4 | EFFECTS OF HIGHER ORDERS OF UNCERTAINTY

Chapter Summary 4: The Spectrum Between Uncertainty and Risk. There has been a bit of discussions about the distinction between "uncertainty" and "risk". We believe in gradation of uncertainty at the level of the probability distribution itself (a "meta" or higher order of uncertainty.) One end of the spectrum, "Knightian risk", is not available for us mortals in the real world. We show how the effect on fat tails and on the calibration of tail exponents and reveal inconsistencies in models such as Markowitz or those used for intertemporal discounting (as many violations of "rationality" aren't violations .

4.1   Meta-Probability Distribution

When one assumes knowledge of a probability distribution, but has uncertainty attending the parameters, or when one has no knowledge of which probability distribution to consider, the situation is called "uncertainty in the Knightian sense" by decision theorisrs(Knight, 1923). "Risk" is when the probabilities are computable without an error rate. Such an animal does not exist in the real world. The entire distinction is a lunacy, since no parameter should be rationally computed witout an error rate. We find it preferable to talk about degrees of uncertainty about risk/uncertainty, using metadistribution, or metaprobability.

The Effect of Estimation Error, General Case

The idea of model error from missed uncertainty attending the parameters (another layer of randomness) is as follows.

Most estimations in social science, economics (and elsewhere) take, as input, an average or expected parameter,

$$\bar{\alpha} = \int \alpha \ \phi(\alpha) \ d\alpha, \tag{4.1}$$

where α is ϕ distributed (deemed to be so a priori or from past samples), and regardless of the dispersion of α, build a probability distribution for x that relies on the mean estimated parameter, $p(X = x) = p\left(x \left| \bar{\alpha} \right. \right)$, rather than the more appropriate metaprobability adjusted probability for the density:
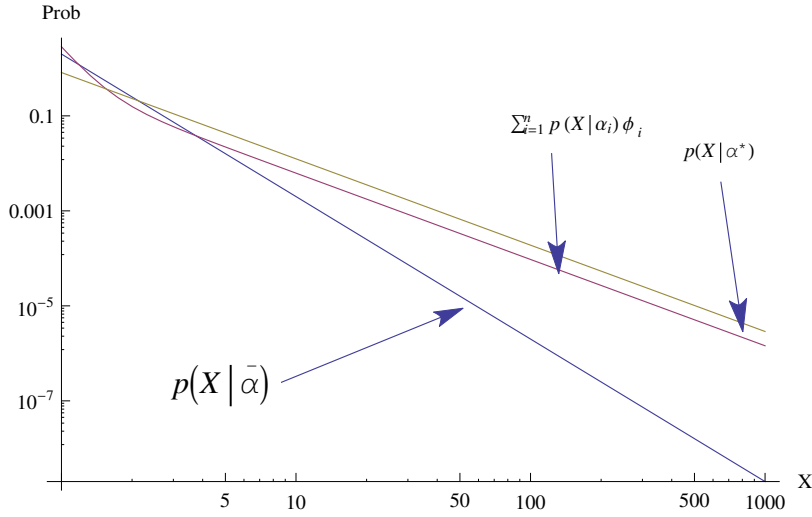
$$p(x) = \int \phi(\alpha) \, d\alpha \tag{4.2}$$

Figure 4.1: Log-log plot illustration of the asymptotic tail exponent with two states.

In other words, if one is not certain about a parameter α, there is an inescapable layer of stochasticity; such stochasticity raises the expected (metaprobability-adjusted) probability if it is $< \frac{1}{2}$ and lowers it otherwise. The uncertainty is fundamentally epistemic, includes incertitude, in the sense of lack of certainty about the parameter.

The model bias becomes an equivalent of the Jensen gap (the difference between the two sides of Jensen's inequality), typically positive since probability is convex away from the center of the distribution. We get the bias ω_A from the differences in the steps in integration

$$\omega_A = \int \phi(\alpha)\, p(x|\alpha)\, \mathrm{d}\alpha - p\left(x | \int \alpha\phi(\alpha)\, \mathrm{d}\alpha\right)$$

With $f(x)$ a function , $f(x) = x$ for the mean, etc., we get the higher order bias $\omega_{A'}$

$$\omega_{A'} = \int \left(\int \phi(\alpha)\ f(x)\ p(x|\alpha)\ d\alpha \right)\, \mathrm{d}x - \int f(x)\ p\left(x | \int \alpha\ \phi(\alpha)\, \mathrm{d}\alpha\right)\, \mathrm{d}x \quad (4.3)$$

Now assume the distribution of α as discrete n states, with $\alpha = (\alpha_i)_{i=1}^n$ each with associated probability $\phi = \phi_i\,\_i{=}1{\hat{}}\mathrm{n}, \ \ \sum_{i=1}^n \phi_i = 1$. Then 4.2 becomes

$$p(x) = \phi_i \left(\sum_{i=1}^n p\left(x\,|\alpha_i\right) \right) \quad (4.4)$$

So far this holds for α any parameter of any distribution.

4.2   Metadistribution and the Calibration of Power Laws

Remark 1. *In the presence of a layer of metadistributions (from uncertainty about the parameters), the asymptotic tail exponent for a powerlaw corresponds to the lowest possible tail exponent regardless of its probability.*

This explains "Black Swan" effects, i.e., why measurements tend to chronically under-estimate tail contributions, rather than merely deliver imprecise but unbiased estimates.

When the perturbation affects the standard deviation of a Gaussian or similar non-powerlaw tailed distribution, the end product is the weighted average of the probabilities. However, a powerlaw distribution with errors about the possible tail exponent will bear the asymptotic properties of the *lowest* exponent, not the average exponent.

Now assume p(X=x) a standard Pareto Distribution with α the tail exponent being estimated, $p(x|\alpha) = \alpha x^{-\alpha-1} x_{\min}^{\alpha}$, where $x_{\min}$ is the lower bound for x,

$$p(x) = \sum_{i=1}^{n} \alpha_i x^{-\alpha_i-1} x_{\min}^{\alpha_i} \phi_i$$

Taking it to the limit

$$\lim_{x\to\infty} x^{\alpha^*+1} \sum_{i=1}^{n} \alpha_i x^{-\alpha_i-1} x_{\min}^{\alpha_i} \phi_i = K$$

where K is a strictly positive constant and $\alpha^* = \min_{1\leq i\leq n} \alpha_i$. In other words $\sum_{i=1}^{n} \alpha_i x^{-\alpha_i-1} x_{\min}^{\alpha_i} \phi_i$ is asymptotically equivalent to a constant times x^{α^*+1}. The lowest parameter in the space of all possibilities becomes the dominant parameter for the tail exponent.

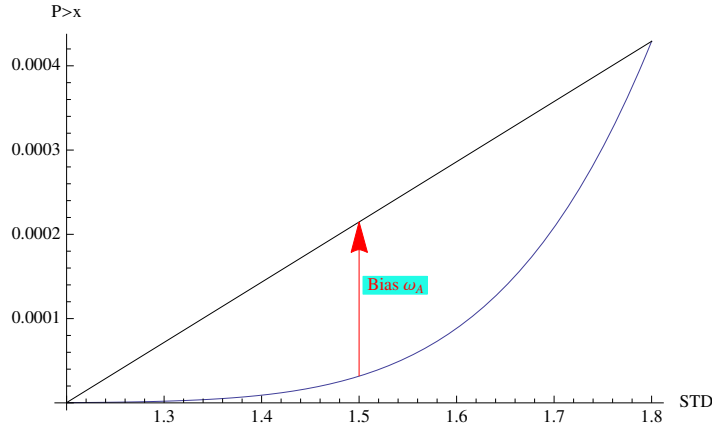*Figure 4.2: Illustration of the convexity bias for a Gaussian from raising small probabilities: The plot shows the STD effect on P>x, and compares P>6 with a STD of 1.5 compared to P> 6 assuming a linear combination of 1.2 and 1.8 (here a(1)=1/5).*

Figure 4.1 shows the different situations: a) $p(x|\bar{\alpha})$, b) $\sum_{i=1}^{n} p(x|\alpha_i) \phi_i$ and c) $p(x|\alpha^*)$. We can see how the last two converge. The asymptotic Jensen Gap ω_A becomes $p(x|\alpha^*) - p(x|\bar{\alpha})$.

Implications

Whenever we estimate the tail exponent from samples, we are likely to underestimate the thickness of the tails, an observation made about Monte Carlo generated α-stable variates and the estimated results (the "Weron effect")[74].

The higher the estimation variance, the lower the true exponent.

The asymptotic exponent is the lowest possible one. It does not even require estimation.

Metaprobabilistically, if one isn't sure about the probability distribution, and there is a probability that the variable is unbounded and "could be" powerlaw distributed, then it is powerlaw distributed, and of the lowest exponent.

The obvious conclusion is to in the presence of powerlaw tails, focus on changing payoffs to clip tail exposures to limit $\omega_{A'}$ and "robustify" tail exposures, making the computation problem go away.

4.3   The Effect of Metaprobability on Fat Tails

Recall that the tail fattening methods in 2.4 and 2.6.These are based on randomizing the variance. Small probabilities rise precisely because they are convex to perturbations of the parameters (the scale) of the probability distribution.

4.4   Fukushima, Or How Errors Compound

"Risk management failed on several levels at Fukushima Daiichi. Both TEPCO and its captured regulator bear responsibility. First, highly tailored geophysical models predicted an infinitesimal chance of the region suffering an earthquake as powerful as the Tohoku quake. This model uses historical seismic data to estimate the local frequency of earthquakes of various magnitudes; none of the quakes in the data was bigger than magnitude 8.0. Second, the plant's risk analysis did not consider the type of cascading, systemic failures that precipitated the meltdown. TEPCO never conceived of a situation in which the reactors shut down in response to an earthquake, and a tsunami topped the seawall, and the cooling pools inside the reactor buildings were overstuffed with spent fuel rods, and the main control room became too radioactive for workers to survive, and damage to local infrastructure delayed reinforcement, and hydrogen explosions breached the reactors' outer containment structures. Instead, TEPCO and its regulators addressed each of these risks independently and judged the plant safe to operate as is."Nick Werle, n+1, published by the n+1 Foundation, Brooklyn NY

4.5   The Markowitz inconsistency

Assume that someone tells you that the probability of an event is exactly zero. You ask him where he got this from. "Baal told me" is the answer. In such case, the person is coherent, but would be deemed unrealistic by non-Baalists. But if on the other hand, the person tells you "I estimated it to be zero," we have a problem. The person is both unrealistic and inconsistent. Something estimated needs to have an estimation error. So probability cannot be zero if it is estimated, its lower bound is linked to the estimation error; the higher the estimation error, the higher the probability, up to a point. As with Laplace's argument of total ignorance, an infinite estimation error pushes the probability toward $\frac{1}{2}$. We will return to the implication of the mistake; take for now that anything estimating a parameter and then putting it into an equation is different from estimating the equation across parameters. And Markowitz was inconsistent by starting his "seminal" paper with "Assume you know E and V" (that is, the expectation and the variance). At the end of the paper he accepts that they need to be estimated, and what is worse, with a combination of statistical techniques and the "judgment of practical men." Well, if these parameters need to be estimated, with an error, then the derivations

need to be written differently and, of course, we would have no such model. Economic models are extremely fragile to assumptions, in the sense that a slight alteration in these assumptions can lead to extremely consequential differences in the results. The perturbations can be seen as follows. Let $\vec{X} = (X_1, X_2, \ldots, X_m)$ be the vector of random variables representing returns. Consider the joint probability distribution $f(x_1, \ldots, x_m)$. We denote the m-variate multivariate Normal distribution by $N(\vec{\mu}, \Sigma)$, with mean vector $\vec{\mu}$, variance-covariance matrix Σ, and joint pdf,

$$f\left(\vec{x}\right) \;=\; (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\left(\vec{x} - \vec{\mu}\right)^T \Sigma^{-1} \left(\vec{x} - \vec{\mu}\right)\right) \qquad (4.5)$$

where $\vec{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, and Σ is a symmetric, positive definite $(m \times m)$ matrix. The weights matrix $\vec{\Omega} = (\omega_1, \ldots, \omega_m)$, normalized, with $\sum_{i=1}^{N} \omega_i = 1$ (allowing exposures to be both positive and negative): The scalar of concern is; $r = \Omega^T.X$, which happens to be normally distributed, with variance

$$v = \vec{\omega}^T.\Sigma.\vec{\omega}$$

The Markowitz portfolio construction, through simple optimization, gets an optimal $\vec{\omega}^*$, obtained by, say, minimizing variance under constraints, getting the smallest $\vec{\omega}^T.\Sigma.\vec{\omega}$ under constraints of returns, a standard Lagrange multiplier. So done statically, the problem gives a certain result that misses the metadistribution. Now the problem is that the covariance matrix is a random object, and needs to be treated as so. So let us focus on what can happen under these conditions:

Route 1: The stochastic volatility route. This route is insufficient but can reveal structural defects for the construction. We can apply the same simplied variance pre-serving heuristic as in 2.4 to fatten the tails. Where a is a scalar that determines the intensity of stochastic volatility, $\Sigma_1 = \Sigma(1 - a)$ and $\Sigma_2 = \Sigma(1 - a)$. Simply, given the conservation of the Gaussian distribution under weighted summation, maps to $v(1 + a)$ and $v(1 - a)$ for a Gaussian and we could see the same effect as in 2.4. The corresponding increase in fragility is explained in Chapter 15.

Route 2: Full random parameters route. Now one can have a fully random matrix —not just the overal level of the covariance matrix. The problem is working with matrices is cumbersome, particularly in higher dimensions, because one element of the covariance can vary unconstrained, but the degrees of freedom are now reduced for the matrix to remain positive definite. A possible technique is to extract the principal components, necessarily orthogonal, and randomize them without such restrictions.

4.6 Psychological pseudo-biases under second layer of uncertainty.

Often psychologists and behavioral economists find "irrational behavior" (or call it under something more polite like "biased") as agents do not appear to follow a normative model and violate their models of rationality. But almost all these correspond to missing a second layer of uncertainty by a tinky-toy first-order model that doesn't get nonlinearities − it is the researcher who is making a mistake, not the real-world agent. Recall that the expansion from "small world" to "larger world" can be simulated by perturbation of

Figure 4.3:   The effect of $H_{a,p}(t)$ "utility" or prospect theory of under second order effect on variance. Here $\sigma = 1, \mu = 1$ and t variable.
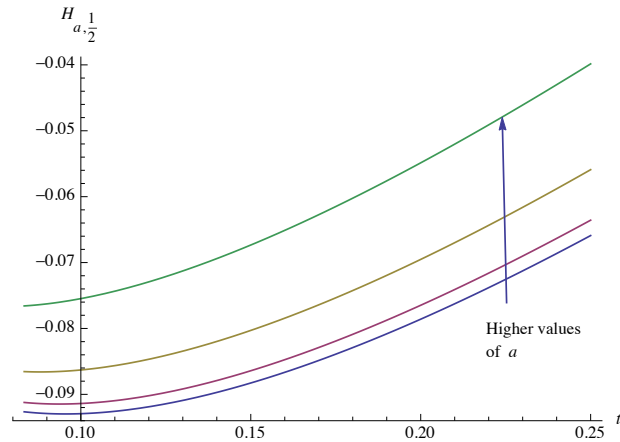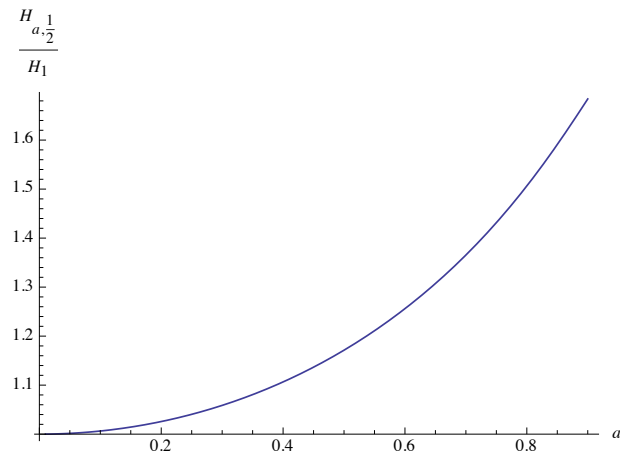

Figure 4.4: The ratio $\frac{H_{a,\frac{1}{2}}(t)}{H_0}$ or the degradation of "utility" under second order effects.

parameters, or "stochasticization", that is making something that appears deterministic a random variable itself. Benartzi and Thaler [3], for instance, find an explanation that agents are victims of a disease labelled "myopic loss aversion" in not investing enough in equities, not realizing that these agents may have a more complex, fat-tailed model. Under fat tails, no such puzzle exists, and if it does, it is certainly not from such myopia.

This approach invites "paternalism" in "nudging" the preferences of agents in a manner to fit professors-without-skin-in-the-game-using-wrong-models.

The problem also applies to GMOs and how "risk experts" find them acceptable; researchers pathologize those who do not partake of the baby models (thin tailed). The point, an extension of the Pinker problem, is discussed in Chapter x.

Let us use our approach in detecting convexity to three specific problems: 1) the myopic loss aversion that we just discussed, 2) time preferences, 3) probability matching.

4.6.1   Myopic loss aversion

Take the prospect theory valuation w function for x changes in wealth.

$$w_{\lambda,\alpha}(x) = x^\alpha \, \mathbb{1}_{x \geq 0} - \lambda(-x^\alpha) \, \mathbb{1}_{x < 0}$$

Where $\phi_{\mu t, \sigma \sqrt{t}}(x)$ is the Normal Distribution density with corresponding mean and standard deviation (scaled by t)

The expected "utility" (in the prospect sense):

$$H_0(t) = \int_{-\infty}^{\infty} w_{\lambda,\alpha}(x)\phi_{\mu t,\sigma\sqrt{t}}(x)\,\mathrm{d}x \tag{4.6}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\pi}} 2^{\frac{\alpha}{2}-2} \left(\frac{1}{\sigma^2 t}\right)^{-\frac{\alpha}{2}} \\
&\left(\Gamma\left(\frac{\alpha+1}{2}\right)\left(\sigma^{\alpha}t^{\alpha/2}\left(\tfrac{1}{\sigma^2 t}\right)^{\alpha/2} - \lambda\sigma\sqrt{t}\sqrt{\tfrac{1}{\sigma^2 t}}\right) {}_1F_1\left(-\tfrac{\alpha}{2};\tfrac{1}{2};-\tfrac{t\mu^2}{2\sigma^2}\right)\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \tfrac{1}{\sqrt{2}\sigma}\mu\Gamma\left(\tfrac{\alpha}{2}+1\right) \\
&\left.\left(\sigma^{\alpha+1}t^{\frac{\alpha}{2}+1}\left(\tfrac{1}{\sigma^2 t}\right)^{\frac{\alpha+1}{2}} + \sigma^{\alpha}t^{\frac{\alpha+1}{2}}\left(\tfrac{1}{\sigma^2 t}\right)^{\alpha/2} + 2\lambda\sigma t\sqrt{\tfrac{1}{\sigma^2 t}}\right) {}_1F_1\left(\tfrac{1-\alpha}{2};\tfrac{3}{2};-\tfrac{t\mu^2}{2\sigma^2}\right)\right)
\end{aligned}
\tag{4.7}
$$

We can see from 4.7 that the more frequent sampling of the performance translates into worse utility. So what Benartzi and Thaler did was try to find the sampling period "myopia" that translates into the sampling frequency that causes the "premium" —the error being that they missed second order effects.

Now under variations of σ with stochatic effects, heuristically captured, the story changes: what if there is a very small probability that the variance gets multiplied by a large number, with the total variance remaining the same? The key here is that we are not even changing the variance at all: we are only shifting the distribution to the tails. We are here generously assuming that by the law of large numbers it was established that the "equity premium puzzle" was true and that stocks *really* outperformed bonds.

So we switch between two states, $(1+a)\,\sigma^2$ w.p. p and $(1-a)$ w.p. $(1-p)$.

Rewriting 4.6

$$H_{a,p}(t) = \int_{-\infty}^{\infty} w_{\lambda,\alpha}(x)\left(p\,\phi_{\mu t,\sqrt{1+a}\,\sigma\sqrt{t}}(x) + (1-p)\,\phi_{\mu t,\sqrt{1-a}\,\sigma\sqrt{t}}(x)\right)\mathrm{d}x \tag{4.8}$$

Result. Conclusively, as can be seen in figures 4.3 and 4.4, second order effects cancel the statements made from "myopic" loss aversion. This doesn't mean that myopia doesn't have effects, rather that it cannot explain the "equity premium", not from the outside (i.e. the distribution might have different returns", but from the inside, owing to the structure of the Kahneman-Tversky value function $v(x)$.

Comment. We used the (1+a) heuristic largely for illustrative reasons; we could use a full distribution for σ^2 with similar results. For instance the gamma distribution with density $f(v) = \frac{v^{\gamma-1}e^{-\frac{\alpha v}{V}}\left(\frac{V}{\alpha}\right)^{-\gamma}}{\Gamma(\gamma)}$ with expectation V matching the variance used in the "equity premium" theory.

Rewriting 4.8 under that form,

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} w_{\lambda,\alpha}(x)\phi_{\mu t,\sqrt{v\,t}}(x)\,f(v)\,\mathrm{d}v\,\mathrm{d}x$$

Which has a closed form solution (though a bit lengthy for here).

4.6.2   Time preference under model error

This author once watched with a great deal of horror one Laibson [37] at a conference in Columbia University present the idea that having one massage today to two tomorrow, but reversing in a year from nowm is irrational and we need to remedy it with some policy. (For a review of time discounting and intertemporal preferences, see [27], as economists temps to impart what seems to be a varying "discount rate" in a simplified model).

Intuitively, what if I introduce the probability that the person offering the massage is full of balloney? It would clearly make me both prefer immediacy at almost any cost and conditionally on his being around at a future date, reverse the preference. This is what we will model next.

First, time discounting has to have a geometric form, so preference doesn't become negative: linear discounting of the form Ct, where C is a constant ant t is time into the future is ruled out: we need something like C^t or, to extract the rate, $(1+k)^t$ which can be mathematically further simplified into an exponential, by taking it to the continuous time limit. Exponential discounting has the form e^{-kt}. Effectively, such a discounting method using a shallow model prevents "time inconsistency", so with $\delta < t$:

$$\lim_{t \to \infty} \frac{e^{-kt}}{e^{-k(t-\delta)}} = e^{-k\delta}$$

Now add another layer of stochasticity: the discount parameter, for which we use the symbol λ, is now stochastic.

So we now can only treat $H(t)$ as

$$H(t) = \int e^{-\lambda t} \phi(\lambda) \, d\lambda$$

It is easy to prove the general case that under symmetric stochasticization of intensity $\Delta\lambda$ (that is, with probabilities $\frac{1}{2}$ around the center of the distribution) using the same technique we did in 2.4:

$$H'(t, \Delta\lambda) = \frac{1}{2}\left(e^{-(\lambda-\Delta\lambda)t} + e^{-(\lambda+\Delta\lambda)t}\right)$$

$$\frac{H'(t, \Delta\lambda)}{H'(t,0)} = \frac{1}{2}e^{\lambda t}\left(e^{(-\Delta\lambda-\lambda)t} + e^{(\Delta\lambda-\lambda)t}\right) = \cosh(\Delta\,\lambda t)$$

Where cosh is the cosine hyperbolic function − which will converge to a certain value where intertemporal preferences are flat in the future.

Example: Gamma Distribution. Under the gamma distribution with support in $\mathbb{R}^+$, with parameters α and β, $\phi(\lambda) = \frac{\beta^{-\alpha}\lambda^{\alpha-1}e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha)}$ we get:

$$H(t, \alpha, \beta) = \int_0^\infty e^{-\lambda t} \frac{\left(\beta^{-\alpha}\lambda^{\alpha-1}e^{-\frac{\lambda}{\beta}}\right)}{\Gamma(\alpha)} \, d\lambda = \beta^{-\alpha}\left(\frac{1}{\beta} + t\right)^{-\alpha}$$

so

$$\lim_{t \to \infty} \frac{H(t, \alpha, \beta)}{H(t-\delta, \alpha, \beta)} = 1$$

Meaning that preferences become flat in the future no matter how steep they are in the present, which explains the drop in discount rate in the economics literature.

Further, fudging the distribution and normalizing it, when

$$\phi(\lambda) = \frac{e^{-\frac{\lambda}{k}}}{k},$$

we get the *normatively obtained* (not empirical pathology) so-called hyperbolic discounting:

$$H(t) = \frac{1}{1 + k\,t}$$