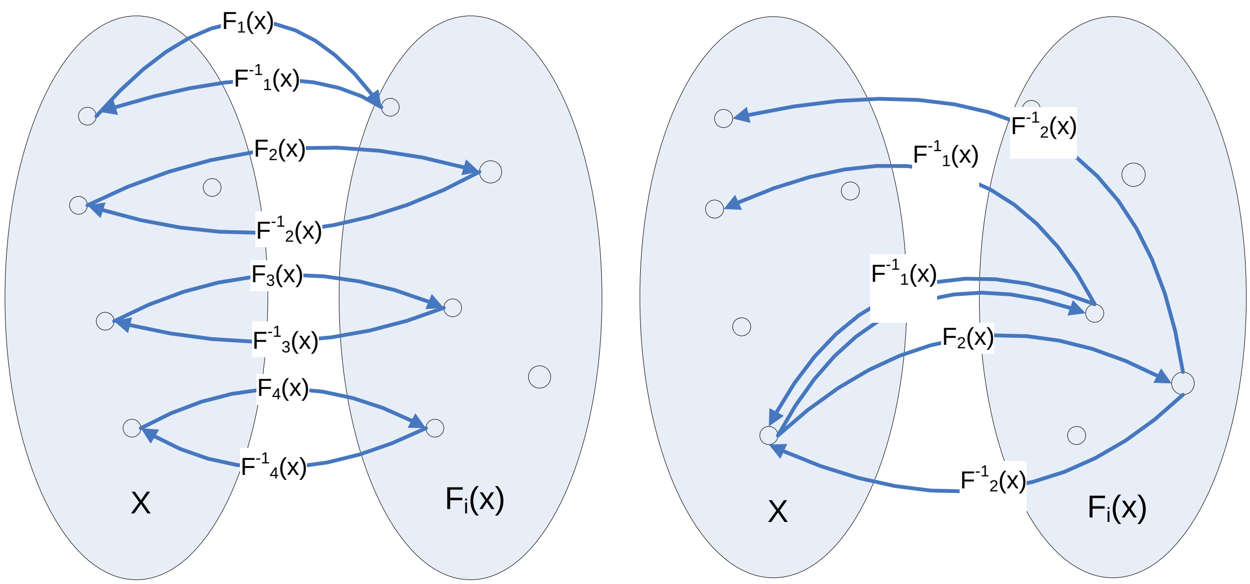
Beyond <u>The Black Swan</u>: A Characterization of Dark Risk and a Backtest Conjecture

The difficulty with modeling markets in a dynamical way is that their essence is free human choice, while the central core of dynamical systems is determinism. "Determinism" means that the past determines the future and as will be demonstrated, vice versa. This implies that zero topological entropy is a requirement for predictability. This will take some unpacking.

<u>Theorem (Kaminski) [1]</u>: A system (X, T) is <u>topologically predictable</u> if for every continuous function f in C(X) we have f in <1, T, f, T^2, f, … > where <F> ϕ C(X) denotes the closed algebra generated by a family F ϕ C(X). (X, T) topologically predictable if and only if every factor of (X,T) is invertible, where a factor is a system (Y, S) and a continuous onto map π: X Ψ Y such that πT = Sπ..

What this means for a class of functions F: X Ψ F(X):



Deterministic, Topologically Predictable System, $F_i(x)$

Topological Unpredictability: $F_1(x)$, $F_2(x)$

In both scenarios, the system consists of a set of past-present states X and future states F(X), defined by rule F governing time evolution from past-present states to future states. Crucially, for each rule governing past to future, there is an inversion of the rule that determines how present-future states relate to the past.

In the deterministic-predictable case on the left, the mapping of past to present-future is clear and one-to-one. This corresponds to an "if A-happened-then-B-will-happen" logic. Further, the present can be used to determine past values. This is validates the premise of back-testing models based on historical data. The if-then logic combined with the ability to invert time evolution rules to return information of the initial state is a powerful concept. It ensures that the system can be understood at all points in time (within a tolerance) and the rules governing evolution are unchanged over time.

Topologically unpredictable systems imply there is no one-to-one correspondence from the past-present to the future: the rules governing the transition from past do not map to unique future states. This is not insurmountable. It just makes prediction more imprecise. The real problem is that the inverse of transition rules (mapping the present-future to past states) are non-unique as well.

This is what makes deterministic models break down. It implies that well-worn "if-then" logic and intuitions don't work, because in this setting knowing where you are now offers little indication of where you were. It is equivalent to saying (in the case of $F_2(x)$ and its inverse: *if A in the past then B or C, but if B now then either A or D, and if C now then either A or E.* Confused yet? As the number of states and evolution rules grow in size, it becomes difficult to understand the dynamics of the system. Roughly speaking, when complications exceed tolerance thresholds, we say that the system has positive entropy. So the presence of non-invertibility implies the presence positive entropy which implies complexity which implies unpredictability.

This is the intuition behind Kaminski's theorem. It implies that a starting criteria to determine unpredictability lies in non-invertibility. At the extreme, each point of a system has multiple pre-images it is totally non-invertible. In a totally non-invertible system, there is always an open set of points whose pre-images (the $F^{-1}(x)$s in the picture)are bounded below by some positive constant. So putting things in reverse, a topologically predictable system has zero topological entropy, every invariant measure on it has zero entropy, and the system is invertible.

Predictability means there is information from the past embedded in the present (determinism), and the relationship between the past, present, and future is simple enough (zero topological entropy) that rules governing how the system evolves over time can be known. Thus there are two senses in which a system is unpredictable. Type i): the transition rules can be known, but knowing them requires infinite time or computational resources which no one has; and Type ii): the transition rules cannot be known because they are indeterminate. Type ii) unpredictability is equivalent to the dark risk coined by Stuart Turnbull.

Incomplete Information Induces Dark Risk

Type i) risk is premised on a partition of unknown future sets and completely known past states. A class of dark risk stems from incomplete information. Incomplete information implies that not only are future states unknown, but a subset S of past states are also unknown. This could be because of bad ticks thrown out, or measurement error, or just because everything about everything isn't captured.

Incomplete information enables a characterization of some types of dark risk. Assume a simple system possessing the following type of determinism: a future state $F_1(X)$ of the system is not affected (it is independent of) by any other past states outside of the past state X_1. This means there are only a finite number of past states that matter in figuring out the transition path from the past to the future. Thus prediction depends on particular past states, and all other historical data is irrelevant from a model perspective. Missing or incomplete data can make it impossible to correctly model the system even if it possesses this dependence structure.

Under incomplete information, rational people under a pretty wide set of circumstances* use sufficient statistics—they include all historical observed information—in their decision making. Thus missing data can make prediction a function of the entire past history of observed values even though better models exist to capture system behavior. This in turn affects model choice in such a way that prediction outputs perform poorly out of sample.

This dark risk is prevalent for a wide class of systems, because differential information is a commonplace kind of incomplete information. For example, since Goldman Sachs traders can have different information than other tier 1 banks, sufficient estimates that drive decisions can differ.

Even if the underlying dynamics of a system possess the determinism described above, there is no substantiation for assuming that the influence of the past decays quickly in the case of incomplete information. Positive entropy implies complexity, but they are not the same things, so entropy measures are not adequate. Needed is a measure of the minimal information such that all past data with predictive content is incorporated in model. The efficacy of such measures is subject to debate.

Implications for Backtesting

Since such measures are equivocal, rough techniques applied to backtests are a possible way forward. Backtesting typically withholds a subsample of the data and uses it to simulate prediction, usually from the most recent past. This subsample choice implicitly assumes that the influence of the distant past decays quickly, or essentially the same information is embedded in the recent past. If this is wrong, backtests may indicate good training period performance against simulation data, but the model has poor trading performance in even the immediate future.

The backtest procedure below provides some insight into the degree of unpredictability, if the conjecture holds. First, create two simulation buckets. One bucket contains data from the recent past while the other bucket contains data from further back in time. A stationary process implies that model simulation is equally good (statistically speaking) for both buckets given appropriate sample size and model selection. If the converse true, it has far-reaching consequences.

Conjecture: If model simulation is equally good for both buckets, then the underlying process is stationary. Quants can wear it out with all their wiles.

[1] Brunon Kaminski, Artur Siemaszko, and Jerzy Szymanski. The determinism and the Kolmogorov property in topological dynamics. Bull. Polish Acad. Sci. Math., 51(4):401-417, 2003.

* Proving this is true requires some special sauce which I prefer to not be contained in an anonymous posting.