WHY HEAVY TAILS?

David Harris, West Virginia University

ABSTRACT

Two key puzzles exist in finance. They are the question of why fat-tailed distributions model returns so well and why the normative models generate empirical contradictions. It is determined that these two phenomena are related. Using both Bayesian and Frequentist methodologies, it is shown that the models of mean-variance finance do not follow from first principles and are not valid scientific models. In the Bayesian framework, mean-variance finance models lead to a mathematical contradiction. In a non-Bayesian framework, valid inference cannot be performed.
JEL Classification: G10, G11, G12

INTRODUCTION

Mean-variance finance has dominated discussions of capital at various times over the last sixty years. It is a common required topic for students in finance and financial economics. It appears in regulations and its ideas have been incorporated into several sets of law including that class of state laws called uniform acts.1

Beginning in 1963 empirical contradictions began appearing in the literature. (Mandelbrot, 1963) These contradictions are problematic for any scientific theory and more extensive lists of contradictions have since appeared. (Fama & French, 2008) (Yilmaz, 2010)

The problem with the list of contradictions is that they do not appear to make mathematical sense. This has created two general concerns. The first is the belief that while the models are true it has to follow that some assumption is missing. The second is that the heavy tails violate the guaranteed coverage built into non-Bayesian statistics creating too many false positives while not simultaneously reducing false negatives.

The difficulty is that the models created through mean-variance finance appear to be tautologies. Surprisingly, it turns out that they are not tautologically true. It is shown that if a Bayesian framework is adopted then the models are false by contradiction. This is impossible by construction in non-Bayesian methods. Instead, given that mean-variance finance is true, the result is that no valid inference can be performed.

In the methodology proposed by Fisher or by Pearson and Neyman the models must be true by assumption. The data is conditioned on the model and as such should be used to show they are false.

The surprising finding is that even with an infinite amount of data the results of the test statistic are uncorrelated with nature. In the Fisherian or Pearson-Neyman framework, hereafter called Frequentist, no inference is possible and so the models, regardless of their truth value, cannot be scientific models.

The intuition for this is that if returns can be thought of as the ratio of a future value and a present value minus one then it can be shown that a particular ratio distribution, the Cauchy distribution, must be the true distribution in nature given only the assumptions of the models. (Geary, 1930) (Gurland, 1948) (White, 1958) (Rao, 1961)

The ratio nature of the reward or alternatively the return for investing has been unnoticed. The impact on mean-variance finance is catastrophic. The Cauchy distribution has the unusual property of having neither a mean nor a variance. The National Institute of Standards and Technology describe the Cauchy distribution thus:

> The Cauchy distribution is important as an example of a pathological case. Cauchy distributions look similar to a normal distribution. However, they have much heavier tails. When studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality. Likewise, it is a good check for robust techniques that are designed to work well under a wide variety of distributional assumptions.
>
> The mean and standard deviation of the Cauchy distribution are undefined. The practical meaning of this is that collecting 1,000 data points gives no more accurate an estimate of the mean and standard deviation than does a single point. (National Institute of Standards and Technology/Sematech, 2012)

Although the Cauchy distribution does have known statistical methods for dealing with it as well as tests of inference, using them would require abandoning mean-variance finance as a valid methodology.

The paper traces the development of thought through the formulation of the modern models and transitions into an exposition of Bayesian and Frequentist methodologies. This is to bring the paper into a proper historical and methodological setting. Following from this methodological perspective the paper presents the results of papers previously unnoticed by economists. (Gurland, 1948) (White, 1958) (Rao, 1961)

The unexpected consequence of changing the distribution from the Normal to the Cauchy is that other seemingly unrelated methods such as econophysics or behavioral finance have at least a partial explanation in standard utility theory. This expands the idea of financial economics from a narrow methodology of portfolio selection and pricing to a tool to discuss the consequence of deferring consumption with the belief that a gain in utility will happen in the future. Gift giving, marriage, child rearing, religion and transformational relationships now become part of the domain of financial economics because they all require deferrals of consumption in anticipation of a future gain in utility.

HISTORICAL BACKGROUND LITERATURE

To begin understanding why there are heavy tailed distributions for returns it is first important to understand why it was mistakenly believed there should be anything else. Many of the antecedent ideas come to us from the beginning of the 19th century. The foremost of these is the classical central limit theorem.

The central limit theorem is so named, neither because of some limit at the center of the distribution, nor due to the presence of the mean at the center, but rather due to its central importance to the field of statistics. (Jaynes, 2003) While it is central to statistics, its importance to economics is even greater. The normal distribution and the expectations operator are everywhere in the modeling of economic processes.

What very few people other than statisticians are aware of is that there is an important restriction in the classical central limit theorem regarding the existence of a mean and a variance. The classical central limit theorem applies to any arbitrary probability distribution with a fixed mean and variance. This requirement, if not met, causes the classical central limit theorem to be inapplicable to real world problems.

This restriction in the normal law of errors, as it was originally called, first appeared in a note by Poisson. In reviewing the theorem, Poisson noted that the distribution $f(x)=\pi(1+x^2)^{-1}$ was a counter example to the theorem, as the distribution has neither a mean nor a variance. Still, Poisson wrote,

> But we shall not take this particular case into consideration; it will suffice to have remarked upon the reason for its singularity and note that we will without doubt not encounter it in practice. (Stigler, 1974)

Independently, Bienaymé wrote an article showing that least squares regression provided the best possible mechanism to fit a line to data, in contrast to a method provided by Cauchy. (Stigler, 1974) He had discovered that the method of ordinary least squares gave the best linear unbiased estimator. This triggered a

$$f(\epsilon)=\frac{k}{\pi}\frac{1}{1+k^2\epsilon^2} \qquad (1)$$

In this specific circumstance using the least squares algorithm, finding a sample mean or a sample variance has the shocking consequence of having no predictive value. Indeed, Sen (1968) notes that such a method would be perfectly inefficient when compared with valid solutions when the Cauchy distribution is present.

The first appearance of the normal distribution in economics and finance appears to be a presentation by Jules Regnault in 1853. (Davis & Etheridge, 2006) He discovered empirically what Bachelier would argue theoretically in 1900. (Bachelier, 2006) In the interim, Edgeworth (1888), following work by Laplace, Jevons and Quetelet, argued for applying the law of errors to investments in general and Bank of England notes in particular. Further, he sought to unite utility theory and probability theory, anticipating von Neumann and Morganstern (1953) by more than a half century.

Although Edgeworth discusses this in reference to the normal distribution, the first direct linkage between utility theory and probability is by Bernoulli in his solution to the St. Petersburg paradox. (von Neumann & Morgenstern, 1953) Between the great statistician and economist Edgeworth and mean-variance finance a wide range of basic problems needed solved first.

To leap from Edgeworth into mean-variance finance one must first pass through the works of Clark (1908), Böhm-Bawerk (1890) (1891), Veblen (1904) (1899), Fisher (1930), Keynes (1936), Pareto and Hicks (Hicks, 1939). By pulling together the work of Böhm-Bawerk, Clark, and Pareto, one should arrive at Fisher's conclusion that the interest rate is the marginal cost of impatience. A careful read of Veblen's work on the leisure class could be read as the first work on behavioral finance. Keynes work creates an idea not possible in the classical school, inefficiency and emotion in markets. The thinking behind efficient markets pushes aside the thinking and observations of Veblen and Keynes until they are independently rediscovered later by others.

Their work stands in contrast to the combined work of Pareto and Hicks. Hicks' work is central to the classical school of thought regarding capital. It is this work that starts Markowitz (1952) down his ground breaking idea of having economists measure both risk and return using the mean and the variance.

While Veblen and Keynes would continue to influence future economists in other areas, the latter more than the former; it is Markowitz who would set in motion Hick's unattained goal of "an economics of risk."' Although Roy (1952) simultaneously discovered the same thing, it is Markowitz's work that is remembered.

Hicks (1939) appears to make two conflicting comments in his book *Value and Capital*. On the one hand, he clearly argues that people include risk in their plans and prices, implying economists should measure risk. However it is also clear from his writing that the tools to measure risk do not exist.

Hicks goes on to state that economists can ignore risk because it is included in the plans and expectations of the actors. By watching actual returns economists implicitly get the risk variable and hence need not try to measure it directly.

It is improbable that Markowitz guessed the impact of his initial writing would have. The transformation was greater than formulating a trade-off scheme between risk and return, it created a way of thinking about and including statistical measures in economic thought and economic processes.

A casual read of this initial work shows a field of economics in a comparatively primitive state. Indeed, without Markowitz, this article and any subsequent work would be impossible. Although earlier writers, such as Regnault, Edgeworth, Hicks and von Neumann bring uncertainty and risk into the discussion, Markowitz and Roy are the first to propose a mechanism of exchange between return and risk.

Unintentionally, Markowitz broke with statistical theory, even though he was calling to embrace it. What Markowitz could not have known in 1952, was that there were three mathematical cases to solve and not one. The case he solved does not apply to finance, though it is a valid solution to many physical processes.

Warnings that something was amiss began in 1963 when Mandelbrot published an article stating that the distribution of financial returns actually observed in nature followed a Cauchy distribution. (Mandelbrot, 1963) What is unfortunate was that throughout Mandelbrot's life he never saw the reason for it. As no one could provide a theoretical foundation for the presence of heavy tails, and given the centrality of the law of errors, the mean-variance methodology appeared more than reasonable. Indeed, it came to be viewed as a tautology.

One need only look to the derivation of the normal distribution and one will note the potential for contradiction immediately. It is dangerous to forget that the distributions are the result of some mathematical process. The most commonly viewed derivation of the normal distribution is, of course, the one by Gauss. Another one, probably better known to economists, is one by Mann and Wald in 1943. (Mann & Wald, 1943) In it they show that for the equation $x_{t+1}=\alpha x_t+\epsilon_{t+1}$, where α and ϵ are unobservable and where the diffusion term is centered on zero and has finite variance, then $\hat{\alpha}-\alpha$ will follow a normal distribution provided that $|\alpha|<1$.

In deriving the distribution of $\hat{\alpha}-\alpha$ they had created a derivation of the normal distribution. Just as any proof that arrives at $a^2+b^2=c^2$ can be thought of as a proof of the Pythagorean Theorem, any proof arriving at the Gaussian distribution can be thought of as a derivation of the Gaussian distribution.

When viewed as a moving average infinity model the implication becomes that as the amount of time goes to infinity the effect of prior errors goes to zero. Financial time series have a value of $|\alpha|<1$. No one wants their savings account to converge to zero simply from the passage of time.

A one unit error at time zero will grow to infinity rather than converge to zero. That the errors explode rather than converge is a warning, but in and of itself, is not sufficient as it does not resolve either the distribution or the relationship between x_{t+1} and α. Three mathematicians, Gurland (1948), White (1958) and Rao (1961) solve this third case, but the only connection between their work and financial time series is in using White's proof as the basis for unit root tests. (Dickey & Fuller, 1979)

The proof herein is simply the adaptation of a standard statistical result to economics and finance. Indeed, instead of viewing a distribution as central to economics, it may be better to see mixtures of distributions as central.

The history, post Markowitz, splits down three paths. The first seeks to confirm or disprove mean-variance finance. The second follows a path using heavy-tailed distributions, some of which have a mean and some of which do not. The third is behavioral finance. This third form investigates whether actual behavior matches the predictions of theory and then attempts to develop theories of behavior based on observation. Although this paper may in fact have a substantial impact on the third path, it is the first two paths that are directly addressed by this paper.

Fortunately a number of individuals have compiled summaries of the contradictions in empirical findings. In particular, Fama and French (2008) in their work *Dissecting Anomalies* and Yilmaz (2010) in his masters thesis provide an extensive list of the anomalies and contradictions observed in the literature. The lists are extensive and intensely problematic if mean-variance models are in fact valid.

It implies there are a very unusual set of behaviors present in the market, given the models are true. These behaviors range from systematic biases in returns to volatility clustering. Further, residuals reflect significant excess kurtosis. Indeed, the "outliers," given that mean-variance is true, are both extreme and the tail past the third, or even the sixth standard deviation is quite dense. A simple review of the data shows that the Bieneymé-Chebyshev Inequality does not seem to hold empirically for the residue.

The second path has been in one of two forms. Either the authors were building upon the ideas set forth by Mandelbrot (1963) or they were physicists who saw time series data that resembled work they were doing in physics. (Ball, 2006) Despite being successful in replicating market behavior in some models and unsuccessful in others, they have suffered from an absence of first principles reasoning. This is due to the absence of a first principles reason for any specific distribution or process to be present or absent. What mean-variance finance provided was a first principles method for discovering what should be present. This article is intended as a first step toward this distant goal.

Although this paper will show that under very mild assumptions that the distribution of returns in both the Bayesian and Frequentist paradigms must converge to a Cauchy distribution this paper should not be construed as arguing that returns follow a Cauchy distribution. Rather, in the blackboard economics generally used in finance and economics, returns must converge to a Cauchy distribution. Follow-on papers argue that adding in very simple economic constraints can have an unexpectedly large impact on the distribution observed in nature and also confirm that the Cauchy distribution is in fact a reasonable likelihood function when compared with the normal distribution.

MODEL ASSUMPTIONS

This paper has a number of relatively simple assumptions that should be non-controversial. In particular, the model assumes that the Böhm-Bawerk and marginalist paradigms, generally accepted for over a hundred years, are valid. The model adopts the mean-variance assumption that future wealth equals current wealth times a reward plus a random shock. The paper further generalizes this and argues that the static model is the same as an auto-regressive of degree one process, without a loss of generality. It assumes that both the Bayesian and the Frequentist models of probability and statistics, when viewed separately, are completely valid understandings of their fields. Finally, it assumes that scientific models have at least two properties, which are that models are mathematically coherent and that measurable inference can be performed on a model. This last assumption is little more than a reworking of Cox's postulates for a narrow purpose. (Jaynes, 2003)

Difference Equations

Key to understanding the various models is the structure of the equations used to make them. Implicitly or explicitly, the models use difference or differential equations. Stochastic economic models can be divided into three groups: static models, discrete time models and continuous time models. The relationship between discrete and continuous time models is through scale invariance. (Donsker, 1951-52) The relationship between static models and discrete time models in economics comes from the proposition that, subject to a model's assumptions, economic models are statements of general economic principles that hold across time.

For example, if a model contains $x_1=f(x_0)$ and $x_t=f(x_{t-1})$, then by induction it can be shown that $x_{t+1}=f(x_t)$, $\forall\ t \in \mathbb{W}$, So static models of the form $\tilde{w}=R\bar{w}+\epsilon$, where $\tilde{w}$ is an uncertain future wealth, R is a parameter, and ϵ is a random variable, could be re-written, without a loss of generality, as $w_{t+1} = Rw + \epsilon_{t+1}$.

The equation $\tilde{w}=R\bar{w}+\epsilon$, it should be noted, is the basis of an ill-posed problem as used by economists. Gauss reminds us that it is only in the limiting form of a well posed mathematical process that any real discussion of the properties of $w_{t+1} = Rw + \epsilon_{t+1}$ can begin. (Jaynes, 2003) If $\tilde{w}$ and w_{t+1} were not treated as being equivalent constructions, then indeed this would have a most peculiar case at least in regards to economics.

While it is quite possible to imagine single gambles which have no economic consequence in the future, this is not what is generally discussed in economics. That said, this does not preclude the existence of multiple limiting models. This proof is one such model, but it is believed that it fully encompasses the range of behaviors possible in a mean-variance finance proof.

Assumption

The equation

$$\tilde{w} = R\bar{w} + \epsilon$$

can be expressed as

$$w_{t+1} = Rw_t + \epsilon_{t+1}$$

without loss of generality. In the equation, ϵ is drawn from a distribution with finite variance and is centered on zero. As well, $\epsilon_t \perp \epsilon_{t+1}, \forall \{t, t+1\}$.

Böhm-Bawerk Theory

A rather unexpected argument has been made that while purely technical and not reasonable within the context of economics is nonetheless a key element for the existence of heavy tailed distributions. As such it needs addressed. The argument is that finance theories do not explicitly require that the marginal actor is trying to make a profit from investing. Technically, this is often true. The assumption is usually implicit.

Regardless, in the late 19th and early 20th century significant work was done on capital and interest rates; this work underpins all modern thinking. In particular, the work of Eugen von Böhm-Bawerk (1890) (1891) on the agio, or premium, theory of interest rates and the writing of James Bates Clark (1908) on marginalism come together in the writings of Irving Fisher (1930) and later in J.R. Hicks (Hicks, 1939).

Of importance to this paper is the idea of an investor requiring an anticipated premium for deferring consumption. This implies that for Frequentist models, that $R>1$ and the center of location of R for Bayesian models is greater than one.

Showing this is true is rather simple. Ignoring issues of uncertainty for a moment, a utility maximizer will prefer a positive return if the alternative is a zero return on nominal money. Under uncertainty some funds may be maintained in money if there is some minimum level of consumption required in following time period or under strong risk aversion.

As an alternative way to view the issue, the counterfactual question , “what if the reward for investing was anticipated to be a loss in every period, ignoring shocks,” is instructive. The capital stock in a finite resource environment would go to zero. This would imply no spears, no seeds, no machinery. This implies extinction so systemic losses are excluded for the parameter R. If $R=1$ then while capital could form it couldn't be partitioned to allow for different prices for different risks. As such, $R>1$ is the only available option with market traded capital.

Assumption

The anticipated return for investing by the marginal actor must be positive.

Bayesian Versus Frequency Based Models

This paper looks at the presence of heavy tailed distributions in both a Bayesian and a Frequentist manner. As economists are rarely trained in Bayesian methods and as Frequentist tools have usually been used to look at mean-variance finance models, some description of the differences is felt necessary.

The paper seeks to show that the outcomes are independent of the school of thought employed. To quote from Egon Pearson (1955):

> Controversies in the field of mathematical statistics seem largely to have arisen because statisticians have been unable to agree on how theory is to provide, in terms of probability statements, the numerical measures most helpful to those who have to draw conclusions from observational data. We are concerned here with the ways in which mathematical theory may be put, as it were, into gear with the common processes of rational thought.

It is incorrect to think that Bayesian and Frequency based models are different ways of solving the same problem. Rather they are ways of using the same data to solve different problems. It is often true, however, that there are no numerical differences in their estimates. The differences then become one of interpretation.

For most of the 250 years of Bayesian statistics, it was called the method of inverse probability. (Fienberg, 2006) Inference was of the form, $\Pr(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{\theta}$ is a parameter or vector of parameters of interest and $\boldsymbol{y}$ is the data. As such, causes were inferred from effects. Effects can be seen in the data, but the cause is often hidden. It was a statistical form of solving the inverse problems so common in economics.

This structural form requires that the data are given as true and therefore are fixed points and not instantiations of a random variate. Conversely, the parameters are random variables, or more precisely beliefs about the parameters are random variables.

An hypothesis is considered a belief, so the idea that $\mu>5$ is one of many possible beliefs about μ. Inference about that belief would be shown as $\Pr(\mu>5|\boldsymbol{y})$. Beliefs about μ change as more information arrives. So as the data set goes from $\boldsymbol{y}$ to $\boldsymbol{y}'$ the belief about $\mu>5$ goes from $\Pr(\mu>5|\boldsymbol{y})$ to $\Pr(\mu>5|\boldsymbol{y}')$. This forces a necessarily subjective view of probability, as different viewers have access to different information. This leads to epistemic probabilities, something quite removed from the Neyman-Pearson concept of aleatory probabilities.

Frequentist, or frequency based statistics, are modeled on the long run probabilities of some event occurring. For this methodology to be used, it implies that the long run model can be known. Rather than looking at past information and testing new information given prior information, frequency based measures look at the long run model and asks, "what is the probability the data looks as it does given the model is true?" That is to say $\Pr(\boldsymbol{y}|\boldsymbol{\theta})$.

In frequency based statistics, the parameters are fixed points and the data is considered random. This is the very opposite of Bayesian inference. As such, an hypothesis is true or false. It is a fixed point and cannot have probabilities of truth or falsehood assigned to it. Since the data are not considered fixed, probability statements can be made about the likelihood of observing the data given the fixed parameter. Because of the connection with modus tollens, this is usually stated in terms of the probability of observing data as extreme or more extreme due to chance.

Using Frequentist statistics, the hypothesis $\mu>5$ is either true or false. Usually, however, if the real concern is whether or not $\mu>5$, then the complementary hypothesis of $\mu\leq5$ is tested instead. Whereas Bayesian tests determine the probability a belief is true, frequency based statistics test the probability the data disconfirms a null hypothesis.

This probability is based upon the long run frequencies given the hypothesis and not the data alone. Whereas Bayesian statistics use only the observed data conditioned on prior knowledge to make decisions, frequency based measures consider the samples that could have been observed according to the model.

These subtle differences can lead to rather sharp differences in the understanding of the same events. Aleatory probabilities are closely related to physical probability in the sense of dice rolls or coin tosses. Bayesian probabilities are subjective, so the tie to physical probability is looser. In a sense, it is one step removed from the physical probabilities.

As a pragmatic illustration of the difference, imagine two possible dice games under perfect competition for customers. One type of dice game is run by an honest casino where everything is fair. In the other type of dice game, con men and ex-magicians run the same game. The players do not know which type of game they are in. The house takes two die, places them in a cup, shakes them in the cup, and turns the cup upside down with the dice still covered by the cup. Players then wager against the house on whether the sum of the digits is even or odd. The house, through a croupier, rolls the dice, but the player chooses "even" or "odd." Players pay a cover charge of one dollar in advance and can play all day for one dollar per dice roll.

Even in such a simple model of probability, the contrasts can be quite stark. The Frequentist methodology has one giant advantage here, the solution will always be unique. The most natural way to approach this question is to have two hypothesis:

1. The casino's policies do not adversely impact the player
2. The casino does cheat the players.

It is possible to either test the count of the wins versus losses or the percentage of times the house wins versus the house loses. For simplicity of presentation, it is easiest to choose the latter method of percentages. Setting π_{house} as the probability of the house winning, the most logical null hypothesis is $\pi_{house} \leq 0.5$, with the alternative hypothesis being $\pi_{house} > 0.5$.

The Bayesian method, however, does not automatically yield a unique answer or set of hypothesis. The hypothesis could be the same as the frequency based method. It could be an infinite number of hypothesis, where each point on the number line is hypothesized as the true value. This would be expressed as $\pi_{house} = i, \forall i \in [0,1]$. It could also be any mutually exclusive and exhaustive set of hypothesis that combine intervals and points.

In addition, the Bayesian method requires the choosing of a prior distribution for the parameter π_{house}. If a uniform prior is used, the result will be numerically identical with the frequency based method, provided of course the same hypothesis are used. The difference would be one of interpretation. However, there is a strong economic argument and therefore statistical argument against the uniform prior. The economist is in possession of information from the model.

With regard to the example, competition should drive out cheating that could be detected by casual players using non-rigorous methods as it is costless to change casinos.

In the absence of cheating by the croupier in favor of the player in exchange for a kick-back of profits, the expected value of $\pi_{house} > 0.5$ in perfect competition. Since it is reasonable to believe the house is monitoring for cheating by the croupiers, with maybe a slight chance being present of cheating by croupiers, the prior probability distribution for the estimate of π_{house} should be centered slightly to the right of 50%. Unlike the Frequentist method, this skews the posterior calculations toward the hypothesis "the casino does cheat players," unless enough data comes in to overcome the prior.

Even if the prior were centered on 50% with a variance of 25%, this additional information would reduce the required number of observations to reach a conclusion. Indeed, Jaynes shows that in similar situations, the number of required observations may be half that required for the unbiased Frequentist estimate. (Jaynes, 2003) As each observation bears a potential cost, cutting the required number of estimates in half can be meaningful.

One other difference is what is considered random by the two schools. The Frequentist school would not consider the dice rolls to be random variables as they are fixed points at the time the player calls out "even" or "odd."' Rather it is what is called out by the player that is random and hence it is the matches that are random. The players are betting they can match a fixed point.

If there are too few matches then it can be said, to some degree of confidence, that the result is not likely due to chance. The Bayesian method, on the other hand, is going from effects to causes. So it sees the parameter of wins as uncertain and the matching as fixed points once they occur. The Bayesian sees nothing random in the matches and non-matches that actually happened; they are the result of casino policy. What is uncertain is which type of game is being played, and hence the true value of the parameter.

HOW BAYESIAN AND FREQUESTIST PARADIGMS AFFECT THE EQUATION

In both models only the vector $\boldsymbol{w}_t$ is observable. The error term, ϵ and the reward for investing, R, are unobservable and of course $\boldsymbol{w}_{t+1}$ is yet to be observed. What differs between the Bayesian and the Frequentist paradigms, is what is a random variable and what is a fixed point.

In the Frequentist model R is a fixed point. It has a degenerate distribution. The vector $\boldsymbol{w}_{t+1}$ and ϵ are random variates. Although R does not have a distribution with density, there is a distribution of $\hat{R}$-R. Indeed, these differences are thought of as errors as the true value is a fixed point.

In the Bayesian model R and ϵ are random variates and the vector $\boldsymbol{w}_t$ is fixed. The future value, $\boldsymbol{w}_{t+1}$ has not been observed and so remains a random variate until seen.

<u>**Frequentist Assumption**</u>

In the equation $\mathrm{w}_{t+1} = Rw_t + \epsilon_{t+1}$, $R > 1$.

<u>**Bayesian Assumption**</u>

In the equation $\mathrm{w}_{t+1} = Rw_t + \epsilon_{t+1}$, *the center of location of R,* μ_R, *is greater than one*.

Scientific Modelling

A definition of what constitutes a scientific model is necessary here. It seems to require at least two parts. The first part is mathematical coherence. This only requires that the models follow the standard rules of mathematics unless some axiom or postulate is added to create differences. Any standard regularity conditions assumed by economists may be included implicitly. Fundamentally, the connections must be logical and consistent with the rules of mathematics.

The second portion is that the variables and/or parameters of interest are measurable and inference about those parameters is possible. If some aspect of the model could not be measured, then it fails the second criteria.

Boundary Conditions

Neither the models of mean-variance finance, nor other economic models with stochastic difference and differential equations tend to include boundary conditions. It is possible that prices could be infinitely negative where a normal distribution is used and there is no upper bound in resources. The consequences of this are not necessarily trivial. There are two potentially large consequences of boundary conditions being absent.

The first is that frequency based statistics tend to explicitly or implicitly depend upon rank statistics in order to perform significance testing when a Cauchy distribution is present. If the Cauchy distribution is truncated on the left at zero, but the center of location and scale parameters are unknown, then the rank measures are shifted an unknown amount. Many estimators depend upon the median being the center of location. With truncation, the median and the mode no longer match. The mode, as the basin of attraction, is now the center of location.

The second has to due with thin tails and market failure. If one posits that a future budget constraint exists, then there exists a positive probability that the constraint will be to the left of the market clearing price causing a market to fail. This both skews the distribution and thins the tails from the tails expected by a Cauchy distribution. Not accounting for bankruptcy on the left and potential market failure on the right results in a truncated, skewed distribution without finite variance and possibly without known analytic properties.

RETURNS

One of the large challenges in financial economics has been explaining and modeling the presence of heavy tails in the distribution of returns. While many difficult models have been proposed, they are based on the fit to the data and not on beliefs about how humans must behave in an economic system. A difficulty in finding a solution is that the Bayesian solution and the frequency based solution are not the same at all, even though they have the same outcome.

Intuition Behind the Proof

Bayesian statistics are a form of case-based reasoning. Each data point is an individual case and the goal is to extract any relevant information from each data point. This happens through the likelihood function. Looking at the data on a point-by-point basis, the question is whether a natural likelihood function exists for R from which to extract information.

The intuition behind the proof for the distribution of returns can be constructed from a far simpler method already used by economists, that is, to divide the realized future value by the present value. Since Bayesian methodology permits viewing one data point at a time for information, it is possible to discuss the sample by discussing it one point at a time through the process called Bayesian updating. Choosing any one observation at an arbitrarily chosen time t and given the earlier equations, a specific generic observation R_t can be defined thus:

At the moment in time just prior to time t the order opening the allocation w_t is placed. Using the assumptions of the models, the actors are price takers. As such the errors will be by the counterparties to the buy and the sell at both points in time. R_t has a distribution from which its realization will be drawn based on the ratio of two future random variates w_t and w_{t+1}. Once observed, $R_1 \dots R_t$ become fixed points from which inference about μ_R can be performed.

For any observation about R_t, note that:

$$w_{t+1} = \mu_R w_t + \epsilon_{t+1} \tag{3}$$

where

$$w_t = \mu_R w_t + \epsilon_t \tag{4}$$

and this resolves to:

$$R_t = \frac{\mu_R^2 w_{t-1} + \mu_R \epsilon_t + \epsilon_{t+1}}{\mu_R w_{t-1} + \epsilon_t} \tag{5}$$

This simplifies to:

$$R_t = \mu_R + \frac{\epsilon_{t+1}}{\mu_R w_{t-1} + \epsilon_t} \tag{6}$$

Since $\mu_R w_{t-1}$ is a constant, R_t is a function of the ratio of two random variables. The question that remains is to determine the distribution of the shock. If the ubiquitous answer in economics is used, which is that ϵ converges to a normal distribution, then by well known theorem the distribution of R about its center of location across time is a Cauchy distribution. (Geary, 1930) (Gurland, 1948)

On the other hand, if the basic tenets of mean-variance finance are accepted then many buyers and sellers exist. Equities are traded in a double auction. If we add the assumption from mean-variance finance that the market is in equilibrium then there would be no "winner's curse," because it would only be rational for individuals to bid the expectation of their personal subjective distribution of valuations. The distribution of appraisals will, by the central limit theorem, converge to normality as the bids are in fact the expected future sample means of each actor's distribution of appraisal values. In that case also, the Cauchy distribution will be present for the returns.

Under very mild assumptions, the likelihood for R should converge to a Cauchy distribution in each static period in a Bayesian framework. This intuition permits the transition from an estimator of $\hat{R}$ to $\hat{R}|w_{t+1} = Rw_t + \epsilon_{t+1}$. The best way to do this is to begin with the Frequentist proof by White.

The Frequency Based Solution

Frequency based statistics are a form of deductive reasoning. The goal is to create a statistical form of modus tollens. An hypothesis is created and then the data is tested as if the hypothesis were true. If the test rejects the hypothesis, then to some degree of confidence, the hypothesis is false. The concern here is the construction of a test which could falsify an hypothesis.

Noting that R is a fixed point, the goal is to construct a test which could be based upon an hypothesized R and an estimator $\hat{R}$. White notes that from prior research, the maximum likelihood estimator for $\hat{R}$ given that $w_{t+1} = Rw_t + \epsilon_{t+1}$ is the least squares estimator, for all possible values of R. (White, 1958) Normalizing the scale parameter to 1, he notes that the limiting distribution of $\hat{R} - R$ is the Cauchy distribution, where ϵ follows any distribution with finite variance and is centered on zero. It is also assumed that $\epsilon_t \perp \epsilon_{t+1}$.

Impact of White's Frequentist Proof on the Bayesian Likelihood Function

A Bayesian solution could follow directly from White's proof for two reasons. First, the form of the proof has a Bayesian interpretation; but secondly, under suitable regularity conditions the asymptotic posterior can be estimated from the Fisher information and the maximum likelihood estimate. (Koop, Poirier, & Tobias, 2007)

While the Bayesian method has made use of the method of maximum likelihood since at least Laplace and Gauss, it is used as a special case of the method of maximum a posteriori. (Jaynes, 2003) Bayesian methods require a prior distribution for the parameters of interest. If that prior distribution is the uniform distribution then the two methods are computationally identical. This is important as it also means the distributions are identical, although White was solving a Frequentist problem. While White was solving a different type of problem, his proof happens to have a Bayesian interpretation.

White solves for the distribution by normalizing the distribution of the difference between the estimated value and the true value of the center of location with the square root of Fisher information. In Bayesian statistics, the square root of Fisher information is known as the Jeffreys' prior. (Lee, 2004) Although the Cauchy distribution has no Jeffreys' prior, the likelihood estimator of R given the difference equation does have one. For all finite samples, it is a constant.

The Jeffreys' prior is an uninformative prior that is invariant under transformation of the data. By multiplying the distribution about the estimate by the Jeffreys' prior, it added no information to the posterior distribution and only the information contained in the likelihood function passed into the posterior.

There is a question then about the likelihood function. White's proof indirectly addresses this. In White's proof it is observed that product of the Jeffreys' prior and the distribution of the error maps to the product of the Jeffreys' prior and the distribution of the ratio of two random variates. This ratio is shown to converge to a Cauchy distribution. This ratio is the likelihood function.

Effectively what White has shown is that the product of the likelihoods, also known as Bayesian updating, has the same distribution as the ratio distribution of a future value and a present value. Since the product of a series of Cauchy distributions is a Cauchy distribution, and White shows that for ϵ of any distribution which admits a mean of zero and finite variance, the distribution of R about the true value is a Cauchy distribution. The predictive distribution of returns is also a Cauchy distribution.

The question that remains is whether an uninformative prior reasonable? At time zero, before humans invented capital there was no information about the value of capital. As time goes to infinity, that value becomes more certain. Since there was no information at time zero about its value, then it is reasonable to use an uninformative prior. As the likelihood function, though not its value, can be assumed to be invariant across time, then it is reasonable to apply a Cauchy likelihood function to the data.

EFFECT ON CURRENT THEORY

The effect of the Cauchy distribution on existing theory depends, of course, on what part of theory is being discussed. For some areas of finance and economics, the use of a mean or a variance was only a convenience and the results would be approximately the same on a distribution free basis. For others, the problems are more extensive.

Mean-Variance Finance

There are three principle normative models in mean-variance finance: the Capital Asset Pricing Model (CAPM), the Arbitrage Pricing Theory (APT), and the Black-Scholes Option Pricing Model and related Itô calculus based methods (OPM). It is simplest to begin the discussion with the CAPM as Black-Scholes can be derived from it. The form of the Black CAPM is the simplest in that it has the fewest number of assumptions. (Black, 1972) The goal is to choose a portfolio of securities while minimizing the portfolio's variance by choosing a desired level of return. The dual problem of maximizing return while choosing the variance would have the same mathematical outcome. The form of the Black CAPM is:

$$\underset{\{\boldsymbol{s}'\}}{Min}\ \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s} \tag{7}$$

subject to:

$$\boldsymbol{s}'\,\mathbf{1}=1 \tag{8}$$

and

$$E(\boldsymbol{s}'\boldsymbol{R} + \boldsymbol{s}'\lambda)=\mu_{portfolio} \tag{9}$$

In these equations: **s**; **Σ** is a covariance matrix; **1** is a vector of ones; **λ** is a vector of normally distributed errors; and *R* is an unobserved true growth rate. Two implicit assumptions of mean-variance finance are brought out here.

<u>Assumption</u>

In models of mean-variance finance, an expected return on investment (or alternatively expected reward) exists.

<u>Assumption</u>

In models of mean-variance finance, a variance of returns exists. For multi-asset models, a positive definite covariance matrix of returns exists.

Bayesian Interpretation

The Bayesian interpretation of this formulation would have the vector of returns to be drawn from Cauchy distributions. The share of the portfolio for any given asset is not stochastic and as such can be treated as a constant for the purposes of forming the expectation.

What does need to be solved is the predictive expectation of $R_i\,|w_1 \ldots w_t$, for each asset i. Given the most general form of the Cauchy distribution, the expected return is:

$$E(R_i|\mu_i, \sigma_i, w_1 \ldots w_t) = \int_{-\infty}^{\infty} \frac{R_i \sigma_i}{\pi(\sigma_i^2 + (R_i - \mu_i)^2)} dR_i \quad (10)$$

$$= \frac{\sigma_i \log(\mu_i^2 - R_i\mu_i + \sigma_i^2 + R_i^2) - 2\tan^{-1}\left(\frac{\mu_i - R_i}{\sigma_i}\right)}{2\pi}\Bigg|_{-\infty}^{\infty} \quad (11)$$

$$= \infty - \infty + \mathbf{0} \quad (12)$$

It follows that $E(R_i \,|\, \mu_i, \sigma_i, w_1 \ldots w_t)$ does not exist, for any i. This contradicts the above assumption that it does exist. The CAPM is false by contradiction.

Similar assumptions about returns are present in the APT and the OPM. Since the mean does not exist, the variance about the mean does not exist. Nothing about the CAPM is mathematically coherent in Bayesian statistics. Since the math is not valid, it cannot be a valid scientific model.

The Frequentist Interpretation

The Frequentist interpretation of the same set of equations is quite different. In the Bayesian interpretation, none of the necessary expectations exist for the model to function. In the Frequentist interpretation they must exist as they are non-random fixed points. The expectation operator only has the effect of getting rid of the diffusion term as the drift term is fixed though unknown. The question isn't whether such a model can be constructed, but rather whether the data can falsify it?

At this point, it is important to be careful how to interpret this model of fixed but unknown points. There are a number of dangerous statistical traps to be found in this construction.

Consider the question of how people find the equilibrium conditions? Whereas Bayesian methods could be interpreted as a tool for the search algorithm, Frequentist methods posit finding the equilibrium as true by assumption.

There are two paths possible. One leads to the idea of fiducial statistics and the other to perfect foreknowledge. While fiducial statistics is a largely discredited topic, research on the field still continues. (Hampel, 2003) The alternative, perfect foreknowledge has a deus ex machina element to it.

The attempt to construct fiducial statistics by R.A. Fisher was based on a very simple observation. In performing a significance test on an hypothesis, say $\mu=5$, it should be possible to perform a significance test for every value on the real number line, not merely at five. This collection of tests does not end up forming a proper density function. As tempting as fiducial statistics is, it turns out to not be valid.

The mechanism to arrive at the equilibrium is unclear; it only matters that it is assumed that the arrival happens. Although this creates some philosophical discomfort, it is necessary discomfort. The methodology requires positing that the model of fixed points is true. The mechanics of the process remain a mystery.

Noting that $\hat{R}_i - R_i$ is drawn from a Cauchy distribution and that R_i is a fixed point, it follows that $\hat{R}$ is drawn from a Cauchy distribution. It was noted earlier that prior proofs have shown that the maximum likelihood estimator for R_i is the least squares estimator.

The least squares estimator is the estimator for the expectation for the slope. The algorithmic solution for the least squares estimator represents the effect of the sample on the test. The question, however, is about the ability to perform inference on the CAPM. Can it be shown as true or false?

What are the properties of any significance test of the CAPM(or any standard mean-variance model), given the mathematical properties of the model(s) are strictly true?

As precision is defined as the reciprocal of the variance, one can find the precision of a test by finding its asymptotic variance about a point. For all R_i the precision of the test for a sample is estimated knowing that $\hat{R}$ is drawn from a Cauchy distribution.

Although a variance is a form of expectation, in order to construct this, the Cauchy principal value will be used instead as no variance about the mean can exist.

$$Var\left(\hat{R}_i - R_i\right) = \lim_{c\to\infty}\int_{R_i-c}^{R_i+c} \frac{\sigma_i}{2\pi}\frac{(\hat{R}_i - R_i)^2}{\sigma_i^2 + \left(\hat{R}_i - R_i\right)^2} d\hat{R}_i \tag{13}$$

$$= \sigma_i\hat{R}_i - \sigma_i^2 \tan^{-1}\left(\frac{\hat{R}_i - R_i}{\sigma_i}\right)\Bigg|_{-\infty}^{\infty} \tag{14}$$

$$=\infty \tag{15}$$

Therefore, at the limit, any significance test is of precision zero even with an infinite amount of data. The CAPM is immeasurable in the Frequentist paradigm. While by construction it must be a valid mathematical model, it is not a valid scientific model as the CAPM and any other mean-variance model cannot be constructed with valid measures as written.

It is important to note that there is a valid methodology when dealing with the Cauchy distribution in both frequency based and Bayesian statistics, but to go to those methods is to assume mean-variance finance is false.

A separate estimation issue occurs when economists estimate the CAPM and related mean-variance finance tools by directly taking market returns, subtracting the risk-free rate and using that difference to form the standard β-based solution.

If the returns are treated as data, then they become random variates and from this random variate is subtracted a constant, the risk-free rate. As a constant by assumption, the risk-free rate cannot have a distribution associated with it at any time t. As in the Bayesian intuitive solution, the test statistic of returns will converge to a Cauchy distribution and be translated by an amount equal by the risk free rate. If the least squares method is used as an approximation, the effect of the algorithm on the interpretation of results needs to become the foremost question.

Fortunately, this is already answered in the literature. Sen finds that the asymptotic relative efficiency of the method of least squares is zero compared to any median based method. (Sen, 1968) To adopt a median based method is to abandon the mean-variance method. If it is used as an approximation, then anyone using Theil's method of regression would gain an immediate advantage over the mean based method as Theil's method, especially if augmented with other median estimation tools, has the highest known efficiency. That being known, it should be possible to form a statistical arbitrage process over mean-variance users and systematically win. Standard economic theory rules that out, so this approximation should be excluded by both statistical theory and economic theory. Why would someone knowingly adopt a perfectly inefficient tool or even a tool is perfectly inefficient on a relative basis when standard tools exist that are efficient?

Finally, there is the log difference approximation. This one is a bit more challenging to address. There are two reasons to use the logarithmic transformation of prices to arrive at an approximation of return. The first is to linearize the data to make it easier to work with. The other is to use it for reproducibility with older studies. Older studies took the differences in the logs of the prices as an approximation due to poor computing power.

There are two real issues with this latter usage. First, the underlying theory makes no sense in logarithmic space. People do not purchase log(*5000 shares*) for log(*\$5.00 per share*). Second, using a distorting approximation simply because the last person did so defeats reason. Things do not gain validity simply from tradition or age. The originators of the practice did it from computational necessity. That constraint no longer exists.

The first case, linearizing the data, is a valid goal. Nonetheless, the use the logarithmic transformation is not problem free. The logarithmic transformation trims the tails so that the distribution is no longer heavy tailed because the reward (or return) on prices is no longer what is being measured. To understand why, note that systematic rewards must be greater than one but also less than e, the base of natural logarithms. Transformed into logs if $p_1/p_0 = 1.05$ then $\log p_1 - \log p_0 \approx .05$. If this is systematically true, then the regression estimator of return will be between zero and one. So by the results of Mann and Wald, it follows that returns will converge to normality as the sample size becomes very large. (Mann & Wald, 1943)

Unfortunately, it is no longer possible to determine what the coefficients mean. The value of the regression constant is now multiplicative, when theory says it should be near zero and additive. The allocations are no longer allocations and it isn't clear what they have become. Finally, β should map onto Theil's regression if the various components of the stock market are strictly independent, but they are not.

It is not invalid to use the logarithmic transformation, but this doesn't support mean-variance finance either. Indeed, it is somewhat difficult to determine what is being supported. There is an information loss in the logarithmic transformation, but it isn't clear what that implies for human behavior acting in markets.

Heavy-Tailed and Econophysics Methods

Although the Cauchy distribution lacks an expectation, certain utility functions will have an expectation. If wealth is drawn from a Cauchy distribution and the utility function is logarithmic utility, then expected wealth is a function of the Bose-Einstein distribution. The indefinite integral for the expected utility of wealth is:

$$\begin{aligned} E\big(U(\tilde{w})\big) = \frac{-i\sqrt{\sigma}}{2}\Bigg\{&\mathrm{Li}_2\left(\frac{w}{\mu + i\sqrt{\sigma}}\right) - \mathrm{Li}_2\left(\frac{w}{\mu - i\sqrt{\sigma}}\right) \\ &+ \log(w)\left[\log\left(1 - \frac{w}{\mu + i\sqrt{\sigma}}\right) - \log\left(1 + \frac{w}{i\sqrt{\sigma} - \mu}\right)\right]\Bigg\} \\ &+ constant \end{aligned} \tag{16}$$

The Li_2 operator is the dilogarithm, a special case of the polylogarithm. Although the polylogarithm can be defined as a series, it can also be defined as the Bose-Einstein distribution divided by the gamma function. This would bring equity securities into Bose-Einstein statistics by simply solving the above problem with reference to logarithmic utility. If the same problem were solved using zero as the lower bound for wealth and infinity as the upper bound the problem simplifies to:

$$\left(U(\tilde{w})\right) = \frac{1}{4} i\sqrt{\sigma}\left[\log^2\left(-\frac{1}{\mu + i\sqrt{\sigma}}\right) - \log^2\left(\frac{1}{i\sqrt{\sigma} - \mu}\right)\right] \tag{17}$$

This brings us to exponential utility. The indefinite integral for the expected utility of wealth, where $U(\tilde{w})= -e^{-\alpha w}$ is:

$$E\left(U(\tilde{w})\right) = -\frac{i\left(e^{2i\alpha\sigma}\mathrm{Ei}\left(\alpha(-w + \mu - i\sigma)\right) - \mathrm{Ei}\left(\alpha(-w + \mu + i\sigma)\right)\right)}{2e^{\alpha(\mu + i\sigma)}} \tag{18}$$

The definite integral from zero to infinity

$$\begin{aligned} E\left(U(\tilde{w})\right) = \frac{1}{2} i &\left(e^{-\alpha(\mu - i\sigma}\left(\mathrm{Ei}\left(\alpha(\mu - i\sigma)\right) + \log(-\mu + i\sigma) - \log(\mu - i\sigma)\right)\right. \\ &\left. - e^{-\alpha(\mu + i\sigma)}\left(\mathrm{Ei}\left(\alpha(\mu + i\sigma)\right) - \log(\mu + i\sigma) + \log(-\mu - i\sigma)\right)\right) \end{aligned} \tag{19}$$

The Ei operator is the exponential integral operator used in neutron transfer and interstellar heat problems. Very quickly, simple models of rational expectations turn into deep physics problems.

Behavioral Finance

It is an observation of behavioral finance that the utility function should be concave in gains and convex in losses. (Thaler & Dawes, 1992) Although a complicated model could be constructed, a simplified model has interesting implications. A function that naturally is convex on the left and concave on the right is the arctangent. The arctangent is also the cumulative density function of the Cauchy distribution. Behavioral finance implies that losses are weighted more than gains. Ignoring that for a second, it should be noted that unweighted arctangential utility is risk neutral.

Unfortunately, there isn't a known analytic solution for a general form of expected arctangential utility, but there is one if the utility of wealth is centered on μ. In that case, if utility is:

$$U(\tilde{w}) = \alpha \tan^{-1}\left(\frac{w - \mu}{\sigma}\right) \tag{20}$$

where α is a weighting over some segment of the function. This allows for piecewise integration to meet the needs of behavioral finance. Under this set of assumptions, the indefinite integral for expected arctangential utility becomes:

$$E\left(U(\tilde{w})\right) = \alpha\left[\tan^{-1}\left(\frac{w - \mu}{\sigma}\right)\right]^2 + constant \tag{21}$$

This is a very simple mathematical function with the interesting property that it is the square of the cumulative density function. In all three of the above cases with definite integrals, it should be possible to construct allocation models using the

Envelope Theorem as they are all functions of the portfolio mode and probable error.

What may not be obvious is that the Cauchy distribution is intimately linked to complex numbers, the logarithm and the trigonometric functions. Another aspect that may not be apparent is the absence of a Taylor expansion. Because the cumulative normal distribution lacks an analytic form, it is common in economics to perform estimates around a point. This is not an issue for the Cauchy distribution, but as the Cauchy distribution lacks moments, if a Taylor expansion were needed, it does not exist. These seemingly mild changes have significant consequences for standard modeling tools.

Macroeconomics

Heavy tailed studies have generally discarded the Cauchy distribution in empirical studies since the broader class of four parameter stable distributions provides a better fit and also have no defined variance as implemented. This gives the possibility to explain certain elements of the Keynesian/Classical split.

Consider a particular, but imaginary, allocation at time t, with w_t being the specific holdings of some firm by a specific household. It is from this point that a specific discussion of an abstracted model can form. So consider a purchase on January 3rd, 2011 of 100 shares of IBM stock at $147.50 per share from a family's endowment of cash. The changes in prices seen in the market are changes in the endowments in the budget constraint of the various households and firms as time unwinds. One year later, the position is closed at $187.00 per share. The cash is moved from the endowment of cash of another household or firm to the current budget constraint of the family in question.

Decisions about what to do with the funds are part of a constrained optimization problem for the household. The question in forming the model by the economist is “what does the model need to do?” A choice of distributions then determines the model, or alternatively, choosing a model determines the nature of the possible distribution in use.

Assume two models exist that are rough equivalents in the sense that they both map onto the observed sample of returns. They will be treated as equivalent for the purpose of explaining returns in the sense that it isn't obvious one model is better than another. One model features a possible set of returns that are drawn from the best fit four parameter stable distribution. The other model features a possible set of returns from a mixture distribution. Both models are supported only on the non-negative real numbers to allow for bankruptcy. The mixture model is a mixture of a Cauchy distribution, as above, with a distribution for constraints on the future budget constraint of counterparties.

Although not directly observable, external constraints determine the frontier of the family's budget. As an example, bank reserve requirements, legal lending limits as a function of bank capital, prudential regulation and loss reserve requirements all play a role in the limitations on the capacity of the family to access liquidity.

The model with the stable distribution implicitly has no form of borrowing constraint. It is conceptually possible to borrow infinite sums. The distribution is skewed, but the sources of skew are not part of the model. Had those sources been separately modeled, then the distribution of returns, subject to those sources, would become a symmetric Cauchy distribution. Skew in the data warns of the possible existence of information not accounted for in the model.

Now consider a relatively simple model that includes planetary product as well as a constraint on what portion of planetary income could be spent on investment activities. With Y being planetary product, γ the growth rate, and ε a normal error, let the difference equation for planetary product be:

$$Y_{t+1} = \gamma Y_t + \varepsilon_{t+1} \tag{22}$$

Let the same equation of value persist, $w_{t+1} = Rw_t + \epsilon_{t+1}$, but with an added constraint:

$$\boldsymbol{w}_{t+1} \leq \alpha Y_{t+1},\ 0 < \alpha < 1 \tag{23}$$

Assume that α is a prudential or political constraint that limits the external costs from over investment. In this model there is no money supply, although it would result in a similar outcome if a money supply were used. In this circumstance, α is a non-market constraint on the budget constraint.

From Bayes law:

$$\Pr(\boldsymbol{w}_{t+1}|\boldsymbol{w}_{t+1} \leq \alpha Y_{t+1}) \propto \Pr(\boldsymbol{w}_{t+1} \leq \alpha Y_{t+1}) \,\big|\, \boldsymbol{w}_{t+1}) \Pr(\boldsymbol{w}_{t+1}) \tag{24}$$

Since the distribution for αY_{t+1} must be a Cauchy distribution, given the assumptions, it follows that the probability of choosing a value for w_{t+1} such that it is also less than or equal to αY_{t+1} is the cumulative density function from 0 to w_{t+1}. Further, since it is truncated at 0, this probability is:

$$\Pr(w_{t+1} \leq \alpha Y_{t+1}|w_{t+1}) = \frac{\pi - 2\tan^{-1}\left(\frac{w_{t+1} - \alpha\mu_Y}{\alpha\sigma_Y}\right)}{\pi + 2\tan^{-1}\left(\frac{\mu_Y}{\sigma_Y}\right)} \tag{25}$$

With the unconditional distribution truncated at zero, w_{t+1} has the density function:

$$\Pr(w_{t+1}) = \frac{\sigma_w}{\left[\frac{\pi}{2} + \tan^{-1}\left(\frac{\mu_w}{\sigma_w}\right)\right]\left[\sigma_w^2 + (w_{t+1} - \mu_w)^2\right]} \tag{26}$$

So the density function for $\Pr(w_{t+1}|w_{t+1} \leq \alpha Y_{t+1})$ is the product of the two terms, divided by the constant of integration. Currently, the constant of integration is unknown. An analytic solution is yet to be found. Nonetheless, numerical methods to estimate it exist.

If the value of physical capital, k_{t+1} is modeled using the same autoregressive of degree one explosive process as elsewhere, then it will be independent of w_{t+1}, even if both processes depend upon k_t. Then it is possible to talk about binding constraints on the capital markets as not permitting prices to reach a free market clearing price in the short run. If the free-market equilibrium clearing price is defined as $k_{t+1} \approx w_{t+1}$ then disequlibrium could be defined when the two values are far apart. Otherwise, in such a model, the present value of cash flows from physical capital, the price of capital and the price of the financial captial representing it should be equal when adjusted for acquisition costs.

The classic Keynesian prescription when far from equilibrium is to lift the constraints would either to be to relax α or to increase Y. However, it isn't clear this is the correct solution.

How the constraint is set should matter. All that is posited is that a constraint exists. It may serve prudential goals to protect the broader society. It could also serve purely political goals to protect elected officials and regulators.

When the regulatory constraint is binding, it is quite possible that the liquidity available would prevent the market price of financial capital from equalling the discounted present value of the physical capital in the system.

Since physical capital is a slowly decaying stock, compared to the speed of capital market trades, it can be completely unaffected by capital market errors. However, if the constraint is systematically binding for some time, the real economy can be impacted as there are two channels through which new physical capital is formed.

The first channel is through the reinvestment of cash flows. The second is through the formation of new capital. Although there are no explicit loans in this model, if one disaggregated the components of w_{t+1}, then both equity IPO's and loans could be made.

Assume that both financial capital and physical capital can be purchased by firms. If the yield on financial capital is higher than the yield on physical capital, to some degree of probability, then the capital stock should fall to meet the market value of the capital stock in the capital markets. In the classical model this would be a very desirable response. Actors who over-built would find the market respond adversely and the excess physical capital would depreciate out of existence. The market would adjust on its own. No activity from the constraint setting body would make sense. In this case, a prudential regulation serves an efficiency purpose.

If the appropriate value of α is uncertain set then the problem of a binding constraint is multi-fold. If a prudential constraint is reached, should it be relaxed if it is actually prudential?

This triggers two possible cases. If it is believed the value of α was correctly set then it should not be changed. If the value was incorrectly set, then it should be altered up or down to the appropriate prudential level.

What if the constraint is a political constraint instead of a prudential one, such as maintaining employment? Then it is quite possible the government should expand spending to increase the value of Y to make the constraint slack in the following period. Alternatively, the terms of the constraint could be relaxed, possibly through open market operations to support the value of capital.

This raises the question of the distribution involved. In the four parameter model the budget constraint and constraints on the budget constraint, such as reserve requirements, are implicit. This can only be a classical style model.

In its Keynesian form the above mixture distribution depends upon the structure of such regulation. Of course if the constraint is non-linear, the above distribution would be a poor fit. Another distribution should be modeled.

Although the above is less than a toy model of the economy, it is an attempt to point out that distributions should not be assumed into existence. Distributions should follow from the models employed.

Regnault and Bachelier

So why did Regnault and Bachelier observe what they observed? They were studying the price movements, over short periods of time, of *rentes*, a fixed income investment. (Bachelier, 2006) While an investment in stocks has an uncertain future value, an investment in bonds does not. Each cash flow discounted from face value for a fixed rate bond will mature with fixed value, assuming that payments are made as agreed. The risk at time zero is in appraising the probability of a failure to pay. Although there are a number of ways to model that probability, one of the simplest is the normal distribution.

Indeed, with all things except the probability of payment being fixed and certain, the only appraisal risk comes in a failure to estimate bankruptcy risk. While the reinvestment risk for the portfolio of cash flows probably does converge to a Cauchy distribution, each specific bond has an upper bound payment, a lower bound at zero and a probability for each intermediate cash flow. An expected value exists and it is usually associated with a finite period of time, not the unbounded life of the equity of a corporation.

CONCLUSION

This paper can be viewed in one of two ways. Either it can be used to unite mathematical, empirical and behavioral finance, or it could be viewed as a problem for any one of the three.

This paper argues that in the world of blackboard economics, the limiting distribution of returns as time goes to infinity is the Cauchy distribution. This paper almost ignores the budget constraint, bankruptcy, bank reserve requirements, taxes and an entire host of other things which fundamentally alter the distribution of returns. Nonetheless, mean-variance finance models are the normative models of economics. They are taught at all levels, they are on doctoral comprehensive exams, they are used in industry, and they underlie regulatory models explicitly or implicitly. This paper requires the abandonment of mean-variance finance.

Generally Accepted Accounting Principles use mean-variance methods. Textbooks carry the CAPM and have students calculate the cost of capital based on the model. Real firms evaluate management and projects on what is an improper algorithm. Regulatory models that use a normal distribution rather than a Cauchy distribution to calculate required capital for financial intermediaries will tend to undercapitalize those institutions. Hedge funds using Itô calculus based methods are using methods uncorrelated with the true model.

The profession has discussed this issue for fifty years. This study appears to be the first to provide a mathematical reason to reject mean-variance finance on its own assumptions. On the other hand, there are general ideas from mean-variance finance that should be discussed under the new distributions.

Markowitz's and Roy's intuition on this issue is invaluable. Markowitz, with good reason, tried to turn the prior paradigm on its head by making risk explicit. Maybe as important, mean-variance finance discusses market efficiency. While a Cauchy distribution eliminates the idea of expectational efficiency, it does not imply markets are strictly inefficient either.

While mean-variance models collapse, not all models using a normal distribution should do so. It has been quite often the case that the normal distribution was used as a mathematical convenience and was unnecessary to support the primary findings of models. In those cases, the use of the normal distribution overly constrains the models that used it as a convenience. The question becomes whether or not the model be shown to hold in a distribution free environment? If so, then the model survives. If it depends upon normality, or at least finite variance, then it does not survive.

The challenge of the next decade is the careful inspection of existing models and the development of new ones. It is reasonable to believe that the behavioral and mathematical models will merge. Both should provide insight and that should trigger growth in the other models.

Divergent time series are unlike their convergent cousins. In convergent series, behavior is driven to a point at the limit. In divergent series, behavior is expansive. Humans are not restricted to tight well-behaved series. Entrepreneurship and catastrophes happen in divergent series. A robust model of human behavior is now possible; one with religion and marriage; stocks and bonds; trading alliances and friendships; even children are in this broader model. When someone defers consumption, anticipating a future gain in utility from that deferral, then the Cauchy distribution appears.

Financial economics, which was traditionally about investment behavior, should include marriage, religion, child rearing and all those transformational processes where consumption is deferred for future gains. Buying a toy for a small child in anticipation of a smile is no different than placing money in a security with the anticipation of happiness later. The application is different but the math is the same.

In both the blackboard world of economics and the real world of application it is time to set mean-variance finance aside and move forward. At the most basic level there are a host of questions to be answered. It is time to tie together the disparate pieces of economics into a common tapestry of methods, tools and ideas.

ENDNOTES

[1]See in particular the Uniform Prudent Investor Act as proposed by the National Conference of Commissioners on Uniform State Laws for the justification for the language of the act. (National Conference of Commissioners on Uniform State laws, 1995)

REFERENCES

Bachelier, L. (2006). Theorie de la Speculation. In L. Bachelier, Theorie de la speculation. English: Louis Bachelier's theory of speculation the origins of modern finance (M. H. Davis, & A. Etheridge, Trans.). Princeton, NJ: Princeton University Press.

Ball, P. (2006, February). Econophysics: Culture Crash. *Nature*, 441(7094), 686-688.

Black, F. (1972). Capital market equilibrium with restricted borrowing. J*ournal of Business*, 45(3), 444-455.

Clark, J. B. (1908). The Distribution of Wealth: A Theory of Wages, Interest and Profits. New York: The MacMillan Company.

Davis, M. H., & Etheridge, A. (2006). Introduction. In L. Bachelier, Theorie de la speculation. English:Louis Bachelier's theory of speculation the origins of modern finance (M. H. Davis, & A. Etheridge, Trans.). Princeton, NJ: Princeton University Press.

Dickey, D. A., & Fuller, W. A. (1979, June). Distribution of the estimators for autoregressive time series with a unit root. J*ournal of the American Statistical Association*, 74(366), 427-431.

Donsker, M. D. (1951-52). An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6, 1-12.

Edgeworth, F. Y. (1888). The Mathematical Theory of Banking. *Journal of the Royal Statistical Society*, 51(1), 113-127.

Fama, E. F., & French, K. R. (2008, August). Dissecting anomalies. *The Journal of Finance*, LXIII(4), 1653-1678.

Fienberg, S. E. (2006). When did Bayesian inference become "Bayesian?". *Bayesian Analysis*, 1(1), 1-40.

Fisher, I. (1930). The Theory of Interest: As Determined by Impatience to Spend Income and Opportunity to Invest It. New York: The MacMillan Company.

Geary, R. C. (1930). The Frequency Distribution of the Quotient of Two Normal Variates. *Journal of the Royal Statistical Society*, 93(3), 442-446.

Gurland, J. (1948, June). Inversion formulae for the distribution of ratios. *The Annals of Mathematical Statistics*, 19(2), 228-237.

Hampel, F. (2003). The proper fiducial argument. Seminar fur Statistik-ETHZentrum.

Hicks, J. R. (1939). Value and Capital: An inquiry into some fundamental principles of economic theory. Oxford: Clarendon Press.

Jaynes, E. T. (2003). Probability Theory: The Language of Science. Cambridge: Cambridge University Press.

Keynes, J. M. (1936). The General Theory of Employment, Interest and Money. New York: Harcourt Brace.

Koop, G., Poirier, D. J., & Tobias, J. L. (2007). Bayesian Econometric Methods (Vol. 7 of Econometric Exercises). New York: Cambridge University Press.

Lee, P. M. (2004). Bayesian Statistics: An Introduction (3rd ed.). New York: Arnold Wiley.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394-419.

Mann, H. B., & Wald, A. (1943). On the statistical treatment of linear stochastic

difference equations. *Econometrica*, 11, 173-200.
Markowitz, H. (1952, March). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91.
National Conference of Commissioners on Uniform State laws. (1995, April 18). Uniform Prudent Investor Act. Retrieved 2013, from Uniform Law Commission: http://www.uniformlaws.org/shared/docs/prudent%20investor/upia_final_94.pdf
National Institute of Standards and Technology/Sematech. (2012, April). NIST/SEMATECH *e-Handbook of Statistical Methods*. Retrieved from http://www.itl.nist.gov/div898/handbook/
Pearson, E. S. (1955). Statistical concepts in the relation to reality. *Journal of the Royal Statistical Society*. Series B (Methodological), 17(2), 204-207.
Rao, M. M. (1961, March). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *The Annals of Mathematical Statistics*, 32(1), 195-218.
Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica*, 20, 431-439.
Sen, P. K. (1968, December). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, 63(324), 1379-1389.
Stigler, S. M. (1974). Studies in the History of Probability and Statistics. XXXIII: Cauchy and the witch of Agnesi: An historical note on the Cauchy distribution. *Biometrika*, 61(2), 375-380.
Thaler, R. H., & Dawes, R. M. (1992). The Winner's Curse: Paradoxes and Anomalies of Economic Life. Princeton: Princeton University Press.
Veblen, T. (1899). The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions. New York: The Macmillan Company.
Veblen, T. (1904). The Theory of the Business Enterprise. New York: C. Scribner's Sons.
von Bohm-Bawerk, E. (1890). Capital and Interst: A Critical History of Economical Theory. (W. Smart, M.A., Trans.) London: MacMillan and Co.
von Bohm-Bawerk, E. (1891). The Positive Theory of Capital. (W. Smart, M.A., Trans.) London: MacMillan and Co.
von Neumann, J., & Morgenstern, O. (1953). Theory of Games and Economic Behavior (3rd ed.). Princeton: Princeton University Press.
White, J. (1958, December). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, 29(4), 1188-1197.
Yilmaz, B. Z. (2010). Completion, pricing and calibration in a Levy market model. The Institute of Applied Mathematics of Middle East Technical University.