Rational Turbulence

Kent Osband[1]

Draft
July 2011

Abstract

Fluids are turbulent when tiny differences in space or time make for gross differences in behavior. The mathematical signature of turbulence is an endless moment or cumulant hierarchy. Bayesian tracking of continuous-time processes turns out to have a similar mathematical structure. As a result, tiny doubts about regime change or tiny errors in estimation or calculation are prone under stress to balloon into gross differences of opinion. In effect, reasonable people are bound to disagree. This finding has profound implications for our understanding of financial markets and other instruments of social learning. In particular it explains forecast degradation, frequent trading, excess volatility, and GARCH behavior without imputing widespread irrationality. Rational learning makes markets turbulent.

.

[1] RiskTick, LLC. I thank Annette Osband for encouragement and editing corrections. To comment or request the latest version, please write rationalturbulence@risktick.com. For more background and elaboration, see Pandora's Risk: Uncertainty at the Core of Finance (Columbia Business School Publishing, 2011), especially Chapters 6, 7 and 11.

Summary

Fluids are turbulent when tiny differences in space or time make for gross differences in behavior. The mathematical signature of turbulence is an endless moment hierarchy, where the change in each moment depends on the moment above. Bayesian tracking of continuous-time processes turns out to have a similar mathematical structure, which is most neatly expressed using cumulants.

Rational turbulence means that tiny doubts about regime change or tiny errors in estimation or calculation are prone under stress to balloon into gross differences of opinion. In effect, reasonable people are bound to disagree. The only exceptions involve natural phenomena with unchanging laws of motion.

This finding has profound implications for our understanding of financial markets and other instruments of social learning. In particular it explains forecast degradation, frequent trading, excess volatility, and GARCH behavior in terms of rational learning. In a sense it salvages the insights of Mandelbrot and other critics of orthodox finance, but without imputing widespread irrationality. Rational learning makes markets turbulent.

1. Introduction

Why do reasonable people evaluating the same evidence disagree? This is a harder question than it seems. The first temptation is to define it away. Perhaps people only think they're being reasonable, and the disagreement shows that at least one of them is not. Perhaps people are misreading the evidence or infusing tangential observations.

Standard Bayesian theory offers a partial solution. it defines agreement as a common relative weighting, or "subjective probability distribution", over beliefs. It defines reasonableness as applying Bayes' rule to update their "prior" beliefs. Under these definitions, reasonable people evaluating the same evidence may disagree strongly if their priors substantially diverge.

This is only a partial solution because the prior's influence is inherently transient. As more evidence accumulates, the weight assigned to the prior shrinks. A classic example is the measurement of a tiny particle through repeated sampling. Over time, all beliefs converge to the sample mean.

On reflection, a crucial driver in this result is the notion of "independent, identically distributed” (i.i.d.) sampling. Every particle of the same type is assumed to have identical statistical properties, which neither time nor sampling significantly alter. In physics, this assumption is so grounded in experience, and so consistent with generations of tests, that it is treated as a fundamental property of matter. While the uncertainty principle is bound to abridge i.i.d. behavior, the deviations are typically irrelevant to sample means.

Outside the natural sciences, i.i.d. assumptions are much less compelling. Each new piece of evidence requires in effect a dual assessment. First, how reliable is it as a sample of past behavior? Second, how relevant is it to future behavior?

For example, in finance, the prices of trillions of dollars of bonds depend critically on the perceived default risks of the issuers. A perceived shift of one percentage point in annual default rates -- that is, a mean change of 1 default per century --- can have a big impact on valuations. Yet for many issuers it is impossible to identify one hundred years of unambiguously relevant observations, much less i.i.d observations. Indeed, given the variations across issuers and the flux in economic and political environments, it is rare to identify even two i.i.d. observations to the standards physics expects.

We can easily modify the Bayesian framework to handle dual assessments of relevance and import. When we do, we find that uncertainty needn't vanish in the limit. Moreover, reasonable observers needn't converge to the same uncertainty. Instead, disagreements are bound to flare up and recede. Every resolution bodes new divergence.

Mathematically, the problem stems from an inherent shortcoming of general Bayesian updating in continuous time. While a single equation can easily summarize ideal learning, no finite processor can accurately solve it. Moreover, small errors can blow up quickly into big ones.

Formally, when all beliefs are rationally updated, the volatility of each cumulant of beliefs is a multiple of the cumulant one order higher. This typically implies an endless cumulant hierarchy, in which tiny differences in outlier opinion can quickly affect the mainstream response to major regime change. The only exceptions involve perfectly Gaussian beliefs, in which case the optimal response is a Kalman filter.

Cumulant hierarchies arise naturally in physics. There they are associated with turbulence: small differences in location or time that make big differences in behavior. Given the parallels in the governing equations, non-Gaussian Bayesian updating can be viewed as creating "rational turbulence".

This paper is organized as follows. Section 2 presents some simple examples of Bayesian disagreement. Section 3 analyzes continuous-time jump processes and reveals a cumulant hierarchy of Bayesian updating. Section 4 extends the analysis to Brownian motion with unstable drift. Section 5 sketches some practical implications for market prices. Section 6 explores the connections between cumulant hierarchies and turbulence. Section 7 concludes with some reflections on the broader implications.

2. Bayesian Disagreement

Let us define “regimes” $\{i\}$ by the sets of parameters that govern their probabilistic behavior, and “beliefs” $\{p_i\}$ by the subjective likelihoods attached to various regimes. Given fresh evidence x, Bayes’ rule says to revise beliefs in proportion to the conditional probabilities $f\left(x|i\right)$ that regimes would generate that evidence:

$$p_{new}(i) \propto f\left(x|i\right) p_{old}(i). \tag{1}$$

For any two regimes and *j*, a bit of algebra establishes that

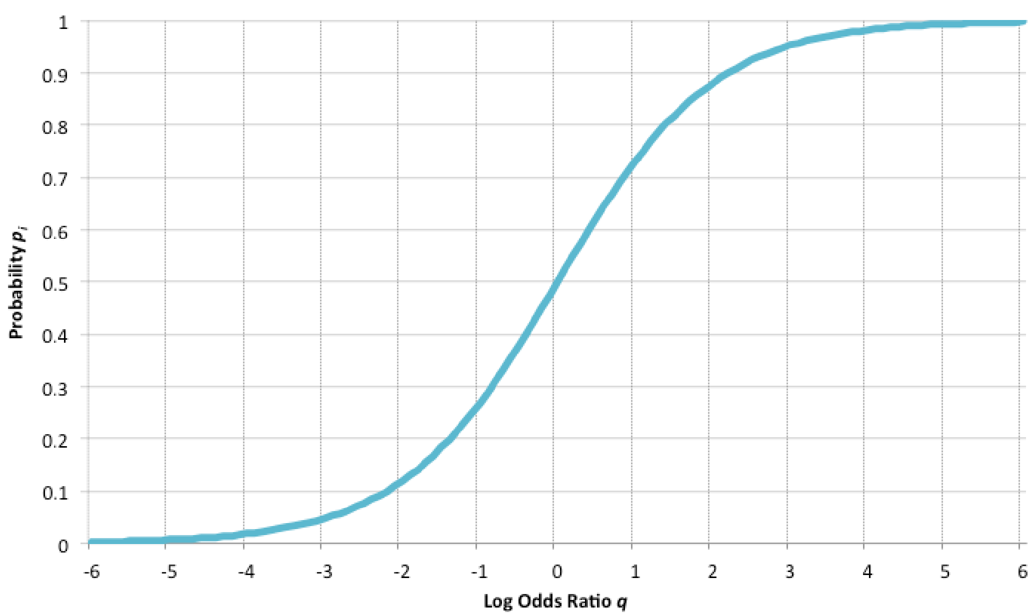
$$\Delta \ln\left(\frac{p(i)}{p(j)}\right) = \ln\left(\frac{f\left(x|i\right)}{f\left(x|j\right)}\right), \tag{2}$$

where Δ indicates the difference between the old value and the new value. The left-hand side, which specifies the updating of beliefs, is known as the “log-odds ratio”. The right-hand side depends only on the conditional probabilities and is known as the “weight of the evidence” (Kullback, 1968).

Equation (2) should apply for every Bayesian-rational observer, regardless of starting priors. Each observer should shift her log odds ratios by the same amounts. Hence, measured in log odds space, observation should neither fan disagreement nor quell it.

However, we usually measure disagreement by differences in ordinary probabilities. Conditional on either regime i or regime j applying, ordinary probability p_i is a logistic function of the log odds $q \equiv \log\left(\frac{p_i}{1-p_i}\right)$; that is, $p_i = \frac{1}{1+e^{-q}}$. The logistic function, as Figure 1 shows, is highly nonlinear. A constant horizontal gap in terms of log odds implies a gyrating vertical gap in terms of probability. The vertical gap is huge when people are highly uncertain (q near zero) and tiny when people are nearly certain (q very large in absolute value).

Figure 1: Probability versus Log Odds



For example, suppose Punch and Judy are both 99.99+% convinced that regime 1 applies, with Punch according a 1 in a million likelihood of regime 2 and Judy according a 1 in a billion likelihood. In most practical applications it would be extremely difficult to distinguish those views or confirm they're not bluster. However, the implied $\log(1000)$ gap in log odds is huge. So if the evidence gradually reveals that regime 2 does apply, the day is bound to come when

Punch is more than 95% convinced while Judy is less than 5% convinced, since $\log\left(\frac{95}{5}\right)-\log\left(\frac{5}{95}\right)<\log(1000)$.

Granted, if evidence continues to accumulate for regime 2, Judy will eventually accept it and her differences with Punch will fade. However, if the regime switches back to 1, Judy will respond sooner than Punch does. While neither Punch nor Judy is inherently irrational, each will forever view each other as overreacting, lagging or catching up.

In general, Bayesian updating fans disagreement when observers become less certain and quells disagreement when observers become more certain. Most Bayesian textbooks focus on the latter case only. Indeed, the observer is often assumed to start out maximally certain and end with nearly full certainty.

As soon as we allow for possible regime change, beliefs can wander back and forth across the likelihood range. When they do, disagreements will repeatedly flare up and fade. A third party oblivious to regime change will find the process irrational and confusing.

Compounding the confusion, the differentials between observers in the log odds they assess are unlikely to stay constant. People don't observe exactly the same evidence, and even when they do they may not accord it the same relevance. Suppose, for example, that observers are handed a muddy coin and asked to ascertain whether it is fair. How relevant is the evidence from clean coins? How relevant is the evidence from other muddy coins, which might not have been muddied in the same way? How relevant is the evidence from the muddy coin itself after it has been flipped a few times and mud knocked off? Reasonable people can disagree. Some may exclude evidence completely, others may count it only partially. In effect, they multiply the right-hand side of equation (2) by a fraction, which might be as low as zero. Differences in that fraction will expand or contract the differences in log odds.

It may seem strange to speak about fractional evidence. At best it seems a reduced form for a more complex Bayesian updating, in which the evidence is either fully relevant or fully irrelevant but the observer isn't sure which. Nevertheless, it is a useful simplification that captures the notion of "take this report with a grain of salt".

In finance, for example, the credit spread of a bond, or premium of its yield over a purportedly risk-free benchmark, reflects the perceived probability of default times its expected fractional loss L on default. To estimate the default risk, we can imagine collecting a sample of T relevant years of observations on servicing, counting the number D of defaults, and using the mean rate D/T. Credit spreads for high-grade bonds typically run less than one percent per annum, implying $T > 100LD$.

Suppose that unexpectedly the bond defaults tomorrow, or that some other directly relevant default occurs. After the dust settles the issuer can borrow again, but at an extra spread to cover the perceived extra risk. How much extra? Our counting game suggests a revised estimator of $\frac{L(D+1)}{T}$, or L/T higher. In practice, credit spreads typically jump after a relevant default by five percentage points or more, implying $T < 20L$. Reconciling these two inequalities requires $D < 0.2$.

3. Continuous-Time Jump Processes

Let us now allow for multiple regimes in continuous time, assuming a simple Poisson jump structure for evidence. The baseline signal is zero, corresponding to smooth operation of a machine or to continuous payment on a debt. At isolated instants the signal is one, corresponding to a machine breakdown or default on debt. In a short interval dt, the probability of breakdown/default is approximately θdt, independently of any previous activity. Our task is to infer the unknown θ.

For θ conditionally fixed, the natural conjugate prior is a gamma distribution with shape parameter D and inverse scale parameter T. The gamma density for θ is proportional to $\theta^D e^{-\theta T}$, with mean $E \equiv D/T$ and variance $V \equiv D/T^2 = E/T$. In every instant dt, the relevant T for the gamma density edges up by dt, while the relevant D jumps to $D+1$ if default occurs and otherwise stays the same. Hence T can be interpreted as “relevant time” and D interpreted as “relevant defaults”, like in the simple counting model we started with. However, now D and T are proxies for a more complex belief structure.

For $D=1$, beliefs are exponentially distributed. For $D<1$, the gamma density is more convex than exponential, with infinite density at the origin and fatter tail. For $D<<1$, a gamma density behaves approximately like a two-point distribution with probability $\frac{D}{D+1}$ of default risk $\frac{D+1}{T}$ and probability $\frac{1}{D+1}$ of default risk 0. Alternatively, we can view a gamma distribution with $D<1$ as a mixture of exponential distributions with inverse scale parameters greater than T. See Bernstein (1928) and Steutel (1969), or the related discussion in Osband (2011, Appendix to Chapter 7).

Given a gamma distribution of beliefs, the mean belief E will shrink at proportional rate $1/T$ or jump by an absolute $1/T$, depending on whether default occurs. We can identify E by probing willingness to pay, or by relying on a competitive market to probe it for us. We can identify E/T by tracking the change in E in calm times. Hence, even though default is an exceedingly rare event, with probability measure 0, it appears we can predict the impact of default with complete certainty, apart from tiny measurement errors in T.

However, this certainty hinges on the gamma distribution assumption. Suppose our prior has a slight admixture, with weight δ, of an even more convex gamma distribution. Let the latter have governing parameters $(\alpha D, \alpha T)$ for $\alpha << 1$, so that the mean is the same as for the baseline but the conditional variance is much greater. Its conditional estimate of risk jumps by $\frac{1}{\alpha T}$ after default.

If a default occurs under this revised prior, the aggregate mean jumps not by $1/T$ but by a fraction $\delta/\alpha - \delta$ more. The difference needn't be small even if delta is. For example, if $\alpha < \delta$, then the aggregate mean jumps by order of magnitude $2/T$ or more.

If the regime is known to be stable, the impact of small doubts will fade over time. Even after the first default, the shape parameters for the gamma mixing distributions shift to $D+1$ and $\alpha D+1$ respectively. The shapes and implied responses are a lot more similar than for the original D to αD distinction. So even though the relative weight on each component will shift over time, it will eventually cease to make much difference. The mean belief will converge to the true mean while the variance dwindles to zero.

However, if regimes are unstable and the regime shifts are not directly identifiable, new doubts will emerge along with new differences among observers. Almost surely we will disagree on the current relevance of past information. Even where we agree on the current regime, we will almost surely disagree over the scale and scope of likely future changes. Under these conditions, significant variance in beliefs will be the norm rather than the exception, although the level of variance will fluctuate.

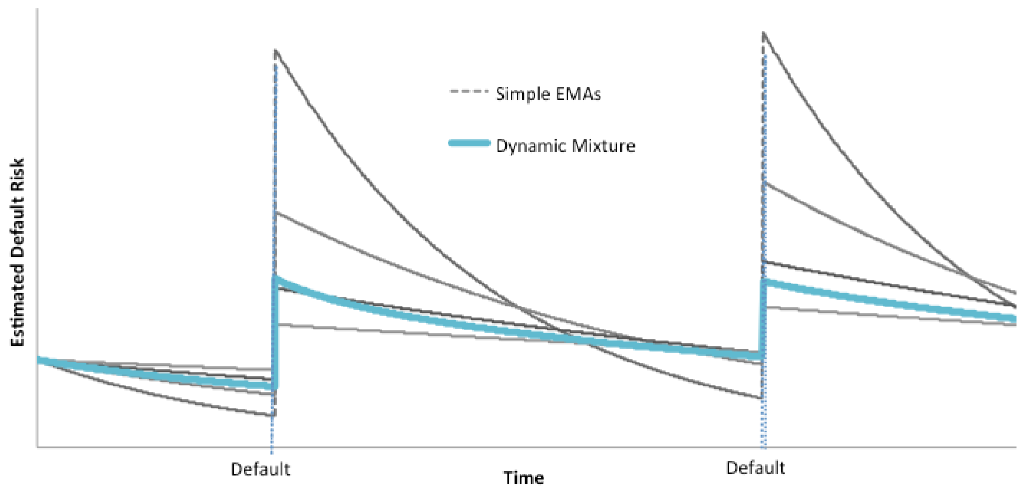
Imagine, for example, that observers perceive every instant dt a likelihood λdt that the past becomes irrelevant, without shifting their mean belief. We might model this as if the driving D and T parameters fade at a proportional decay rate λ. Since $\frac{dT}{dt} = 1 - \lambda T$, this will stabilize T near $1/\lambda$, while $\frac{dD}{dt} = -\lambda D$ apart from the occasional unit jump in *D*. It follows that dE equals $+\lambda$ on default and $-\lambda E dt$ on servicing.

This is an extremely easy system to analyze, once we know λ. It amounts to forming an exponentially-weighted moving average (EMA) of past observations, where default is counted as 1, servicing is counted as 0, and new evidence is weighted by λ. But which λ is best? Different observers will have different ideas.

Decades of evidence may be needed to winnow down the uncertainty, and in the meantime the optimal λ may change.

Figure 2, taken from Osband (2011, Chapter 7) illustrates the estimation impact of differences in λ.The higher λ is, the faster observers adjust to new information, and the more likely they are to overreact. Agreement between observers is the rare exception; significant disagreement is rule.

Figure 2: Estimation of Default Risk Using EMAs



If observers allow for various possible λ, the dynamic mixture of beliefs will tend to be more stable than the average constituent EMA. However, a dynamic mixture loses the simplicity of the EMA. The effective decay rate – that is, the single EMA it best maps to – changes over time, in ways that are hard to predict without knowing the constituents. And empirically the constituents are impossible to discern without leaving small doubts – doubts that are occasionally bound to grow big.

To understand these phenomena better, let us develop some general formulas for Bayesian updating of evidence on jump processes. Let $p(\theta)$ denote the perceived likelihood of an instantaneous default risk θ, let $\langle \ \rangle$ denote the expectation with respect to beliefs, and let $E \equiv \langle \theta \rangle$ again denote the mean risk given beliefs. If default occurs, Bayes' rule resets the likelihood to $\theta p(\theta)$ and then divides by E to normalize to one, for a discrete jump $dp(\theta)$ of $\dfrac{\theta - E}{E} p(\theta)$

If there's no default, the likelihood shifts approximately to $\left(1-\theta dt\right) p(\theta)$ and then divides by $1-Edt$. This equals $\left(1-\left(\theta-E\right)dt\right) p(\theta)$ plus second-order terms. It follows that $\frac{dp(\theta)}{dt}=-\left(\theta-E\right) p(\theta)$.

We can unite these expressions by defining dJ as the surprise element in the news. Since approximately Edt is expected over an instant dt,

$$dJ=\begin{cases}-Edt & \text{if payment,}\\ +1 & \text{if default.}\end{cases} \tag{3}$$

We can then write

$$dp(\theta)=p(\theta)\frac{\theta-E}{E}dJ. \tag{4}$$

Equation (4) is a neat updating rule. However, we can rarely apply it exactly, as it requires a continuum of likelihoods to adjust infinitely fast. To simplify to a countable number of updates per infinitesimal instant, let's focus on the moments $M_n \equiv \left\langle \theta^n \right\rangle$. Multiplying both sides of (4) by θ^n and integrating, we see that

$$dM_n=\frac{M_{n+1}-EM_n}{E}dJ \quad \text{for } n=1,2,... \tag{5}$$

At first glance, (5) is a great relief. It not only achieves the desired simplification but also links the moments recursively. Unfortunately the recursion works in the wrong direction. No moment can be updated until we know the moment of next higher order. It's an infinite ladder, and each higher moment is more sensitive to small doubts and calculation errors.

To summarize (5) in a single equation, let us look at the moment-generating function $\left\langle e^{b\theta}\right\rangle$. (To guarantee existence it is better to use the characteristic function $\left\langle e^{b\theta\sqrt{-1}}\right\rangle$, but the complex numbers offer no additional insight over the reals). Multiplying each side of (4) by $e^{b\theta}$ and aggregating over θ, we find that

$$d\left\langle e^{b\theta}\right\rangle=\frac{\left\langle \theta e^{b\theta}\right\rangle-E\left\langle e^{b\theta}\right\rangle}{E}dJ. \tag{6}$$

Expanding as a power series and equating terms in b^n confirms the equivalence with (5). We can also redo in terms of the cumulant generating function (CGF) $\mathrm{K}(b) \equiv \log\left\langle e^{b\theta}\right\rangle$. Since $\mathrm{K}'(b) = \dfrac{\left\langle \theta e^{b\theta}\right\rangle}{\left\langle e^{b\theta}\right\rangle}$, (6) transforms into

$$d\mathrm{K}(b) = \begin{cases} \left(E - \mathrm{K}'(b)\right)dt & \text{if payment,} \\ \log\left(\dfrac{\mathrm{K}'(b)}{E}\right) & \text{if default.} \end{cases} \tag{7}$$

At first glance this looks cryptic. To decode, let's reformulate in terms of cumulants – that is, the coefficients $\{\kappa_n\}$ in the expansion $\mathrm{K}(b) = \sum_{n=1}^{\infty} \kappa_n \dfrac{b^n}{n!}$. Since $\kappa_1 \equiv M_1 \equiv E$, the first line of (7) simplifies to

$$d\kappa_n = -\kappa_{n+1}dt \quad \text{if payment, for } n = 1, 2, ... \tag{8}$$

In other words, absent default, each cumulant of beliefs gets updated in direct proportion to the cumulant of next higher order.

The second line of (7) is more complex. It equals $Q - \dfrac{Q^2}{2} + \dfrac{Q^3}{3} - \cdots$ where $Q \equiv \dfrac{1}{\kappa_1}\sum_{m=1}^{\infty}\dfrac{\kappa_{m+1}b^m}{m!}$. Matching terms in b^n, it follows that

$$d\kappa_n = \frac{\kappa_{n+1}}{\kappa_1} - \sum_{j=1}^{n-1}\frac{\kappa_{j+1}\kappa_{n-j+1}}{2\kappa_1^2} + \sum_{k=1}^{n-2}\sum_{j=1}^{n-k-1}\frac{\kappa_{j+1}\kappa_{k+1}\kappa_{n-j-k+1}}{3\kappa_1^3} - \ldots \quad \text{if default, for } n = 1, 2, ... \tag{9}$$

Hence $d\kappa_n$ given default depends on all cumulants of order 1 through $n+1$.

The only distribution with a finite number of nonzero cumulants is the Gaussian (Marcinkiewicz 1938). Since default risks can never be negative, their perceived likelihoods cannot be Gaussian. Hence (8) must involve an infinite number of nontrivial adjustments.

When beliefs are gamma-distributed, most of these adjustments are redundant. The gamma CGF is $-D\log\left(1 - b/T\right)$, with cumulants $\kappa_n = (n-1)!DT^{-n}$. Equation (8) confirms that $d\kappa_n$ shifts by $-n!DT^{-n-1}dt$ absent default. When default is observed, both sides of (7) equal $-\log\left(1 - b/T\right)$, so that $d\kappa_n = (n-1)!T^{-n}$.

No other distribution appears to be so easily updated in terms of a finite number of sufficient statistics. Moreover, anticipations of regime-switching can easily blemish the gamma shape of beliefs. Hence, approximate updating must be the norm and exact updating the rare exception.

In Monte Carlo simulations, Bayesian updating during calm periods without default tends to restore a gamma shape to beliefs. Blemishes fade away. Default on the other hand greatly magnifies blemishes, as (9) suggests.

4. Brownian Motion

The news from continuous-time jump processes is chunky. Servicing provides little new information; the sporadic defaults tell us nearly everything. Is that what's driving the unpredictability?

For more perspective, let's turn to a model of Brownian motion. A particle x has unknown drift μ and volatility σ, so that absent parameter change its position at time t is normally distributed around the origin with mean μt and variance $\sigma^2 t$. We're trying to identify the parameters through continuous observations on changes dx.

Let's start by estimating volatility. Because cumulative squared Brownian motion is deterministic, in principle we can simply measure it quickly and divide by the time elapsed to determine σ^2. For a more intuitive explanation, note that the squared change between two instants h apart will have mean $\mu^2 h^2 + \sigma^2 h$ and variance $2\sigma^4 h^2$. For small h the standard deviation is less than 1.5 times the true mean, so by progressively halving h and averaging over twice as many non-overlapping samples we can drive the estimation error on σ^2 to zero over as short an interval as we like.

The mean is much harder to estimate precisely. The net change over time h will have mean μh and variance $\sigma^2 h$, implying a standard deviation that is $\frac{\sigma}{\mu\sqrt{h}}$ times the true mean. Taking intermediate measures and averaging them provide no net gain. The only way to improve precision is to lengthen the total time of measurement, and even that may not work if the drift changes over the observation period.

In practice, we lack sufficiently precise and rapid measurement technologies to estimate σ precisely. Nevertheless, σ is so much easier to estimate than μ that the rest of this section assumes the former is known and focuses wholly on estimation of the latter.

Let $p(\mu)$ denote the perceived probability that the true drift is μ, let $E \equiv \langle \mu \rangle$ denote the current consensus estimator of drift, and let dx denote the observed change over an infinitesimal time dt .To incorporate regime switching, let $\lambda_{\mu\nu}$ denote the instantaneous probability of shifting to another drift ν . Liptser and Shirayev (1977) derived an optimal updating rule of

$$dp(\mu) = p(\mu)\left(\mu - E\right)\frac{dx - E\,dt}{\sigma^2} + \sum_{\nu\neq\mu}\lambda_{\nu\mu}\,p(\nu)dt - \sum_{\nu\neq\mu}\lambda_{\mu\nu}\,p(\mu)dt. \tag{10}$$

For intuitive justification of (10), note that the density f of dx satisfies the following relation:

$$\begin{aligned} f\left(dx|\mu\right) &\propto \exp\left(-\frac{\left(dx-\mu dt\right)^2}{2\sigma^2 dt}\right) \\ &\propto \exp\left(\frac{\mu}{\sigma^2}dx - \frac{\mu^2}{2\sigma^2}dt\right) \cong 1 + \frac{\mu}{\sigma^2}dx, \end{aligned} \tag{11}$$

where the last step follows from a second-order Taylor series expansion. Bayesian updating using (1) shows that $p(\mu)+dp(\mu) \cong Cp(\mu)\left(1+\mu/\sigma^2\,dx\right)$ for some factor C . Since the revised weights must sum to one, $C = \left(1+E/\sigma^2\,dx\right)^{-1} \cong 1 - E/\sigma^2\,dx$. A bit more algebra establishes that

$$\begin{aligned} dp(\mu) &\cong p(\mu)\frac{\mu-E}{\sigma^2}dx\left(1-E/\sigma^2\,dx\right) \\ &\cong p(\mu)\frac{\mu-E}{\sigma^2}\left(dx - E\,dt\right), \end{aligned} \tag{12}$$

where the last line uses the approximation $\left(dx\right)^2 \cong \sigma^2 dt$, which we invoked previously to estimate variance.

The last two terms in (10) adjust for anticipated regime-switching, with the first summation incorporating probabilistic inflows and the second summation incorporating probabilistic outflows. For tidier notation, define $\lambda_{\mu\mu} \equiv -\sum_{\nu\neq\mu}\lambda_{\mu\nu}$, allowing these terms to be compressed into $\left\langle \lambda_{\bullet\mu}\right\rangle dt$, where the expectation is taken over the states feeding into μ . This is equivalent to $dp^{shift}(\mu)$ for a process p^{shift} that includes the regime-switching components only.

The updating rule (10) for tracking Brownian motion is very similar to the updating rule (4) for tracking jump processes. In each case we multiply current conviction p by the idiosyncrasy $\theta - E$ or $\mu - E$ of the belief and by the surprise

dJ or $dx - Edt$. We then divide by the expected variance E or σ^2 of the observation and adjust for dp_{shift}, the anticipated regime switching.

Given these similarities, it is not surprising that a similar cumulant hierarchy applies. For the cumulant generating function $\mathrm{K}(b) \equiv \log\left\langle e^{b\mu}\right\rangle$, $\mathrm{K}'(b) = \frac{\left\langle \mu e^{b\mu}\right\rangle}{\left\langle e^{b\mu}\right\rangle}$, so

$$\begin{aligned}\frac{d\left\langle e^{b\mu}\right\rangle}{\left\langle e^{b\mu}\right\rangle} &= \left(\frac{\left\langle \mu e^{b\mu}\right\rangle - \left\langle e^{b\mu}\right\rangle E}{\left\langle e^{b\mu}\right\rangle}\right)\frac{dW}{\sigma} + \frac{\left\langle \sum_{\nu} \lambda_{\mu\nu} e^{b\nu}\right\rangle}{\left\langle e^{b\mu}\right\rangle} dt \\ &= \frac{\mathrm{K}'(b) - E}{\sigma} dW + d\mathrm{K}^{shift}(b),\end{aligned} \tag{13}$$

where K^{shift} is the cumulant generating function for p^{shift}. Itō's lemma indicates that

$$d\mathrm{K}(b) = \frac{\mathrm{K}'(b) - E}{\sigma} dW + d\mathrm{K}^{shift}(b) - \frac{\left(\mathrm{K}'(b) - E\right)^2}{2\sigma^2} dt. \tag{14}$$

Equating higher-order terms of the corresponding Taylor expansions indicates that

$$d\kappa_n = \frac{\kappa_{n+1}}{\sigma} dW + \left(\kappa_n^{shift} - \frac{1}{2\sigma^2}\sum_{j=1}^{n-1}\binom{n}{j}\kappa_{j+1}\kappa_{n-j+1}\right)dt, \quad \text{for } n = 1, 2, \ldots \tag{15}$$

When beliefs are Gaussian, this collapses to system of two equations. The mean and variance get updated in what amounts to a Kalman filter. Otherwise the cumulant hierarchy is never-ending, with the property that

$$\text{volatility}(\kappa_n) = \frac{\kappa_{n+1}}{\sigma}, \quad \text{for } n = 1, 2, \ldots \tag{16}$$

Hence, non-normality at one cumulant quickly percolates down to cumulants below. Under stress, small disagreements about outliers can balloon into major disagreement s about means. Small differences in the λ parameters that fade relevance can also make a big impact,

Granted, absent regime shifts, all cumulants will eventually shrink toward zero, thanks to the last term in (14). Intuitively, given repeated i.i.d. observations on Brownian motion, rational learners eventually identify its parameters. However, when the drift is subject to change, uncertainty need not vanish, and no cumulant need converge to zero.

5. Practical Implications

The cumulant updating hierarchy has striking implications for capital markets. They include forecast degradation, frequent trading, excess volatility, and GARCH behavior. This section briefly sketches the connections:

Forecast Degradation: Imagine a forecaster has closely studied past market reaction to news, and correctly anticipates future news. Her knowledge can't be perfectly accurate, as measuring tools, processing abilities and storage capacities are limited. However, it might be sufficiently accurate that no one else can currently identify any discrepancy. Her short-term forecasts of market reactions should excel. Nevertheless, the cumulant hierarchy is bound to degrade the accuracy of her longer-term forecasts.

Forecast decay after default is especially quick. From (9), default tends to reshuffle higher cumulants. If related defaults occur soon after the first, credit spreads will likely surge throughout the asset class. Conversely, an extended spell without default will quell fears. But how much and how quickly is extremely difficult to predict.

Frequent Trading: In standard finance treatments, rational investors possessing similar information should have similar views on fair values. That doesn't justify a lot of trading. Even if investors differ substantially in risk aversion, the portions of their portfolios that are risky should be similar in composition. However, once we take into account the cumulant hierarchy, along with discretionary judgments on the relevance of evidence, near-agreement becomes more the exception than the rule.

Behavioral finance attributes frequent trading to human foolishness. It blames traders for getting over-excited or reading too much information into noise. However, it rarely considers the flip side: inattention, fear and regret that might dampen trading too much. Moreover, whichever way foolishness tilts, behavioral finance rarely considers its evolutionary stability. In the long run, more rational traders should dominate markets, in wealth if not in numbers, unless wealth itself breeds wrong-headedness. The cumulant hierarchy offers a neater and more plausible explanation.

Excess Volatility: Stock markets are supposed to reflect the net present value of future dividend streams. Shiller (1981) showed that with perfect foresight the S&P 500 stock market index would have been much less volatile than observed. He attributed the excess volatility to a form of investor foolishness, which he later termed "irrational exuberance". For a vigorous defense, see Shiller (2006).

However, as Kurz (1996) cogently observed, learning under uncertainty can readily explain the excess volatility. Let us decompose the log(arithm of) price into the sum of the log price/dividend ratio and the log current dividend. If the price-to-dividend ratio is fixed, the volatility of log price should match the volatility

σ of log dividends. Kurz called this “exogenous” volatility. However, Kurz noted a second source of volatility, which he called “endogenous” volatility. A surge in dividends, all else being equal, should make rational learners more optimistic about longer-term dividend growth, which in turn would justify a higher price-to-dividend ratio. Conversely, sagging dividends should make learners more pessimistic about growth trends, crimping the price-to-dividend ratio. As a first approximation, exogenous and endogenous volatility are perfectly correlated, so log price volatility should always exceed log dividend volatility.

The essence of the effect is captured in the first-order expression of (16), which indicates that the volatility of the mean belief $E \equiv \kappa_1$ about current drift is directly proportional to the variance $\text{var} \equiv \kappa_2$ of beliefs. Since the price-to-dividend ratio typically varies more than linearly with E, disagreement induces excess volatility in market prices compared with dividends.

How big is the effect? Adapting a two-state regime-switching model first analyzed by David (1997), Osband (2011, pp. 258-261) shows that

$$\frac{\sigma_{endogenous}}{\sigma} = \frac{p(1-p)S^2}{\lambda_{12} + \lambda_{21} - \mu_1 - \mu_2 + E}, \tag{17}$$

where μ_i denotes the dividend growth in regime i less the risk-free discount rate, and where $S \equiv \frac{\mu_1 - \mu_2}{\sigma}$ denotes the signal-to-noise ratio. For regimes that last longer than two years, the denominator will usually be less than one. Hence, as long as S exceeds one and observers aren’t too certain (p near zero or one), price volatility will significantly exceed dividend volatility. Analysis of equilibrium uncertainty shows that excess learning-induced volatility will typically be at least half of dividend volatility, unless the regimes are too similar for uncertainty to matter much.

GARCH Behavior: GARCH stands for Generalized Autoregressive Conditional Heteroskedasticity. It expresses the notion that volatility follows a random walk with changing volatility and mean reversion. Pioneered by Engle (1982), who received a Nobel Prize for his work, GARCH behavior is nowadays considered the norm for financial volatility. However, there is a puzzling disconnect between GARCH behavior and the standard capital asset pricing model (CAPM) that is supposed to drive market prices. Neither orthodox finance nor behavioral finance appears to explain GARCH. GARCH is usually invoked purely descriptively, the way Ptolemaic epicycles were once invoked to describe the observed motion of the planets.

In fact, GARCH behavior follows naturally from the second-order expression of (15). Letting $\text{skew} \equiv \kappa_3 \cdot \text{var}^{-3/2}$ denote the skewness of beliefs, we can express the percentage change in variance as

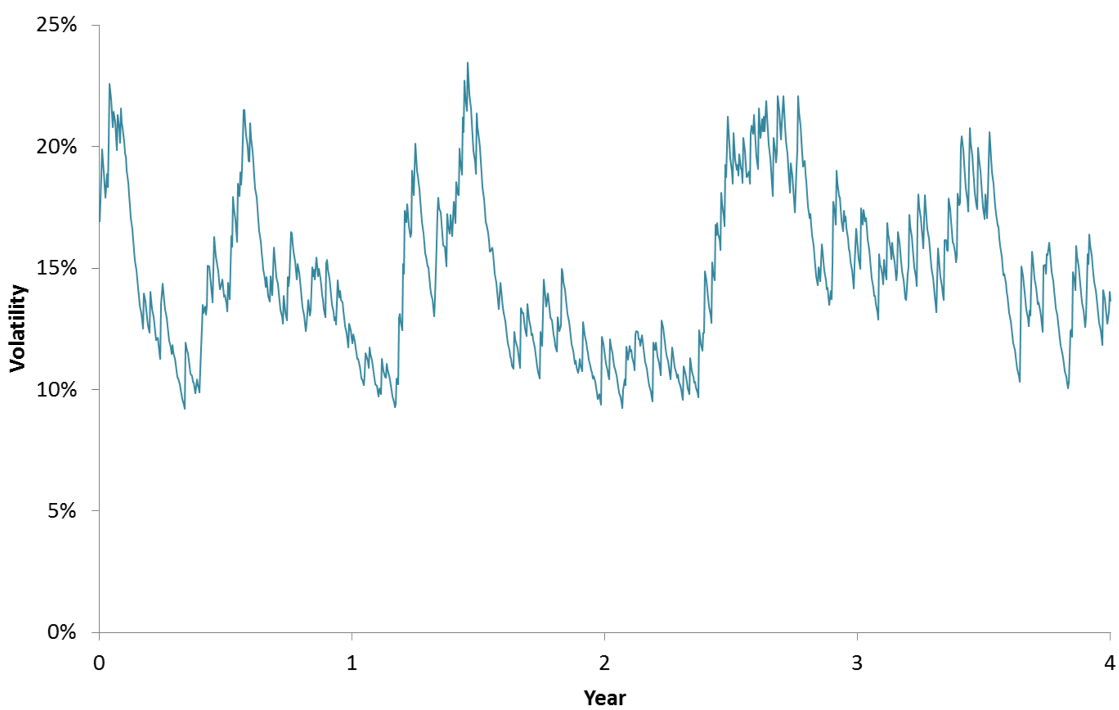
$$\frac{d\,\mathrm{var}}{\mathrm{var}} = \left(\frac{\mathrm{var}^{shift}}{\mathrm{var}} - \frac{\mathrm{var}}{\sigma^2} \right) dt + \frac{\mathrm{skew}\sqrt{\mathrm{var}}}{\sigma} dW \qquad (18)$$

Given their infrequent announcement, dividends are best viewed as having relatively stable variance. In that case, (18) describes most of the change in price volatility, apart from scaling factors. It shows that volatility follows a random walk, that its volatility changes in proportion to $\mathrm{skew}\sqrt{\mathrm{var}}$, and that there is a mean reverting component proportional to $-\mathrm{var}$. Volatility can fail to be volatile only if beliefs are fixed or perfectly Gaussian, which as we have seen are highly exceptional cases.

To illustrated the potential magnitude of the effects, Osband (2011, pp. 260-261) posits that one regime lasts15 years on average with a dividend that matches the discount rate, while the other regime lasts three months on average with dividends halving annually. For a dividend volatility of 3%, Monte Carlo simulations show that observed price volatility — estimated as an EMA with six weeks' average lag — rarely falls below 8% and in 5% of cases exceeds 20%.

Figure 3 charts a four-year simulation of price volatility. Note both the huge excess volatility and the pronounced GARCH behavior. For comparison, observed dividend volatility for this simulation never breached 2% at the lows or 4% at the highs.

Figure 3: Monte Carlo Simulation of Price Volatility with Dividend Volatility of 3%



6. Cumulant Hierarchies and Turbulence

How should we interpret excess market volatility that's highly volatile, hard to forecast, and ripe for misinterpretation? The closest physical analogue would appear to be the turbulence of the weather or the seas. Hence, it is tempting to describe financial markets as turbulent.

Benoit Mandelbrot, the discoverer of fractals, was the first to identify price behavior as turbulent. The introductory essay in Mandelbrot (1997) dates this identification to 1972. However, Mandelbrot did not link turbulence to learning or any other governing equation of rational market behavior. Instead, he emphasized the fractal geometry of market prices, which he interpreted as *prima facie* evidence of irrationality. His 2006 book with Richard Hudson summarized this perspective. It's titled The (Mis)Behavior of Markets: A Fractal View of Financial Turbulence.

Nassim Taleb's 2007 best seller, The Black Swan: The Impact of the Highly Improbable, applauds Mandelbrot's insight. Like Mandelbrot, Taleb links turbulence to widespread investor irrationality, but with more focus on investors' disturbed response to rare events.

In some ways, Taleb's argument runs parallel to the argument presented here. Both link small doubts about unknown risks to market turbulence. However, Taleb denies that Bayesian analysis is applicable to the analysis. He distinguishes between a "Mediocristan", where ordinary inferences apply, and an "Extremistan" where they do not. More formally, Douady and Taleb (2010) contend that events may be too rare or ill-conditioned to be described by Bayesian priors.

In contrast, this article identifies Bayesian priors with willingness to pay on small bets. It denies a fundamental difference in analytic approach between "Mediocristan" and "Extremistan", even where uncertainty is rife. Most importantly, this article ties turbulence to rational learning.

To justify that inference, this section shows that the cumulant hierarchy of Bayesian updating has direct counterparts in physics, and that these counterparts are closely associated with turbulence. The main connection runs through the Navier-Stokes equation, which applies Newton's law of conservation of momentum to the motion of a continuous fluid. There are also parallels with the Bogoliubov hierarchy (also known as the Bogoliubov-Born-Green-Kirkwood-Yvon or BBGKY hierarchy), which describes the dynamics of systems composed of many discrete particles.

When rewritten in terms of moments, the Navier-Stokes equation indicates that the motion of each moment depends on moments one order higher. In general, the system of moment equations cannot be closed without potentially misleading approximations. William McComb's (1990) esteemed text The Physics of Fluid

Turbulence introduces this problem in the very first chapter and gives it center stage:

> "The problem of closing the moment hierarchy is usually referred to as the 'closure problem' and is the underlying problem of turbulence theory."

Why is the closure problem so fundamental to turbulence? Imagine two drops of fluid very near each other in space or time. The surrounding fluid distribution and forces will be very similar for each drop. Only very high moments or cumulants will register any noticeable differences. Nevertheless, each of these moments impacts the moment below, which impacts the next moment and so on. Given fast enough transmission relative to the resistance of the fluid – a comparison summarized in the Reynolds number – the two drops may exhibit strikingly different behavior. The loss of smoothness over space and time is the physical essence of turbulence.

The Bogoliubov hierarchy, first derived in 1945 and published in 1946, identifies a related closure problem. It shows that in a system with many interacting particles, the evolution of the one-particle probability density depends on the two-particle probability density, the evolution of the two-particle probability density depends on the three-particle probability density, and so on. Lopping off higher-particle evolution equations is generally bound to introduce significant errors down below.

Note that moment hierarchies degrade longer-term forecasts. Forecasters using slightly different information may vehemently disagree. The motion will be significantly more volatile on average than the smooth (laminar) motion or diffusion associated with simpler equations. The volatility will be highly volatile. In these respects, fast-moving fluids, large systems of strongly interacting particles, and markets behave more similarly to each other than to more placid systems. A common description as "turbulent" seems apt.

7. Conclusions

When fluids are turbulent, tiny differences in space or time make for gross differences in behavior. The mathematical signature of turbulence is an endless moment hierarchy, where the change in each moment depends on the moment above. Bayesian tracking of continuous-time processes turns out to have a similar mathematical structure, which is most neatly expressed using cumulants.

Rational turbulence means that tiny doubts about regime change or tiny errors in estimation or calculation are prone under stress to balloon into gross differences of opinion. In effect, reasonable people are bound to disagree. The only exceptions involve natural phenomena with unchanging laws of motion.

This finding has profound implications for our understanding of financial markets and other instruments of social learning. In particular it explains forecast degradation, frequent trading, excess volatility, and GARCH behavior in terms of rational learning. In a sense it salvages the insights of Mandelbrot and other critics of orthodox finance, but without imputing widespread irrationality. Rational learning makes markets turbulent.

References

Bernstein, Sergei N. 1928. “Sur les fonctions absolument monotones.” *Acta Mathematica* 52: 1–66.

David, Alexander. 1997. “Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility.” *Journal of Financial and Quantitative Analysis* 32(4): 427–462.

Douady, Raphael and Nassim N. Taleb. 2010. “Statistical Undecidability.” Working Paper Series, Social Sciences Research Network, October 12.

Kullback, Solomon. 1968. Information Theory and Statistics. Mineola, NY: Dover.

Kurz, Mordecai. 1996. “Rational Beliefs and Endogenous Uncertainty.” *Economic Theory* 8: 383–397.

Liptser, Robert S. and Albert N. Shirayev. 1977. *Statistics of Random Processes*, vols. 1 and 2. New York: Springer-Verlag. Translation of *Statistika Sluchainykh Protsessov*. 1974. Moscow: Nauka.

Mandelbrot, Benoit B. 1997. Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. New York: Springer-Verlag.

Mandelbrot, Benoit B. and Richard L. Hudson. 2007. The (Mis)Behavior of Markets: A Fractal View of Financial Turbulence. New York: Basic Books.

Marcinkiewicz, Józef. 1938. “Sur une propriete de la loi de Gauss.” *Mathematische Zeitschrift* 44: 612–618.

McComb, William D. 1990. The Physics of Fluid Turbulence. Oxford, UK: Clarendon.

Osband, Kent. 2011. Pandora’s Risk: Uncertainty at the Core of Finance. New York: Columbia Business School Publishing.

Shiller, Robert J. 1981. “Do Stock Market Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71(3): 421–436.

Shiller, Robert J. 2006. Irrational Exuberance. New York: Doubleday.

Steutel, Frederick W. 1969. “Note on Completely Monotone Densities.” *Annals of Mathematical Statistics* 40(3): 1130–1131.

Taleb, Nassim N. 2007. The Black Swan: The Impact of the Highly Improbable. New York: Random House.