risks

MDPI

Article

The Regress of Uncertainty and the Forecasting Paradox

Nassim Nicholas Taleb [1,2] and Pasquale Cirillo [3,*]

[1] Maroun Semaan Faculty of Engineering and Architecture, American University of Beirut, Riad El-Solh, Beirut 1107 2020, Lebanon; nt48@aub.edu.lb
[2] Universa Investments L.P., 2601 South Bayshore Drive, Miami, FL 33133, USA
[3] ZHAW School of Management and Law, Theaterstrasse 17, 8401 Winterthur, Switzerland
* Correspondence: pasquale.cirillo@zhaw.ch; Tel.: +41-58-934-44-36

Abstract

We show that epistemic uncertainty–our iterated ignorance about our own ignorance–inevitably thickens statistical tails, even under perceived thin-tailed environments from past realizations. Any claim of precise risk carries a margin of error, and that margin itself is uncertain, in an infinite regress of doubt. This "errors-on-errors" mechanism rules out thin-tailed certainty: predictive laws must be heavier-tailed than their in-sample counterparts. The result is the Forecasting Paradox: the future is structurally more extreme than the past. This insight collapses branching scenarios into a single heavy-tailed forecast, with direct implications for risk management, scientific modeling, and AI safety.

Keywords: uncertainty; risk; risk management; forecasting; tails; machine learning; finance

check for updates

Academic Editor: Paolo Giudici

Received: 27 October 2025
Revised: 27 November 2025
Accepted: 3 December 2025
Published: 10 December 2025

Citation: Taleb, Nassim Nicholas, and Pasquale Cirillo. 2025. The Regress of Uncertainty and the Forecasting Paradox. *Risks* 13: 247. https://doi.org/10.3390/risks13120247

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. The Regress Argument and the Structure of Ignorance

One of the central problems of knowledge concerns the gap between representation and truth (that is, between the map of reality and reality), the source of model risk and a fundamental subject of statistical inference (Rüschendorf et al. 2023). While a long and distinguished tradition of philosophical inquiry into probability—including Laplace, Ramsey, Keynes, De Finetti, von Mises, and Jeffreys (Childers 2013; de Finetti 1974, 1975; Laplace 1814; von Plato 1994)—has grappled with the nature of belief and evidence, its deepest epistemological questions have often remained at the periphery of applied probabilistic decision making and risk management. Yet epistemology is a central, not peripheral, element of the field (Taleb and Pilpel 2004, 2007). Any rigorous attempt at inference must eventually confront the foundational queries: *How do you know what you know? How certain are you of that knowledge?* This paper argues that the structure of our answer to these questions has direct, unavoidable, and mathematically potent consequences. We go further by considering higher orders of this doubt: *How certain are you of your certainty?*

In philosophy and decision theory this concern has long been articulated through the distinction between risk and uncertainty (and "potential surprises"), already emphasized by Knight (1921) and Shackle (1968), and later formalized by scholars like Halpern (2005). In the policy sciences, the same challenge is framed as "deep uncertainty"—a condition in which analysts cannot agree on models, probability distributions, or even on the structure of causal relationships (Gigerenzer 2002; Lempert et al. 2003; Marchau et al. 2019). In climate economics, Weitzman (2009), among others like Millner et al. (2010) and Lemoine and Traeger (2014), highlighted how ambiguity about sensitivity parameters leads to fat-tailed climate and welfare risks (leading to the so-called "dismal theorem"). In machine learning, the problem resurfaces in the tendency of modern AI systems to be overconfident: despite

good predictive accuracy (on the short run), they often lack calibrated and reliable estimates of their own uncertainty (Amodei et al. 2016; Gal and Ghahramani 2016; Kendall and Gal 2017; Ovadia et al. 2019). These strands of literature, spanning philosophy, economics, risk management, and computer science, converge on the same central insight: *our ignorance about our ignorance structurally reshapes predictive distributions*.

Our investigation begins with a simple, tautological observation about estimation. Any estimate, whether statistical or otherwise, is by definition an imprecise proxy for a true value. As such, it must necessarily harbor an error rate; otherwise, it would be a certainty. Here, however, we enter a chain of nested doubts: an epistemic regress. The estimated error rate is itself an estimate. The methods used to determine it are themselves imprecise. Therefore, the estimate of the error rate must also have an error rate. This recursion has no logical endpoint. There is no unimpeachable final layer, no bedrock of perfect knowledge upon which our models can rest. By a regress argument, our inability to account for this cascade of "errors on errors" fosters a deep model uncertainty whose full, recursive structure is often truncated at the first order and remains largely underappreciated in the standard literature (Draper 1995; Viertl 1997).

The consequences of this regress have a direct and powerful effect on probability itself. Simply, uncertainty about a probability inflates the probability of tail events.

Assume a forecaster tells you the probability of a catastrophic event is *exactly* zero (thus violating Cromwell's rule (Lindley 1991)). The crucial epistemological question is not whether the statement is true, but "How do you know?" If the answer is, "I estimated it," then the claim of absolute certainty collapses. An estimate implies an error. The "true" probability must lie in a range above zero, however small (even if the error is small, it cannot be symmetric around zero, as probabilities are bounded below). Whenever our knowledge is not perfect, and thus with the exclusion of statements that are logically true or false, epistemic uncertainty structurally forbids claims of impossibility and forces the higher probabilities to rare events (and lower probabilities to frequent events) than our base models would suggest.

A similar concern has been emphasized in applied risk management, where the underestimation of tail probabilities has direct consequences. In finance, model risk is often discussed in terms of misspecified dynamics or wrong volatility inputs (Derman 1996; McNeil et al. 2015); in epidemiology and climate science, ensemble forecasting practices attempt to deal with disagreement across models. Yet these approaches typically treat the ensemble itself as given, avoiding the "ensemble of ensembles" problem. Our approach highlights that this hierarchy of uncertainty is the in the deep structure of model risk: uncertainty propagates recursively, and the structure of this regress thickens tails across domains. Practitioners often halt this regress after the first step, treating their own statements of uncertainty–such as confidence intervals–as if they were themselves known with certainty. This is a profound mistake that leads to the severe, often monumental underestimation of risk from higher moments (Embrechts et al. 2003; Taleb 2021).

This challenge becomes acute in forecasting (Petropoulos et al. 2022), where uncertainty multiplies. Standard methods for projecting the future, such as the counterfactual analysis of Lewis (Lewis 1973) or its business-school version, "scenario analysis," suffer from a combinatorial explosion. Each step forward in time branches into multiple possible outcomes, which in turn branch further, creating an intractable tree of futures that grows at a rate of at least 2^n for n steps. One of the principal contributions of this paper is to show that this seemingly intractable branching can be tamed. We present a method to structure the regress of errors analytically, collapsing the exploding tree of counterfactuals into a single, well-defined probability distribution. By parameterizing the rate of "error on error," we can vary our degree of epistemic doubt and perform sensitivity analyses on our own ignorance.

In what follows, we show how, by taking this epistemic argument to its mathematical conclusion, fatter tails necessarily emerge from layered uncertainty. We begin with a maximally "tame," thin-tailed world represented by the Gaussian distribution. By systematically perturbing its scale parameter, that is, by introducing explicit, nested layers of doubt about its "true" value, we show analytically how tail risk increases until the distribution becomes robustly heavy-tailed. Our argument is that this is not a mere mathematical curiosity, but a reflection of the true state of the world. There has been a philosophical tradition bundling as diverse thinkers as de Finetti, Bradley, Hume, Nietzsche, Russell, and Wittgenstein, who questioned claims to objective truth (Bradley 1914; de Finetti 2006; Nietzsche 1873; Russell 1958), stating that our knowledge is and will always be an imperfect approximation—"*adequatio intellectus et rei*" (Aquinas 1256)—or, to cite Levi (1967)'s apt phrase, "Gambling with Truth". This paper puts, to use the expression, some engineering-style "plumbing" around such claims. Real-world risk is therefore necessarily fatter-tailed than in our idealized models.

The aim of this work lies in finding an (overlooked) unifying epistemological principle behind the different mathematical and philosophical traditions, connecting the dots of unrelated scholarly groups. By casting the infinite regress of uncertainty into a tractable analytical structure, we show that many disparate phenomena–financial risk, climate uncertainty, AI overconfidence, epidemiological forecasts–are all manifestations of the same mechanism: each layer of doubt structurally thickens the tails of predictive laws. The binomial-tree representation is for instance a simple device to visualize this compounding mechanism, but the lasting contribution is the general epistemic framework that collapses branching scenarios into a universal law of thickened tails.

The rest of this discussion is structured as follows. We begin in Section 2 by rigorously defining the terminology related to the tails of distributions. In Section 3, we introduce our core model of layered uncertainty, which we then analyze under the assumptions of a constant error rate and a decaying error rate. In Section 4, we use a Central Limit Theorem argument to provide a first generalization of the results, and provide a tractable analytical approximation in Section 5. In Section 6, we show that this is a universal mechanism, proving that scale uncertainty generates fat tails for any light-tailed baseline. In Section 7, we discuss the profound consequences of our findings, formalizing the Forecasting Paradox and exploring its applications in finance, AI safety, and scientific modeling. Section 8 summarizes our work and proposes some future research paths.

2. Basic Terminology for Thick-Tailedness

Let X be a random variable with cumulative distribution function F and survival function $\bar{F}$. Following Nair et al. (2022), and to clarify some analyses further down, it is useful to distinguish rigorously among various notions of heavy-tailedness.

A *light-tailed* distribution has a finite moment generating function (mgf) in a neighborhood of the origin, i.e.,

$$M_X(t) = \mathbb{E}[e^{tX}] < \infty \quad \text{for some } t > 0.$$

Equivalently, its survival function satisfies

$$\limsup_{x\to\infty} \frac{\log \bar{F}(x)}{x} < 0,$$

so the tail decays at least exponentially fast. Gaussian, Exponential, and Gamma laws are light-tailed.

A *heavy-tailed* distribution is one with $M_X(t) = \infty$ for all $t > 0$. Equivalently,

$$\limsup_{x\to\infty} e^{cx}\,\bar{F}(x) = \infty \quad \text{for every } c > 0,$$

so its tail decays strictly slower than any exponential e^{-cx}. Within heavy tails, finer classes are distinguished: long tails, subexponential tails and fat tails.

F is *long-tailed* if for every fixed $t \geq 0$,

$$\frac{\bar{F}(x+t)}{\bar{F}(x)} \longrightarrow 1 \qquad (x \to \infty).$$

For a nonnegative (absolutely continuous) X with unbounded support, long-tailedness implies that the hazard rate $q(x) \to 0$ and that the mean residual life $m(x) = \mathbb{E}[X - x \mid X > x]$ diverges as $x \to \infty$. Long-tailed distributions satisfy the so-called *explosion principle*: once a large value is observed, the tail beyond it remains asymptotically undiminished over any fixed further stretch (Nair et al. 2022).

F is said *subexponential* if

$$\bar{F}^{*2}(x) \sim 2\bar{F}(x) \quad (x \to \infty),$$

where $\bar{F}^{*2}$ denotes the survival function of the convolution $F * F$, i.e., the law of $X_1 + X_2$ for i.i.d. $X_1, X_2 \sim F$. The implication is simple: in the extreme, the maximum dominates the sum. This is the *catastrophe principle*: a single large shock outweighs many moderate ones. If you think of a financial portfolio, it is like saying that one single large loss accounts for most of the total loss of the portfolio over a given time horizon (Embrechts et al. 2003).

Finally, F is *fat-tailed* (or regularly varying) if

$$\mathbb{P}(X > x) = x^{-\alpha} L(x), \qquad \alpha > 0,$$

with $L(x)$ being a slowly varying function at infinity. Formally, this means that for any constant $t > 0$, $\lim_{x\to\infty} L(tx)/L(x) = 1$. Intuitively, this implies that the tail of the distribution, asymptotically, behaves like a power law (Kleiber and Kotz 2003). The most important property of fat-tailed distributions is that moments of order $k \geq \alpha$ do not exist (i.e., they diverge) (Cirillo 2013), making their empirical counterparts unreliable for inference (Taleb et al. 2022).

One can show that Fat tails $\subset$ Subexponential tails $\subset$ Long tails $\subset$ Heavy tails. In words: every fat tail is also subexponential, long, and heavy, but not vice versa (Nair et al. 2022).

Important: In the following, borrowing from the parlance of practitioners, and unless otherwise stated or strictly necessary, we will often use "fatter-tailed" relative to a baseline to mean larger excess kurtosis and asymptotically larger upper-tail probabilities, i.e., riskier than the comparison case, without committing to whether the law is strictly fat-tailed, subexponential, or long-tailed.

3. Layering Uncertainties

Take a rather standard probability distribution, say the Normal, which falls in the thin-tailed class (Embrechts et al. 2003; Nair et al. 2022). Assume that its dispersion parameter, the standard deviation σ, is to be estimated following some statistical procedure to get $\hat{\sigma}$. Such an estimate will nevertheless have a certain error, a rate of epistemic uncertainty, which can be expressed with another measure of dispersion: a dispersion on dispersion, paraphrasing the "volatility on volatility" of option operators (Derman 1996; Dupire 1994; Taleb 2021). This makes particular sense in the real world, where the asymptotic assumptions usually

made in mathematical statistics (Shao 1998) do not hold (Taleb 2021), and where every model and estimation approach is subsumed under a subjective choice (de Finetti 2006).

Let $\phi(x;\mu,\sigma)$ be the probability density function (pdf) of a normally distributed random variable X with known mean μ and unknown standard deviation σ. To account for the error in estimating σ, we can introduce a density $f_1(\hat{\sigma};\bar{\sigma}_1,\sigma_1)$ over $\mathbb{R}^+$, where σ_1 represents the scale parameter of $\hat{\sigma}$ under f_1, and $\bar{\sigma}_1 = \sigma$ its expected value. We are thus assuming that $\hat{\sigma}$ is an unbiased estimator of σ, but our treatment could also be adapted to the weaker case of consistency (Shao 1998). In other words, the estimated volatility $\hat{\sigma}$ is the realization of a random quantity, representing the true value of σ with an error term.

The unconditional law of X is thus no longer that of a simple Normal distribution, but it corresponds to the integral of $\phi(x;\mu,\hat{\sigma})$ across all possible values of $\hat{\sigma}$ according to $f_1(\hat{\sigma};\bar{\sigma}_1,\sigma_1)$. This *is* known as a scale mixture of normals (Andrews and Mallows 1974; West 1987), and in symbols one has

$$g_1(x) = \int_0^\infty \phi(x;\mu,\hat{\sigma}) f_1(\hat{\sigma};\bar{\sigma}_1,\sigma_1)\,\mathrm{d}\hat{\sigma}. \tag{1}$$

Depending on the choice of f_1, that in Bayesian terms would define an a priori, $g_1(x)$ can take different functional forms.

Now, what if σ_1 itself is subject to errors? As observed before, there is no obligation to stop at Equation (1): one can keep nesting uncertainties into higher orders, with the dispersion of the dispersion of the dispersion, and so forth. There is no reason to have certainty anywhere in the process.

For $i = 1,\dots,n$, set $\bar{\sigma}_i = \mathbb{E}[\hat{\sigma}_i]$, with $\bar{\sigma}_1 = \sigma$, and for each layer of uncertainty i define a density $f_i(\hat{\sigma}_i;\bar{\sigma}_i,\sigma_i)$, with $\hat{\sigma}_1 = \hat{\sigma}$. Generalizing to n uncertainty layers, one then gets that the unconditional law of X is now

$$g_n(x) \quad = \quad \int_0^\infty \int_0^\infty \cdots \int_0^\infty \phi(x;\mu,\hat{\sigma}_n)\, f_1(\hat{\sigma}_1;\bar{\sigma}_1,\sigma_1)\cdots f_n(\hat{\sigma}_n;\bar{\sigma}_n,\sigma_n)\, d\hat{\sigma}_1\, d\hat{\sigma}_2 \cdots d\hat{\sigma}_n,$$

where $\hat{\sigma}_n$ denotes the "effective" volatility after n nested perturbations.

This approach is clearly parameter-heavy and computationally demanding, as it requires the specification of all the subordinated densities f_i for the different uncertainty layers and the resolution of a possibly very complicated integral.

Let us consider a simpler version of the problem, by playing with a basic multiplicative process à la Gibrat (1931), in which the estimated σ is perturbed at each level of uncertainty i by dichotomic alternatives: overestimation or underestimation. We take the probability of overestimation to be p_i, while that of underestimation is $q_i = 1 - p_i$.

Let us start from the true parameter σ, and let us assume that its estimate is equal to

$$\hat{\sigma} = \begin{cases} \sigma(1+\epsilon_1) & \text{with probability } p_1, \\ \sigma(1-\epsilon_1) & \text{with probability } q_1, \end{cases}$$

where $\epsilon_1 \in [0,1)$ is an error rate (for example it could represent the proportional mean absolute deviation (Taleb 2025)).

Equation (1) thus becomes

$$g_1(x) = p_1\phi(x;\mu,\sigma(1+\epsilon_1)) + q_1\phi(x;\mu,\sigma(1-\epsilon_1)).$$

Now, just to simplify notation, but without any loss of generality, hypothesize that, for $i = 1,\dots,n$, overestimation and underestimation are equally likely, i.e., $p_i = q_i = \frac{1}{2}$. Clearly one has that

$$g_1(x) = \frac{1}{2}(\phi(x;\mu,\sigma(1+\epsilon_1)) + \phi(x;\mu,\sigma(1-\epsilon_1))).$$

Assume now that the same type of uncertainty affects the error rate ϵ_1, so that we can introduce $\epsilon_2 \in [0,1)$ and define the element $(1 \pm \epsilon_1)(1 \pm \epsilon_2)$. Figure 1 gives a tree representation of the uncertainty over a few layers.

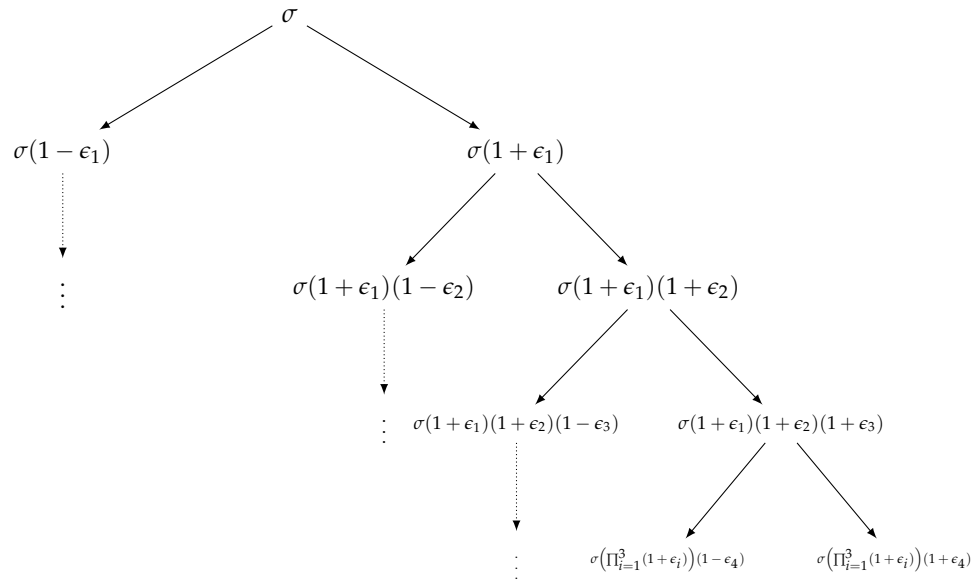


Figure 1. Tree representation of the layers of uncertainty. The diagram shows the progression to the fourth layer along a single path to illustrate the recursive mechanism, with vertical dots indicating the omitted branches.

With two layers of uncertainty the law of X thus becomes

$$g_2(x) = \frac{1}{4}\Big[\phi\big(x;\mu,\sigma(1+\epsilon_1)(1+\epsilon_2)\big) + \phi\big(x;\mu,\sigma(1-\epsilon_1)(1+\epsilon_2)\big) + \phi\big(x;\mu,\sigma(1+\epsilon_1)(1-\epsilon_2)\big) + \phi\big(x;\mu,\sigma(1-\epsilon_1)(1-\epsilon_2)\big)\Big].$$

While at the n-th layer, we recursively get

$$g_n(x) = 2^{-n}\sum_{i=1}^{2^n}\phi\Big(x;\mu,\sigma M_i^{(n)}\Big), \tag{2}$$

where $M_i^{(n)}$ is the i-th entry of the vector

$$\mathbf{M}^{(n)} = \left[\prod_{j=1}^{n}(1+\epsilon_j T_{i,j})\right]_{i=1}^{2^n},$$

with $T_{i,j} \in \{-1,1\}$ enumerating all sign patterns. For example, for $n=3$, one convenient ordering is

$$\mathbf{T} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{M}^{(3)} = \begin{bmatrix} (1-\epsilon_1)(1-\epsilon_2)(1-\epsilon_3) \\ (1-\epsilon_1)(1-\epsilon_2)(1+\epsilon_3) \\ (1-\epsilon_1)(1+\epsilon_2)(1-\epsilon_3) \\ (1-\epsilon_1)(1+\epsilon_2)(1+\epsilon_3) \\ (1+\epsilon_1)(1-\epsilon_2)(1-\epsilon_3) \\ (1+\epsilon_1)(1-\epsilon_2)(1+\epsilon_3) \\ (1+\epsilon_1)(1+\epsilon_2)(1-\epsilon_3) \\ (1+\epsilon_1)(1+\epsilon_2)(1+\epsilon_3) \end{bmatrix}.$$

Once again, it is important to stress that the various error rates ϵ_i are not sampling errors, but rather projections of error rates into the future. They are, to repeat, of epistemic nature.

Equation (2) can be analyzed from different perspectives. In what follows we will discuss two relevant hypotheses regarding the error rates ϵ_i: constant error rates and decaying error rates. We omit increasing error rates, as they trivially only exacerbate the situation. Finally, we answer an important question for practitioners: do we really need $n \to \infty$? The answer is no, showing that we are not just philosophizing.

Remark 1 (Asymmetric layer probabilities)**.** *In the previous derivations we assumed, for simplicity, that upward and downward shocks are equally likely, i.e., $p_i = q_i = \frac{1}{2}$. This symmetry is not necessary: one can allow for general probabilities p_i and $q_i = 1 - p_i$ at each layer.*

Asymmetry in $\{p_i\}$ acts as a systematic tilt of the scale mixture: when $p_j > q_j$ the distribution puts more weight on larger realizations of the effective scale (i.e a larger volatility); when $p_j < q_j$ it tilts towards a smaller scale. Because tail exceedances of a Normal law are increasing functions of the scale, this tilt directly transfers to tail probabilities of X.

The effect is multiplicative and compounds across layers. Even a small, persistent tilt ($p_j - q_j = \delta > 0$) amplifies tail thickness across layers (approximately $\propto \exp\{2\delta \sum_j \epsilon_j\}$ for the variance factor when ϵ_j are small).

All in all, allowing $p_i \neq q_i$ does not change the mechanism—layered uncertainty still thickens the tails relative to the thin-tailed baseline—but it provides a practical dial. Tilting toward overestimation ($p_i > q_i$) further inflates exceedance probabilities at any fixed high threshold; tilting toward underestimation ($p_i < q_i$) attenuates them relative to the symmetric case, but as long as there is non-zero mass on higher-volatility paths, the resulting mixture is asymptotically thicker-tailed than the original fixed-σ Normal for sufficiently large thresholds.

3.1. Hypothesis 1: Constant Error Rate

Assume that $\epsilon_1 = \epsilon_2 = \cdots = \epsilon_n = \epsilon$, i.e., we have a constant error rate at each layer of uncertainty. What we can immediately observe is that matrix $\mathbf{M}$ collapses into a standard binomial tree for the dispersion at level n, so that

$$g_n(x) = 2^{-n} \sum_{j=0}^{n} \binom{n}{j} \phi\Big(x; \mu, \sigma(1+\epsilon)^j (1-\epsilon)^{n-j}\Big).$$

Because of the linearity of the sum, when ϵ is constant, we can use the binomial distribution to weigh the moments of X, when taking n layers of epistemic uncertainty. One can easily check that the first four raw moments read as

$$\begin{aligned}
\mu_1' &= \mu, \\
\mu_2' &= \mu^2 + \sigma^2\Big(1+\epsilon^2\Big)^n, \\
\mu_3' &= \mu^3 + 3\mu\sigma^2\Big(1+\epsilon^2\Big)^n, \\
\mu_4' &= \mu^4 + 6\mu^2\sigma^2\Big(1+\epsilon^2\Big)^n + 3\sigma^4\Big(1+6\epsilon^2+\epsilon^4\Big)^n.
\end{aligned}$$

From these, one can then obtain the following key moments:

$$\begin{aligned}
&\text{Mean}: \ \mu, \qquad \text{Variance}: \ \sigma^2\Big(1+\epsilon^2\Big)^n, \\
&\text{Skewness}: \ 0, \qquad \text{Excess kurtosis}: \ 3\left(\frac{\left(1+6\epsilon^2+\epsilon^4\right)^n}{\left(1+\epsilon^2\right)^{2n}} - 1\right).
\end{aligned}$$

It is then interesting to measure the effect of n on the thickness of the tails of X. The obvious effect, as per Figures 2 and 3, is the rise of tail risk.

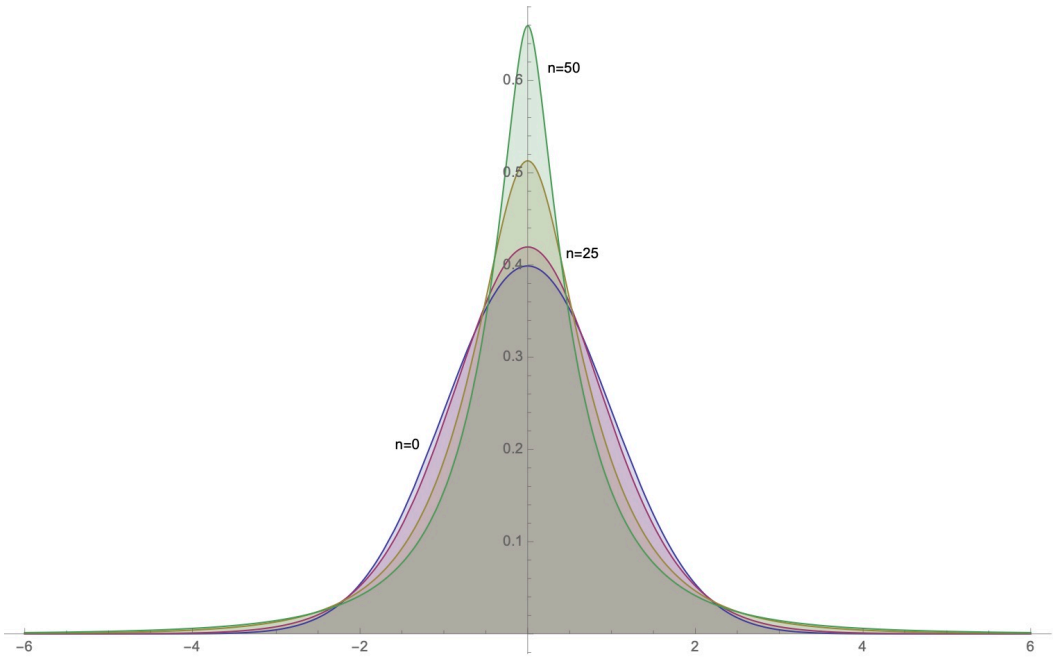


Figure 2. Examples of the density of X (with $\mu = 0$ and $\sigma = 1$), when $\epsilon = 0.1$, for different values of n. Larger n corresponds to visibly fatter tails.

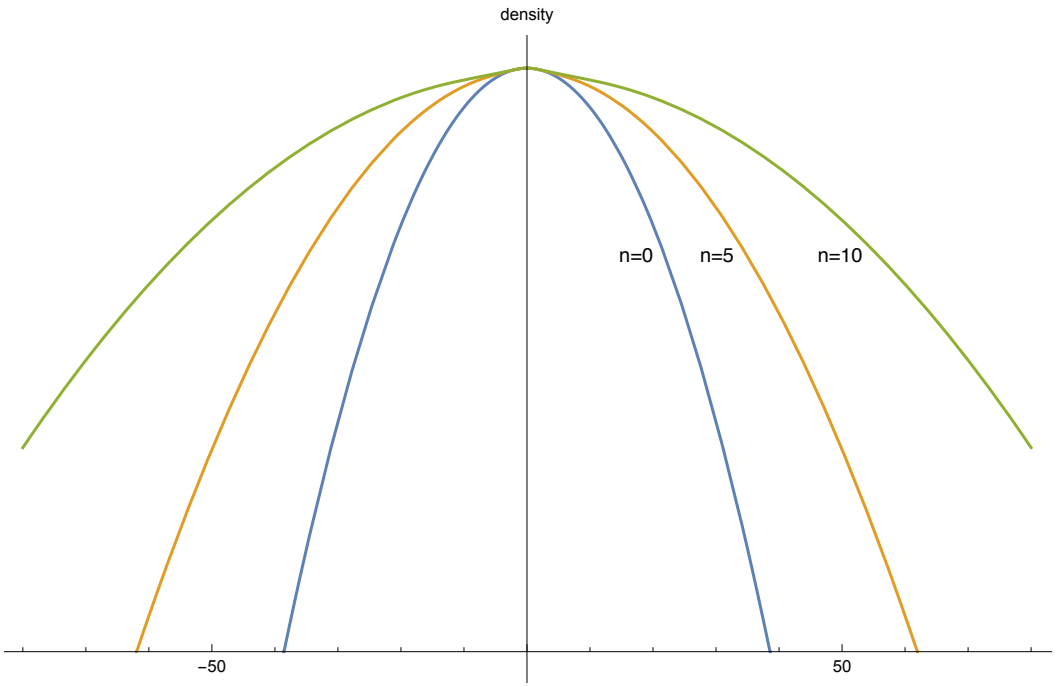


Figure 3. Log-plot of the density of X for $\epsilon = 0.1$ and different values of n. As n grows the tails decay much more slowly than Gaussian; over finite ranges the curves may appear approximately straight on a log-scale (Cirillo 2013).

Fix n and consider the exceedance probability of X over a given threshold K, i.e., the tail of X, when ϵ is constant. One clearly has

$$\mathbb{P}(X \geq K) = \sum_{j=0}^{n} 2^{-n-1} \binom{n}{j} \operatorname{erfc}\left(\frac{K-\mu}{\sqrt{2}\sigma(1+\epsilon)^j(1-\epsilon)^{n-j}} \right),$$

where $\text{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \mathrm{d}t$ is the complementary error function.

The results in Tables 1 and 2 quantify the dramatic impact of layered uncertainty. Even with a small epistemic error of 1% ($\epsilon = 0.01$ in Table 1), the probability of a 10-sigma event with 25 layers of uncertainty becomes over 3300 times larger than what the baseline Normal model would suggest.

Table 1. Ratio of the exceedance probability of X over that of a Normal, for $\epsilon = 0.01$.

Layers (n)	Threshold K		
	$K = 3$	$K = 5$	$K = 10$
5	1.017	1.156	7.57
10	1.035	1.327	45.19
15	1.052	1.515	221.53
20	1.069	1.720	922.24
25	1.086	1.943	3347.0

Table 2. Ratio of the exceedance probability of X over that of a Normal, for $\epsilon = 0.1$.

Layers (n)	Threshold K		
	$K = 3$	$K = 5$	$K = 10$
5	2.747	146.10	1.09×10^{12}
10	4.437	805.99	8.998×10^{15}
15	5.985	1980.80	2.214×10^{17}
20	7.383	3529.41	1.210×10^{18}
25	8.642	5321.36	3.623×10^{18}

When the epistemic error is more substantial, say 10% ($\epsilon = 0.1$ in Table 2), the effect becomes explosive. The probability of the same 10-sigma event is magnified by a factor of 3.6×10^{18}—a number so vast it transforms an event previously dismissed as impossible into a conceivable threat. This demonstrates that ignoring the regress of uncertainty is not a minor oversight, but a source of massive, quantifiable underestimation of risk.

3.2. Hypothesis 2: Decaying Error Rates

As observed before, one may have (actually one needs to have) *a priori* reasons to stop the regress argument and take n to be finite. For example one could assume that the error rates vanish as the number of layers increases, so that $\epsilon_i \geq \epsilon_j$ for $i < j$, and ϵ_i tends to 0 when i approaches a given n. In this case, one can show that the higher moments tend to be capped, and the tail of X less extreme, yet riskier than what one could naively think.

Take a value $\kappa \in [0, 1]$ and fix ϵ_1. Then, for $i = 2, \ldots, n$, hypothesize that $\epsilon_i = \kappa \epsilon_{i-1}$, so that $\epsilon_n = \kappa^{n-1} \epsilon_1$. As to X, without loss of generality, set $\mu = 0$. With $n = 2$, the variance of X becomes

$$\sigma^2 \left(1 + \epsilon_1^2\right) \left(1 + \kappa^2 \epsilon_1^2\right).$$

For $n = 3$ we get

$$\sigma^2 \left(1 + \epsilon_1^2\right) \left(1 + \kappa^2 \epsilon_1^2\right) \left(1 + \kappa^4 \epsilon_1^2\right).$$

For a generic n the variance is

$$\sigma^2 \prod_{i=0}^{n-1} \left(1 + \epsilon_1^2 \kappa^{2i}\right) = \sigma^2 \left[-\epsilon_1^2; \kappa^2\right]_n, \tag{3}$$

where $[a; q]_n = \prod_{i=0}^{n-1} \left(1 - a q^i\right)$ is the q-Pochhammer symbol.

Going on computing moments, for the fourth central moment of X, one gets for example

$$3\sigma^4 \prod_{i=0}^{n-1} \left(1 + 6\epsilon_1^2 \kappa^{2i} + \epsilon_1^4 \kappa^{4i}\right).$$

For $\kappa = 0.9$ and $\epsilon_1 = 0.2$, we get a variance of $1.23\sigma^2$, with a significant yet relatively benign convexity bias. And the limiting fourth central moment is $9.88\sigma^4$, more than 3 times that of a simple Normal, which is $3\sigma^4$. Such a number, even if finite—hence the corresponding scenario is less extreme than the constant-rate case—yet still definitely suggests a tail risk that must not be ignored.

For values of κ in the vicinity of 1 and $\epsilon_1 \downarrow 0$, the fourth moment of X converges towards that of a Normal, closing the tails, as expected.

3.3. Do We Really Need Infinity?

It is critical to note that the thickening of the tails does not require $n \to \infty$. The impact of epistemic uncertainty is immediate and significant even for very small values of n. A practitioner does not need to contemplate an infinite regress; acknowledging just one or two layers of doubt is enough to materially alter the risk profile. It is therefore totally understandable to put a cut-off somewhere for the layers of uncertainty, but such a decision should be taken *a priori* and motivated, in the philosophical sense. One should thus explicitly acknowledge model risk.

Let's consider a practical example. Assume a baseline model of a standard Normal distribution, $X \sim \mathcal{N}(0, 1)$, which represents the naive view with no parameter uncertainty ($n = 0$). Now, let a risk manager introduce a single layer of uncertainty ($n = 1$), acknowledging that their estimate of $\sigma = 1$ could be wrong by, say, $\epsilon = 20\%$.

- **For $n = 0$ (the naive model):** The probability of a 3-sigma event is $\mathbb{P}(X > 3) \approx 0.00135$, or about 1 in 740. The excess kurtosis is 0.
- **For $n = 1$ (one layer of doubt):** The distribution becomes a 50/50 mixture of $\mathcal{N}(0, 1.2^2)$ and $\mathcal{N}(0, 0.8^2)$. The probability of a 3-sigma event is now:

$$\begin{aligned}\mathbb{P}(X > 3) &= 0.5 \cdot \mathbb{P}(Z > 3/1.2) + 0.5 \cdot \mathbb{P}(Z > 3/0.8) \\ &= 0.5 \cdot \mathbb{P}(Z > 2.5) + 0.5 \cdot \mathbb{P}(Z > 3.75) \approx 0.00315,\end{aligned}$$

about 1 in 317.

Simply acknowledging one layer of 20% uncertainty about the volatility has more than doubled the probability of a 3-sigma event. The excess kurtosis has jumped from 0 to approximately 0.44. The effect is not asymptotic; it is a direct and quantifiable consequence of the very first step into the regress argument. Ignoring even a single layer of uncertainty thus constitutes a significant underestimation of risk.

4. A Central Limit Theorem Argument

We now discuss a central limit theorem argument for epistemic uncertainty as a generator of fatter tails and risk. To do so, we introduce a more convenient representation of the normal distribution, which will also prove useful in Section 5.

Consider again the real-valued normal random variable X, with mean μ and standard deviation σ. Its density function is thus

$$\phi(x; \mu, \sigma) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}. \tag{4}$$

Without any loss of generality, let us set $\mu = 0$. Moreover let us re-parametrize Equation (4) in terms of a new parameter $\lambda = \frac{1}{\sigma^2}$, commonly called "precision" in Bayesian statistics (Bernardo and Smith 2000). The precision of a random variable X is nothing more than the reciprocal of its variance, and, as such, it is just another way of looking at variability (actually Gauss (1809) originally defined the Normal distribution in terms of precision). From now on, we will therefore assume that X has density

$$\phi(x;\lambda) = \frac{\sqrt{\lambda}e^{-\frac{1}{2}(\lambda x^2)}}{\sqrt{2\pi}}. \tag{5}$$

Imagine now that we are provided with an estimate of λ, i.e., $\hat{\lambda}$, and take $\hat{\lambda}$ to be close enough to the true value of the precision parameter. The assumption that λ and $\hat{\lambda}$ are actually close is not necessary for our derivation, but we want to be optimistic by considering a situation in which the person who estimates $\hat{\lambda}$ knows what they are doing, using an appropriate method, checking statistical significance, etc.

We can thus write

$$\lambda = \hat{\lambda}(1+\epsilon_1), \tag{6}$$

where ϵ_1 is now a first-order random error term such that $\mathbb{E}[\epsilon_1] = 0$ and $\sigma^2(\epsilon_1) < \infty$. Apart from these assumptions on the first two moments, no other requirement is put on the probabilistic law of ϵ_1.

Now, imagine that a second order error term ϵ_2 is defined on $1+\epsilon_1$, and again assume that it has zero mean and finite variance. The term ϵ_2 may, as before, represent uncertainty about the way in which the quantity $1+\epsilon_1$ was obtained. Equation (6) can thus be re-written as

$$\lambda = \hat{\lambda}(1+\epsilon_1)(1+\epsilon_2). \tag{7}$$

Iterating the error on error reasoning we can introduce a sequence $\{\epsilon_i\}_{i=1}^n$ such that $\mathbb{E}[\epsilon_i] = 0$ and $\sigma^2(\epsilon_i) \in [c, \infty)$, $c > 0$, so that we can write

$$\lambda = \hat{\lambda}\prod_{i=1}^{n}(1+\epsilon_i). \tag{8}$$

For $n \to \infty$, Equation (8) represents our knowledge about the parameter λ, once we start from the estimate $\hat{\lambda}$ and we allow for epistemic uncertainty, in the form of multiplicative errors on errors. The lower value $c > 0$ for the variances of the error terms is meant to guarantee a minimum level of epistemic uncertainty at every level, and to simplify the application of the central limit argument below.

Now take the logs on both sides of Equation (8) to obtain

$$\log(\lambda) = \log(\hat{\lambda}) + \sum_{i=1}^{n}\log(1+\epsilon_i). \tag{9}$$

If we assume that, for every $i = 1, \ldots, n$, $|\epsilon_i|$ is small with respect to 1, we can introduce the approximation $\log(1+\epsilon_i) \approx \epsilon_i$, and Equation (9) becomes

$$\log(\lambda) \approx \log(\hat{\lambda}) + \epsilon_1 + \cdots + \epsilon_n. \tag{10}$$

To simplify treatment, let us assume that the error terms $\{\epsilon_i\}_{i=1}^n$ are independent from each other (in case of dependence, we can refer to one of the generalizations of the CLT, see, e.g., Feller (1968)). For n large, a straightforward application of the Central Limit Theorem (CLT) of Laplace–Liapounoff tells us that $\log(\lambda)$ is approximately distributed as a Normal$(\log\hat{\lambda}, S^2)$, where $S^2 = \sum_{i=1}^n \sigma^2(\epsilon_i)$. This clearly implies that $\lambda \sim$ Lognormal$(\log\hat{\lambda}, S^2)$, for $n \to \infty$. Notice that, for n large enough, we could also

assume $\hat{\lambda}$ to be a random variable (with finite mean and variance), but still the limiting distribution of λ would be a Lognormal (Feller 1968).

Remark 2 (A variant without small-error)**.** *Set* $\xi_i := \log(1+\epsilon_i)$ *with* $\mathbb{E}[\xi_i] = m_i$ *and* $\mathrm{Var}(\xi_i) = v_i \in (0,\infty)$*. If* $\{\xi_i\}$ *are independent (or weakly dependent) and satisfy Lindeberg's or Lyapunov's condition, then* $\left(\sum_{i=1}^n(\xi_i - m_i)\right)/\sqrt{\sum_{i=1}^n v_i} \Rightarrow \mathcal{N}(0,1)$ *(Feller 1968). Hence* $\log\lambda = \log\hat{\lambda} + \sum_{i=1}^n \xi_i \Rightarrow \mathcal{N}\big(\log\hat{\lambda} + \sum m_i,\ S^2\big)$ *with* $S^2 := \sum v_i$*, so* λ *is asymptotically Lognormal. The first-order derivation in the text is the special case* $\xi_i \approx \epsilon_i$*.*

Epistemic doubt has thus a very relevant consequence from a statistical point of view. Using Bayesian terminology, the different layers of uncertainty represented by the sequence of random errors $\{\epsilon_i\}_{i=1}^n$ correspond to eliciting a Lognormal prior distribution on the precision parameter λ of the initial Normal distribution. This means that, in case of epistemic uncertainty, the actual marginal distribution of the random variable X is no longer a simple Normal, but a Compound Normal–Lognormal distribution, which we can represent as

$$g(x) = \int_0^\infty \underbrace{\frac{\sqrt{\lambda}}{\sqrt{2\pi}}\exp\Big(-\tfrac{1}{2}\lambda x^2\Big)}_{\text{Normal}} \cdot \underbrace{\frac{1}{\lambda S\sqrt{2\pi}}\exp\left(-\frac{(\log\lambda - \log\hat{\lambda})^2}{2S^2}\right)}_{\text{Lognormal}} \mathrm{d}\lambda. \tag{11}$$

Despite its apparent simplicity, the integral in Equation (11) cannot be solved analytically, and no moment generating function can be defined (the normal-lognormal compound is indeed heavy-tailed). However, its moments can be obtained explicitly, providing a clear picture of how epistemic uncertainty inflates risk.

Let us derive the moments of X. Since the distribution is symmetric around $\mu = 0$, all odd moments are zero. For the even moments $\mu_{2m} = \mathbb{E}[X^{2m}]$, we can use the law of total expectation:

$$\mu_{2m} = \mathbb{E}[\mathbb{E}[X^{2m}|\lambda]], \qquad \mathbb{E}[X^{2m}|\lambda] = \frac{(2m)!}{m!\,2^m}\lambda^{-m}.$$

For $\lambda \sim \text{Lognormal}(\log\hat{\lambda}, S^2)$ and $k \in \mathbb{R}$, $\mathbb{E}[\lambda^k] = \hat{\lambda}^k \exp(k^2S^2/2)$. Hence

$$\mu_{2m} = \frac{(2m)!}{m!\,2^m}\,\hat{\lambda}^{-m}\,\exp\left(\frac{m^2S^2}{2}\right).$$

By rescaling x we may normalize $\hat{\lambda} = 1$, yielding the simple forms

$$\mathrm{Var}(X) = e^{S^2/2}, \quad \mu_4 = 3e^{2S^2}, \quad \mu_6 = 15e^{9S^2/2},$$

and an excess kurtosis of $3\big(e^{S^2} - 1\big)$.

5. An Analytical Approximation

The impossibility of solving Equation (11) can be bypassed by introducing an approximation to the Lognormal distribution on λ. The idea is to use a Gamma distribution to mimic the behavior of the lognormal prior on precision, also looking at tail behavior.

Both the Lognormal and Gamma distribution are skewed distributions on $\mathbb{R}_+$ and share a convenient property: their coefficient of variation (CV) is constant with respect to location. For Lognormal(μ_L, S^2) the CV equals $\sqrt{e^{S^2} - 1}$; for $Y \sim \text{Gamma}(\alpha, \beta)$ with density

$$\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}, \qquad \alpha > 0,\ \beta > 0,$$

the CV is $1/\sqrt{\alpha}$.

From the point of view of extreme value theory, the Lognormal is subexponential (hence heavy-tailed) and in the Gumbel maximum domain of attraction; conversely, the Gamma is *light-tailed* (hence not subexponential), but still in the Gumbel domain (de Haan and Ferreira 2006; Embrechts et al. 2003). As shown in Figure 4, in the bulk they can be hard to distinguish (Johnson et al. 1994; McCullagh and Nelder 1989), but asymptotically the Lognormal tail dominates: for Gamma(α, β) (shape–rate) the hazard $h(x) \to \beta$ as $x \to \infty$, whereas for the Lognormal $h(x) \to 0$.

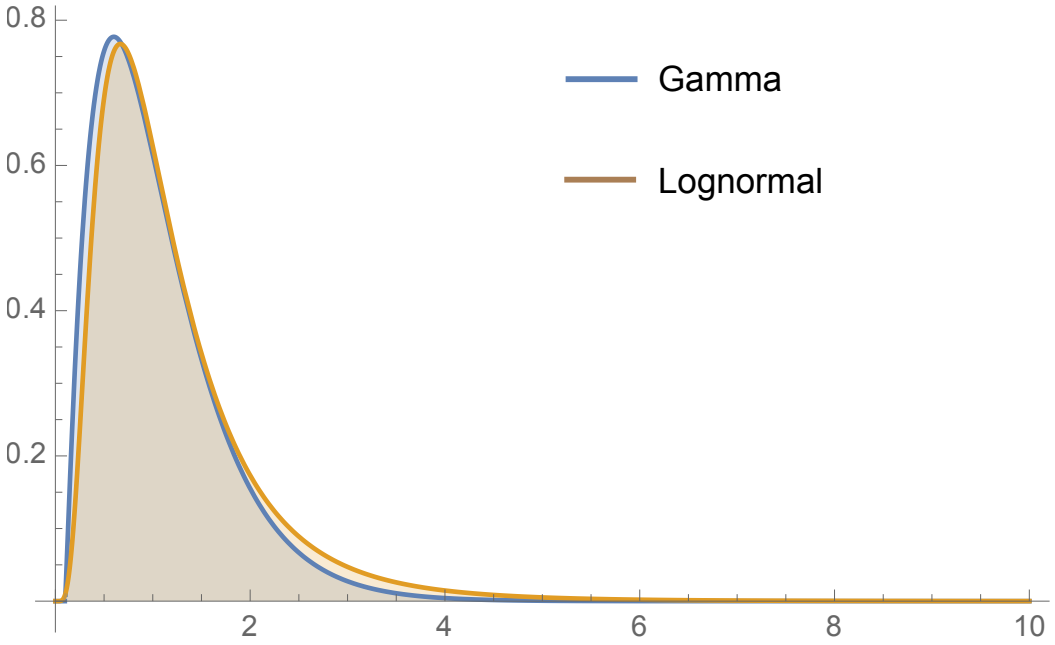


Figure 4. Comparison of the density functions of a Gamma and a Lognormal sharing the same coefficient of variation of 0.7.

Concerning the presence of the precision parameter λ: moving from the Lognormal to the Gamma has a great advantage. A Normal distribution with known mean (for us $\mu = 0$) and Gamma-distributed precision parameter has an explicit closed form.

Let us rewrite Equation (11) by substituting the lognormal density with an approximating Gamma(α, β) prior on λ:

$$\begin{aligned} g(x;\alpha,\beta) &= \int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda x^2} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\, \mathrm{d}\lambda \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\sqrt{2\pi}} \int_0^\infty \lambda^{\alpha+\frac{1}{2}-1} \exp\left(-\lambda\left(\beta + \frac{1}{2}x^2\right)\right) \mathrm{d}\lambda. \end{aligned} \tag{12}$$

The integral above can be solved explicitly, yielding

$$g(x;\alpha,\beta) = \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)\sqrt{2\pi\beta}} \left(1 + \frac{x^2}{2\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)}. \tag{13}$$

This is the density of a Student's t with

$$\nu = 2\alpha \quad \text{degrees of freedom, and scale} \quad s^2 = \beta/\alpha,$$

i.e.,

$$g(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\, s} \left(1 + \frac{x^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}.$$

Matching the coefficient of variation of Lognormal$(\log \hat{\lambda}, S^2)$ fixes

$$\alpha = \frac{1}{e^{S^2} - 1}, \qquad \beta = \frac{e^{-S^2/2}}{e^{S^2} - 1} \hat{\lambda}^{-1}.$$

Interestingly, the Student's t distribution in Equation (13) is fat-tailed on both sides (Embrechts et al. 2003), especially for small values of α. It is worth noting an apparent paradox in this result: we are mixing a Normal distribution (light-tailed) with a Gamma prior on its precision (which is also light-tailed), yet the result is robustly fat-tailed. This phenomenon highlights the power of the scale-mixing mechanism. The key is that the Gamma distribution is placed on the *precision* $\lambda = 1/\sigma^2$. A light-tailed prior on precision that allows for values arbitrarily close to zero implies a heavy-tailed prior on the variance σ^2. It is precisely the possibility of drawing a near-zero precision (and thus a near-infinite variance) that generates the fat tails in the final predictive distribution.

Since α decreases in S^2 (the sum of the variances of the epistemic errors), the more doubts we have about the precision parameter λ, the more the resulting Student's t is fat-tailed, thus increasing tail risk. This is in line with the findings of Section 3.1.

Therefore, starting from a simple Normal distribution, by considering layers of epistemic uncertainty, we have obtained a thick-tailed predictive distribution with the same mean ($\mu = 0$), but capable of generating more extreme scenarios, and its tail behavior is a direct consequence of imprecision and ignorance.

Remark 3 (Sampling variance versus epistemic uncertainty)**.** *One might observe that fattened tails already arise when a Normal law with unknown variance is integrated over the sampling distribution of the sample variance. Indeed, if* $\hat{S}^2$ *is computed from n Gaussian observations,*

$$\frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1},$$

and mixing the Normal density over this χ^2*-induced law produces the classical Student's t distribution. This captures sampling uncertainty.*

Here we target a different phenomenon. Sampling variability vanishes as $n \to \infty$ *under a correctly specified model; epistemic uncertainty does not. Model risk, specification error, and uncertainty about how* $\hat{S}^2$ *is constructed and used persist even with large amounts of data (unless we consider ourselves some version of the demon of Laplace (1814)). These generate a distinct form of scale uncertainty—the regress of errors on errors—which our framework makes explicit. The resulting tail-thickening is therefore of a different nature: it reflects limits of knowledge, not of sample size.*

Remark 4 (Conservative upper bound via Normal–Gamma)**.** *Although the lognormal prior on* λ *has fatter tails than its Gamma approximation at the* prior *level, the resulting* compound *laws invert this ordering. The Normal–Gamma mixture is a Student's t with power-law tails, asymptotically fatter than the Normal–Lognormal mixture (whose tail decays like* $\exp\{-c(\log x)^2\}$*). Thus the Normal–Gamma route provides a tractable and conservative upper bound for extreme tails.*

6. A Universal Mechanism

Using a simple example based on a Gaussian distribution, the preceding sections showed that layers of epistemic doubt about the standard deviation, modeled as a multiplicative regress, naturally lead to a heavy-tailed, compound distribution. But is this phenomenon specific to the Normal distribution, or does it reflect a more fundamental statistical law?

In this section, we demonstrate the latter. We generalize the mechanism and show that epistemic uncertainty about a model's scale parameter is a universal generator of fatter tails. This principle creates a mathematical bridge from the philosophical concept of layered ignorance to the practical necessity of using heavy-tailed distributions for any realistic modeling, forecasting, or risk management endeavor. The mathematical foundation for this principle is the theory of scale mixtures, which states that mixing any light-tailed distribution with a heavy-tailed scale factor produces a new, heavy-tailed distribution, as formalized in the following proposition.

Proposition 1. *Let $Z \geq 0$ be independent of $S \geq 0$ and define $X := SZ$.*

(i) If S is subexponential on $\mathbb{R}_+$ and Z is light-tailed in the mgf sense, i.e., there exists $t_0 > 0$ with $M_Z(t_0) = \mathbb{E}[e^{t_0 Z}] < \infty$ and $\mathbb{P}(Z > 0) > 0$, then X is subexponential on $\mathbb{R}_+$. Moreover, for all $x > 0$,

$$\bar{F}_X(x) = \mathbb{E}\Big[\bar{F}_S\Big(\frac{x}{Z}\Big)\Big].$$

(ii) If S is regularly varying with index $\alpha > 0$ and $\mathbb{E}[Z^{\alpha+\delta}] < \infty$ for some $\delta > 0$, then X is regularly varying with the same index α and

$$\mathbb{P}(X > x) \sim \mathbb{E}[Z^\alpha]\,\mathbb{P}(S > x) \qquad (x \to \infty).$$

Proof. For (i): The identity $\bar{F}_X(x) = \mathbb{E}[\bar{F}_S(x/Z)]$ follows by conditioning on Z. Closure of the subexponential class under multiplication by an independent light–tailed factor (with $M_Z(t_0) < \infty$ for some $t_0 > 0$) is standard; see (Foss et al. 2013, Thm. 3.27 and Cor. 3.28). For (ii): this is Breiman's lemma (Breiman 1965) under the stated moment condition on Z. □

Remark 5. *If the baseline is symmetric $Z_0 \overset{d}{=} -Z_0$, apply the Proposition to $Z := |Z_0|$ to obtain the right-tail result for $|X| = S|Z_0|$, and then transfer it to two-sided tails by symmetry.*

Proposition 1 thus guarantees that our mechanism–representing ignorance via a random scale factor–produces heavy tails for any light-tailed phenomenon, not just the Normal (as in our initial toy model). Our CLT argument leading to a Lognormal (subexponential) distribution for the scale factor is nothing but a specific instance of this universal principle.

Notice that Proposition 1 is not an abstract curiosity; it reveals the hidden structure of many well-known statistical models. Whenever a simple model is made more realistic by treating its rate, volatility, or scale parameter as unknown and variable, a fatter-tailed distribution emerges. The following examples illustrate this principle across different domains:

- **Exponential → Pareto**: An Exponential distribution with a fixed rate λ is light-tailed. If we express uncertainty about λ by assuming it follows a Gamma distribution, the resulting mixture distribution for the variable is a Lomax (Pareto Type II) distribution, which is fat-tailed.
- **Poisson → Negative Binomial**: A Poisson distribution is used for counts with a fixed rate μ. If we acknowledge that this rate is uncertain and model it with a Gamma distribution, the resulting compound distribution is the Negative Binomial. This distribution is famous for its "overdispersion" and fatter tail compared to the Poisson itself, making it far more suitable for modeling real-world count data.
- **Normal → Student's t**: As shown in Section 5, a Normal distribution (light-tailed) mixed with an Inverse-Gamma prior on its variance (equivalent to a Gamma prior on precision) results in a Student's t-distribution, the canonical example of a fat-tailed distribution in statistics.

These examples reveal a profound pattern: many canonical fat-tailed distributions can be interpreted as simple, light-tailed models whose scale has been randomized to account for uncertainty. The regress of uncertainty finds its mathematical expression in the operation of scale mixing. This provides a direct and robust path from epistemic doubt to the necessity of heavy-tailed models.

Remark 6 (Mathematical Relations to Other Approaches)**.** *Our approach shares mathematical tools with hierarchical Bayesian modeling, as both involve integrating over parameter uncertainty (Bernardo and Smith 2000). However, it departs from that framework in a fundamental epistemological sense. A standard Bayesian analysis captures first-order uncertainty–a prior on a parameter–and may extend to higher orders through hierarchical layers. Such hierarchies, however, are typically truncated once considered sufficient for the practical problem at hand, thereby asserting a limit to uncertainty. This is particularly true under the so-called objective Bayesianism (Berger et al. 2024; Williamson 2010).*

By contrast, our framework proceeds in the opposite direction, moving from less uncertainty to more. Any layer in which the analyst assumes fixed hyperparameters represents a claim to perfect knowledge, contravening the principle of regress. Our aim is not to propose a deeper hierarchical model, but to formalize the very structure of this infinite regress. We demonstrate that its necessary implication–regardless of the specific distributions employed–is the systematic thickening of predictive tails. This result provides both a philosophical and mathematical justification for why predictive distributions must be structurally fatter-tailed than those produced by any model with a finite, arbitrarily imposed level of uncertainty.

Even if one prefers to summarize all higher-order doubt in a single prior on a scale parameter, the regress argument implies that such a prior must place non-negligible mass on large scales if it is to represent epistemic uncertainty honestly; the ensuing scale mixture then necessarily yields a predictive distribution with fatter tails than any model based on a fixed scale.

Likewise, the iterative layering of uncertainty on a scale parameter is, in a narrow mathematical sense, reminiscent of a one-dimensional Renormalization Group (RG) flow (Zinn-Justin 2021). In statistical physics, the RG formalism provides a powerful framework for understanding how a system's properties evolve across physical scales by systematically "coarse-graining," that is, averaging out small-scale fluctuations. Nevertheless, the emergence of heavy-tailed distributions in our toy model, starting from thin-tailed assumptions, is not parallel to an RG flow approaching a fixed point under a multiplicative cascade (Kesten 1973; Kolmogorov 1962; Mandelbrot 1974).

Our framework operates on a different and more general plane. The RG describes objective dynamics of physical systems that exhibit properties such as scale invariance or criticality; its flow is ontic, reflecting how measurable quantities change with physical scale. By contrast, the epistemic regress concerns the structure of subjective knowledge. It requires no assumptions about physical systems and applies to the parameters of any model, regardless of domain. Here, the flow does not occur through physical scales but through successive layers of epistemic uncertainty—the analyst's own deepening recognition of ignorance. Our contribution is not to reproduce the mathematics of scale invariance but to identify the regress of uncertainty as a universal principle of reasoning and forecasting. Whereas the RG characterizes how matter organizes across scales, the epistemic regress captures how belief transforms through successive orders of doubt. The RG flows through matter; the regress flows through mind. Its fixed points are not limits of energy, but the unavoidable limits of understanding.

Remark 7 (Uncertainty about the mean (location))**.** *Throughout the paper we have focused on epistemic uncertainty about scale (or precision), because this is what drives tail thickening. For completeness, note that the mean μ is almost always estimated in practice, typically by a sample mean $\bar{X}$, and can itself be treated as a random quantity. A simple specification in the Gaussian case is*

$$\mu \mid \hat{\mu} \sim \mathcal{N}(\hat{\mu}, \tau^2),$$

for some epistemic variance $\tau^2 > 0$*. Conditioning on* σ^2 *and integrating over* μ *then yields*

$$X \mid \sigma^2 \sim \mathcal{N}(\hat{\mu}, \sigma^2 + \tau^2),$$

so location uncertainty widens dispersion and increases tail probabilities at any fixed threshold, but, under such light-tailed specifications, it preserves the light tail classification (and, in the Gaussian–Gaussian case, the kurtosis can be lowered owing to an artifact of its computation). Therefore it affects risk metrics through a pure scale effect, rather than through the kind of structural tail thickening that drives the forecasting issues identified by Taleb et al. (2022).

More generally, let Z *be a light-tailed baseline in the mgf sense (that is,* $M_Z(t) < \infty$ *for some* $t > 0$*), and let the independent location noise* μ *also be light-tailed (*$M_\mu(t) < \infty$ *for some* $t > 0$*). Then the predictive law* $X = Z + \mu$ *remains light-tailed, because* $M_X(t) = M_Z(t)M_\mu(t)$ *is finite for some* $t > 0$*. A heavy-tailed distribution for the location parameter can of course induce heavy tails in* X*, but this is unsurprising: the heaviness is injected directly through* μ*, not by the regress mechanism itself.*

This behaviour contrasts with uncertainty about scale. As shown in the scale analysis above, a regress of multiplicative errors on a scale (or precision) parameter naturally leads to subexponential mixing laws (e.g., Lognormal for precision in our CLT argument), and even light-tailed approximations such as a Gamma prior on precision produce fat-tailed predictive distributions (Student's t in our Normal case). In applications, both location and scale uncertainty can be included, but it is epistemic scale uncertainty that really governs the behaviour of extremes (de Haan and Ferreira 2006).

Remark 8 (Fat-tailed baselines)**.** *In many applications, the baseline is not necessarily thin-tailed. The universal mechanism applies unchanged. For example, let* $Z \sim t_\nu$ *and let* S *denote an epistemically uncertain scale factor. Then* $X = SZ$ *remains regularly varying with the same tail index* ν *(under the moment condition* $\mathbb{E}[S^{\nu+\delta}] < \infty$*). Exceedance probabilities at any fixed high threshold inflate asymptotically by the factor* $\mathbb{E}[S^\nu]$*, while the corresponding high quantiles inflate by the factor* $(\mathbb{E}[S^\nu])^{1/\nu}$*. Thus, even when practitioners already adopt heavy-tailed t models, epistemic scale uncertainty might preserve the tail index but introduces a second layer of practical significance, especially for risk metrics and long-horizon pricing.*

7. The Forecasting Paradox and Its Consequences

The principle that layered epistemic uncertainty thickens tails is not a mere theoretical curiosity: it has profound and actionable consequences across any domain that relies on statistical modeling and forecasting. Ignoring the regress of errors on errors leads to a systemic and dangerous underestimation of risk, brittleness in automated systems, and a false sense of certainty in scientific projections. Our framework thus provides a formal basis for a crucial principle of forecasting.

Principle 1 (The Forecasting Paradox)**.** *The predictive distribution for out-of-sample data must be treated as having fatter tails than the descriptive model of in-sample data.*

The reason is now clear. A model fitted to past data (in-sample) is conditioned on a point estimate of its parameters, such as $\hat{\sigma}$. However, a forecast for the future (out-of-sample) cannot treat this estimate as perfect. One must account for the uncertainty surrounding $\hat{\sigma}$ itself. This is achieved by integrating over all possible values of the parameter, weighted by their probabilities—the very scale mixture operation that our paper proves is a generator of fat tails. Therefore, a responsible forecast must reflect this added layer of epistemic uncertainty, making it structurally more conservative than a historical description.

From an epistemic point of view, the paradox is thus conditional: once one accepts that there is non-degenerate uncertainty about scale, any coherent predictive law must have tails that are thicker than those of the in-sample descriptive fit, even if the degree of thickening is left as a tunable "dial of doubt".

7.1. Applications in Finance and Quantitative Risk Management

The consequences for risk management are immediate. Standard models like Value-at-Risk (VaR) or Expected Shortfall (ES) often rely on fitting a distribution to historical data and then extrapolating its quantiles (Klugman et al. 1998; McNeil et al. 2015). In our opinion, this is insufficient and risky.

Let us discuss a practical example. An analyst wants to compute the one-day 99% Value-at-Risk (VaR) for a portfolio, based on a historical sample of daily returns. The standard procedure is as follows:

1. **The Standard (and Flawed) Approach:** The analyst calculates the historical standard deviation, obtaining a point estimate $\hat{\sigma} = 1.5\%$ per day. Assuming a mean-zero Gaussian distribution for simplicity (Hull 2023), the 99% VaR is calculated as:

$$\text{VaR}_{\text{descriptive}} = 2.33 \times \hat{\sigma} = 2.33 \times 1.5\% = 3.495\%$$

   This value is often directly used as a forecast of tomorrow's risk (McNeil et al. 2015). This is a critical epistemological error in our view.
2. **The Robust Approach (The Regress of Uncertainty):** A prudent analyst, following our principle, must acknowledge that $\hat{\sigma} = 1.5\%$ is merely an estimate. To account for this, we model the volatility for the next day not as a point, but as a random variable. For simplicity, assume a minimal epistemic error of $\pm 10\%$, leading to a 50/50 mixture of two scenarios: a "high-volatility" world where $\sigma_H = 1.65\%$ and a "low-volatility" world where $\sigma_L = 1.35\%$.
3. **The Quantifiable Consequence:** The 99% VaR of this new, fatter-tailed distribution is not the average of the individual VaRs. It is the value x for which the probability of the loss being less than x is 99% under the mixture distribution. Formally, we must solve for $\text{VaR}_{\text{predictive}}$ in the equation:

$$0.5 \cdot \Phi\left(\frac{\text{VaR}_{\text{predictive}}}{\sigma_L}\right) + 0.5 \cdot \Phi\left(\frac{\text{VaR}_{\text{predictive}}}{\sigma_H}\right) = 0.99$$

   where Φ is the standard Normal cumulative distribution function. Solving this numerically yields:

$$\text{VaR}_{\text{predictive}} \approx 3.56\%$$

   This value is higher than the naive estimate of 3.495% due to convexity. The increase of 6.5 basis points, on a $1 billion portfolio, translates into $650,000 of additional underestimated risk capital for a single day. This gap represents the quantifiable cost of ignoring even just the *first layer* of epistemic uncertainty.

This example provides a concrete demonstration of the Forecasting Paradox: the risk of the future is structurally greater than the risk observed in the past because the future must also contain our uncertainty about the parameters of the past. Standard methods systematically under-price this fundamental uncertainty.

To consider another situation, we all know that the Black–Scholes model's assumption of constant volatility is famously flawed (Embrechts et al. 2003; McNeil et al. 2015; Taleb 2021). Models incorporating stochastic volatility (the "volatility of volatility") are a step in the right direction (Dupire 1994; McNeil et al. 2015). Our framework provides an

epistemological justification for this, suggesting that the hierarchy does not stop. Uncertainty about the parameters of the volatility model itself must be considered, especially for long-dated options where such uncertainty compounds over time.

If $C_{BS}(K, T; \sigma)$ denotes the Black–Scholes European call price at maturity T with strike K under volatility σ, and volatility itself is epistemically uncertain with predictive law $\Pi(\mathrm{d}\sigma)$, then a natural predictive price is the mixture

$$\tilde{C}(K, T) = \int C_{BS}(K, T; \sigma)\, \Pi(\mathrm{d}\sigma).$$

Because C_{BS} is increasing and non-linear in σ, a non-degenerate distribution Π induces a non-linear correction term which tends to become more pronounced as T grows. Computation of $\tilde{C}$ is straightforward via Gauss–Hermite quadrature over $\log \sigma$ or via low-variance Monte Carlo. In this way, the epistemic variance parameter becomes an explicit control knob for how volatility uncertainty impacts long-maturity prices.

Finally, in market risk as well as in credit risk modeling, instead of relying on a handful of ad-hoc historical or imaginary scenarios (Hull 2023; McNeil et al. 2015), our method allows for the systematic generation of a universe of scenarios. The parameters of model (n and ϵ_i, or S^2) can be seen as "dials of doubt." By adjusting them, an institution can explore the consequences of varying degrees of model uncertainty, moving from simple sensitivity analysis to a full-blown distributional understanding of future risks.

In line with models à la CR+ (Credit Suisse 1997), let defaults N follow a Poisson law with uncertain rate Λ. It is known that, if epistemic uncertainty is represented by a Gamma law for Λ, then the predictive distribution of N is Negative Binomial, naturally recovering the overdispersion observed in many real portfolios. For portfolio loss $L = \sum_{i=1}^{N} Y_i$, where the severities Y_i may themselves carry scale uncertainty, compound Negative-Binomial mixtures yield closed-form probability generating functions and efficient Panjer-recursion evaluation of the entire predictive loss distribution. This replaces ad-hoc "stress scenarios" with a principled and tractable distributional forecast. Clearly, the errors on errors regress naturally allows for even more general settings.

7.2. Machine Learning and the Challenge of AI Safety

Modern machine learning, particularly deep learning, has achieved remarkable performance on many tasks, but this success is often undermined by a dangerous overconfidence (Guo et al. 2017; Liu et al. 2023). These models frequently produce point estimates with little to no reliable measure of their own uncertainty, a failure that becomes especially acute when they are faced with data different from what they were trained on (Aliferis and Simon 2024).

Our framework is a form of uncertainty quantification. While methods like Bayesian Neural Networks (BNNs) place priors over model weights, our work suggests a deeper requirement: hierarchical models that account for uncertainty in the hyperparameters of those priors. A safe AI should not only provide a prediction but also deliver a meta-assessment of its own certainty, derived from its nested model of self-doubt.

For high-stakes applications like autonomous driving or medical diagnostics, an AI that is "99.9% safe" is a liability if that certainty is brittle. The regress argument implies that true robustness comes from acknowledging the potential for errors in the model's own construction. This leads to predictive distributions that are more cautious, assigning non-trivial probability to rare but catastrophic failures, which is the first step toward mitigating them.

Consider a practical example from an autonomous vehicle's perception system. An AI model detects an object on the road and reports its classification with a very high point-estimate confidence: "99.5% probability this is a harmless plastic bag." A standard,

overconfident system would take this at face value, potentially leading to the decision to drive over it.

Our framework insists that this 99.5% figure is merely an estimate subject to its own regress of uncertainty. A robust AI must ask: What is my uncertainty about this confidence score? If the lighting is unusual or the object's shape is ambiguous (i.e., the data is partially out-of-distribution), the "true" confidence is not a point but a distribution with fatter tails towards lower probability. The system might calculate that there is a non-trivial probability (say, 1%) that the object's "true" classification confidence is actually below 50%. Faced with this ambiguity, the robust decision is not to trust the point estimate but to default to a safer action: slow down and avoid the object. This shows how formalizing the regress of uncertainty translates directly into more cautious and reliable AI behavior in safety-critical applications, in line with a rational precautionary principle (Taleb 2025; Taleb et al. 2014).

A natural question is how, in practice, to model this "model uncertainty about the model" in machine learning systems. There is naturally no single canonical construction, and the right implementation is architecture- and domain-dependent, but the regress argument gives a general recipe. At any input x, instead of committing to a single predictive law $p(y \mid x, \hat{\theta})$ with fixed parameters $\hat{\theta}$, one specifies a family of predictive laws $p(y \mid x, \theta)$ and a distribution over θ that captures epistemic uncertainty, including uncertainty about the predictive variance itself. Concretely, one may treat the network's predictive variance (or log-variance) as a random quantity, with a parametric meta-uncertainty law, and estimate its hyperparameters by ensembles, bootstrap resampling, or hierarchical Bayesian fitting (Gal and Ghahramani 2016; Ovadia et al. 2019). The operational object is then the mixture

$$p_{\text{mix}}(y \mid x) = \int p(y \mid x, \theta)\, \mathrm{d}\Pi(\theta),$$

which directly encodes the regress of uncertainty: tail probabilities and safety-relevant events are computed under p_{mix}, not under a single (thin-tailed) surrogate. Our results imply that as soon as this meta-uncertainty acts on a scale or dispersion parameter and is non-degenerate, p_{mix} will be systematically heavier-tailed than the corresponding fixed-parameter baseline, yielding more conservative behaviour in safety-critical regimes without requiring a specific neural architecture or loss function.

7.3. Humility in Scientific Modeling: Climate and Epidemiology

The same principle applies to complex scientific modeling where parameters are deeply uncertain (Taleb et al. 2022). During a risk event like a pandemic (Cirillo and Taleb 2020), different models produce a wide range of forecasts for variables like infection rates or hospitalizations. This variance is a direct effect of parameter uncertainty. A consolidated forecast should not be a simple average, but a formal mixture of the models' outputs, which, as we have shown, results in a fatter-tailed distribution that better communicates the true range of possibilities to policymakers.

In climate science, long-term climate projections depend on parameters like "climate sensitivity," for which we have only uncertain estimates (IPCC 2023; Millner et al. 2010; Weitzman 2009). The regress argument applies perfectly: there is an error on the estimate, an error on the error estimate, and so on. A true risk assessment of climate change must therefore be based on the fat-tailed distribution that arises from this layered uncertainty, forcing us to confront the non-negligible probability of extreme warming scenarios, even more extreme than commonly discussed (Hausfather and Peters 2020; IPCC 2023; Lempert et al. 2003; National Research Council 2013).

8. Conclusions and Future Directions

Ultimately, our approach provides a mathematical backbone to the philosophical wisdom of epistemic humility. As Rescher observed, taking estimates as objective truth is the most consequential error a decision-maker can commit (Rescher 1983). Our analysis formalizes the central lesson of *The Black Swan* (Taleb 2021): the future must be treated as capable of generating rarer and more consequential events than those observed in historical records. "Black swan" surprises arise precisely from the realization, often too late, that one was victim of the "retrospective distortion" and has ignored the layers of uncertainty affecting one's own estimates and representations.

The core novelty of the paper is to treat uncertainty about uncertainty as an explicit, primary object of analysis and to show that layered epistemic doubt systematically thickens predictive tails. This mechanism is not tied to a particular baseline distribution: it operates across domains, models, and time horizons. It explains why descriptive fits to the past and the responsible predictive laws for the future cannot share the same tail behavior. It also clarifies the distinct role of different parameter classes: location uncertainty widens dispersion, but it is scale uncertainty that governs tail thickness. In doing so, the framework unifies diverse disciplines—finance, climate science, medicine, engineering, epidemiology, and machine learning—under a single epistemic principle: acknowledging errors on errors forces the future to be heavier-tailed than the past.

A second contribution is practical. The framework transforms scenario analysis from a collection of discrete narratives into a disciplined predictive mixture with explicit dials of doubt. Forecasts, prices, and capital requirements should be computed from a distribution that integrates parameter and meta-uncertainty, rather than from point estimates or ad-hoc stress scenarios. In risk management, tail metrics such as VaR and ES become mixture quantities that naturally capture convexity and reveal the capital shortfall implicit in ignoring meta-uncertainty. For long-maturity derivatives, volatility uncertainty compounds with horizon and produces robust uplifts in pricing. In credit risk, rate uncertainty leads directly to overdispersion and full predictive loss distributions. And in safety-critical ML systems, uncertainty about the model's own uncertainty must be propagated so that decisions respond to tail probabilities from the mixture, not to brittle confidence points.

At the same time, we have been explicit about scope and limits. The analytics rely on standard idealizations—small-error linearizations, independence or weak dependence across layers—not because the mechanism requires them, but because they make its structure transparent. The qualitative conclusion remains unchanged under richer dynamics, alternative priors, or regime changes: whenever scale is treated as uncertain rather than fixed, the tails thicken. Likewise, although mean uncertainty does not change the asymptotic tail class, it remains relevant for finite-horizon decisions and should be incorporated in applied work.

The results open several directions for future research. One is to develop joint regress mechanisms that propagate uncertainty simultaneously across multiple parameters and model components, clarifying when and how scale effects dominate. Another is temporal: characterizing how layered uncertainty evolves with forecasting horizon and how it interacts with temporal aggregation and persistence. A third avenue concerns evaluation: designing backtests and diagnostics that are calibrated to mixture tails rather than to thin-tailed surrogates. A final direction lies in methodology for ML and AI safety, given the potential impact: benchmarking out-of-distribution robustness when meta-uncertainty is treated explicitly, thereby turning epistemic humility into an operational safety buffer.

In short, the regress of uncertainty is not merely a philosophical caveat but also an organizing principle that yields a tractable predictive law. Embedding it into practice replaces a fragile sense of precision with a reproducible form of prudence: a forecast that

acknowledges what we do not know and, by doing so, better equips us to face the shocks that matter the most.

Author Contributions: Conceptualization, N.N.T.; methodology, N.N.T. and P.C.; formal analysis, N.N.T. and P.C.; investigation, N.N.T. and P.C.; original draft preparation, N.N.T. and P.C.; review and editing, N.N.T. and P.C.; visualization, P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We thank two anonymous referees for their careful comments and suggestions, which helped us improve the quality of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Aliferis, Constantin, and Gyorgy J. Simon. 2024. Overfitting, Underfitting and General Model Overconfidence and Under-Performance: Pitfalls and Best Practices in Machine Learning and AI. In *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences*. Edited by Gyorgy J. Simon and Constantin Aliferis. Berlin and Heidelberg: Springer.

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. *arXiv* arXiv:1606.06565. [CrossRef]

Andrews, David F., and Colin L. Mallows. 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B* 36: 99–102. [CrossRef]

Aquinas, Thomas. 1256. *De Veritate*. English translation in *Truth*. 1994. Indianapolis: Hackett Publishing Company.

Berger, James O., José M. Bernardo, and Dongchu Sun. 2024. *Objective Bayesian Inference*. Singapore: World Scientific.

Bernardo, José M., and Adrian F. M. Smith. 2000. *Bayesian Theory*. Hoboken: Wiley.

Bradley, Francis H. 1914. *Essays on Truth and Reality*. Oxford: Clarendon Press.

Breiman, Leo. 1965. On some limit theorems for tail probabilities. *Journal of the American Statistical Association* 60: 185–88.

Childers, Timothy. 2013. *Philosophy and Probability*. Oxford: Oxford University Press.

Cirillo, Pasquale. 2013. Are your data really Pareto distributed? *Physica A: Statistical Mechanics and Its Applications* 392: 5947–62. [CrossRef]

Cirillo, Pasquale, and Nassim Nicholas Taleb. 2020. Tail risk of contagious diseases. *Nature Physics* 16: 606–13. [CrossRef]

Credit Suisse. First Boston International. 1997. CreditRisk+: A Credit Risk Management Framework. Available online: https://globalriskguard.com/resources/credit/creditrisk.pdf (accessed on 2 December 2025).

de Finetti, Bruno. 1974. *Theory of Probability*. Hoboken: Wiley, vol. 1.

de Finetti, Bruno. 1975. *Theory of Probability*. Hoboken: Wiley, vol. 2.

de Finetti, Bruno. 2006. *L'invenzione della verit'a*. Milan: R. Cortina.

de Haan, Laurens, and Ana F. Ferreira. 2006. *Extreme Value Theory. An Introduction*. New York: Springer.

Derman, Emanuel. 1996. Model Risk. *Risk* 9: 139–45.

Draper, David. 1995. Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 45–97. [CrossRef]

Dupire, Bruno. 1994. Pricing with a Smile. *Risk* 7: 18–20.

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 2003. *Modelling Extremal Events*, 2nd ed. Berlin and Heidelberg: Springer.

Feller, William. 1968. *An Introduction to Probability Theory and Its Applications*, 3rd ed. Hoboken: Wiley, vol. 1.

Foss, Serguei, Dmitry Korshunov, and Stan Zachary. 2013. *An Introduction to Heavy-Tailed and Subexponential Distributions*, 2nd ed. Berlin and Heidelberg: Springer.

Gal, Yarin, and Zoubin Ghahramani. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Paper presented at the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, June 19–24; pp. 1050–59.

Gauss, Carl F. 1809. *Theoria Motus Corporum Coelestium*. Washington, DC: Biodiversity Heritage Library.

Gibrat, Robert. 1931. *Les In'egalit'es 'economiques*. Paris: Recueil Sirey.

Gigerenzer, Gerd. 2002. *Reckoning with Risk: Learning to Live with Uncertainty*. London: Penguin.

Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On Calibration of Modern Neural Networks. Paper presented at the 34th International Conference on Machine Learning (ICML), Sydney, Australia, August 6–11, vol. 70, pp. 1321–30.

Halpern, Joseph Y. 2005. *Reasoning About Uncertainty*. Cambridge, MA: MIT Press.

Hausfather, Zeke, and Glen P. Peters. 2020. Emissions—The "business as usual" story is misleading. *Nature* 577: 618–20. [CrossRef]

Hull, John. 2023. *Risk Management and Financial Institutions*. Hoboken: Wiley.

IPCC. 2023. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by Hoesung Lee and José Romero. Geneva: IPCC.

Johnson, Norman, Samuel Kotz, and Narayanaswamy Balakrishnan. 1994. *Continuous Univariate Distributions*. Hoboken: Wiley, vol. 1.

Kendall, Alex, and Yarin Gal. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Advances in Neural Information Processing Systems (NeurIPS)* 30: 5574–84.

Kesten, Harry. 1973. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131: 207–48. [CrossRef]

Kleiber, Christian, and Samuel Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken: Wiley.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 1998. *Loss Models: From Data to Decisions*. Hoboken: Wiley.

Knight, Frank H. 1921. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin Company.

Kolmogorov, Andrey N. 1962. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics* 13: 82–85. [CrossRef]

Laplace, Pierre-Simon. 1814. *Essai Philosophique sur les Probabilit'es*. Translated by Eric T. Bell. 1952. Garden City: Dover.

Lemoine, Derek, and Christian P. Traeger. 2014. Watch Your Step: Optimal Policy in a Tipping Climate. *American Economic Journal: Economic Policy* 6: 137–66. [CrossRef]

Lempert, Robert J., Steven W. Popper, and Steven C. Bankes. 2003. *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. Santa Monica: RAND Corporation.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge, MA: MIT Press.

Lewis, David. 1973. *Counterfactuals*. Hoboken: Wiley.

Lindley, Dennis. 1991. *Making Decisions*. Hoboken: Wiley.

Liu, Yang, Yuling Yao, Jean-François Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. 2023. Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment. *arXiv* arXiv:2308.05374.

Mandelbrot, Benoit B. 1974. Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics* 62: 331–58. [CrossRef]

Marchau, Vincent A. W. J., Warren E. Walker, Pieter J. T. M. Bloemen, and Steven W. Popper, eds. 2019. *Decision Making Under Deep Uncertainty: From Theory to Practice*. Cham: Springer.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*. Boca Raton: Chapman and Hall/CRC.

McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management*. Princeton: Princeton University Press.

Millner, Antony, Simon Dietz, and Geoffrey Heal. 2010. *Ambiguity and Climate Policy. NBER Working Paper No. 16050*. Cambridge, MA: National Bureau of Economic Research.

Nair, Jayakrishnan, Adam Wierman, and Bert Zwart. 2022. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge: Cambridge University Press.

National Research Council. 2013. *Climate and Social Stress: Implications for Security Analysis*. Washington, DC: The National Academies Press.

Nietzsche, Friedrich W. 1873. *Uber Wahrheit und Lüge im Aussermoralischen Sinne*. English translation in *The Portable Nietzsche* in 1996. New York: Viking Press.

Ovadia, Yaniv, Elad Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, and Balaji Lakshminarayanan. 2019. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift. *Advances in Neural Information Processing Systems (NeurIPS)* 32: 13991–4002.

Petropoulos, Fotios, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K. Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J. Bessa, Jakub Bijak, John E. Boylan, and et al. 2022. Forecasting: Theory and practice. *International Journal of Forecasting* 38: 705–871. [CrossRef]

Rescher, Nicholas. 1983. *Risk—A Philosophical Introduction to the Theory of Risk Evaluation and Management*. Lanham: University Press of America.

Russell, Bertrand. 1958. *The Will to Doubt*. New York: Philosophical Library.

Rüschendorf, Ludger, Steven Vanduffel, and Carole Bernard. 2023. *Model Risk Management: Risk Bounds Under Uncertainty*. Singapore: Cambridge University Press.

Shackle, George L. S. 1968. *Expectations, Investment and Income*. Oxford: Clarendon Press.

Shao, Jun. 1998. *Mathematical Statistics*. Berlin and Heidelberg: Springer.

Taleb, Nassim Nicholas. 2021. *Incerto*. New York and London: Random House and Penguin.
Taleb, Nassim Nicholas. 2025. *The Statistical Consequence of Fat Tails*, 3rd ed. Cambridge: STEM Academic Press.
Taleb, Nassim Nicholas, and Avital Pilpel. 2004. I problemi epistemologici del risk management. In *Economia del rischio*. Edited by Daniele Pace. Milano: Giuffrè.
Taleb, Nassim Nicholas, and Avital Pilpel. 2007. Epistemology and Risk Management. *Risk and Regulation* 13: 6–7.
Taleb, Nassim Nicholas, Rupert Read, Raphael Douady, Joseph Norman, and Yaneer Bar-Yam. 2014. The Precautionary Principle (with Application to the Genetic Modification of Organisms). *arXiv* arXiv:1410.5787. [CrossRef]
Taleb, Nassim Nicholas, Yaneer Bar-Yam, and Pasquale Cirillo. 2022. On single point forecasts for fat-tailed variables. *International Journal of Forecasting* 38: 413–22. [CrossRef]
Viertl, Reinhard. 1997. On Statistical Inference for Non-Precise Data. *Environmetrics* 8: 541–68. [CrossRef]
von Plato, Jan. 1994. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. Cambridge: Cambridge University Press.
Weitzman, Martin L. 2009. On Modeling and Interpreting the Economics of Catastrophic Climate Change. *Review of Economics and Statistics* 91: 1–19. [CrossRef]
West, Mike. 1987. On scale mixtures of normal distributions. *Biometrika* 74: 646–48. [CrossRef]
Williamson, Jon. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.
Zinn-Justin, Jean. 2021. *Quantum Field Theory and Critical Phenomena*, 5th ed. Oxford: Oxford University Press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.