A Precautionary Tale of Uncertain Tail Fattening

Martin L. Weitzman

Abstract   Suppose that there is a probability density function for how bad things might get, but that the overall rate at which this probability density function slims down to approach zero in the tail is uncertain. The paper shows how a basic precautionary principle of tail fattening could then apply. The worse is the contemplated damage, the more should a decision maker consider the bad tail to be among the relatively fatter-tailed possibilities. A rough numerical example is applied to the uncertain tail distribution of climate sensitivity.

1 Introduction Via Climate Change

A striking feature of the economics of climate change is that worst-case scenarios are both highly unsure and non-negligible. Deep structural uncertainty about what might conceivably go awry with the planet is coupled with essentially unlimited downside liability on the ultimate extent of possible global damages. This can be a recipe for producing "fat tails" in the extremes of critical probability distributions. There is a race being run in the extreme tail between how rapidly probabilities are declining and how rapidly damages are increasing. Who wins this race, and by how much, depends on how fat (with probability mass) the extreme tails are. It is difficult to judge how fat the tail of really bad climate change might be because it represents events that are very far outside the realm of ordinary experience. Motivated by the example of climate change, this paper proposes a framework for gaining insight into some of the basic issues involved in conceptualizing the tail probabilities of catastrophic events.

The basic idea of this paper is along the following lines. Suppose that there is some probability density function (PDF) for how bad things might get, but that this PDF itself is uncertain. In particular, some aggregate measure of the overall rate of tail slimming of

M. L. Weitzman (✉)
Department of Economics, Harvard University, Cambridge, MA 02138, USA
e-mail: mweitzman@harvard.edu

the relevant PDF is itself a random variable. A Bayesian decision maker is then confronted with a probability distribution over the degree of tail slimming of members of some family of probability distributions. The basic insight of the paper is that the Bayesian aggregated posterior-predictive PDF is then increasingly fat in its extremes (here the extreme "badness" of the possible outcome). The deeper into the tail is the event being analyzed (the worse is the possible outcome), the more relevant are the fatter-tailed probability assessments about how slowly the tail might be slimming down. For the reduced-form aggregated PDF, the posterior-predictive tail gets relatively fatter as it gets longer. A kind of precautionary principle then applies for the probabilities of extreme events, since a decision maker should then effectively act as if the relatively fatter-tailed possibilities are correct. In perhaps overly colorful language, if you know that a situation is bad, then it is more likely to be much worse than you originally thought.

In previous work I made use of an explicit decision-making framework with a particular structure to argue for the presence of a "dismal theorem" capable of making cost-benefit outcomes very sensitive to the possibility of fat-tailed disasters.[1] This paper lacks an explicit decision-making framework and uses a different structure, but the "big picture" message is similar. There can exist plausible circumstances that fatten significantly the tails of catastrophic outcomes—and these fattened tails should influence policy towards greater precaution.

Technically, the mathematical model relies on some neat aggregation properties of linearly-related members of the same family of hazard functions. Each member of such a family differs from each other member by a multiplicative constant representing the relative degree of "slimming down" of the probabilities as the distribution goes deeper into the tail. Some general insights into the process of posterior-predictive tail fattening are derived and interpreted under this "proportional hazards" specification. By further imposing a gamma distribution on families of hazard functions, the paper is able to obtain closed-form analytical solutions.

A numerical example of this approach is applied to empirically derived PDFs of climate sensitivity as a kind of case study. While extremely crude and subject to many possible criticisms, this application illustrates concretely some of the basic ideas of the paper and conveys some rough sense of the magnitude of tail fattening that might be involved in an important application.

2 The Hazard Function as a Tail Slimming Rate

Let X be a random variable quantifying something *bad*, like global warming temperatures. Essentially, X represents the magnitude of any undesirable event. Depending on the context, X might stand for mean planetary surface temperature changes, earthquake energy, tsunami heights, dollars of loss, fatalities, injuries, environmental degradation, or some other measure of something bad. The bigger is X, the worse is the situation. Really large values of X represent really bad outcomes. Huge values of X indicate catastrophes.

This paper is concerned with the behavior of the upper tail of the PDF of X. More specifically, the paper is interested in the rate at which the PDF of X declines towards zero as X increases. In what follows, the parameter θ ($0 < \theta < \infty$) is conceptualized as some aggregate measure of the *overall rate of slimming down* of the extreme tail, yet to be specified. In this section, for the time being, θ is quasi-fixed in the background and merely carried along

---

in the analysis as a "silent parameter." (Later, θ will be treated as a random variable with its own PDF $g(\theta)$.).

Because the primary object of study of this paper is the distribution of the upper tail of the random variable X, it is then natural (and more elegant) to condition all probabilities on X already being in the upper tail. Let τ represent the "beginning" of the upper tail region (for all θ). Of course τ is arbitrary to at least some extent, but such a sharp distinction between the pre-tail lower part of the probability distribution and the post-tail upper part of the probability distribution is a simplifying assumption that is analytically useful for obtaining crisp results. Let the probability that X exceeds x (for given θ) be denoted $P_\theta[X \geq x]$. (The square-bracket notation $P[A]$ stands for the probability of event A, while the square-bracket notation $P[A \mid B]$ stands for the probability of event A conditioned on event B having occurred.) Since the paper is ultimately concerned with an uncertain family of tail PDFs, I simplistically assume that $P_\theta[X \geq \tau]$ is parametrically specified as some value $\overline{q}$ for all θ. Any desired value of $\overline{q}$ ($0 < \overline{q} < 1$) can later be plugged into the model. Throughout the rest of the paper we are dealing with the tail region, so x is any number satisfying $x \geq \tau$ and the paper works with *conditional* tail probability distributions of the form $P_\theta[X \geq x \mid X \geq \tau]$. If for some reason we desired to know it, the *unconditional* probability distribution $P_\theta[X \geq x]$ could be recovered from $P_\theta[X \geq x \mid X \geq \tau]$ by the relationship $P_\theta[X \geq x] = \overline{q}\, P_\theta[X \geq x \mid X \geq \tau]$, where $\overline{q} \equiv P_\theta[X \geq \tau]$ is a quasi-constant of the analysis that is treated as being fixed in the background.

Tail "fatness" and tail "thinness" have mirror-image properties because fatness is the inverse of thinness. They differ only in polarity. For the mathematical purposes of this paper, it is more elegant to use tail thinness as the basic primitive simply because it generates neater aggregation formulas than tail fatness. Otherwise, there is no substantive difference between the two concepts.

To proceed further in this paper requires a formal definition of tail thinness. While several possible definitions exist, a natural measure of tail thinness is what I will call the tail "slimming rate"

$$h_\theta(x) \equiv \lim_{\epsilon \downarrow 0} \left[\frac{P_\theta[x \leq X \leq x + \epsilon \mid X \geq x]}{\epsilon} \right], \tag{1}$$

which quantifies how rapidly the tail is slimming down locally within a neighborhood of $X = x$.

The *tail-conditional exceedance probability distribution* $\overline{P}_\theta(x)$ is defined here as the function

$$\overline{P}_\theta(x) \equiv P_\theta[X \geq x \mid X \geq \tau], \tag{2}$$

which automatically incorporates the normalization $\overline{P}_\theta(\tau) = 1$ that corresponds to the condition of X being in the tail. (In reliability theory, $X - \tau$ would have the interpretation of tail-conditional time to failure and the function $\overline{P}_\theta(x)$ would be called the tail-conditional survival probability distribution.)

Incorporating (2) into (1), writing out explicitly the formula for conditional probability, and then taking the limit, shows that the definition of the "slimming rate" $h_\theta(x)$ given by (1) is equivalent to the well known "hazard function"

$$h_\theta(x) = \frac{p_\theta(x)}{\overline{P}_\theta(x)}, \tag{3}$$

where the tail-conditional PDF corresponding to (2) is $p_\theta(x) = -\overline{P}'_\theta(x)$. [Conditioning or not conditioning on $x \geq \tau$ makes no difference to the value of $h_\theta(x)$ when properly

calculated because the normalization constant $\overline{q}$ cancels from the numerator and denominator of formula (3)].

Throughout this paper I use the terms "hazard function" and "slimming rate" interchangeably.[2] One could just as well start with the hazard function (3) as the primary concept and derive the slimming rate (1) as its basic property, which is the more traditional route.

There are three major advantages to working with the hazard function $h_\theta(x)$ as a measure of the degree of tale slimming. First, it has all of the right intuitive properties and is not inferior in this regard to any other single function as a measure of how rapidly a tail is slimming down. Second, the hazard function is familiar from probability theory, with well-known properties that have been studied intensively and widely applied. Third, as this paper will show, using $h_\theta(x)$ can bring to bear an elegant mathematical structure for rigorously analyzing an important set of issues involving precautionary tail fattening.

What does it signify in the present context when the hazard function (or slimming rate) $h_\theta(x)$ is relatively small or relatively big? When $h_\theta(x)$ is small (indicative of relatively fatter tails) the bad tail is slimming down slowly, which means that if you know a situation far out in the tail is bad, then it is likely much worse than you originally thought. Conversely, when $h_\theta(x)$ is big (indicative of relatively thinner tails) the bad tail is slimming down rapidly, which means that if you know a situation far out in the tail is bad, then it is likely not much worse than you originally thought.

3 A Model of Tail Slimming Uncertainty

From (3), any given hazard function $h_\theta(x)$ implies a unique tail-conditional exceedance probability distribution $\overline{P}_\theta(x)$ via the relationship

$$\overline{P}_\theta(x) = \exp\left(-y_\theta(x)\right), \tag{4}$$

where the *cumulative hazard function* $y_\theta(x)$ is defined as

$$y_\theta(x) \equiv \int_\tau^x h_\theta(z)\,dz. \tag{5}$$

Note that the cumulative hazard function $y_\theta(x)$ is strictly monotone increasing in x provided, as will be assumed, that the hazard function $h_\theta(x)$ is positive. Note also that $y_\theta(\tau) = 0$, which corresponds to $\overline{P}_\theta(\tau) = 1$, and that in order for $\overline{P}_\theta(\infty) = 0$ to hold, it must be true that $y_\theta(\infty) = \infty$.

Because (3) is mathematically identical with (4) and (5), hazard functions and probability distributions are two sides of the same coin. Usually the probability distribution is conceptualized as being given first, while the hazard function is then defined in terms of it by (3). Here the order in which they are introduced is reversed. Because this paper is focused on uncertain tail thinness, it begins with the tail hazard function as a given primitive measure of tail thinness, while the corresponding tail probability distribution is then defined in terms of the tail hazard function by (4), (5).

---

[2] The "hazard function" standardly refers to a situation where the random variable X has the interpretation of "time to failure" and the exceedance probability distribution has the interpretation of "survival probability distribution." Whether the random variable X is interpreted as the magnitude of damages or interpreted as the time to failure is mathematically irrelevant because the two concepts are formally isomorphic.

If the tail hazard function (aka tail slimming rate) were *known* to be $h_\theta(x)$ (i.e., if the value of the parameter θ was *known*), then the problem this paper is trying to address would not exist. The problem here is that the rate of tail slimming is *not* known exactly because observations of low-probability high-impact extreme events are rare and difficult to interpret—or they do not even exist for many situations of interest (in which case they must somehow or other be extrapolated from what *is* known).

One cannot proceed fruitfully when the hazard function specification remains in the general form $h_\theta(x)$. To go further with the analysis, one simply must specify more precisely how θ is supposed to represent tail thinness. It is difficult to obtain crisp neat results without tying θ crisply and neatly to some intuitive measure of the overall rate of tail slimming. Crisp neat results can be obtained if the tail slimming rates (aka tail hazard functions) $h_\theta(x)$ are assumed to be multiplicative shifts of each other, with θ representing the multiplicative shift parameter. This multiplicative-shift specification is now introduced and henceforth adopted throughout the rest of the paper.

The main given primitive in this setup is some "parent" hazard function $h(x)$, which can be *any positive function* satisfying the regularity requirement

$$\int_\tau^\infty h(x)\,dx = \infty. \tag{6}$$

A family of hazard functions is then generated by imposing the specification that each "child" hazard function is of the linear form

$$h_\theta(x) = \theta \times h(x), \tag{7}$$

which is an instance of what is called in the reliability literature the "proportional hazards" assumption.

The positive parameter θ in (7) represents the *relative thinness* of the tail probabilities of a child hazard function *within* the family of all child hazard functions whose parent is $h(x)$. By assuming the multiplicatively-separable form (7), one can obtain fairly crisp and neat results having an intuitive interpretation. Assumption (7) follows a long tradition in economics of analyzing changes via parametric shifts of critical curves or functions. But the "proportional hazards" assumption (7) is an imposed restriction nevertheless, which amounts to assuming a uniform degree of *relative* tail thinness for each child distribution of the same parent hazard function. I note for the low-probability extreme-event situations relevant to this paper that there might typically be considered to be a lot of variability in the shift parameter θ because the appropriate degree of relative tail thinness might typically be considered to be highly uncertain.

The cumulative parent hazard function is

$$y(x) = \int_\tau^x h(z)\,dz. \tag{8}$$

From (5) and (8), specification (7) implies that the cumulative hazard function for given θ is of the form

$$y_\theta(x) = \theta \times y(x). \tag{9}$$

Plugging (9) into (4) gives

$$\overline{P}_\theta(x) = \exp\left(-\theta\, y(x)\right). \tag{10}$$

Before continuing on with developing the abstract model of this section, I want first to give insight into the general approach by way of three specific examples based on three especially simple functional forms for $h(x)$. These three examples illustrate concretely how particular hazard-function families satisfying the proportional hazards condition (7) conjugate with their implied probability-distribution families via (4). Because the probability-distribution families in these examples are more or less familiar, the examples should serve to strengthen intuition about the general approach.

Example 1 The prototype example here is the exponential distribution. This example serves as a template throughout the rest of this paper because, as the paper will show, tail-fattening behavior in the exponential case readily translates into tail-fattening behavior in the more general case (7). For the exponential family the given primitive parent hazard function is

$$h^e(x) = 1. \tag{11}$$

From (7), the child hazard functions in the exponential case are of the form $h_\theta^e(x) = \theta$. The cumulative hazard function (9) here is $y_\theta^e(x) = \theta (x - \tau)$. From (10), the tail-conditional exceedance probability distributions for the exponential family are of the form

$$\overline{P}_\theta^e(x) = \exp(-\theta (x - \tau)). \tag{12}$$

Example 2 The second example is the Pareto (or power) distribution. For the Pareto family the given primitive parent hazard function is

$$h^p(x) = \frac{1}{x}. \tag{13}$$

From (7), the child hazard functions in the Pareto case are of the form $h_\theta^p(x) = \theta/x$. The cumulative hazard function (9) here is $y_\theta^p(x) = \theta (\ln x - \ln \tau)$. From (10), the tail-conditional exceedance probability distributions for the Pareto (or power) family are of the form

$$\overline{P}_\theta^p(x) = \left(\frac{x}{\tau}\right)^{-\theta}. \tag{14}$$

Example 3 An interesting third example begins with the given primitive parent hazard function being of the form

$$h^w(x) = x^{\gamma-1} \tag{15}$$

for some non-negative constant γ. When $\gamma = 1$, then (15) implies $h^w(x) = h^e(x) = 1$. When $\gamma = 0$, then (15) implies $h^w(x) = h^p(x) = 1/x$. This third example is interesting because the family of probability distributions having parent hazard function (15) subsumes as special cases both the exponential family (12) for $\gamma = 1$ (each member of which is a prototype thin-tailed distribution) and the Pareto family (14) for $\gamma = 0$ (each member of which is a prototype fat-tailed distribution). From (7), the child hazard functions here are of the form $h_\theta^w(x) = \theta x^{\gamma-1}$. The cumulative hazard function (9) here is $y_\theta^w(x) = \theta (x^\gamma - \tau^\gamma)/\gamma$ for $\gamma > 0$ (and $y_\theta^w(x) = \theta (\ln x - \ln \tau)$ for $\gamma = 0$). From (10), the tail-conditional exceedance probability distributions for this family are of the form

$$\overline{P}_\theta^w(x) = \exp\left(-\theta \left(\frac{x^\gamma - \tau^\gamma}{\gamma}\right)\right), \tag{16}$$

which is describing a family of probability distributions having similar form and properties to a corresponding family of Weibull distributions.

4 Insights and Implications

I return now to the broader situation where $h(x)$ represents *any* parent hazard function (not just the special examples of last section where $h^e(x) = 1$ or $h^p(x) = 1/x$ or $h^w(x) = x^{\gamma-1}$).

In the last section, θ was treated as given and fixed. Throughout the rest of this paper θ is viewed as a random variable having prior PDF $g(\theta)$.

Making use of (4) and (9), the *posterior-predictive* tail-conditional exceedance probability distribution, denoted $\widehat{P}(x)$, is then

$$\widehat{P}(x) \equiv \int_0^\infty \overline{P}_\theta(x)\, g(\theta)\, d\theta = \int_0^\infty \exp(-\theta\, y(x))\, g(\theta)\, d\theta. \tag{17}$$

From differentiating (17) and (8) with respect to x, the corresponding posterior-predictive hazard function (aka posterior-predictive tail slimming rate) is

$$\widehat{h}(x) \equiv \frac{-\widehat{P}'(x)}{\widehat{P}(x)} = \left[\frac{\int_0^\infty \theta\, \exp\left(-\theta\, y(x)\right)\, g(\theta)\, d\theta}{\int_0^\infty \exp\left(-\theta\, y(x)\right)\, g(\theta)\, d\theta} \right] \times h(x). \tag{18}$$

A more elegant way of seeing what (18) is trying to tell us comes from decomposing it into three simpler parts. Viewed this way, (18) is equivalent to

$$\widehat{h}(x) = \widehat{\theta}(x) \times h(x), \tag{19}$$

where

$$\widehat{\theta}(x) = \psi(y(x)), \tag{20}$$

and the function $\psi(y)$ is defined by the equation

$$\psi(y) = \frac{\int_0^\infty \theta\, \exp\left(-\theta\, y\right)\, g(\theta)\, d\theta}{\int_0^\infty \exp\left(-\theta\, y\right)\, g(\theta)\, d\theta}. \tag{21}$$

In analyzing the above trio of Eqs. (19), (20), (21), notice that (19) is of the identical form as the child hazard function (7) (with both sharing the same parent $h(x)$), except that $\widehat{\theta}(x)$ in (19) replaces θ in (7). This signifies that, for any given x, we are allowed the mental convenience of conceptualizing the appropriate posterior-predictive rate of tail slimming *as if* it were the child hazard function having the relative thinness parameter value $\widehat{\theta}(x)$. The next issue to be investigated is the behavior of $\widehat{\theta}(x)$ and what it depends upon.

From (20), $\widehat{\theta}(x)$ depends on x in the particular nested form $\widehat{\theta}(x) = \psi(y(x))$. From (8), the function $y(x)$ is strictly monotone increasing in x with derivative $y'(x) = h(x) > 0$. Therefore, if we can understand the basic properties of $\psi(y)$ as y increases, then we will understand (up to a monotone transformation) the basic properties of $\widehat{\theta}(x)$ as x increases.

To see in sharp relief the behavior of $\psi(y)$ as a function of y, rewrite (21) as the weighted average

$$\psi(y) = \int_0^\infty \theta\, \omega(\theta; y)\, d\theta, \tag{22}$$

where the non-negative weights are

$$\omega(\theta; y) \equiv \frac{\exp\left(-y\theta\right)\, g(\theta)}{\int_0^\infty \exp\left(-y\theta\right)\, g(\theta)}, \tag{23}$$

which sum to one because $\int_0^\infty \omega(\theta; y)\, d\theta = 1$.

The behavior of $\psi(y)$ as a function of y can then be understood by examining the behavior of the aggregation weights $\omega(\theta; y)$ as joint functions of y and θ. It is readily apparent that, as y increases for given θ, the exponential term $\exp(-y\theta)$ in (23) places less ω-weight on relatively high values of θ and places more ω-weight on relatively low values of θ. Therefore, from (22), $\psi(y)$ is a declining function of y, and consequently $\widehat{\theta}(x) = \psi(y(x))$ is also a declining function of x, beginning for $y = 0$ or $x = \tau$ at the mean value of θ

$$\int_0^\infty \theta \, g(\theta) \, d\theta = \psi(0) = \widehat{\theta}(\tau), \tag{24}$$

and approaching asymptotically as $y \to \infty$ or $x \to \infty$

$$\inf \{\theta \mid g(\theta) > 0\} = \psi(\infty) = \widehat{\theta}(\infty). \tag{25}$$

Further insight into the behavior of $\psi(y)$ can be gained by differentiating (22), (23) with respect to y, which, after rearranging terms, gives

$$\psi'(y) = -\int_0^\infty (\theta - \psi(y))^2 \, \omega(\theta; y) \, d\theta. \tag{26}$$

Note from (26) that the rate at which $\psi(y)$ declines is proportional to the weighted variance of the distribution of θ. Other things being equal, a weighted-mean-preserving spread of θ accelerates the decline of $\psi(y)$ with respect to y.

The reason why $\psi(y)$ decreases in y, or why $\widehat{\theta}(x) = \psi(y(x))$ decreases in x, runs deep in the nature of the underlying problem. Other things being equal, tails that slim down at a relatively faster rate also have tail exceedance probabilities that are relatively smaller. As one moves further out in the tail, therefore, the posterior-predictive distribution is increasingly dominated by the relatively higher survival probabilities of the relatively fatter sub-distributions corresponding to lower values of θ. In a sense, the thinner-tailed child distributions are discounting their own relevance out of existence, leaving the field to their fatter-tailed siblings. A corollary is that the rate of slimming down of the posterior-predictive distribution decreases further out in the tail. The posterior-predictive tail becomes fatter as it gets longer, reflecting the higher survival rates of its fatter-tailed sub-distributions. Equations like (22) and (26) are merely formulas that express these basic ideas concisely.

The results of this section indicate that as x increases by going deeper into the tail, then, for any given parent hazard function $h(x)$, the posterior-predictive hazard function $\widehat{h}(x)$ declines monotonically relative to $h(x)$ from the "average" child hazard function corresponding to (24) towards the fattest possible child hazard function corresponding to (25). In this sense the posterior-predictive tail gets ever fatter for ever larger x within the class of child hazard functions derived from any given parent hazard function. To say more precisely than this what exactly *is* the posterior-predictive probability distribution requires further restricting the PDF $g(\theta)$ to some analytically tractable form.

5 Gamma Tail Fattening

The gamma probability distribution is ideally suited for further analysis here because of the neat way it integrates to simplify various complicated formulas involving proportional hazard functions. Henceforth I assume the PDF $g(\theta)$ is of the gamma form

$$g(\theta) = \frac{1}{b^a \, \Gamma(a)} \, \theta^{a-1} e^{-\theta/b} \qquad (27)$$

with positive parameters a and b. I presume that the reader has (or can acquire) a basic knowledge of the gamma PDF. In this and the next section, repeated use is made of the formula

$$\int_0^\infty z^{\alpha-1} e^{-z/\beta} \, dz = \beta^a \, \Gamma(\alpha) \qquad (28)$$

for various values of $\alpha > 0$ and $\beta > 0$.

Plug (27) into (17). Then use the formula (28) for $\alpha = a$ and $\beta = 1/[1/b + y(x)]$. Rearrange terms to derive the closed-form expression for the posterior-predictive tail-conditional exceedance probability distribution

$$\widehat{P}_\Gamma(x) = [1 + b \, y(x)]^{-a}, \qquad (29)$$

where $y(x)$ is the cumulative parent hazard function (8).

It is striking to compare Eq. (29) with Eq. (10). The child probability distribution (10) is *exponential* in $y(x)$. The posterior-predictive distribution (29) is *polynomial* in $y(x)$. The effect of imposing a gamma PDF (27) is to fatten the tail of the posterior-predictive distribution by moving $y(x)$ from an exponential distribution class into a polynomial distribution class. For the exponential parent (11), (12), the effect is to change the exceedance probability from the form $\overline{P}_\theta^e(x) = \exp(-\theta \, (x - \tau))$ to the fatter-tailed form $\widehat{P}_\Gamma^e(x) = [1 + b(x - \tau)]^{-a}$. For the Pareto parent (13), (14), the effect is to change the exceedance probability from the form $\overline{P}_\theta^p(x) = (x/\tau)^{-\theta}$ to the fatter-tailed form $\widehat{P}_\Gamma^e(x) = [1 + b(\ln x - \ln \tau)]^{-a}$.

Further insight can be obtained by noting that the mean of the gamma PDF (27) is $\mu = ab$, while the variance of the gamma PDF (27) is $\sigma^2 = ab^2$. Inverting these expressions, (29) can be rewritten as

$$\widehat{P}_\Gamma(x) = [1 + y(x) \, \sigma^2/\mu]^{-\mu^2/\sigma^2}. \qquad (30)$$

It can readily be shown that expression (30) increases with σ. The greater the variance of θ for a given mean, the greater is the posterior-predictive exceedance probability. In the limit $\sigma \to 0$, the expression (30) becomes $\exp(-\mu \, y(x))$, which is of the same form as (10). With no variance, the posterior-predictive exceedance probability reverts to the unfattened child distribution having the point value $\theta = \mu$.

6 A Numerical Application to Climate Sensitivity

There are so many sources of uncertainty in climate change that a person almost does not know where or how to begin cataloging them. In this Sect. 1 make an extremely crude attempt to apply the model of this paper to a particular numerical example concerning one particular aspect of climate-change uncertainty. For specificity, I focus here on the uncertainty of so-called "equilibrium climate sensitivity."

"Equilibrium climate sensitivity" (hereafter denoted X to conform with previous notation) is a key macro-indicator of the eventual temperature response to greenhouse gas changes. It is defined as the global average surface warming that follows a sustained doubling of atmospheric carbon dioxide (CO_2), after the climate system has reached a new equilibrium.[3]

---

[3] In scientific jargon, X is a so-called "fast equilibrium" concept based upon "fast feedbacks" (geologically speaking). The concept omits slower-acting feedbacks, such as changes in land albedo, changes in biological

Calculating the actual time trajectory of temperatures is a complicated task that requires sophisticated computer modeling based on general circulation models with hundreds of parameters and variables. The human mind being what it is, however, there is a compelling need to reduce such a complicated dynamic reality to some comprehensible aggregate indicator. This is a simplistic reduction that overlooks important temporal and spatial aspects of climate change. Temporally, really high atmospheric warming would take a very long time to equilibrate because the oceans must first absorb tremendous amounts of heat (which itself might be considered a scary proposition). Spatially, regional climate effects are far more unpredictable than global average warming. Despite many complications, the concept of equilibrium climate sensitivity can still serve as a useful aggregate proxy for the overall severity of the climate change problem.

The economics of climate change consists of a very long chain of sometimes tenuous inferences fraught with big uncertainties in every link. It should be understood clearly that under the rubric of "equilibrium climate sensitivity" I am trying to aggregate together a large suite of uncertainties. Empirically, it is not the fatness of the tail of the climate sensitivity PDF *alone*, or the reactivity of damages to high temperatures *alone*, or the degree of relative risk aversion *alone*, or the rate of pure time preference *alone*, or any other factor *alone*, that counts, but rather the *combination* of all such factors in determining the upper-tail fatness of the PDF of the relevant measure of overall expected welfare. So climate sensitivity as a random variable X is to be understood here as a prototype example or a metaphor, which is being used primarily to illustrate much more generic issues in the economics of highly uncertain extreme climate change. The insights and results of this paper are not intended to stand or fall on the narrow issue of accurately modeling uncertain climate sensitivity per se.

The Intergovernmental Panel on Climate Change in its IPCC-AR4 (2007) Executive Summary explains climate sensitivity this way: "The equilibrium climate sensitivity is a measure of the climate system response to sustained radiative forcing. It is not a projection but is defined as the global average surface warming following a doubling of carbon dioxide concentrations. It is *likely* to be in the range 2–4.5 °C with a best estimate of 3 °C, and is *very unlikely* to be less than 1.5 °C. Values substantially higher than 4.5 °C cannot be excluded, but agreement of models with observations is not as good for those values." Using the IPCC definition of "likely" a fair interpretation might be that $P[X \geq 4.5\,°C] \approx 17\,\%$.[4] Overall, the IPCC statement might be construed as saying that the upper tail of the PDF of X has a disturbingly large amount of total probability mass (and that the IPCC also "thinks" that, whatever is the total probability mass in the tail, there is a disturbingly large amount of uncertainty about how this probability mass is actually *distributed* throughout the tail region $X \geq 4.5\,°C$).

For further background and motivation on the unsure degree of fatness of the upper-tail PDF of X, along with some data, this paper relies on a recent study that conducted personal interviews with 14 leading climate scientists, using formal methods of expert PDF elicitation (Zickfeld et al. 2010). The 14 experts were listed by name in the study, but the association of numbers with names was not published in order to protect the anonymity of respondents.

---

sinks or sources, temperature-induced releases of carbon, and the like. So-called "earth system sensitivity" includes slower-acting feedbacks and is presumably larger, perhaps significantly so. For a time horizon on the scale of a century or so, it is not implausible that "earth system sensitivity" might be the more relevant concept. Greater details are available, e.g., in Hansen et al. (2008). See also the useful survey article of Knutti and Hegerl (2008).

[4] The study of Zickfeld et al. (2010), for example, reads the IPCC definition as signifying $P[X \geq 4.5\,°C] = 17\,\%$.

Table 1 Expert estimates of tail probabilities for climate sensitivity (in °C)

$i =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$s_i^{25\%}$	4.25	5.1	4	3.6	4.5	4.9	4.2	3.8	4	4.1	5	4.5	3.9	5
$s_i^{5\%}$	6	6.1	8	4.4	6.5	7	5.5	5	5.5	5.4	7	8	4.5	7.2
$L_i^{1/5}$	1.75	1	4	.8	2	2.1	1.3	1.2	1.5	1.3	2	3.5	.6	2.2

There are so many serious problems and difficulties with the way I am using numbers from this study that, in the interest of brevity, I simply move along without attempting to justify each step in detail. The only possible overall justification of my quick and dirty approach is that the underlying issue is important and superior data for the purposes of this paper simply do not exist. What I am presenting here, then, should merely be seen as a suggestive numerical example that illustrates concretely some of the basic ideas of this paper and conveys, however crudely, some rough sense of the magnitude of tail fattening that might be involved in an important application. I claim nothing more than this.

In the part of the (Zickfeld et al. 2010) study most relevant to this paper, the 14 scientific experts were asked to enumerate their 25 and 5 % upper confidence levels for high values of equilibrium climate sensitivity, denoted here, respectively, by $s^{25\%}$ and $s^{5\%}$. In the notation of this paper, expert i (for $i = 1, 2, \ldots, 14$) believes that $P[X \geq s_i^{25\%}] = .25$ and $P[X \geq s_i^{5\%}] = .05$. Results of this part of the (Zickfeld et al. 2010) survey of expert opinions are reproduced in Table 1 above.[5] The last row of Table 1 gives what might be called empirical estimates of the "tail one-fifth-life" in degrees centigrade, $L_i^{1/5} \equiv s_i^{5\%} - s_i^{25\%}$.

From now on I concentrate just on the tail one-fifth-life estimates $L_i^{1/5}$ $(= s_i^{5\%} - s_i^{25\%})$ in the last row of Table 1. This is my primitive data. (I do not know how representative is this sample, but will later simply presume it is a representative sample of all scientific opinions.) I next attempt to force the $\{L_i^{1/5}\}$ numbers of Table 1 into the framework model of this paper. I arbitrarily assume that the tail begins at $\tau = 4.5\,°C$ and condition all further tail probabilities on being in this tail region. In the interest of plowing ahead, I make an audacious assumption that the one-fifth-life estimate $L_i^{1/5}$ *applies uniformly all throughout the tail region* $X \geq 4.5\,°C$. This amounts to postulating a thin-tailed exponential PDF, in the spirit of Example 1, throughout the entire tail region. Thus, I am effectively presuming that

$$P_i[X \geq x \mid X \geq \tau] = \exp(-\theta_i\,(x - \tau)). \tag{31}$$

The next question to be addressed is: where do the $\{\theta_i\}$ come from? I assume that the $\{\theta_i\}$ are a random sample of draws from the gamma PDF (10), which represents the population distribution of scientific thought concerning θ.[6] We do not observe realizations of θ directly. The primitive data are considered to be the realizations of $L_i^{1/5}$ from the last row of Table 1, including their assumed-independent additive observation errors. The random variable $L^{1/5}$ is related to the random variable θ by the relationship

$$L^{1/5} = \frac{\ln 5}{\theta}. \tag{32}$$

---

[5] I am indebted to Karen Zickfeld for providing the spreadsheet upon which the numbers in my Table 1 are based. These numbers were not given explicitly in the published paper (Zickfeld et al. 2010).

[6] This is not the only way of representing or interpreting expert opinion. For a discussion of some of the relevant issues, see Clemen and Winkler (1999).

Table 2 A comparison of population and sample quartiles for $L^{1/5}$

Quartile of $L^{1/5}$	25 %	50 %	75 %
Population	$L^{1/5} = 1.18$	$L^{1/5} = 1.57$	$L^{1/5} = 2.14$
Sample	$1.0 \le L^{1/5} \le 1.2$	$1.5 \le L^{1/5} \le 1.75$	$2.1 \le L^{1/5} \le 2.2$

Making use of (28), calculations with the gamma PDF (27) applied to (32) for $\alpha = a - 1$ and $\beta = b$ yields

$$E\left[L^{1/5}\right] = \ln 5 \, E\left[\frac{1}{\theta}\right] = \frac{\ln 5}{(a-1)\,b} \tag{33}$$

while for $\alpha = a - 2$ and $\beta = b$ it yields

$$V\left[L^{1/5}\right] = (\ln 5)^2 \, V\left[\frac{1}{\theta}\right] = \frac{(\ln 5)^2}{(a-1)^2 \, (a-2)\,b}. \tag{34}$$

The sample mean of $\{L_i^{1/5}\}$, denoted m, is an unbiased estimate of the population mean (33). The (corrected by dividing the sum of squared deviations by $n - 1 = 13$) sample variance of $\{L_i^{1/5}\}$, denoted v, is an unbiased estimate of the population variance (34). Using these two sample moments in place of the two population moments allows (33) and (34) to be inverted, yielding the values

$$a = 2 + \frac{m^2}{v} \tag{35}$$

and

$$b = \frac{v \, \ln 5}{(v + m^2)\, m}. \tag{36}$$

From the last row of Table 1, the sample mean of $\{L_i^{1/5}\}$ is calculated to be $m = 1.8036$ and the (corrected, unbiased) sample variance of $\{L_i^{1/5}\}$ is calculated to be $v = .9302$. From (35) and (36), the corresponding parameters for the gamma PDF (27) are $a = 5.497$ and $b = .1984$.

The calculated values $a = 5.497$ and $b = .1984$ can be used to check how good is the fit of the gamma *population* PDF (27) of θ for generating the observed *sample* values of $\{L_i^{1/5}\}$. In Table 2 above are enumerated the population quartiles generated by the gamma PDF (27) for $a = 5.497$ and $b = .1984$, as transformed by (32) into corresponding population quartile values of $L^{1/5}$. The last row of Table 2 gives the sample quartiles observed in the last row of Table 1.

As is evident from Table 2, there is a good quartile fit to the story that $L_i^{1/5} = \ln 5/\theta_i$, where values of $\{\theta_i\}$ are generated as iid draws from the gamma PDF (27). In this story the gamma PDF parameters are $a = 5.497$ and $b = .1984$, which have been determined by the first two sample moments of $\{L^{1/5}\}$.

For the exponential tail exceedance probabilities of the form $\overline{P}_\theta^e(x) = \exp(-\theta\,(x - \tau))$ in (12), the cumulative parent hazard function (8) is $y(x) = x - \tau$. Plugging values $a = 5.497$ and $b = .1984$ into (29) for $y(x) = x - \tau$ turns (29) into

$$\widehat{P}_\Gamma^e(x) = [1 + .1984\,(x - 4.5)]^{-5.497}. \tag{37}$$

Table 3 Comparison of tail-conditional exceedance probabilities

x	5 °C	6 °C	7 °C	8 °C	9 °C	10 °C	11 °C	12 °C
$\overline{P}^e_\mu(x)$ (%)	58	19	6.5	2.2	.74	.25	.08	.03
$\widehat{P}^e_\Gamma(x)$ (%)	59	23	11	7.0	3.0	1.7	1.1	.66

I now want to compare the posterior-predictive tail-conditional exceedance probability distribution (37) with an "average" tail-conditional exceedance probability distribution of the underlying exponential form (12). The mean value of the gamma PDF (27) is $\mu = ab$. From (24), taking $\theta = \mu$ ensures the normalization $\widehat{\theta}(\tau) = \mu$, which means that both of the two hazard rates from the two different probability distributions are initialized to be equal at the beginning of the tail region $\tau = 4.5$ °C. Picking $\theta = \mu$ therefore ensures a kind of level playing field in the sense that the initial rate of tail slimming of this "representative" child probability distribution is the same as the initial rate of tail slimming of the posterior-predictive probability distribution.

Using parameter values $a = 5.497$ and $b = .1984$ gives $\mu = 1.0908$. For purposes of comparison, I take this average value of $\theta = \mu = 1.0908$ as the "representative" child distribution here. Then (10) becomes

$$\overline{P}^e_\mu(x) = \exp(-1.0908\,(x - 4.5)). \tag{38}$$

Table 3 above compares tail-conditional "representative" tail probabilities from (38) with tail-conditional posterior-predictive tail probabilities from (37).

The representative tail-conditional exceedance probability distribution $\overline{P}^e_\mu(x)$ given by formula (38) is of the classical thin-tailed exponential form. The posterior-predictive tail-conditional exceedance probability distribution $\widehat{P}^e_\Gamma(x)$ given by formula (37) is of the fat-tailed polynomial form. The distinction between the two forms does not much show itself in Table 3 for relatively low values of climate sensitivity x at the beginning of the tail region $x \geq 4.5$ °C. (Note that $\overline{P}^e_\mu(4.5) = \widehat{P}^e_\Gamma(4.5) = 100$ % by definition.) However, Table 3 shows that the degree of posterior-predictive tail fattening is quite pronounced at higher values of x located further out in the tail. Rough as this numerical example is, it might be construed as a crude warning that a kind of precautionary principle could apply to at least some critical unknowns in climate change. The results of Table 3 are hinting that when probabilities of extreme climate sensitivity are being considered in situations where the rate of tail slimming is unknown, it might be wise to err on the side of caution by accepting fatter-tailed estimates as being more appropriate for decision making. This is the kind of message that a policy maker might take away from Table 3.

The numerical example of Table 3 illustrates the operation of a basic underlying tail-fattening mechanism. Of course the underlying data are far from ideal, several challengeable assumptions are being made, and the example is stretched, to put it mildly. Overall, this numerical application is far more illustrative of underlying principles than it is directly informative for policy advice. Nevertheless, basic principles are important and it is useful to have an actual numerical example indicating how a principle might apply to a real-world situation, even if it is a stretched application.

7 Concluding Comments

To this point, I think the paper largely speaks for itself. When there is some PDF for how bad things might get, but the overall degree of tail slimming of this PDF is itself uncertain, then the deeper into the tail is the event being analyzed, the more relevant are the fatter-tailed probability assessments about how slowly the tail might be slimming down. A kind of precautionary principle then applies for the probabilities of extreme events, since a decision maker should then effectively act as if the relatively fatter-tailed underlying PDFs are appropriate.

To model and derive this version of a Bayesian precautionary principle, several assumptions were made throughout the paper. Some of these assumptions might legitimately be challenged. Based on a linear family of hazard functions that quantifies the overall degree of tail slimming, the paper shows that uncertainty about the overall degree of tail slimming ends up fattening tails. It remains an open research question to what extent the kind of relatively tractable results this paper derives can be broadened to encompass significantly more general settings. In a sense the paper is more about a class of examples than it is a general theory of tail fattening under uncertainty. I attempt to flag the limited scope of this paper by characterizing it as a "precautionary tale" rather than trying to pretend it is some broader-sounding general theory.

The remaining remarks of this concluding section are a few brief reflections on what it all might mean.

The basic nature of uncertainty has long been a controversial subject. Situations where it is difficult to assign numerical probabilities are often characterized as having ambiguity. Ambiguous situations often end up putting relatively more weight on the worst-case scenarios, and in so doing effectively promote the inclusion in decision making of some version of a so-called precautionary principle, which emphasizes avoiding the more scary or unpleasant situations.[7]

The economics of climate change constitutes one vast laboratory for trying out various ideas about uncertainty. The primary purpose of this paper has been to show that a kind of precautionary principle (in the form of posterior-predictive tail fattening) can emerge naturally from a purely Bayesian setup focused on the uncertain rate of tail slimming of bad events. Furthermore, this Bayesian posterior-predictive tail-fattening approach has a sufficiently operational component to generate some numerical implications in at least one, admittedly greatly over-simplified, data-based situation pertaining to climate change.

Whether this Bayesian reductionism is a good thing or not depends on one's viewpoint. For me it is useful because the Bayesian framework has an appealing overall rational consistency and it is interesting to examine what it might have to say on the subject of this paper. For those who believe that the principles of ambiguity aversion override Bayesian principles in situations with uncomfortably non-quantifiable probabilities, so that a precautionary principle does not need to rely on a Bayesian foundation, there will be plenty of legitimate criticisms that can be leveled at this paper and my interpretations of what it might mean. In any event I think that an attempt to apply various theories of uncertainty to actual examples—here the economics of climate change—is bound to be a good thing generally because without such applications it is difficult to see what the theories might mean in practice and the theories can therefore tend to become ingrown.

---

[7] For more on various aspects of this subject, see Etner et al. (2012), Henry and Henry (2002), and Millner et al. (2010).

The one thing that seems nearly certain is that a debate over the meaning and significance of a precautionary principle will continue, and that it will continue to have relevance for the economics of climate change.

References

Clemen RT, Winkler RL (1999) Combining probability distributions from experts in risk analysis. Risk Anal 19(2):187–203

Etner J, Jeleva M, Tallon JM (2012) Decision theory under ambiguity. J Econ Surv 26(2):234–270

Hansen J, Sato M, Karecha P, Beerling D, Masson-Delmotte V, Pagani M, Raymo M, Royer DL, Zachos JC (2008) Target atmospheric CO_2: where should humanity aim? Open Atmos Sci J 2:217–231

Henry C, Henry M (2002) Formalization and application of the precautionary principle. Columbia University Department of Economics discussion paper series 0102-22

IPCC-AR4. Climate Change 2007 (2007) The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Available online at http://www.ipcc.ch

Knutti R, Hegerl g (2008) The equilibrium sensitivity of the earth's temperature to radiation changes. Nat Geosci 1:735–743

Millner A, Dietz S, Heal G (2010) Ambiguity and climate policy. NBER working paper 16050, forthcoming in Environmental and Resource Economics

Millner A (2012) On welfare frameworks and catastrophic climate risks. Mimeo, forthcoming in Journal of Environmental Economics and Management

Weitzman ML (2009) On modeling and interpreting the economics of catastrophic climate change. Rev Econ Stat 91(1):1–19

Weitzman ML (2011) Fat-Tailed uncertainty in the economics of climate change. Rev Environ Econ Policy 5(2):275–292

Zickfeld K, Morgan MG, Frame DJ, Keith DW (2010) Expert judgments about transient climate response to alternative future trajectories of radiative forcing. Proc Natl Acad Sci 106(38):16129–16135