Grey Swans: Fifty Shades of Grey

Plausible Stress Testing

Gary Nan Tie

September 4, 2014

Abstract: For the purposes of risk management and stress testing we characterize a spectrum of plausible extreme events, that we dub 'Grey Swans', by introducing a probabilistic method involving the concentration of measure phenomenon. As a result, stress tests can be triaged according to severity, probability and now, information based plausibility.

Executive Summary:

Businesses are exposed to multiple risk factors such as insurance perils or changes in macroeconomic variables like interest rates, credit spreads, inflation, GDP growth, equity prices and FX. For a non-life insurance company's underwriting risk factor exposure, consider a line of business list that aggregates similar perils. Suppose that we have a data set of historical experience or forward looking multi-period simulations of this list and that we can estimate the marginal distribution of each risk factor. Risk management seeks to mitigate deviations from benchmark risk exposures. Deviations are concentrated around the 'business as usual' benchmark in terms of probability. Extreme deviations because of their rarity are more difficult to quantify. To address this situation, stress tests posit extreme risk exposures to quantify their impact, ignoring their likelihood and dependency assumptions to mitigate model risk. Now it is tempting to immediately jump to a 'perfect storm', that is a list of worst case risk exposures. Although possibly informative it's improbable and implausible nature may not lend itself to feasible mitigative actions. We seek a middle ground between large deviations and a 'perfect storm'. Indeed, suppose we start with a list of downside risk exposures, such as our worst actual experience, and then one by one replace an actual risk exposure with the worst ever seen, culminating in the 'perfect storm' list. The starting list of downside risk exposure is certainly plausible because it actually happened, but as we pile on worst ever risk exposures, the less plausible the list of risk exposures becomes. In terms of plausibility we introduce a spectrum of extreme events that we call 'Grey Swans' for the purposes of stress testing. This enables management to focus on mitigating the most plausible and most probable lists of extreme risk exposures. The new dimension we are introducing, beyond severity and probability, is the notion of plausibility. The gist of which is how different in terms of pattern is a proposed stress test list from known downside benchmark lists of risk exposures. Like the notion of probability, plausibility can be made mathematically rigorous. We use information theory. So between a downside benchmark and Black Swan lists of risk factor exposures there is a spectrum of Grey Swans, with different degrees of plausibility, for risk managers to consider when stress testing. In conclusion, stress tests can be triaged by severity, probability and now plausibility.

Grey Swans: Plausible Stress Testing

Introduction:

Suppose we observe or can simulate a large set A of random n-vectors and that we can estimate the marginal distributions. For example, insurance line of business losses, including investment losses. For a suitable notion of (probability) measure ν and (convex) distance that is information based, we will describe a parameterized family of enlargements $\{A_t\}_{t\geq 0}$ of A such that $s > t$ implies $A_s \supset A_t \supset A$ and $\nu(1 - A_t) \to 0$ as $t \to \infty$. Think of A_t as a set of plausible yet unrealized n-vectors close/similar to A and complement A_t^c as possible but unlikely less plausible events different from A. Indeed the smaller the parameter t, the closer to A, the more plausible the enlargement A_t of A. Consider a feature or statistic of A, such as the n-vector $y \in A$ whose aggregate loss is the greatest amongst those in A. For each enlargement A_t, we will find $x \in A_t$ similar in pattern to $y \in A$ such that $\sum_{i=1}^n x_i \geq \sum_{i=1}^n y_i$ i.e. a plausible n-vector x close to worst experienced y in A whose aggregate loss is even larger. So A is a set of known losses, enlargement A_t is a set of unrealized plausible losses, complement A_t^c are the possible but unlikely implausible events containing 'Black Swans', and $x \in A_t$ (similar in pattern to $y \in A$) is a plausible potential loss greater than anything in known A, that we call a 'Grey Swan'. Moreover, we can find different shades of grey by varying the degree of plausibility t.

Application of Talagrand's Concentration of Measure Inequality:

By Sklar's representation theorem the distribution function F_X of a random n-vector X can be represented as a copula:

$$F_X(x) = C(F_1(x_1), F_2(x_2), \cdots, F_n(x_n)), \quad x \in \mathbb{R}^n.$$

The probability density function of X with copula C and marginal distribution functions $F_1, F_2, \cdots, F_n$ admits the representation:

$$f_X(x) = (\prod_{i=1}^n f_i(x_i))c(F_1(x_1), \cdots, F_n(x_n)), \quad x \in \mathbb{R}^n,$$

where copula density $c(u) = \frac{\partial^n}{\partial u_1 \cdots \partial u_n} C(u), \quad u \in [0,1]^n$
i.e. f_X is obtained by reweighting the pdf corresponding to independence using the copula density c.

Let A be a large sample of n-vectors from X. View A as a subset of the product space of marginals of X. In this probability space we will enlarge

A in each dimension and show that the degree of plausibility depends on information content. Before we recall Talagrand's concentration inequality we will need to introduce some notation:

For n-vectors x and y, the Hamming distance $d_H(x,y)$ is the number of indices i such that $x_i \neq y_i$.

$d_H(x,A) \stackrel{\triangle}{=} \inf\{d_H(x,y) \mid y \in A\}$

Let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n) \geq 0$,

$d_\alpha(x,y) \stackrel{\triangle}{=} \sum_{i:x_i \neq y_i} \alpha_i$

(so $d_H(x,y) = d_{(1,\cdots,1)}(x,y)$)

$d_\alpha(x,A) \stackrel{\triangle}{=} \inf\{d_\alpha(x,y) \mid y \in A\}$

Talagrand's convex distance $d_T(x,A) \stackrel{\triangle}{=} \sup\{d_\alpha(x,A) \mid \|\alpha\|_2 = 1\}$

Note that $d_T(x,A) \geq d_{(\frac{1}{\sqrt{n}},\cdots,\frac{1}{\sqrt{n}})}(x,A) = \frac{1}{\sqrt{n}} d_H(x,A)$.

An equivalent geometric definition of Talgrand's convex distance is as follows: let $P(x,y)$ denote the set of zero-one n-vectors which are one on coordinates i for which $x_i \neq y_i$.

$P(x,A) \stackrel{\triangle}{=} \cup_{y \in A} P(x,y)$.

Then $d_T(x,A) = \min\{\|z\|_2 \mid z \in \text{conv } P(x,A)\}$, where conv$S$ denotes the convex hull of a set $S \subset \mathbb{R}^n$. Hence $d_T(x,A) \leq d_H(x,A) \leq \sqrt{n} d_T(x,A)$.

Let t be a positive real number and Ω be an n-dimensional product space. Let $A \subset \Omega$ and let $A_t = \{x \in \Omega \mid d_T(x,A) \leq t\}$. Note that $s > t > 0 \Rightarrow A_s \supset A_t \supset A$.

Theorem (Talagrand's concentration inequality)

$$Pr[A](1 - Pr[A_t]) \leq e^{-t^2/4}$$

In particular, if $Pr[A] \geq \frac{1}{2}$ (or any fixed constant) and t is "large" then all but a very small proportion of Ω is within "distance" t of A.

We now use Talagrand's concentration inequality to define Grey Swans, giving two examples.

Insurance example:

For ease of exposition we continue in the insurance context.

Let $A_i = \{z \in \Omega_i \mid \exists\, a \in A \text{ with } a_i = z\}$. Choose $y \in A$ such that aggregate loss $\sum_{i=1}^{n} y_i$ is maximized. Rank the component losses $y_{[1]} \leq y_{[2]} \leq \cdots \leq y_{[n]}$.

Let $m_i = \max\{z \in A_{[i]}\}$ and let $x = (m_1, m_2, \cdots, m_t, y_{[t+1]}, \cdots, y_{[n]}) \in \Omega$. Then $d_H(x, A) \leq t$, so for $1 \leq t \leq n$

$x \in A_t$ and $\sum_{i=1}^{n} x_i \geq \sum_{i=1}^{n} y_i$.

The larger the t, the more extreme and less plausible the x.

x is a plausible potential extreme n-vector similar in pattern to known $y \in A$ whose aggregate loss is greater than anything experienced in the sample A, that we call a Grey Swan. For example, if $n = 6$, degree of plausibility $t = 4$ and $Pr[A] \geq \frac{1}{2}$ then $Pr[A_4] \geq 0.9634$ and $x = (m_1, m_2, m_3, m_4, y_5, y_6)$ is a Grey Swan, extreme but plausible. $m = (m_1, m_2 \cdots, m_6)$ on the other hand is an implausible perfect storm stress test outside of A_4. So in this example the random sample A generated by X occupies more than 50% of the product space of marginals and the plausible enlargement A_4 of A more than 96.34%. The copula density distorts the independent product space pdf to induce the actual dependence structure of the pdf of X. Nonetheless A is a subset of A_4 and Grey Swan $x \in A_4$ has aggregate loss greater than anything experienced in A yet is similar to $y \in A$. Moreover we have a spectrum of less extreme, more plausible Grey Swans corresponding to degrees of plausibility $1 \leq t \leq 4$. 'Fifty Shades of Grey'!

Banking example:

In the context of banking, let X be a random vector of macroeconomic variables, for example X = (10-year Treasury yield, S&P 500, VIX, GDP, CPI, \$/€). We want to design plausible stress tests of portfolio value $V(x)$. Order each component macrovariable so that V is componentwise monotone. The sample A from X is a large collection of economic scenarios. Suppose we can estimate each marginal distribution well so that $Pr[A] \geq \frac{1}{2}$. Choose $y \in A$ so that $V(y)$ is say the 30th percentile, i.e. y is a stress scenario. Let $x \notin A$ be a modified y where we have replaced 3 components with 'lower' values so that $V(x) \leq V(y)$. The choice of which components and values will depend on our utility for their severity and probability. This $x \in A_3$, similar in pattern to $y \in A$, is a plausible (triple whammy) stress scenario, that we call a Grey Swan. Talagrand's inequality implies $Pr[A_3] \geq 78.92\%$. So this enlargement of A captures a majority of the possible combinations of the six macrovariables. Thus a scenario outside of A_3, like a perfect storm, would not be consistent with the majority of economic patterns similar to those in A. In this sense, it is plausible to have up to 3 extreme events occurring at once but not more. So in designing stress tests, not only is there severity and probability of events to consider, there is also the plausibility of their pattern. Note that there is flexibility in choosing the feature/statistic $V(y)$ and that this approach to plausible stress testing is a form of pattern recognition.

In practice, our application of Talagrand's inequality to strings of extreme events, has a 'curse of dimensionality'. To see this, recall that if $Pr[A] \geq \frac{1}{2}$ then $1 - Pr[A_t] \leq 2e^{-t^2/4}$. Now $(0.8)^3 = 0.512$, $(0.9)^6 = 0.531$ and $(0.95)^{13} = 0.513$. Note too that

$$\prod_{i=1}^{n} p_i \leq \left(\frac{p_1 + p_2 + \cdots + p_n}{n} \right)^n, \quad 0 < p_i < 1.$$

So the higher the dimension n of the product space, the higher the average probability mass of A across the marginals for the bound above to apply. The issue is how well can one estimate $Pr[A]$ in higher dimensions. The less (more) we know about A, the greater (fewer) the number of Grey Swans to be entertained in stress tests.

Stress test results whilst informative often do not lend themselves to feasible mitigative actions because of their improbable worst case nature. In the spectrum of scenarios between 'business as usual' and a 'perfect storm' we have introduced degrees of plausibility. So stress tests can be triaged by severity, probability and now, information based plausibility. This enables management to focus on mitigating the most plausible and most probable lists of extreme events.

References:

Talagrand, Michel (1995). "Concentration of measure and isoperimetric inequalities in product spaces." Publications Mathématiques de l'Institut des Hautes Études Scientifique, Dec. 1995, Vol 81, Issue 1, pp. 73-205.