages into two types: those up to endgame catastrophe λ, and the expected damages from endgame catastrophe alone. We thus make further definitions.

Intermediate dismal expected damages are the dismal damages before the ultimate limit λ of damages is met:

$$\mathrm{DT}_{1,\lambda}(q) := E_{T|q}[d(T)\,|\,l \leq d(T) \leq \lambda, q]P(l \leq d(T) < \lambda\,|\,q).$$

Endgame catastrophe dismal expected damages are expected damages from catastrophe.

$$\mathrm{DT}_{2,\lambda}(q) := \lambda P(d(T) \geq \lambda\,|\,q).$$

All together,

$$D_\lambda(q) = \mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q) + \mathrm{DT}_{2,\lambda}(q) \tag{12}$$

To calculate expected damages, we must consider all three of ‘conventional expected damages’, ‘intermediate dismal expected damages’ and ‘endgame catastrophe dismal 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.

From this presentation, we see more clearly a simplification in the way we think about the Dismal Theorem. Once we have fixed a value for λ, and calculated $P(d(T) \geq \lambda\,|\,q)$, the behaviour of the remainder of the tail of the PDF of s no longer makes a difference to expected damages. If we know that we cannot neglect the term $\mathrm{DT}_{2,\lambda}(q)$, the probability-weighted expected damages from endgame catastrophe, then it is certainly true that the Dismal Theorem holds (in the language of Definition 4.2). This is a much weaker condition to check than the Dismal Theorem as stated by Weitzman. 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) 25°C – if we believe 'endgame catastrophe' to correspond to a smaller temperature change. However, such a distribution is formally not fat tailed at all. Moreover, for our presentation, the question of whether or not the Dismal Theorem holds can depend on the concentration q in question.[34] Indeed, once one has noted the existence of λ, one realises that it is always with us. Even if climate sensitivity can be modelled with reasonable certainty, there will exist q large enough that the bound λ will have to bite.

The question remains, of which out of $\mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$ is more significant. We do not attempt to answer this question, but we develop its implications for the outcome of the model; see Section 5. It is worthwhile, however, to include here a statement of the dependence of these terms on q and λ. The proof is given in Appendix C.

Proposition 4.3.

1. *$\mathrm{DT}_{2,\lambda}(q)$ is monotone increasing with q, and tends to λ as $q \to \infty$.*
2. *$\mathrm{CT}(q)$ and $\mathrm{DT}_{1,\lambda}(q)$ both increase with q for small q, and both tend to zero as $q \to \infty$.*
3. *If Theorem 4.1 holds (i.e. if $D_\lambda(q) \to \infty$ as $\lambda \to \infty$) then $\mathrm{DT}_{1,\lambda}(q) \to \infty$ as $\lambda \to \infty$.* □

Thus, for large q, 'conventional damages' $\mathrm{CT}(q)$ are decreasing with q; this corresponds to emission concentrations where the overwhelming probability is that damages will fit into the 'dismal' category. Similarly, for very large q, 'dismal damages up to catastrophe' $\mathrm{DT}_{1,\lambda}(q)$ are decreasing with q; this corresponds to emission concentrations where the overwhelming probability is that damages will fit into the 'endgame catastrophe' category. It does not necessarily follow that $\mathrm{CT}(q)$ and $\mathrm{DT}_{1,\lambda}(q)$ posses unique maxima but this seems likely to be the case. We conclude that $\mathrm{DT}_{2,\lambda}(q)$ is the dominating term for large enough q. However, for intermediate values of q, any of the three terms may dominate.

It formally need not be the case that $\lim_{\lambda\to\infty} \mathrm{DT}_{2,\lambda}(q) = \infty$, although it often will. As λ increases, the importance of $\mathrm{DT}_{1,\lambda}(q)$ relative to $\mathrm{DT}_{2,\lambda}(q)$ may either increase or decrease. See Examples C.1 and C.2 for further discussion of limiting behaviour here. Note moreover that we need not be in the limiting case; even if $\mathrm{DT}_{2,\lambda}(q)$ is dominating for absolutely vast values of λ, it is perfectly consistent that $\mathrm{DT}_{1,\lambda}(q)$ or even $\mathrm{CT}(q)$ dominates for the value of λ that we consider legitimate. In particular, the fact that the Dismal Theorem holds does *not* now imply that cost-benefit analysis is best characterised as buying insurance against very rare absolute catastrophe (at least if catastrophe is interpreted as the 'endgame' limit to damages λ). It may well be the term $\mathrm{DT}_{1,\lambda}(q)$ that dominates and that our analysis is driven by the risk of very large damages from

[34] For example, suppose that s is indeed bounded above by 25°C, and that the 95^{th} percentile is 6°C. Suppose that damages reach the limit l of what is well understood at 2°C and their endgame catastrophic limit λ at 6°C. One may calculate that, for CO_2 concentrations of less than 300ppm, there is zero probability that global warming exceeds 2 degrees; the situation is not dismal as the dismal terms are all zero. On the other hand, for a stabilisation concentration of 566ppm, there is a 5% probability that global warming exceeds 6 degrees; if damages λ are sufficiently large, say reducing NPV utility by 50%, then the term $\mathrm{DT}_{2,\lambda}(q)$ alone represents 2.5% of NPV utility; expected damages satisfy the Dismal Theorem.

a temperature change of around, say, 4°C; for the stabilisation concentration 566ppm this temperature lies within the 'most likely range' characterised by Solomon et al. (2007) and so cannot be characterised as 'rare'.

It is worthwhile to say a few more words about the 'endgame catastrophe' parameter λ. Weitzman (2009d) defines it in analogy with a 'virtual statistical life', letting it be the disutility associated with the 'end of civilisation'. In this thesis we do not wish to tie its value in with even that level of specificity. 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. Weitzman (2009d) argues that uncertainty in climate sensitivity 'can outweigh the effects of discounting'. This has led some to conclude that the discount rate ceases to be a relevant factor in 'dismal' analysis. However, the comment refers to the fact that his incorporation of uncertainty in climate sensitivity can make *more* difference to the analysis of climate change than does the choice of a low discount rate; this is the main point of his critique (Weitzman, 2007) of Stern (2006). In fact, just as $d(T)$ is a net present value damage function, so λ is the net present value bound on catastrophe. Thus, *ceteris paribus*, λ will take a smaller value if we have a higher pure rate of time preference.

The level of damages we assign to λ need not be consistent across situations, and may indeed be 'reference dependent'.[35] For example, we might decide now that λ corresponds to the disutility from the death of 95% of the world's population; that from our present viewpoint and in terms of present disutility, we simply cannot distinguish between this and the death of 96% of all people. However, if we were to find ourselves in the situation that 90% of the world's population were already dead, we would indeed distinguish between these cases.[36] Thus, if one were to follow a pathway in which substantial damages were accrued, damage valuations would be time inconsistent. This does not alter the model while the model is a one-period net present value one; it would be an extra aspect one could build into a multi-period model.

Although this model is explicitly designed to fit in with the case of climate change, it could be applicable in other situations. For example, some are concerned that the development of nanotechnology may bring serious risks. One could model the amount of nanotechnology in use with the parameter q; this would bring economic benefits, but also health and environmental risks. The level of the risk associated with a given quantity q of technology might be determined by some sensitivity parameter s reflecting the underlying 'riskiness' of the technology – whose distribution may well be fat-tailed. Other issues with

[35]The model thus has certain features in common with the 'Prospect Theory' of Kahneman and Tversky (1979).

[36]I am grateful to Peter Neary for raising this example with me.

possibly catastrophic outcomes, such as asteroid impacts or nuclear war, could similarly fit into this paradigm.

5 Applying the 'prices versus quantities' model to the dismal situation

We now apply the 'prices versus quantities' model of Section 3.3 to a model with damages which may be dismal, as in Section 4.2. However, first, in Section 5.1, we note the implications of our slight modification of the model of Section 3.3, namely that we work with derivatives of expected damages, rather than assuming that the uncertainty has a certain form. In Section 5.2 we lay out the analysis that will subsequently repeatedly be used. As a benchmark, in Section 5.3 we explore the non-dismal case: expected damages which can be well approximated by $D_l(q)$. We then consider the case where damages are dismal. In Section 5.4 we develop the general intuition by looking at the shape of the functions, given the assumptions of Section 3.1 and 4.2. In Section 5.5 we provide the analysis relating to the general case, and discuss its interpretations.

5.1 Immediate consequences of working with derivatives of expected damages

In the original 'prices versus quantities' paper, Weitzman makes fairly restrictive assumptions on the way in which damages $d(T)$ are affected by uncertainty s (these are laid out in Appendix B). Our slight modification, as laid out in Section 3.3, has been to take the expectation of damages first. This may seem trivial; it certainly has no effect on the model as given in that section. However, we allow for much more general forms of damage uncertainty; an important role is now played by the variance and higher moments of s, as we outline in this section.

First, it is instructive to write the associated integrals out in full. Write $f_s(s)$ for the PDF of climate sensitivity, and recall that temperature change T is given by $\frac{s \log q}{\log 2}$. Then

$$D(\hat{q}) = \int_{s=0}^{\infty} d\left(s\frac{\log \hat{q}}{\log 2}\right) f_s(s)ds \tag{13}$$

$$D'(\hat{q}) = \int_{s=0}^{\infty} \frac{\partial}{\partial q} d\left(s\frac{\log \hat{q}}{\log 2}\right) f_s(s)ds \tag{14}$$

$$D''(\hat{q}) = \int_{s=0}^{\infty} \frac{\partial^2}{\partial q^2} d\left(s\frac{\log \hat{q}}{\log 2}\right) f_s(s)ds. \tag{15}$$

(We may take the derivative inside the integral since s and q are independent.) The nature of the distribution of s means that values and derivatives of $d(T)$ from a wide range of possible values of T are relevant in determining $D(\hat{q}), D'(\hat{q})$ and $D''(\hat{q})$. Local approximations are not adequate to determine either the expected damages, or the expected slope of marginal damages.

To illustrate this point further, briefly suppose that $d(T)$ is globally quadratic in T. It follows, since T is linear in s, that d is quadratic in s; write $\tilde{d}(q,s) = d(T(q,s))$. Now, if we write $s_0 = E_s[s]$, then

$$\tilde{d}(q,s) = \tilde{d}(q,s_0) + \tilde{d}_2(q,s_0)(s-s_0) + \tilde{d}_{22}(q,s_0)(s-s_0)^2.$$

It follows that

$$D(q) = E_s[\tilde{d}(q,s)] = \tilde{d}(q,s_0) + \tilde{d}_{22}(q,s_0)\mathrm{var}(s).$$

The non-linearity of d in s means that the expectation over s of $\tilde{d}(q,s)$ is different from the damages $\tilde{d}(q,s_0)$ at the expectation of s – an example of Jensen's inequality. Now,

$$D''(q) = \tilde{d}_{11}(q,s_0) + \tilde{d}_{2211}(q,s_0)\mathrm{var}(s).$$

One might ask whether the fourth-order derivative $\tilde{d}_{2211}(q,s_0)$ is significant, especially given the initial assumption that $d(T)$ is quadratic in T. However, calculations prove that it is so.[37] The important intuition is that T increases linearly with s – so that, if d is quadratic in T, then d is quadratic in s and $\frac{\partial^2 d}{\partial q^2}$ is *still* quadratic in s.

Thus, the variance of s is an important factor in determining the magnitude of $D''(\hat{q})$ in this case. In particular, if we assume that $b(T)$ is of degree at least 2 in T, and if accept the argument of Roe and Baker (2007) that the variance of s is infinite, then the Dismal Theorem applies, to both $D(q)$ and $D''(q)$.

5.2 Technical results we will need

The following sections often call on the same technical results; we present them all now.

One might wish to work with a damage function $d : \mathbb{R}_{\geq 0} \to \mathbb{R}$ which is not differentiable at isolated points. Recall that the truncated damage function $d_\lambda(T)$ is in general not differentiable at the point where damages reach the catastrophic level λ (see Section 4).[38] If $d(T)$ is twice differentiable away from these isolated points, then the expected damage function $D(q)$ is twice differentiable everywhere. A small increase in q leads to a small increase in the probability of passing a threshold temperature and so the discontinuity is smoothed out.

Suppose then that $d(T)$ is smooth except at $T = h$, a jump marginal disutilities. Write $d'(h^-)$ and $d'(h^+)$ for the limiting derivatives before and after this kink.[39]

[37] $\tilde{d}(q,s)$ being quadratic in s is of course equivalent to $d(T)$ being quadratic in T. If $d(T) = d_1T + d_2T^2$ then $\tilde{d}_{2211}(q,s_0) = 4\frac{1-\log q}{(\log 2)^2 q^2} d_2$. We may calculate the multiple of d_2 and find it to lie between 0.3 and 4.3, for the values of q which we consider relevant. If q is large, of course, then $4\frac{1-\log q}{(\log 2) q^2}$ is negative. If we assume that $\mathrm{var}(s)$ is large then we infer that $D''(\hat{q})$ is negative. This issue is discussed in Section 5.4.

[38] One might alternately wish to model damages that jump when a certain temperature is passed – the intuition here is that runaway climate change is triggered at a particular temperature. The proof we provide for Proposition 5.1 allows for such cases, as it is analytically no more difficult.

[39] It does not matter how we define the values $\frac{\partial}{\partial q} d(T(q,s))|_{T=h}$ and $\frac{\partial^2}{\partial q^2} d(T(q,s))|_{T=h}$ in the integrals (16) and (17) because the integral over a point is zero in any case.

Proposition 5.1. *Assume that d is twice differentiable everywhere except at one point h. Then*

$$D'(q) = \int_{s=0}^{\infty} \frac{\partial}{\partial q} d(T(q,s)) f_s(s) ds \tag{16}$$

and

$$\begin{aligned} D''(q) &= \int_{s=0}^{\infty} \frac{\partial^2}{\partial q^2} d(T(q,s)) f_s(s) ds \\ &\quad + \frac{h}{q \log q} \Big[d'(h^+) - d'(h^-) \Big] \frac{\partial}{\partial q} P(T(q,s) \geq h \,|\, q). \end{aligned} \tag{17}$$

The proof is given in Appendix D; here we discuss the intuition. Away from h, one may simply interchange the order of differentiation and integration, since the parameters q and s are independent. For (16), although the derivative in the integrand is not continuous where $T = h$, the integral is still well-defined and extra terms all cancel out. Meanwhile, for (17), one has a new term. It is the marginal effect of additional quantity on the probability of passing the tipping point at h, and the change in marginal damages one encounters from doing so.

Clearly we will often have to take derivatives with respect to q, of functions of temperature $T(q,s) = \frac{s \log q}{\log 2}$. It is useful to lay out now what the first and second derivatives will look like.

Lemma 5.2. *Suppose that $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ twice differentiable, and that $q \geq 1$. Then*

$$\begin{aligned} \frac{\partial}{\partial q} g(T(q,s)) &= \frac{1}{q \log q} T g'(T) \\ \frac{\partial^2}{\partial q^2} g(T(q,s)) &= \frac{1}{(q \log q)^2} \big[T^2 g''(T) - \log q T g'(T) \big]. \end{aligned}$$

The proof is provided in Appendix D. The power function case provides particularly attractive results:

Corollary 5.3. *If $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a power function with exponent n, and if $q > 1$, then*

$$\frac{\partial}{\partial q} g(T(q,s)) = \frac{n}{q \log q} g(T) \tag{18}$$

$$\frac{\partial^2}{\partial q^2} g(T(q,s)) = \frac{n(n-1-\log q)}{(q \log q)^2} g(T). \tag{19}$$

This follows immediately from Lemma 5.2. Thus, we may compare the partial derivatives with respect to q of a function $g(T)$, with the levels of $g(T)$. Such a comparison enables us to better see the relative effects of the various factors which influence the slope of marginal damages, and so it is desirable in the general case. To facilitate this, we make the following definitions, whose relevance will subsequently become clear:

Definition 5.4. *Suppose $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is twice differentiable, that $g(0) = 0$, that $g'(T) \geq 0$, and that $g'(T) > 0$ for $T > 0$.*

1. *For* $T > 0$, *the* relative slope *is defined to be*

$$c_1^g(T) := \frac{Tg'(T)}{g(T)}.$$

2. *For* $T > 0$, *the* relative convexity *is defined to be*

$$c_2^g(T) := \frac{Tg''(T)}{g'(T)}.$$

The relative convexity, $c_2^g(T) = \frac{Tg''(T)}{g'(T)}$ is reminiscent of the Arrow-Pratt coefficient of relative risk aversion (we do not need a minus sign and have introduced differing terminology to use it in wider contexts); the relative slope, $c_1^g(T) = \frac{Tg'(T)}{g(T)}$, is closely related. The straightforward example is the power function case: if $g(T) = T^n$ with $n > 0$ then both terms are constant: $c_1^g(T) = n$ and $c_2^g(T) = n - 1$. Conversely, if $c_1^g(T) = n$ with $n > 0$ and if $g(0) = 0$, then $g(T) = T^n$ and $c_2^g(T) = n - 1$.

These definitions allow us to re-write

Corollary 5.5. *Suppose that* $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *twice differentiable, and that* $q \geq 1$. *Then, wherever* $g(T)$ *and* $g'(T)$ *are non-zero, we have*

$$\frac{\partial}{\partial q} g(T(q,s)) = \frac{c_1^g(T)}{q \log q} g(T) \tag{20}$$

$$\frac{\partial^2}{\partial q^2} g(T(q,s)) = \frac{c_1^g(T)\,(c_2^g(T) - \log q)}{(q \log q)^2} g(T). \tag{21}$$

Note the similarity between these expressions and those of Corollary 5.3. Indeed, we have seen that in the case of a power function with exponent n, we have $c_1^g(T) = n$ and $c_2^g(T) = n - 1$. Thus this is precisely a generalisation of Corollary 5.3. To understand how these expressions may be understood, we consider the terms defined in Definition 5.4 more thoroughly.

Lemma 5.6. *Suppose that* g *and* h *are both monotone increasing and satisfy* $g(0) = h(0) = 0$, *while* $g'(T), h'(T) \geq 0$. *Additionally assume that* $g'(T), h'(T) > 0$ *for* $T > 0$. *Suppose that* $\phi := g \circ h^{-1}$ *so that* $g = \phi \circ h$. *Then, for* $T > 0$, *the following hold.*

1. $c_1^g(T) \geq (>)\ c_2^h(T)$ *if and only if* $\phi'(h(T)) \geq (>)\ \frac{\phi(h(T))}{h(T)}$.

2. $c_2^g(T) \geq (>)\ c_2^h(T)$ *if and only if* $\phi''(h(T)) \geq (>)\ 0$.

The proof of this lemma is elementary and standard.[40]

[40] For 1, note $\frac{Tg'(T)}{g(T)} = \frac{T\phi'(h(T))h'(T)}{\phi(h(T))} > \frac{Th'(T)}{h(T)} \Leftrightarrow \phi'(h(T)) > \frac{\phi(h(T))}{h(T)}$. For 2 we may see that $\frac{Tg''(T)}{g'(T)} = \frac{T\phi''(h(T))h'(T)}{\phi'(h(T))} + \frac{Th''(T)}{h'(T)}$. Now, $h'(T) \geq 0$ by assumption and $\phi'(h(T)) > 0$ by definition, so that $\frac{Tg''(T)}{g'(T)} \geq \frac{Th''(T)}{h'(T)}$ if and only if $\phi''(h(T)) \geq 0$, i.e. if and only if the transformation from h to g is convex at $h(T)$.

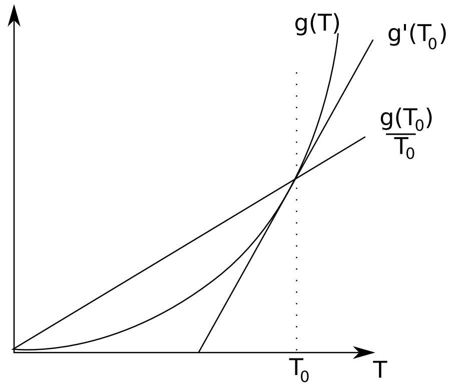

Figure 1: The relative slope $c_1^g(T)$ is the ratio $g'(T_0)$ to $\frac{g(T_0)}{T_0}$.

A function ϕ satisfies $\phi'(x) > \frac{\phi(x)}{x}$ if $\phi''(\tilde{x}) > 0$ for all $\tilde{x} \in [0, x]$, but the condition is more general; it says that ϕ is increasing faster at x than it has done on average between 0 and x. In particular, $c_1^g(T) \geq 1$ if and only if $g'(T) \geq \frac{g(T)}{T}$. It is perhaps clearest to write

$$c_1^g(T) = \frac{g'(T)}{\frac{g(T)}{T}};$$

recalling that $g(0) = 0$, its role as 'relative slope' is now clear. For illustration, see Figure 1. Here, $c_1^g(T)$ evaluated at T_0 is the ratio of slopes of the straight lines $g'(T_0)T$ and $\frac{g(T_0)}{T_0}T$.

If $g'(0) = 0$ then similar observations can be made for $c_2^g(T)$; indeed, we note in that case that $c_2^g(T) = c_1^{g'}(T)$. Thus the term gives the change in slope at T, relative to the overall change in slope. However, we will not always assume that $g'(0) = 0$. In general, it is clear that $c_2^g(T) \geq 0$ if and only if $g''(T) \geq 0$, i.e. if and only if g is convex at the point T. Similarly, $c_2^g(T) \geq 1$ if and only if $g(T)$ is a concave transformation of a quadratic function. Using this second coefficient $c_2^g(T)$ as a measure of the relative convexity of the function $g(T)$ is in any case familiar from the Arrow-Pratt context. In fact, when g is the damage function $d(T)$, Weitzman (2009c) defines $c_2^d(T)$ as the 'coefficient of relative temperature-risk aversion'.[41]

5.3 Benchmark: the non-dismal case

As a benchmark, we consider the non-dismal case: what is the slope of marginal damages in the case that 'catastrophic' damages are negligible, and

[41] Our emphasis here is different from that of Weitzman (2009c), and so we do not use that nomenclature. In a similar vein to classical justifications of the use of CRRA utility, he is seeking to derive the functional form that temperature dependent utility $U(\{X_t\}, T)$ should have, via axioms regarding the coefficient of relative 'consumption-risk' aversion and this new coefficient of 'temperature-risk' aversion. The emphasis is thus on subjective perceptions of temperature risk. Here, we view damages $d(T)$ as being more objectively defined, once consumption risk aversion is given.

what are the implications for 'prices versus quantities'? We continue to assume that climate sensitivity s is highly uncertain, but assume that $D(q) = E_{T|q}[d(T) \,|\, q]$ may be calculated without recourse to some exogenous bound on damages.[42]

The easiest case to consider is when $d(T)$ is a power function with exponent n; then

$$d(T(q,s)) = d\left(\frac{s \log q}{\log 2}\right) = s^n d\left(\frac{\log q}{\log 2}\right)$$

and so

$$D(q) = d\left(\frac{\log q}{\log 2}\right) E_s[s^n]. \tag{22}$$

Expected damages are thus a product of two terms: one a function of q, and the other the n^{th} moment of s. For a right-skewed distribution, the n^{th} moment $E_s[s^n]$ of s is disproportionately large. For such a distribution, there is significant probability density corresponding to higher values of s, and so large values of s, which lead to very large values s^n, are significant in the calculation of the integral $E_s[s^n] = \int_s s^n f_s(s) ds$.

Now, differentiating expected damages with respect to quantities q affects only the first term, the function of q. If expected damages depend on the n^{th} moment of s, then this holds also for expected marginal damages, and for the slope of expected marginal damages; differentiating does not reduce the relevant exponent. We conclude that, in the case where damages are a power function, uncertainty in s affects expected marginal damages, and the slope of expected marginal damages, in the same way and to the same extent as it affects expected damages.

Continuing our assumption that $d(T)$ is a power function, we may apply Corollary 5.3 and write $D'(q)$ in terms of $D(q)$:

$$D'(q) \quad = \quad \frac{n}{q \log q} D(q) \tag{23}$$

and $D''(q)$ in terms of either $D'(q)$ or $D''(q)$:

$$D''(q) \quad = \quad \frac{n - 1 - \log q}{q \log q} D'(q) \tag{24}$$

$$D''(q) \quad = \quad \frac{n(n - 1 - \log q)}{(q \log q)^2} D(q). \tag{25}$$

This repeats the point that the effects of uncertainty in s on $D'(q)$ and $D''(q)$ follow from the effects of uncertainty in s on $D(q)$, and to the same extent. Moreover, we see now that uncertainty in n has an even greater effect on $D'(q)$ and $D''(q)$ than it does on $D(q)$.

[42] To be precise; it remains the case that there will exist some upper bound l to the level of damages we can well understand, and some λ which is the absolute limit to damages. However, as the situation is not 'dismal', we may assume that $D_\lambda(q) \approx D_l(q)$ for all $\lambda > l$ and thus that $D(q) \approx D_l(q)$. Therefore it is unnecessary to truncate the damage function $d(T)$.

q	1	2	3	4	5	6
1.24 (350 ppm)	3.81	7.61	11.42	15.22	19.03	22.83
1.59 (450 ppm)	1.36	2.71	4.07	5.42	6.78	8.14
1.94 (550 ppm)	0.77	1.55	2.32	3.10	3.87	4.65
2.30 (650 ppm)	0.52	1.05	1.57	2.09	2.62	3.14

Table 1: $\frac{n-1-\log q}{q \log q}$ for possible values of n and q.

q	1	2	3	4	5	6
1.24 (350 ppm)	-3.08	22.81	77.65	161.46	274.22	415.95
1.59 (450 ppm)	-0.85	1.97	8.47	18.65	32.51	50.04
1.94 (550 ppm)	-0.40	0.40	2.40	5.60	10.00	15.60
2.30 (650 ppm)	-0.23	0.09	0.96	2.38	4.34	6.86

Table 2: $\frac{n(n-1-\log q)}{(q \log q)^2}$ for possible values of n and q.

Each of the moderating coefficients $\frac{n-1-\log q}{q \log q}$ and $\frac{n(n-1-\log q)}{(q \log q)^2}$ grows with n and declines with q. To see the size of the effect taking place here, these coefficients are calculated in Tables 1 and 2, for possible values of n and q (see Section 3.2 for discussion of these values of q, and Appendix A.1 for the value of n). Naturally, if damages are linear in temperature then expected damages are concave in quantity, due to the logarithmic effect of quantity on temperature. This logarithmic relationship will dominate for any n when q is large enough; the calculations here show that, for the values of q we are interested in, n need only be 2 for the relationship to be convex. Of course, the coefficient in (25) both increases more dramatically with n and decreases more dramatically with q. However, even in the case of the coefficient in (24), the range encompasses over an order of magnitude. A given preconception about the value of $D(q)$ or $D'(q)$ does not mean we know or understand $D''(q)$. The point is that we cannot simply assume that marginal damages are flat in quantity. That may well be the case for the concentrations and polynomial degrees which economic analyses have typically worked with, but as argued in Section 2.2, these calibrations may not be correct.

These moderating coefficients require estimates of $D(q)$ in order to provide estimates of $D''(q)$. Such estimates are made in Appendix A.2, and possible values for $D''(q)$ are given there in Table 8. One should note that, although the moderating coefficient in Table 2 strongly decreases with q, the expected damages $D(q)$ strongly increase with q (see Table 5). Thus it is ambiguous whether $D''(q)$ will increase or decrease with q.[43] We do not formally model uncertainty in the function $d(T)$ here. However, we argue that if it is possible

[43]One may calculate that whether $D''(q)$ increases or decreases with q depends on the sign of $n(n-1)(n-2) - 3n(n-1)\log q + 2n(\log q)^2$. This quadratic in $\log q$ may be checked to posses a real, positive solution whenever $n > 1$. Thus $D''(q)$ will eventually increase with q for large enough q, but there will be and interval of $\mathbb{R}_{\geq 1}$ where $D''(q)$ is decreasing with q.

that n is large then $D''(q)$ is much greater than has typically been modelled. An example of this effect is also provided in Appendix A.2.

We now turn to the case of general damage functions $d(T)$. We start with an informal discussion. By assumption, we are in the non-dismal case; we may approximate $D(q)$ by integrating $d(T)$ against $f_{T|q}(T)$ over some bounded interval $[0, \bar{T}]$. Assume that there exists $n \in \mathbb{N}$ such that $d(T)$ may be approximated on $[0, \bar{T}]$ by a polynomial function

$$d(T) \approx \sum_{i=1}^{n} d_i(T) \text{ for } T \in [0, \bar{T}]$$

where d_i is a monomial of degree i.[44] (We know that there is no constant term because by assumption $d(0) = 0$). Now it follows from (22) that

$$D(q) = \sum_{i=1}^{n} E_s[s^i] d_i \left(\frac{\log q}{\log 2} \right).$$

Thus, the relative importance for $D(q)$ of higher degree terms $d(T)$ is increased, if high moments of s are relatively important, i.e. if s is right skewed. Moreover, from Lemma 5.2, we know

$$\begin{aligned} D'(q) &= \frac{1}{q \log q} \sum_{i=1}^{n} i E_s[s^i] d_i \left(\frac{\log q}{\log 2} \right) \\ D''(q) &= \frac{1}{(q \log q)^2} \sum_{i=1}^{n} i(i - 1 - \log q) E_s[s^i] d_i \left(\frac{\log q}{\log 2} \right) \end{aligned}$$

Thus, the relative importance of high moments of s is further increased when we take derivatives. Now, incorporating uncertainty in s is even more important for $D''(q)$ than it is for $D(q)$. However, the effect of this 'importance' is not clear because we do not know that all the monomial terms in the sum $d(T) = \sum_{i=1}^{n} d_i(T)$ are positive.

For further formalism in this general case, we turn to Corollary 5.5. Recall from there that we may express

$$D'(q) = \frac{1}{q \log q} \int_{T=0}^{\infty} c_d^1(T) d(T) f_{T|q}(T) dT \tag{26}$$

$$D''(q) = \frac{1}{(q \log q)^2} \int_{T=0}^{\infty} c_1^d(T) \left[c_2^d(T) - \log q \right] d(T) f_{T|q}(T) dT \tag{27}$$

where $c_1^d(T)$ and $c_2^d(T)$ are as defined in Definition 5.4. We compare these with

$$D(q) = \int_{T=0}^{\infty} d(T) f_{T|q}(T) dT. \tag{28}$$

[44] This will hold if, for example, $d(T)$ is analytic; then $d(T)$ may be approximated by a sufficiently long truncation of its Taylor series.

q	$\log q$	$\frac{1}{q\log q}$	$\frac{1}{(q\log q)^2}$
1.24 (350 ppm)	0.21	3.81	14.48
1.59 (450 ppm)	0.46	1.36	1.84
1.94 (550 ppm)	0.66	0.77	0.60
2.30 (650 ppm)	0.83	0.52	0.27

Table 3: Calculations of $\log q$, of $\frac{1}{q\log q}$ and of $\frac{1}{(q\log q)^2}$ for the relevant concentrations

We follow the spirit of the power function case, in which we compared $D''(q)$ with $D(q)$ via the table (Table 2) of possible coefficients. The question comes down to an understanding of the modifying factors

$$\frac{1}{q\log q}c_1^d(T) \text{ and } \frac{1}{(q\log q)^2}c_1^d(T)\left[c_2^d(T)-\log q\right].$$

Recall that $c_1^d(T)=\frac{Td'(T)}{d(T)}$ expresses the relative slope of $d(T)$ at T. It is is greater than 1 if the slope of $d(T)$ at T is greater than the average increase of damages from 0 to T, implying that there must be some degree of overall convexity. Meanwhile, $c_2^d(T)=\frac{Td''(T)}{d'(T)}$ expresses relative convexity of $d(T)$ at T. The greater the slope and convexity of the function $d(T)$, the greater both of these terms, and so the greater $D'(q)$ and $D''(q)$.

To understand the scale better, it is useful to have a measure of the size of the terms in q. See Table 3 for calculations at our relevant concentrations. For low eventual concentrations, the initial factor increases the relevance of the term; for high concentrations, it decreases it, but not by more than a factor of at most 4. Meanwhile, the significance of subtracting $\log q$ also depends on the concentration of interest, but $\log q$ is always less than 1. We recall that $c_2^g(T)$ is merely guaranteed to be positive if $g''(T)>0$; for a guarantee that $(c_2^g(T)-\log q)\geq 0$, we need to know that $g(T)$ is a convex transformation of $T^{\log q+1}$. Being a convex transformation of a quadratic function is certainly sufficient, but the overall magnitude of $D''(q)$ will be brought down unless $g(T)$ is more convex still than that. The table (Table 2) of calculations of $\frac{n(n-1-\log q)}{(q\log q)^2}$ illustrates the range of possibility here.

Meanwhile, if $c_d^1(T)$ and $c_d^2(T)$ are increasing with T, this increases the relative importance of high values of $d(T)$ in the integrals (26) and (27), compared with (28). But moreover this increases the importance of the probability density accorded to higher values of T; if the PDF $f_{T|q}(T)$ is highly right-skewed, with a significant tail, then the values of $D'(q)$ and $D''(q)$ will increase significantly. This is the same reasoning as that applied to the importance of higher moments of s at the beginning of this section.

5.4 The sign of slope of marginal damages in dismal case

Now, we return to the dismal case. In this section we discuss the way in which $D'_\lambda(q)$ must vary with q; doing so enables us to better understand expressions we

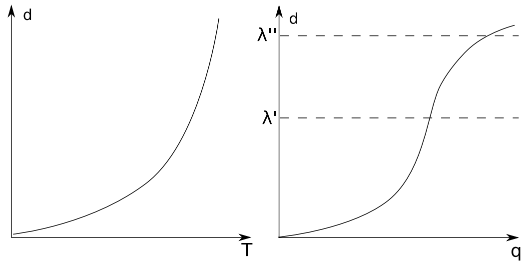

Figure 2: How damages d vary with T and with q

will obtain for $D''_\lambda(q)$. For now, we make no assumptions about the functional forms of damages or probabilities, other than those made in Section 3.1.

We recall that we say damages are 'dismal' if one cannot ignore the terms $\mathrm{DT}_\lambda(q)$ in expression (11) for $D_\lambda(q)$. At least one of the 'dismal' expected damage terms $\mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$ must be significant, and expected damages cannot be calculated without recourse to the exogenous 'endgame catastrophe' bound λ. The close relationship between the Dismal Theorem for expected damages $D(q)$, and for expected marginal damages and the slope of marginal damages $D'(q)$ and $D''(q)$, is illustrated by expressions (23) and (25) derived in Section 5.3. If $D(q)$ is 'infinite' when we do not bound damages, then this seems to also be the case for $D'(q)$ and $D''(q)$, although there are moderating coefficients.

The damage function $d(T)$ has been assumed to be increasing and convex as a function of T. However, it does not necessarily follow that damages $d(T(q,s))$ are convex as a function of q for all q. Recall that equilibrium temperature T is logarithmic in long-term quantities q. For any polynomial function $d(T)$, this logarithmic concavity will eventually 'win'. On the other hand, sufficient convexity of $d(T)$ at $T = 0$ ensures that $d(T(q,s))$ is convex in q for q close enough to 1.[45] The picture is that of Figure 2. For analytic details and a more precise statement of when these observations are valid, see Lemma D.1. Of course, damages $d_\lambda(T)$ are eventually truncated at λ; it is ambiguous whether $d_\lambda(T(q,s))$ is truncated before it has become concave (so $\lambda = \lambda'$) or after (so $\lambda = \lambda''$). The more convex we consider damages $d(T)$ to be, and the lower the quantity q corresponding to endgame catastrophe λ, the less likely it is that such concavity actually manifests itself before the truncation of the function.

We may note a further phenomenon when we obtain expected damages $D_\lambda(q)$ by taking the expectation of $d_\lambda(T(q,s))$ over s. Since damages $d_\lambda(T(q,s))$ are truncated at λ, it follows that expected damages $D_\lambda(q)$ are bounded above by λ. The function $D_\lambda(q)$ thus tends to λ as $q \to \infty$; see Figure 3. This asymptote necessarily implies that the function must possess concavities. Thinking about this another way, marginal damages $\frac{\partial}{\partial q} d_\lambda(T(q,s))$ are zero if q is sufficiently large (how large q need be for this to hold depends on s). Thus expected marginal damages $D'_\lambda(q)$ must decrease with q once q is large enough, and so $D''_\lambda(q) < 0$ for sufficiently large q. See Figure 4(a).

Thus, concavities in the expected damage function $D_\lambda(q)$ arise for larger

[45] Recall that $q \geq 1$ and that $T(1,s) = 0$.

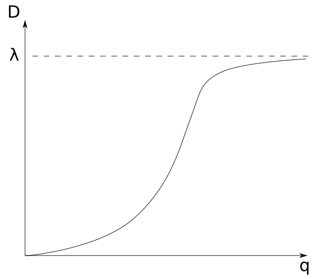

Figure 3: Expected damages D as a function of q.

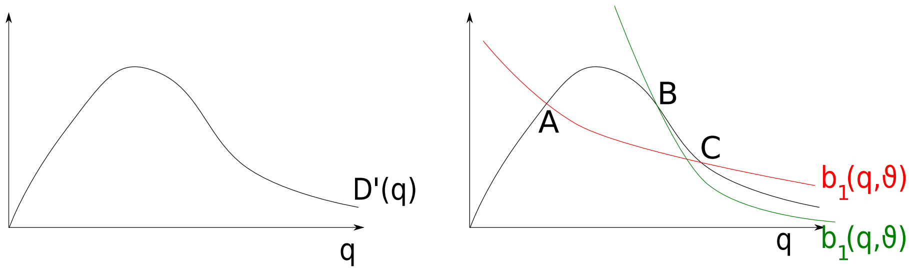

Figure 4:(a),(b). Marginal damages $D'(q)$, with possible marginal benefit functions

values of q, as a result of concavity in the damage response to quantities (an effect that is significant if damages $d(T)$ are not very strongly convex and the quantities in question are large), and as a result of the truncation of damages at the endgame catastrophe level λ (an effect that is significant if expected damages from endgame catastrophe are dominating the analysis).

As to what this analysis has to say about the question of prices versus quantities, we make the following observations. If $D''_\lambda(q) < 0$ then it is certainly clear that (7), the comparative advantage of prices over quantities, is positive. Consider Figure 4(b), where two possible marginal private benefit curves have been drawn. If marginal benefits meet marginal damages as at point **B** in Figure 4(b), then prices do indeed have a comparative advantage over quantities. However, recall the second order condition (5): that $E_\theta[b_{11}(q,\theta)] - D''_\lambda(q) < 0$. If $D''_\lambda(q)$ is sufficiently large and negative that this does not hold, then we cannot possibly have found the global optimal quantity. The situation would be that of point **C** of Figure 4(b); the obvious conclusion is that we should have been aiming for **A** instead.

Thus concavity in $D_\lambda(q)$ may correspond to a strong preference for prices; this is the situation if damages are mild and the optimal concentration is high. On the other hand, concavity in $D_\lambda(q)$ may correspond to the risk of endgame catastrophe being high. Unless we believe in dramatic private benefits from emissions, the quantity which has been chosen seems unlikely to be optimal in such circumstances.

5.5 The dismal case: calculations and interpretations

We now formalise the discussion of Section 5.4. We take all assumptions from Section 3.1, and truncate damages $d(T)$ at λ as in Section 4. We assume that damages satisfy the Dismal Theorem in the sense of Definition 4.2. Thus, as discussed in Section 4.2, we may split up expected damages into 'conventional expected damages' $\mathrm{CT}(q)$ and the two 'dismal expected damage' terms $\mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$; the dismal terms $\mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$ are significant in this sum.

By applying the results already derived, we may show the following.[46]

Theorem 5.7. *Suppose the assumptions of Sections 3.1 hold. We write*

$$D_\lambda(q) = \mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q) + \mathrm{DT}_{2,\lambda}(q)$$

as in Section 4.2. Write $T_\lambda := \sup\{T \mid d(T) < \lambda\}$. *Then*

$$D'_\lambda(q) \quad = \quad \frac{1}{q\log q}\int_{T=0}^{T_\lambda} c_1^d(T)d(T)f_{T|q}(T)dT \tag{29}$$

and

$$\begin{aligned} D''_\lambda(q) \quad = \quad & \frac{1}{(q\log q)^2}\int_{T=0}^{T_\lambda} \left[c_2^d(T) - \log q\right] c_1^d(T)d(T)f_{T|q}(T)dT \\ & - \frac{c_1^d(T_\lambda)}{(q\log q)^2}\frac{T_\lambda f_{T|q}(T_\lambda)}{1-F_{T|q}(T_\lambda)}\mathrm{DT}_{2,\lambda}(q). \end{aligned} \tag{30}$$

The proof is in Appendix D.

As a special case, we may assume all functions in question are power functions. It is easier to see the interpretations in that case, and then look back to the general case and see that they translate across.

Corollary 5.8. *Suppose that damages* $d(T)$ *are power functions with exponent* $n \geq 1$. *Suppose there exist constants* $\alpha > 0$ *and* $m > 0$ *such that the cumulative distribution function* $F_s(s)$ *of* s *satisfies* $F_{T|q}(T) = 1 - \alpha T^{-m}$ *in a neighbourhood of* T_λ. *Then*

$$D'_\lambda(q) = \frac{n}{q\log q}\left[\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)\right]$$

and

$$D''_\lambda(q) = \frac{n}{(q\log q)^2}\left[(n-1-\log q)(\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)) - m\mathrm{DT}_{2,\lambda}(q)\right]. \tag{31}$$

The proof of this corollary is immediate form Theorem 5.7. We now discuss interpretations of these results

[46] Here, $c_1^d(T)$ is defined using $d(T)$ and not the truncated version $d_\lambda(T)$; the term $c_1^d(T_\lambda)$ thus refers to the left-hand side derivative.

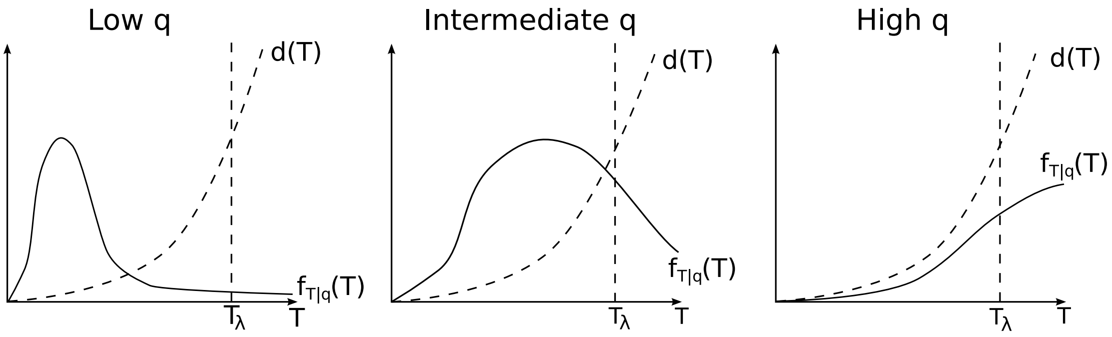

Figure 5: The effect of increasing q on $f_{T|q}(T)$

Marginal damages: Interpretations. We have shown that $D'_\lambda(q)$ is given by

$$\frac{1}{q\log q}\int_{T=0}^{T_\lambda} c_1^d(T)d(T)f_{T|q}(T)dT. \tag{32}$$

Marginal damages only depend on the behaviour of the damage function up to the temperature at which endgame catastrophe takes place. This is clearly sensible; beyond endgame catastrophe there are no marginal damages. The effect is stark in the power function case:

$$D'_\lambda(q) = \frac{n}{q\log q}\big[\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)\big].$$

The second dismal term $\mathrm{DT}_{2,\lambda}(q)$ is irrelevant here: once catastrophe has happened, there is no further damage. Thus the fact that we must curtail damages at endgame catastrophe has a moderating effect on marginal damages – the significance of this depends on the significance of $\mathrm{DT}_{2,\lambda}(q)$. If $\mathrm{DT}_{1,\lambda}(q)$ is significant, then marginal damages are still 'dismal'.

Comparing the integrand of (32), namely $c_1^d(T)d(T)f_{T|q}(T)$, with the integrand of the defining integral for $\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)$, namely $d(T)f_{T|q}(T)$, we see that a damage function with greater relative slope gives rise to greater marginal damages. In particular, an increasing $c_1^d(T)$ puts additional weight on the damage corresponding to higher temperatures; if $f_{T|q}(T)$ accords relative likelihood to such temperatures, then $D'_\lambda(q)$ is correspondingly increased. The convexity of the temperature response to concentrations, embodied in the additional factor $\frac{1}{q\log q}$, moderates the scale of marginal damages, but not too greatly; see Table 3 for calculations of the relevant values.

Consider the effects of increasing q: this changes $\frac{1}{q\log q}$, but also changes $f_{T|q}(T)$. The effect is illustrated in Figure 5, against the backdrop of $d(T)$ and the position of T_λ. As q increases from low values to intermediate ones, the probability of higher values of T in $[0, T_\lambda]$ increase, and so do the relative importance of damages from such temperatures. However, eventually q becomes so large that the probability that $T > T_\lambda$ becomes significant. The probability of achieving a temperature in the range $[0, T_\lambda]$ having diminished, the integral is corresponding decreased. Once one takes into account the additional fact that $\frac{1}{q\log q} \to 0$ as $q \to \infty$, we see that marginal damages certainly tend to

0 as q becomes very large; see also Figure 4(a). Whether this occurs within the temperature range we consider relevant depends on the form of $f_{T|q}$ and on the choice of T_λ. This effect, again, is transparent in the power function case. Recall from Proposition 4.3 that, although $\mathrm{CT}(q)$ and $\mathrm{DT}_{1,\lambda}(q)$ increase with q for small q, they both tend to zero as $q \to \infty$.

Finally, recall from Proposition 4.3 that, if Weitzman's Dismal Theorem (Theorem 4.1) holds, then $\mathrm{DT}_{1,\lambda}(q) \to \infty$ as $\lambda \to \infty$. Recall from Lemma 5.6 that $c_1^d(T) \geq 1$ if $d''(T) \geq 0$, which we assumed to hold on $[0, T_\lambda]$. Thus

$$D'_\lambda(q) \geq \frac{1}{q \log q} \int_{T=0}^{T_\lambda} d(T) f_{T|q}(T) dT = \frac{1}{q \log q} \big[\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)\big]$$

and so $D'_\lambda(q) \to \infty$ as $\lambda \to \infty$. If we cannot sufficiently curtail either the range of possible values of climate sensitivity s, or the level λ of damages that could possibly be accrued, then we similarly cannot curtail marginal damages in this way.

In conclusion, marginal damages are independent of expected damages from 'endgame catastrophe'. They are greater for a damage function with greater or increasing relative slope, and satisfy the same 'dismal' property as expected damages. Although they may in practice be small for very large quantities q, this effect seems not too important for the concentrations we consider relevant.

The slope of marginal damages: interpretations. Now, we turn to $D''_\lambda(q)$. First we consider damages up to endgame catastrophe, namely

$$\frac{1}{(q \log q)^2} \int_{T=0}^{T_\lambda} c_1^d(T) \big[c_2^d(T) - \log q\big] d(T) f_{T|q}(T) dT.$$

This is familiar from Section 5.3; recall from the discussion there that significant convexity in $d(T)$, with $c_1^d(T)$ and $c_2^d(T)$ large, increases $D''_\lambda(q)$; if $c_1^d(T)$ and $c_2^d(T)$ are increasing with T then this increases the relative importance of uncertainty in s. It is possible that $c_2^d(T) - \log q$ is negative; the possibility that $D''_\lambda(q)$ is therefore negative corresponds to the underlying concavity in the temperatures response to emissions. If q is large enough, then this concavity will dominate, as discussed in Section 5.4. However, as illustrated in Table 3, for the concentrations in which we are interested, $\log q$ lies between 0.21 and 0.83. If $d(T)$ is a convex transformation of a quadratic function, it follows that this factor is positive. As shown in Table 4 of Appendix A.1, the net present value damage function is indeed a convex transformation of a quadratic function for all but the most conservative estimates of per-period damages.

As we increase q, we affect $f_{T|q}(T)$ in the way pictured in Figure 5. The integral will tend to increase while q is small, and eventually decrease again when q is large and the probability of endgame catastrophe becomes overwhelming. Meanwhile, the factor $\frac{1}{(q \log q)^2}$ tends to zero as q grows; as calculated in Table 3, it is not too small for the values of q in which we are interested; for large q, we have $D''_\lambda(q) \to 0$, as discussed in Section 5.4. Thus for very large q, the contribution to the slope of marginal damages conventional and intermediate dismal expected damages is small. However, this need not be the case for the values of q and n which we consider reasonable.

In the power function case, these damages are represented via a straightforward multiple of $\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)$, namely $\frac{n(n-1-\log q)}{(q \log q)^2}$. This is familiar from Corollary 5.3, and was discussed in the non-dismal context in Section 5.3; see Table 2 there for calculations of its value for possible n and the range of q in which we are interested. The distinction between the situation here, and that of Section 5.3, is that here we know that $\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q) \to 0$ as $q \to \infty$ (see Proposition 4.3); the fact that $D''_\lambda(q) \to 0$ as $q \to \infty$ is especially clear in this case.

We now consider the final term of (30), namely

$$-\frac{c_1^d(T)}{(q \log q)^2} \frac{T_\lambda f_{T|q}(T_\lambda)}{1 - F_{T|q}(T_\lambda)} \mathrm{DT}_{2,\lambda}(q).$$

This is negative because catastrophic damages are flat and so do not add to marginal damages; hence the greater the importance of endgame catastrophe, the smaller the marginal damages. Catastrophic damages hence reduce the slope of marginal damages.

It is substantially easier to understand the power function case, to which we turn first:

$$-\frac{nm}{(q \log q)^2} \mathrm{DT}_{2,\lambda}(q). \tag{33}$$

Here, n is the exponent of the damage function and $-m$ the exponent of $1-F_s(s)$. Consider a small increase in q. This marginally increases the probability that the catastrophe point is passed, giving rise to the 'm' term in the coefficient. When m is smaller, this effect is less significant. A small value of m corresponds to very slowly declining probabilities. If an increase in quantities makes very little difference to the probability of passing catastrophe, then the effect of catastrophic damages $\mathrm{DT}_{2,\lambda}(q)$ on the slope of marginal damages is reduced. As ever, the larger the quantities q are, the less significant the effect of quantities on temperatures is; this moderates the probability of passing the catastrophe point. This fact is reflected by one of the copies of '$q \log q$' in the denominator.

Given that we pass the catastrophe point, damages cease to increase. Thus marginal damages are reduced and the slope of marginal damages becomes negative. The extent of this effect is governed by how steeply damages were increasing before catastrophe: the 'n' term. The modulation of this effect by the logarithmic response of temperatures to quantities explains the second copy of '$q \log q$' in the denominator.

In the general case, '$c_1^d(T)$' plays the role of 'n' in the explanation above; the effect of passing the point beyond which marginal damages are zero depends on the relative slope of marginal damages at this point. Interpretations of $\frac{T_\lambda f_{T|q}(T_\lambda)}{1-F_{T|q}(T_\lambda)}$ are more delicate. It is helpful to work with $\frac{1}{T}$ rather than T, since it is the tail of the function as $T \to \infty$ which we are considering. Let $x = \frac{1}{T}$ and define $G : \mathbb{R}_{\geq 0} \to \mathbb{R}$ via

$$G(x) := 1 - F_{T|q}\left(\frac{1}{x}\right) \text{ for } x \neq 0 \text{ and } G(0) = 0.$$

Since $\lim_{T\to\infty}(1 - F_{T|q}(T)) = 0$ this gives a function continuous on $[0, \infty)$. Now $G(x) = P\left(T \geq \frac{1}{x} \,|\, q\right) = P\left(\frac{1}{T} \leq x \,|\, q\right)$. In other words, if we consider our random variable to be not T, but $\frac{1}{T}$, then G is simply its cumulative distribution function.

The advantage of this re-writing is that now $G'(x) = f_{T|q}\left(\frac{1}{x}\right)\frac{1}{x^2}$, and thus, if we write $T_\lambda = \frac{1}{x_0}$, it follows that

$$\frac{T_\lambda f_{T|q}(T_\lambda)}{1 - F_{T|q}(T_\lambda)} = \frac{\frac{1}{x_0} f_{T|q}(\frac{1}{x_0})}{1 - F_{T|q}(\frac{1}{x_0})} = \frac{x_0 G'(x_0)}{G(x_0)} = c_1^G(x_0) = c_1^G\left(\frac{1}{T_\lambda}\right).$$

The relative slope $c_1^G\left(\frac{1}{T}\right)$ of G compares the slope $G'(x)$ at x with the overall increment in G between $\tilde{x} = 0$ and $\tilde{x} = x$, namely $\frac{G(x)}{x}$. A large value for the relative slope corresponds to a function $G(x)$ increasing more rapidly at x than it has done on average between 0 and x. A large value of $c_1^G\left(\frac{1}{T}\right)$ thus signifies that quantities significantly affect the probability that we pass the point of endgame catastrophe; a small value signifies that changing the quantity makes little difference here. We return as ever to a power function as our sample case: if $F_{T|q}(T) = 1 - \alpha T^{-m}$ for T in a neighbourhood of λ then $c_1^G(\frac{1}{T_\lambda}) = m$. Now we may present the final term of (30) more attractively, as

$$-\frac{c_1^d(T_\lambda) c_1^G\left(\frac{1}{T_\lambda}\right)}{(q \log q)^2} \mathrm{DT}_{2,\lambda}(q),$$

which is evidently a generalisation of the power function case (33).

The sign of the slope of marginal expected damages is now determined by the relative importance of the positive and negative terms. If $\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)$ is more significant than $\mathrm{DT}_{2,\lambda}(q)$, then $D''_\lambda(q)$ will tend to be positive, but this may be mitigated by the effects of the moderating terms discussed above. In the power function case, we can compare the multiples on $\mathrm{CT}(q) + \mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$ directly. From the presentation in (31) we see that the question of which coefficient is greater is decided by comparing $n - 1 - \log q$ with m. Regarding the discussion in Section 4.1 of functional forms that give rise to unbounded expected damages, we see that for this stricter form of the Dismal Theorem to hold, we must have $n \geq m$. It is thus *ex ante* ambiguous which out of $n - 1 - \log q$ and m is greater. Answering these questions thus requires an understanding of the endgame catastrophe damage level λ, the temperature T_λ at which it takes place, and the way in which the probability density function $f_{T|q}(T)$ declines in a neighbourhood of this temperature, as well as a more detailed understanding of the way in which damages accrue at higher temperatures.

In conclusion, the slope of marginal damages may be either positive or negative in the dismal case. There are factors that tend to make it positive and large: damages having a large and increasing relative slope and relative convexity, and there being a significant probability that temperatures are in the higher parts of the range $[0, T_\lambda]$. A larger eventual concentration q moderates this effect. However, significant expected damages from the endgame catastrophe state tend to turn the situation around, especially if reducing the target concentration has a perceptible impact in reducing the probability of this scenario.

The relative importance of 'dismal' and 'conventional' terms As an example of a damage function which is not a power function, we suppose now that the exponent of T for dismal damages is different from its exponent for conventional damages. That is, if T_l is the temperature change giving rise to damage level l, model

$$d(T) = \begin{cases} \alpha T^{n_1} & T \leq T_l \\ \alpha T_l^{n_1 - n_2} T^{n_2} & T > T_l \end{cases}$$

where α satisfies $\alpha T_l^{n_1} = l$ (and the coefficient of the higher end expression for $d(T)$ has been arranged so that the function is continuous at T_l). Suppose that 'endgame catastrophe expected dismal damages' are negligible. Then Proposition 5.1, Corollary 5.3 (and Lemma D.2 from Appendix D) provide

$$\begin{aligned} D''_\lambda(q) &= \frac{n_1(n_1 - 1 - \log q)}{(q \log q)^2} \mathrm{CT}(q) + \frac{n_2(n_2 - 1 - \log q)}{(q \log q)^2} \mathrm{DT}_{1,\lambda}(q) \\ &\quad + \frac{[n_2 - n_1] c_1^G \left(\frac{1}{T_l} \right)}{(q \log q)^2} l P(T \geq T_l \,|\, q) \qquad (34) \end{aligned}$$

where the function G is as defined earlier in this section.

Thus, the relative importance of conventional and dismal expected damages is determined by the relative sizes of n_1 and n_2. If damages accrue faster beyond the conventional limit, then 'intermediate dismal expected damages' are even more significant in policy choice than in the determination of optimal policy level, and *vice versa*. The difference in the coefficient may be significant for only a small change in exponent. Looking back to Table 2, we see that at a stabilisation concentration of 450 ppm, if $n_1 = 2$ but $n_2 = 3$ then the coefficient jumps from about 2 to over 8. Additionally, $D''_\lambda(q)$ is increased if raising the equilibrium concentration has a large effect on the probability that we exceed the limit l to conventional damages.

Of course, changing n_2 changes $\mathrm{DT}_{1,\lambda}(q)$ too. The point we make here is not just that a higher value of n_2 increases $D''_\lambda(q)$, but that it increases the *relative* importance of $\mathrm{DT}_{1,\lambda}(q)$ compared with $\mathrm{CT}(q)$. Moreover, one may see from the discussion earlier in this section that these observations hold more generally than for just damage functions of this particular form.

The obvious next step would be to introduce uncertainty in n_2. However, we must realise that $\mathrm{DT}_{1,\lambda}(q)$ depends also on n_2. While we fix $\mathrm{DT}_{2,\lambda}(q)$ as negligible, $\mathrm{DT}_{1,\lambda}(q)$ increases with n_2. However, this constraint cannot be held forever. And in general, increasing n_2 will increase $\mathrm{DT}_{2,\lambda}(q)$ at the expense of $\mathrm{DT}_{1,\lambda}(q)$; see Section 5.4. Uncertainty in n_2 will be more formally explored in future work.

6 Conclusion

To conclude: there is very great uncertainty in the damages that will result from a given atmospheric concentration of carbon dioxide. It follows that expected damages are not bounded in the usual way. One way to deal with very large and insufficiently improbable damages is to place an exogenous 'endgame catastrophe' bound on them. This bound 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 right level might be. If such a bound is required, it will also be a key influence on the level of marginal damages, and on the slope of marginal damages. These bounds introduce concavities into the net present value damage functions and so into the expected NPV damage functions. Weitzman (1974) showed that prices being better corresponds to concavities, or to convexities less extreme than the concavity in the private benefit function.

The expected NPV is convex when damages increase steeply with temperature, when 'endgame catastrophe' damages are not too significant in expectation, and when the probability of 'endgame catastrophe' is only very slightly affected by a decrease in equilibrium temperature. 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 temperatures. Concavities in the damage function thus correspond 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. In such circumstances, one should carefully think about 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.

For a more practical policy analysis, one must work with a model of (at least) two periods, incorporating both 'dismal' aspects, and the effects of learning and updating.[47] This will be the subject of future research. The first period will be the 'policy commitment period', the second will be 'thereafter'. Before Period 1, one decides on the optimal total quantity and distributes it between periods. But after Period 1, one updates this allocation. If a price tool had been chosen and a quantity obtained different from the one desired, then the Period 2 optimum shifts. Moreover, the knowledge of the cost of emission cuts in Period 1 will improve estimates of these costs in Period 2. Of course, polluting utilities anticipate this updating and so their incentives are altered. Additionally, the passing of time resolves some of the uncertainty in climate sensitivity.

If policy is updated sufficiently frequently, then the intuition from Section V of Weitzman (1974) becomes relevant. A price tool allows the allocation to each period to be efficiently decided by the market, rather than being determined *ex ante*. This effect is moderated by correlation in the cost of emission cuts across time periods. Issues of updating and learning are important too. However, to provide investment certainty to industry, longer commitment periods may be preferable. If there is flexibility in when the emission reductions take place within each period, the advantage of the price tool is mitigated; a more relevant concern may become certainty in the quantity of emission reductions each period achieves.

Further research will thus show whether cap and trade is always the pre-

[47]Simple multi-period models are developed by Newell and Pizer (2003) and Hoel and Karp (2001, 2002).

ferred instrument. The present work has shown that, when designing very long term strategies, we should guide policy more carefully towards a designated stabilisation concentration than a long-term carbon price.

A Calculations and calibrations

Here we outline and justify the numbers and functions that will be used, and calculate sample values for $D(q)$ and its derivatives.

A.1 Power Law NPV damage functions

As argued in Section 2.2, there is little thorough work on the functional form of per-period damages from global warming, or the level of damages corresponding to higher temperature changes. Accordingly it is not possible to estimate an accurate net present value (NPV) damage function, which would require knowing the damages along a complete pathway of temperature changes. However, to provide clear analytic results in Section 5 we have sometimes assumed that the NPV damage function is a power function:

$$d(T) = \gamma T^n \tag{35}$$

It is useful to see what values of γ and especially n correspond to the per-period damage functions typically in use.

First, we show how to convert from a per-period damage function to a NPV one, using a temperature and gross world product pathway. Let $\{X_t\}$ be the GWP pathway, and let U_t be the time-t utility function. As in Section 3.1, we assume that period-t temperature change T_t affects utility as

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

Let $T = \lim_{t\to\infty} T_t$ and assume that the long-run damage function $d(T)$ acts as

$$U(\{X_t\}, \{T_t\}) = (1 + d(T))U(\{X_t\}),$$

where U is the NPV utility function. Then

$$d(T) = \frac{U(\{X_t\}, \{T_t\})}{U(\{X_t\})} - 1 = \frac{\sum_{t=0}^{\infty} \beta^t (1 + d_t(X_t))U_t(X_t)}{\sum_{t=0}^{\infty} \beta^t U_t(X_t)} - 1, \tag{36}$$

where $(1-\beta)$ is the pure rate of time preference. To estimate the functional form induced by a selection of per-period climate change damages, I used DICE2007 to generate a collection of possible temperature pathways up to the year 2205.[48]

[48]To obtain a range of otherwise consistent temperature pathways, I changed the climate sensitivity parameter and left all other model inputs constant. I am grateful to Alexander Golub and Oleg Lugovoy for helping me with this, and to William Nordhaus for making the model publicly available.

(μ,ν)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
n	0.93	1.42	2.13	3.06	4.16
γ	2.35×10^{-3}	1.41×10^{-3}	6.82×10^{-4}	2.66×10^{-4}	8.69×10^{-5}

(μ,ν)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
n	2.21	2.98	3.95	5.07	6.29
γ	3.90×10^{-4}	1.79×10^{-4}	6.79×10^{-5}	2.26×10^{-5}	6.98×10^{-6}

Table 4: Estimated exponents n of the net present value damage function, given the exponent of the per-period damage function l

I assumed temperatures were constant thereafter.[49] I took the GWP figure for each period from DICE, and assumed that growth continued beyond 2205 at the same rate as from 2195 to 2205. I worked with the elasticity of marginal utility equal to 2, and with pure rate of time preference 0.015 (the standard value for DICE).

I worked with per-period damage functions of the form

$$d_t(T_t) = \begin{cases} \gamma T^{\mu} & T \leq 2 \\ \frac{\gamma}{2^{\nu}} T^{\nu} & T > 2 \end{cases} \tag{37}$$

where γ is chosen such that $d(2)$ is equal to Nordhaus' damage function at $T = 2$. I let μ be 1 or 2 and used a range of values for ν between 1 and 6, and such that $\nu \geq \mu$.[50] Using (36) I thus estimated the NPV damage accruing from my range of long-term temperature changes. I then assumed that damages had form (35) and used a regression to estimate γ and n for each choice of (μ, ν). The results are given in Table 4.[51]

As ν gets large, the exponent n of the NPV damage function increases more than in proportion to ν, the coefficient of the per-period damage function. Aggregating damages over a linear temperature pathway from 0 to T may be understood roughly as integrating

$$\int_{\tilde{T}=0}^{2} \gamma_1 \tilde{T}^{\mu} d\tilde{T} + \int_{\tilde{T}=2}^{T} \gamma_{\tilde{T}} \tilde{T}^{\nu} d\tilde{T}$$

for some constants $\gamma(\tilde{T})$; if T is large we would expect the exponent n in (35) to be greater than ν. However, this effect is moderated by: the low exponent

[49] In the climate model incorporated in DICE, temperatures tend to peak around this period and then come back down. However, Solomon et al. (2009) analysed a range of scenarios in which CO_2 emissions drop to zero at some time between 2000 and 2200 and concluded that 'global average temperatures increase while CO2 is increasing and then remain approximately constant (within 0.5 C) until the end of the millennium.' I have taken this perspective. In any case discounting makes damages beyond 2200 small.

[50] I did not simply work with $\mu = \nu$ throughout because high exponent functions take on very small values for low values of T; I wished to examine the effect of greater damages for greater temperature changes without this unnatural shrinking low temperature damages. I changed the exponent at $T = 2$ because this is commonly referred to as the threshold for dangerous climate change.

[51] The goodness of fit term R^2 was also calculated in each case, and was never less than 0.97.

		(μ, ν) of the per-period damage function					
q	Model for s	(1,2)	(1,4)	(2,2)	(2,4)	(2,5)	(2,6)
1.24	constant	0.0014	0.0003	0.0004	0.0001	0.0000	0.0000
(350ppm)	lognormal	0.0015	0.0004	0.0005	0.0002	0.0001	0.0001
1.59	constant	0.0042	0.0028	0.0021	0.0014	0.0011	0.0009
(450 ppm)	lognormal	0.0044	0.0046	0.0026	0.0036	0.0058	0.0125
1.94	constant	0.0070	0.0083	0.0047	0.0059	0.0069	0.0084
(550 ppm)	lognormal	0.0074	0.0138	0.0058	0.0150	0.0360	0.1199
2.30	constant	0.0097	0.0166	0.0077	0.0143	0.0215	0.0343
(650 ppm)	lognormal	0.0101	0.0274	0.0095	0.0363	0.1125	0.4911

Table 5: Calculated values of NPV damages $D(q)$

for lower temperature damages; the nonlinearity of the temperature pathway (a faster temperature increase earlier tends to increase n relative to ν); the discount rate (a greater discount rate tends to decrease n relative to ν); and the value of ν itself (greater damages from large increases in temperature reduce the effective discount rate). I selected the NPV damage functions generated by per-period damage functions where $(\mu, \nu) = (1, 2)$, $(1, 4)$, $(2, 2)$, $(2, 4)$, $(2, 5)$ and $(2, 6)$ as a fairly representative sample.

A.2 Sample values for $D(q)$ and its derivatives

To calculate $D(q)$ and its derivatives, we require a probability density function for s. To incorporate a significant tail, but avoid dismal results, I used a lognormal distribution with log mean 1.09 and log standard deviation 0.4, as in Golub et al. (2009).[52] The calculations are given in Tables 5, 7 and 8.

Additionally, in Table 5 we contrast expected damages when s is lognormal, with estimated damages when s is taken to be constant at 3.22 (this is the expectation of the PDF for s which we have used). The numbers in Table 5 are for the function $D(q)$, which acts multiplicatively on the utility of consumption, as in Section A.1. Readers may be interested in what these numbers mean in terms of the more intuitive % NPV GWP loss; the converted values are presented in Table 6.

The presentation illustrates the numerical importance of incorporating uncertainty in climate sensitivity; some of the answers differ by more than an order of magnitude. The difference is particularly important at high concentrations (when the risk of higher temperatures under uncertain s is more significant) and for more steeply convex damage functions (under which high temperatures are more damaging). Note again that all the NPV damage functions in use here were generated using the DICE per-period damages function up to 2°C as a starting point.

In Table 7 we record possible values for marginal damages (and return to working only with $D'(q)$ and the lognormal model for s). Unsurprisingly, we

[52] All moments of the lognormal distribution exist, but the moment generating function does not. As I only work with polynomial damage functions, this situation is not 'dismal'.

q	Model for s	(μ,ν) of the per-period damage function (1,2)	(1,4)	(2,2)	(2,4)	(2,5)	(2,6)
1.24	constant	0.14	0.03	0.04	0.01	0.00	0.00
(350 ppm)	lognormal	0.15	0.04	0.05	0.02	0.01	0.01
1.59	constant	0.42	0.28	0.21	0.14	0.11	0.09
(450 ppm)	lognormal	0.44	0.46	0.26	0.36	0.58	1.23
1.94	constant	0.70	0.83	0.47	0.58	0.68	0.83
(550 ppm)	lognormal	0.73	1.36	0.58	1.47	3.48	10.70
2.30	constant	0.96	1.63	0.76	1.41	2.11	3.32
(650 ppm)	lognormal	1.00	2.67	0.94	3.50	10.11	32.94

Table 6: Calculated values of NPV damages as a % of NPV GWP

q	(μ,ν) of the per-period damage function (1,2)	(1,4)	(2,2)	(2,4)	(2,5)	(2,6)
1.24 (350ppm)	0.008	0.005	0.004	0.002	0.002	0.002
1.59 (450ppm)	0.009	0.019	0.008	0.019	0.040	0.107
1.94 (550ppm)	0.008	0.033	0.010	0.046	0.142	0.584
2.30 (650ppm)	0.008	0.044	0.011	0.075	0.299	1.617

Table 7: Calculated values of $D'(q)$

find higher marginal damages result from an assumption that the damage function has a higher degree; this also leads to a greater role being played by the quantity q. One may look ahead to our estimates of $b_1(q,\theta)$, provided in Table 9. Unfortunately there are no estimates of the cost of emission cuts to achieve a stabilisation concentration of 350 ppm, but we see that $b_1(q,\theta)$ is at most 0.011 at 450 ppm. Thus a case for stabilisation at 450ppm is made by all damage functions that incorporate a significant increase in the rate of damage from climate change after 2 degrees of warming.

In Table 8 we finally present the estimates of $D''(q)$. Our estimate of $b''(q)$, given in Section A.3, is that it has order of magnitude at most 0.01. Thus, conservative estimates of the damage function imply that, in the long term, a price policy is better. However, as soon as one lets the per-period damage function be steeper after temperatures have hit 2 degrees, the picture reverses; substantially so, for more pessimistic damage functions.

We may now perform some simple calculations to illustrate the dramatic effect of uncertainty in n on $D''(q)$. Suppose our stabilisation target concentration

q	(μ,ν) of the per-period damage function (1,2)	(1,4)	(2,2)	(2,4)	(2,5)	(2,6)
1.24 (350ppm)	0.006	0.035	0.015	0.026	0.031	0.043
1.59 (450ppm)	0.000	0.041	0.008	0.065	0.196	0.697
1.94 (550ppm)	-0.002	0.035	0.004	0.081	0.374	2.090
2.30 (650ppm)	-0.002	0.028	0.002	0.084	0.507	3.773

Table 8: Calculated values of $D''(q)$

q	State of world θ High	Medium	Low
1.59 (450ppm)	0.011	0.006	0.004
1.94 (550ppm)	0.009	0.005	0.003
2.30 (650ppm)	0.006	0.004	0.002

Table 9: $b_1(q, \theta)$ for possible values of q and θ

is 450 ppm, and that we have been modelling per-period damages as quadratic, but now realise that there is probability 0.9 that $(\mu, \nu) = (2, 2)$ and probability 0.1 that $(\mu, \nu) = (2, 6)$. The impact on expected damages $D(q)$ is perceptible: $D(q)$ increases from 0.0026 to 0.0036. However, the impact on the slope of expected marginal damages is dramatic: $D''(q)$ increases from 0.008 to 0.077. The moderately small probability of a steep damage function increases the slope of marginal damages by an order of magnitude.

As argued in Section 5.3, $D''(q)$ may either increase or decrease with q. We see this here in Table 8; for less steep damage functions $D''(q)$ decreases with q (the effect of concavity of temperature in quantities is dominating), but when $d(T)$ is steep, then $D''(q)$ increases rapidly with q for the range of q in which we are interested.

A.3 The cost of mitigating greenhouse gas emissions

Bole (2009) provides a survey of recent estimates of the carbon price associated with various stabilisation levels. The means and standard deviations found for the carbon price required to stabilise concentrations at around 450, 550 and 650 ppm CO_2-eq are respectively 52.1 (19.2), 48.2 (48.2) and 17.3 (13.5) in \$2008/tCO_2. Unfortunately, she finds only two modellers estimating the carbon price associated with more than two stabilisation levels; additionally, there are only five estimates of the cost of stabilising at concentrations between 450 and 500ppm, and no estimates for lower levels. One study that does calculate a wide range is van Vuuren et al. (2007).[53] From Figure 8 of van Vuuren et al. (2007) we have estimated values for $b_1(q, \theta)$ where the state of the world θ regarding emission cut costs is high, medium or low (represented by the choice of scenario modelled – A1b, B2 or B1 from the IPCC Scenario set); the results are given in Table 9.[54] From this, we may do a basic calculation and use the mean value theorem to obtain a value of $b_{11}(\tilde{q}, \theta)$ that holds for some $\tilde{q}$ between 1.59 and 1.94, and a value for some $\tilde{q}$ between 1.94 and 2.30. These are given in Table 10. These calculations are highly back-of-envelope in nature, but they enable

[53]The other is Popp (2006), but as the costs calculated there are net of damages from climate change, they are not the figures we seek.

[54]The work of van Vuuren et al. (2007) is done in terms of % NPV GDP; we wish to work in fractions of utility. The relationship is $(1 - b(q, \theta))U(X) = U((1 + f(q, \theta))X)$, where $f(q, \theta)$ is the cost function calculated by van Vuuren et al. (2007) (once we have converted from percentages to fractions). Thus $b_1(q, \theta) = \left[-\frac{XU'((1+f(q,\theta))X)}{U(X)}\right] f_1(q, \theta)$. We work with utility with CRRA 2, so if $f(q, \theta)$ is sufficiently small (as it is in all these examples) it follows that $b_1(q, \theta) \approx f_1(q, \theta)$.

$\tilde{q}$	State of world θ High	Medium	Low
$\tilde{q} \in [1.59, 1.94]$	0.007	0.004	0.002
$\tilde{q} \in [1.94, 2.30]$	0.008	0.005	0.002

Table 10: $b_{11}(q, \theta)$ for possible values of q and θ

us to estimate the order of magnitude of $b_{11}(q, \theta)$: if costs are high, it is of the order 0.01; for low costs, it is more of the order 0.001. Although the curvature appears (surprisingly) to increase with q, we are far from having enough data to make such a judgement.

B Derivation of 'prices versus quantities' results

Here we derive the results stated in Section 3.3.

Assume that the second order condition (5) holds, and that we have identified a target quantity $\hat{q}$, via first order condition (4), namely $D'(\hat{q}) = E_\theta[b_1(\hat{q}, \theta)]$. Suppose that, in a neighbourhood $Q \subset \mathbb{R}$ of q, the following second degree Taylor approximations are appropriate:

$$b(q, \theta) \approx b(\hat{q}, \theta) + b_1(\hat{q}, \theta)(q - \hat{q}) + \frac{1}{2} b''(\hat{q})(q - \hat{q})^2. \tag{38}$$

$$D(q) \approx D(\hat{q}) + D'(\hat{q})(q - \hat{q}) + \frac{1}{2} D''(\hat{q})(q - \hat{q})^2. \tag{39}$$

If a price p is fixed, then industry will then pick the quantity $\tilde{q}(p, \theta)$ which satisfies

$$b_1(\tilde{q}(p, \theta), \theta) = p. \tag{40}$$

Now,

Lemma B.1. *Suppose the second order condition (5) is satisfied. Let $\hat{p}$ be the optimal choice of price instrument. Suppose that $\tilde{q}(\hat{p}, \theta) \in Q$ for all θ, so that approximations (38) and (39) may be used. Then $\hat{p} = D'(\hat{q})$.*

Proof. If p satisfies $\tilde{q}(p, \theta) \in Q$ for all θ, it follows from (38) that

$$\tilde{q}(p, \theta) = \hat{q} + \frac{p - b_1(\hat{q}, \theta)}{b''(\hat{q})}. \tag{41}$$

This is linear in prices p, and the derivative $\tilde{q}_1(p, \theta) = \frac{1}{b''(\hat{q})}$ is non-zero and independent of θ. We assume that $\tilde{q}(\hat{p}, \theta) \in Q$ for the optimum price $\hat{p}$, and thus, when we optimise $E_\theta[b(\tilde{q}(p, \theta), \theta) - D(\tilde{q}(p, \theta))]$ with respect to p, the solution $\hat{p}$ satisfies

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

where we have applied the defining equation for $\tilde{q}(p,\theta)$. Now, by assumption, $D'(q)$ is linear in q, so $E_\theta D'(\tilde{q}(\hat{p},\theta)) = D'(E_\theta[\tilde{q}(\hat{p},\theta)])$. We may see from (41) that $E_\theta[\tilde{q}(\hat{p},\theta)] = \hat{q} + \frac{\hat{p}-D'(\hat{q})}{b''(\hat{q})}$. Applying assumption (39), it follows

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

or

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

Thus, $D'(\hat{q}) = \hat{p}$ as long as $b''(\hat{q}) \neq D''(\hat{q})$. However, the latter implies the failure of the second order condition (5), so this provides the required result. □

Corollary B.2. *Given the assumptions of Lemma B.1, it follows that:*

1. *the quantity one expects from a price tool is equal to the quantity that one would have set:*
$$E_\theta[\tilde{q}(\hat{p},\theta)] = \hat{q};$$
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 $E_\theta[b_1(\hat{q},\theta)] = D'(\hat{q})$; the result now follows from Lemma B.1. Now 1. follows from (41) and from $\hat{p} = D'(\hat{q}) = E_\theta[b_1(\hat{q},\theta)]$. □

We now prove Theorem 3.1. Recall that we do not assume that $\hat{q}$ is the true solution to the first order condition (4). Instead, $\hat{q}$ has been fixed externally. If a price policy $\hat{p}$ is used, it satisfies $p_1 = E_\theta[b_1(\hat{q},\theta)]$. By (41), as shown above, it follows that $E_\theta[\tilde{q}(\hat{p},\theta)] = \hat{p}$. As in Section 3.3, we write $\tilde{\tilde{q}}(\theta)$ for $\tilde{q}(\hat{p},\theta)$.

Proof of Theorem 3.1. 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 (41), that

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

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 (38) and (39) into the definition (6) of $\Delta(\hat{q})$ to obtain

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

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

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

We conclude

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

as required. □

C Proofs and further examples for the dismal theorem

Proof of Proposition 4.3. To understand how the damage terms vary with q, we must understand how the probability that the level of damages falls within a certain range varies with q. Intuitively, it is clear that an increase in q reduces the probability of low damages. For clarity, it is useful to write $P(d(T) < l \,|\, q)$ in terms of the cumulative distribution function $F_s(s)$ of s in order to see how the probability varies with q. We let T_l be the temperature increase corresponding to damage level l: namely $T_l := \sup\{T \,|\, d(T) < l\}$; then

$$\begin{aligned} P(d(T) < l \,|\, q) \;=\; P(T < T_l \,|\, q) \;&=\; P\left(\frac{s \log q}{\log 2} < T_l\right) \\ &=\; F_s\left(\frac{T_l \log 2}{\log q}\right). \end{aligned}$$

This clearly is monotone decreasing with q, and tends to 0 as q tends to ∞; unless $l = 0$ we have $P(d(T) < l \,|\, 0) = 0$.

It is straightforward to note now that, since $\mathrm{DT}_{2,\lambda}(q) = \lambda P(d(T) \geq \lambda \,|\, q)$, we have $\mathrm{DT}_{2,\lambda}(q)$ monotone increasing with q and tending to λ as $q \to \infty$. This proves 1.

The behaviour of $\mathrm{CT}(q)$ and $\mathrm{DT}_{1,\lambda}(q)$ with respect to q is more ambiguous. Considering $\mathrm{CT}(q)$ first, note that $\mathrm{CT}(0) = 0$ and that $\mathrm{CT}(q) > 0$ for $q > 0$; conventional damages are increasing with q for small q. Note that $\mathrm{CT}(q) = E_{T|q}[d(T) \,|\, d(T) < l, q] P(d(T) < l \,|\, q)$. The expectation $E_{T|q}[d(T) \,|\, d(T) < l, q]$ is monotone increasing with q, but bounded above by l; the term $P(d(T) < l|q)$ is monotone decreasing with q, and tends to 0 as $q \to \infty$. Thus $\lim_{q\to\infty} \mathrm{CT}(q) = l \cdot 0 = 0$.

The story is similar for $\mathrm{DT}_{1,\lambda}(q)$. First note that $\mathrm{DT}_{1,\lambda}(0) = 0$ and that in general $\mathrm{DT}_{1,\lambda}(q) \geq 0$. Thus $\mathrm{DT}_{1,\lambda}(q)$ is (weakly) increasing for small q. On the other hand, expected damages are all bounded above by λ: by definition

$$\mathrm{DT}_{1,\lambda}(q) = E_{T|q}[d(T) \,|\, l \leq d(T) < \lambda, q] P(l \leq d(T) < \lambda \,|\, q),$$

and the first of these terms is bounded above by λ. The probability that the temperature change is in this region declines for large enough q, namely $P(l \leq d(T) < \lambda|q) \to 0$ as $q \to \infty$. Thus $\mathrm{DT}_{1,\lambda}(q)$ tends to 0 as q tends to infinity. This completes 2.

For 3, we note that $\mathrm{DT}_{1,\lambda}(q)$ is, by definition

$$\mathrm{DT}_{1,\lambda}(q) = \int_{T=0}^{\bar{T}_\lambda} d(T) f_{T|q}(T) dT$$

where $\bar{T}_\lambda = \min\{T|d(T) = \lambda\}$. Thus, $\lim_{\lambda\to\infty} D_\lambda(q) = \lim_{\lambda\to\infty} \mathrm{DT}_{1,\lambda}(q)$; if Weitzman's version of the dismal theorem applies, then $\mathrm{DT}_{1,\lambda}(q)$ tends to infinity with λ. □

Example C.1 ($\mathrm{DT}_{2,\lambda}(q)$ may be bounded as λ grows.)**.** *Let $n > 0$. Suppose that, for T large enough that damages are above level l, the PDF $f_{T|q}(T)$ is equal to αT^{-n-1} for some constant α. Suppose, also, that $d(T) = \beta T^n$ for some constant β. Write T_l for the temperature change giving rise to damages l, and T_λ for the temperature change giving rise to damages λ. Now*

$$\begin{aligned}\mathrm{DT}_{1,\lambda}(q) &= \int_{T=T_l}^{T_\lambda} \alpha T^{-n-1}\beta T^n dT \\ &= \alpha\beta\big[\log T_\lambda - \log T_l\big] \to \infty \text{ as } T_\lambda \to \infty\end{aligned}$$

and moreover note that $T_\lambda \to \infty$ as $\lambda \to \infty$. Thus $D_\lambda(q) \to \infty$ as $\lambda \to \infty$; the 'Dismal Theorem' of Weitzman's formulation holds. However,

$$\mathrm{DT}_{2,\lambda}(q) = \lambda \int_{T_\lambda}^{\infty} \alpha T^{-n-1} dT = \lambda \frac{\alpha}{n} T_\lambda^{-n} = \frac{\alpha\beta}{n}$$

since by definition, $\beta T_\lambda^n = \lambda$. This, we see, is independent of λ (although $D_\lambda(q)$ is not). Thus, in the limit as λ becomes large, $\mathrm{DT}_{1,\lambda}(q)$ dominates.

Example C.2 ($\mathrm{DT}_{1,\lambda}(q)$ and $\mathrm{DT}_{2,\lambda}(q)$ both unbounded as $\lambda \to \infty$; we calculate conditions under which each dominates.)**.** *Let $m > 1$ and $n > m - 1$. Write T_l for the temperature change giving rise to damages l, and T_λ for the temperature change giving rise to damages λ. Suppose that, for $T > T_l$, we have $f_{T|q}(T)\alpha T^{-m}$ for some constant α. Suppose, also, that $d(T) = \beta T^n$ for some constant β. Now*

$$\mathrm{DT}_{1,\lambda}(q) = \int_{T=T_l}^{T_\lambda} \alpha T^{-m}\beta T^n dT = \frac{\alpha\beta}{n-m+1}\big[T_\lambda^{n-m+1} - T_l^{n-m+1}\big].$$

However, noting that $\beta T_\lambda^n = \lambda$, we see

$$\mathrm{DT}_{2,\lambda}(q) = \lambda \int_{T_\lambda}^{\infty} \alpha T^{-m} dT = \frac{\lambda\alpha}{m-1} T_\lambda^{-m+1} = \frac{\alpha\beta}{m-1} T_\lambda^{n-m+1}.$$

As T_λ gets big, the term T_l^{n-m+1} becomes negligible and so which out of $\mathrm{DT}_{1,\lambda}(q)$ *and* $\mathrm{DT}_{2,\lambda}(q)$ *dominate is decided by which is greater out of $\frac{1}{n-m+1}$ and $\frac{1}{m-1}$. Thus, in the limit*

$$\begin{aligned}\frac{\mathrm{DT}_{1,\lambda}(q)}{\mathrm{DT}_{2,\lambda}(q)} \to 0 \text{ as } \lambda \to \infty \;&\Leftrightarrow\; 2(m-1) < n \\ \frac{\mathrm{DT}_{1,\lambda}(q)}{\mathrm{DT}_{2,\lambda}(q)} \to \infty \text{ as } \lambda \to \infty \;&\Leftrightarrow\; 2(m-1) > n.\end{aligned}$$

D Proofs for Section 5 analysis

Proof of Proposition 5.1. For the result as stated, we assume that the function d is continuous at h. However, we may prove the result without this assumption and in fact it is no more trouble to do so. We thus assume that $d : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is twice differentiable everywhere except at point h, at which it is not differentiable and may not even be continuous.

Write

$$s_h(q) = \frac{h \log 2}{\log q};$$

this is the climate sensitivity which gives rise to a temperature increase of precisely h, when the quantity is q.

Define $d^1 : [0, h] \to \mathbb{R}$ by $d^1(T) = d(T)$ for $T < h$ and $d^1(h) = d(h^-)$. Similarly, define $d^2 : [h, \infty] \to \mathbb{R}$ by $d^2(T) = d(T)$ for $T > h$ and $d^2(h) = d(h^+)$. Now

$$D_\lambda(q) = \int_{s=0}^{s_h(q)} d^1(T(q,s)) f_s(s) ds + \int_{s=s_h(q)}^{\infty} d^2(T(q,s)) f_s(s) ds;$$

the set $\{s_h(q)\}$ has measure 0 and so it does not matter how the function is defined there. But we may re-write this as

$$\begin{aligned} D_\lambda(q) &= \int_{s=0}^{s_h(q)} d^1(T(q,s)) f_s(s) ds + \int_{s=s_h(q)}^{\infty} [d^2(T) - d(h^+) + d(h^-)] f_s(s) ds \\ &\quad + \big[d(h^+) - d(h^-)\big] P(s \geq s_h(q)). \end{aligned}$$

Now, if we apply Leibniz rule, we find

$$\begin{aligned} D'_\lambda(q) &= \int_{s=0}^{s_h(q)} \frac{\partial}{\partial q} d^1(T(q,s)) f_s(s) ds + s'_h(q) d^1(h) + \int_{s=0}^{s_h(q)} \frac{\partial}{\partial q} d^2(T(q,s)) f_s(s) ds \\ &\quad - s'_h(q) \big[d^2(h) - d(h^+) + d(h^-)\big] - [d(h^+) - d(h^-)] \frac{\partial}{\partial q} P(T \geq h|q). \end{aligned}$$

The integrals may be written as a single integral, and we gather the remaining terms and simplify, to obtain

$$\begin{aligned} D'_\lambda(q) &= \int_{s=0}^{\infty} \frac{\partial}{\partial q} d(T(q,s)) f_s(s) ds \\ &\quad + \big[d(h^+) - d(h^-)\big] \frac{\partial}{\partial q} P(T(q,s) \geq h|q). \end{aligned} \tag{42}$$

If $d(h^+) = d(h^-)$ this provides the result required.

To prove (17) we apply result (42) to the integral $\int_{s=0}^{\infty} \frac{\partial}{\partial q} d(T(q,s)) f_s(s) ds$ and obtain

$$\begin{aligned} D''_\lambda(q) &= \int_{s=0}^{\infty} \frac{\partial^2}{\partial q^2} d(T(q,s)) f_s(s) ds \\ &\quad + \left[\frac{\partial}{\partial q} d^2(T(q,s))|_{s=s_h(q)} - \frac{\partial}{\partial q} d^1(T(q,s))|_{s=s_h(q)} \right] \frac{\partial}{\partial q} P(T \geq h|q) \\ &\quad + \big[d(h^+) - d(h^-)\big] \frac{\partial^2}{\partial q^2} P(T \geq h|q). \end{aligned}$$

Note, by Lemma 5.2 (proved below), that

$$\begin{aligned} \frac{\partial}{\partial q} d^2(T(q,s))|_{s=s_h(q)} &= d'(h^+) \frac{h}{q \log q} \\ \frac{\partial}{\partial q} d^1(T(q,s))|_{s=s_h(q)} &= d'(h^-) \frac{h}{q \log q}. \end{aligned}$$

Thus we finally present

$$\begin{aligned} D''_\lambda(q) &= \int_{s=0}^{\infty} \frac{\partial^2}{\partial q^2} d(T(q,s)) f_s(s) ds + \frac{h}{q\log q}\left[d'(h^+) - d'(h^-)\right]\frac{\partial}{\partial q}P(T \geq h|q) \\ &\quad + \left[d(h^+) - d(h^-)\right]\frac{\partial^2}{\partial q^2}P(T \geq h|q). \end{aligned}$$

If $d(h^+) = d(h^-)$ then this is the result required. □

Proof of Lemma 5.2. Recall $T = \frac{s\log q}{\log 2}$. Now

$$\frac{\partial}{\partial q} g(T(q,s)) = g'\left(\frac{s\log q}{\log q}\right)\frac{s}{q\log 2} = g'(T)\frac{T}{q\log q}$$

and, differentiating with respect to q again and applying this result,

$$\begin{aligned} \frac{\partial^2}{\partial q^2} g(T(q,s)) &= g''(T)\left(\frac{T}{q\log q}\right)^2 - g'(T)\frac{s}{q^2\log 2} \\ &= \frac{1}{(q\log q)^2}\left[T^2 g''(T) - \log q T g'(T)\right] \end{aligned}$$

as required. □

We now formalise the discussion from Section 5.4 of when the convexity of damages in temperature does and does not dominate the concavity of temperatures in quantities.

Lemma D.1. *Suppose that $d(T)$ is twice continuously differentiable.*

1. *If $d''(0) > \frac{\log 2}{s} d'(0)$, then damages $d(T(q,s))$ are convex in q for small enough q (given $q > 1$).*
2. *If $\frac{d''(T)}{d'(T)} \to 0$ as $T \to \infty$, then damages $d(T(q,s))$ are concave in q for large enough q.*

Note that the hypotheses certainly hold if $d(T)$ is polynomial and satisfies $d'(0) = 0$.

Proof. From Lemma 5.2, and using $T = \frac{s\log q}{\log 2}$, we see:

$$\frac{\partial^2}{\partial q^2} d(T(q,s)) = \frac{T^2}{(q\log q)^2}\left[d''(T) - \frac{\log 2}{s} d'(T)\right].$$

Now, if

$$d''(0) > \frac{\log 2}{s} d'(0)$$

then damages $\frac{\partial^2}{\partial q^2} d(T(q,s))|_{q=1} > 0$, and moreover this will continue to hold in a small enough neighbourhood of $q = 1$; this is sufficient for $d(T(q,s))$ being convex in q for small enough $q > 1$. On the other hand, $\frac{d''(T)}{d'(T)} \to 0$ as $T \to \infty$ is certainty sufficient for $d(T(q,s))$ being concave in q for large enough q. □

To prove Theorem 5.7 we apply Proposition 5.1 and Corollary 5.5. However, as we see from the statement of Proposition 5.1, we also need to be able to differentiate the cumulative distribution function $F_{T|q}(T)$ of temperature with respect to quantity q. This is not just a simple application of Lemma 5.2 because q affects the distribution itself $F_{T|q}$ itself. Instead we must first translate expressions back in terms of s; let $F_s(s)$ be the cumulative distribution function of s.

Lemma D.2. *If $F_{T|q}(T)$ is the cumulative distribution function of T given q, and $f_{T|q}(T)$ is the associated probability density function, then*

$$\frac{\partial}{\partial q}\big[1 - F_{T|q}(T)\big] = f_{T|q}(T)\frac{T}{q\log q}$$

Proof. For any fixed temperature h

$$\begin{aligned} F_{T|q}(h) &= P(T(q,s) \leq h \,|\, q) = P\left(\frac{s\log q}{\log 2} \leq h \,|\, q\right) \\ &= P\left(s \leq \frac{h\log 2}{\log q} \,|\, q\right) = F_s\left(\frac{h\log 2}{\log q}\right). \end{aligned} \tag{43}$$

and so

$$f_{T|q}(h) = f_s\left(\frac{h\log 2}{\log q}\right)\frac{\log 2}{\log q}. \tag{44}$$

and

$$f'_{T|q}(h) = f'_s\left(\frac{h\log 2}{\log q}\right)\left(\frac{\log 2}{\log q}\right)^2. \tag{45}$$

Thus, we differentiate (43) and apply (44) to get

$$\begin{aligned} \frac{\partial}{\partial q}\big[1 - F_{T|q}(h)\big] &= f_s\left(\frac{h\log 2}{\log q}\right)\frac{h\log 2}{q(\log q)^2} \\ &= f_{T|q}(h)\frac{h}{q\log q}, \end{aligned} \tag{46}$$

as required. □

We may now prove our main result.

Proof of Theorem 5.7. Write $s_{q,\lambda}$ for the critical value of climate sensitivity s such that we just attain damages of level λ from a quantity q. Write $T_\lambda = \sup\{T \,|\, d(T) < \lambda\}$.

The case of $D'_\lambda(q)$ is a simple application of Corollary 5.5 and Proposition 5.1; note that

$$\frac{\partial}{\partial q}d(T(q,s)) = \frac{c_1^d(T)}{q\log q}d(T)$$

for climate sensitivity $s < s_{q,\lambda}$ and is zero thereafter, so

$$\begin{aligned} D'_\lambda(q) &= \int_{s=0}^{s_{q,\lambda}} \frac{c_1^d(T)}{q \log q} d(T(q,s)) f_s(s) ds \\ &= \frac{1}{q \log q} \int_{T=0}^{T_\lambda} c_1^d(T) d(T) f_{T|q}(T) dT. \end{aligned}$$

Similarly, we recall from Corollary 5.5 that

$$\frac{\partial^2}{\partial q^2} d(T(q,s)) = \frac{c_1^d(T)(c_2^d(T) - \log q)}{(q \log q)^2} d(T(q,s))$$

for $s < s_{q,\lambda}$; marginal damages are zero thereafter, and so

$$\int_{s=0}^{\infty} \frac{\partial^2}{\partial q^2} d(T(q,s)) f_s(s) ds = \frac{1}{(q \log q)^2} \int_{s=0}^{T_\lambda} c_1^d(T)(c_2^d(T) - \log q) d(T)) f_{T|q}(T) dT.$$

As for the second term, note that $d'(T_\lambda^+) = 0$ and that $hd'(T_\lambda^-) = c_1^d(T_\lambda) d(T_\lambda) = c_1^d(T_\lambda)\lambda$. Finally, we know from Lemma D.2 that

$$\frac{\partial}{\partial q} P(d(T) \geq \lambda | q) \quad = \quad \frac{f_{T|q}(T) T}{q \log q}$$

Multiplying the factors together, we obtain

$$\frac{c_1^d(T_\lambda)}{(q \log q)^2} \frac{f_{T|q}(T_\lambda) T_\lambda}{1 - F_{T|q}(T_\lambda)} \lambda P(D(q) \geq \lambda \,|\, q) = \frac{c_1^d(T_\lambda)}{(q \log q)^2} \frac{f_{T|q}(T_\lambda) T_\lambda}{1 - F_{T|q}(T_\lambda)} \mathrm{DT}_{2,\lambda}(q)$$

as required. □

References

F. Ackerman, E.A. Stanton, C. Hope, S. Alberth, J. Fisher, B. Biewald, and S.E. 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.

Frank Ackerman, Elizabeth A. Stanton, Chris Hope, and Stephane Alberth. Did the Stern Review underestimate US and global climate damages? *Energy Policy*, 37(7):2717 – 2721, 2009.

M.R. Allen and D.J. Frame. Atmosphere: call off the quest. *Science*, 318(5850): 582, 2007.

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

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

T. Bole. Balancing the carbon market: Overview of carbon price estimates. Netherlands Environmental Assessment Agency (PBL), Report no. 500102029, March 2009.

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

S. Dietz. High Impact, Low Probability: An Empirical Analysis of Risk in the Economics of Climate Change. Technical Report Working Paper 10, Centre for Climate Change Economics and Policy, September 2009.

J. Freeman and A.T. Guzman. Sea Walls are Not Enough: Climate Change and US Interests. *Berkeley Program in Law & Economics, Working Paper Series*, page 231, 2009.

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

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.

R.W. Hahn and R.N. Stavins. Why cap-and-trade should (and does) have appeal to politicians. Commentary available on `VoxEU.org`, April 2010.

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

J. Hansen, M. Sato, P. Kharecha, D. Beerling, R. Berner, V. Masson-Delmotte, M. Pagani, M. Raymo, D.L. Royer, and J.C. Zachos. Target atmospheric CO2: Where should humanity aim? *The Open Atmospheric Science Journal*, 2(1):217–231, 2008.

S.V. 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*, 5:23432349, 2009.

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

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

C. Hope. The marginal impact of CO 2 from PAGE2002: An integrated assessment model incorporating the IPCC's five reasons for concern. *Integrated Assessment*, 6(1):19–56, 2006.

A. Indermühle, TF Stocker, F. Joos, H. Fischer, HJ Smith, M. Wahlen, B. Deck, D. Mastroianni, J. Tschumi, T. Blunier, et al. Holocene carbon-cycle dynamics based on CO_2 trapped in ice at Taylor Dome, Antarctica. *Nature*, 398(6723):121–126, 1999. CO_2 data available at http://www.ncdc.noaa.gov/paleo/metadata/noaa-icecore-2419.html.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

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

W. Knorr. Is the airborne fraction of anthropogenic CO2 emissions increasing. *Geophys. Res. Lett*, 36, 2009.

M. Meyer. Risk aversion and the choice between prices and quantities. Research Paper No. 779, Graduate School of Business, Stanford University, 1984.

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

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

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

W.D. Nordhaus. *A question of balance*. Yale University Press, 2008.

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

William D. Nordhaus. An analysis of the dismal theorem. Levine's Working Paper Archive 814577000000000116, UCLA Department of Economics, February 2009.

Robert S. Pindyck. Modeling the impact of warming in climate change economics. Working Paper 15692, National Bureau of Economic Research, January 2010.

D. Popp. Comparison of climate policies in the ENTICE-BR model. *The Energy Journal*, Special issue on Endogenous Technological Change and the Economics of Atmospheric Stabilisation:163–174, 2006.

R. Prinn, S. Paltsev, A. Sokolov, M. Sarofim, J. Reilly, and H. Jacoby. The influence on climate change of differing scenarios for future development analyzed using the MIT integrated global system model. MIT Joint Program on the Science and Policy of Global Change 163, MIT, September 2008.

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

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, 2010.

Shell. Energy scenarios to 2050. Available from http://www.shell.com/scenarios, 2008.

A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, J. M. Melillo, and H. D. Jacoby. Probabilistic forecast for twenty-first-century climate based on uncertainties in emissions (without policy) and climate parameters. *Journal of Climate*, 22(19):5175–5204, 2009.

S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M.M.B. 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.

S. Solomon, G.K. Plattner, R. Knutti, and P. Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences*, 106(6):1704, 2009.

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

N. Stern. Imperfections in the economics of public policy, imperfections in markets, and climate change. *Fondazione Eni Enrico Mattei (FEEM): Sustainability Indicators and Environmental Valuation Working Papers*, 2009.

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

R.S.J. Tol. The economic effects of climate change. *Journal of Economic Perspectives*, 23(2):29–51, Spring 2009.

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

D.P. van Vuuren, M.G.J. den Elzen, P.L. Lucas, B. Eickhout, B.J. Strengers, B. van Ruijven, S. Wonink, and R. van Houdt. Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Climatic Change*, 81(2):119–159, 2007.

M. Weitzman. Reactions to the Nordhaus Critique. Available at `http://www.economics.harvard.edu/faculty/weitzman/papers_weitzman`, March 2009a.

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

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

M.L. Weitzman. Additive Damages, Fat-Tailed Climate Dynamics, and Uncertain Discounting. Economics Discussion Papers, No 2009-26, 2009b.

M.L. Weitzman. What is the “damages function” for global warming - and what difference might it make?, 2009c.

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