The Distribution of Returns

David E. Harris

University of Providence, Great Falls, USA
Email: david.harris@ugf.edu

Abstract

The distribution of the returns on investment depends on the rules in the economic system. The article reviews various return distributions, ranging from equity securities in equilibrium, to antiques bought at auction, to debt instruments with uncertain payouts. A general methodology is provided to construct distributions of returns.

Keywords

Statistical Distributions, Cauchy Distribution, Stock Market Returns, Accounting Ratios, Ratio Distributions

1. Introduction

This article derives the distribution of returns for a variety of assets, liabilities and accounting ratios by asserting two things. The first assertion is that for equity securities that returns are not data, rather prices are data. Returns are transformations of data. The second is that the form of calculating returns and the rules in the economic system determine the distribution that has to be involved. The exception to this would be for instruments like certificates of deposit or fixed annuities where the return is stated, and cash flows follow from the return.

In addition to the form of the calculation, the number of potential buyers and sellers matters; whether errors are independent; the method transactions happen in, such as double auctions, English or Dutch-style auctions and so forth; whether markets are assumed to be in equilibrium or far from equilibrium as was the case in the 2008 financial crisis; how liquidity enters the market; the terminal state of the asset such as a surviving firm, cash, bankruptcy and so forth; and any other constraints such as asymmetric information.

Assuming the normality or lognormality of returns is nearly ubiquitous in economics. Either of these distributions solves a key problem in mean-variance models in that a covariance matrix needs to exist in the data generation process.

Normality is also nice in that it appears in nature quite often. While Markowitz's initial paper appeared in 1952, the first empirical objection was by Mandelbrot in 1963 [1] [2].

Mandelbrot's paper asserted that the data appeared to be some form of Paretian distribution [2]. This was followed on by papers by Eugene Fama [3] [4]; and Fama and Roll in [5] [6] to produce alternative rules and assumptions. Empirical rejection of the models was performed by Fama and MacBeth [7]. Empirical contradictions for the various forms of the Capital Asset Pricing Models are cataloged by Fama and French in [8] and for the Black-Scholes Option Pricing Model by Yilmaz in his master's thesis [9].

The problem created by empirical rejection was that it does not provide a usable solution for economics. While a segment of economics adopted the Fama-French model as an alternative hypothesis, this model is not explicitly grounded in theory [10]. Pearson-Neyman decision theory would encourage the adoption of this model, but Fisherian Likelihood-based methods would merely have rejected mean-variance finance and not provided an alternative solution.

The challenge to economics is that it is constantly attempting to solve inverse problems. When someone is observed purchasing an orange, economists assume that it is the solution to that person's problem. Economists can only see the solutions people use, not the problems which underlie them. The goal of economics is to reverse engineer a generalized solution making process to produce testable predictions. Empirically falsifying a theory without a replacement left financial economics in an untenable position.

The difficulty, as will be shown in Sections 3 and 10.4, is that the distributions involved lack a first moment and in logarithmic form lack a covariance matrix. Nothing necessary for mean-variance finance to function survives, but in addition, neither does the Fama-French model as it exists today.

A population study of data in the Center for Research in Security Prices was conducted in [11] to test the central claims here and mean-variance finance was rejected with posterior odds of more than $10^{8500000}:1$, given a prior probability of $1:999999$. As the Paretian alternative was given only one chance in a million of being correct prior to seeing the data and with it being rejected so resoundingly, the mean-variance family of models are dead without doubt.

Fortunately, knowing the distributions involved permits economic model building. Although this article discusses probability density functions, they should really be thought of as Bayesian likelihood functions. As shown in [12], admissible non-Bayesian solutions do not exist for most of the distributions involved here. The reason is less than obvious.

In the early days of statistics, a fundamental problem was observed. As a statistic is any function of the data and since there are an infinite number of functions, then which function should be used to estimate a parameter? Why would $\sum\left(\sin\left(x_i\right)\right)$ be a bad parameter estimator for the location of the mean of the normal distribution? How does one know that the sample average is a good estimator and that a better estimator does not exist? For the Gaussian, the mean,

median and mode are co-located. Why not choose the mode over the mean?

Fortunately, this problem was solved by Abraham Wald [13]. His solution was clever and it came with unexpected results. Using Frequentist axioms, Wald discovered that all Bayesian solutions were admissible and that non-Bayesian solutions were admissible when they either matched the Bayesian solution in every sample or at the limit. The converse is not true.

To understand why this might be the case, consider solving a problem using either a subjective Bayesian method or an objective non-Bayesian method. If the sample size is large enough, then the two methods must converge if the non-Bayesian method is valid. A subjective understanding of the world must converge to the objective understanding of the world if the sample size were large enough. On the other hand, if the non-Bayesian method did not arrive at nearly the same place, then it does not match reality. Since non-Bayesian methods are conditioned on their model in use, that model in use cannot be true. Instead, the non-Bayesian model was a misunderstanding of the universe.

Because the distributions involved lack a sufficient statistic, it is necessary that non-Bayesian methods will lose information while Bayesian methods will not as the Bayesian likelihood function is always minimally sufficient. It appears that this relationship between sufficiency and admissibility was first discovered by E.T. Jaynes [14].

To avoid these problems in model building, it is enough to note that for independent draws of a variable that if the density function is

$$\Pr(x|\mu) = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}, \forall x \in \chi. \tag{1}$$

where χ is the sample space, then the Bayesian likelihood function is

$$\Pr(x|\mu) = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}, \forall \mu \in \Theta, \tag{2}$$

where Θ is the parameter space. This allows economics to make a robust movement to Bayesian decision theory which is well developed and strongly resembles methods already used by economists. It also differs in that likelihood functions are not probability densities and need not sum to unity.

The prior literature on this topic is split as the statistical and mathematical literature generally do not agree with the economic literature. Although work has carried on in a limited form in the economic literature since Mandelbrot and Fama, most of it has been an attempt to find the best fit distribution. While this has held some sway in the applied component of finance, this is a non-normative viewpoint. Overwhelmingly the literature has been moved to some version of a normal distribution with heteroskedasticity. Unfortunately, attempts to maintain or hold onto normality lead to incorrect solutions.

The statistical literature on this topic is settled, though somewhat sparse. The literature began with correspondence from Poisson to Laplace noting an exception to the central limit theorem. Poisson observed that the distribution $f(x) = \left[\pi\left(1+x^2\right)\right]^{-1}$ was a counterexample to the theorem [15].

The next appearance of this exception is in a battle in the literature between Augustin Cauchy and Irénée-Jules Bienaymé. Bienaymé had written an article showing that ordinary least squares was optimal. Cauchy had just published a method of regression and took this as a personal attack. He searched for a circumstance where the method of least squares would always produce an invalid solution. He arrived at the same solution as Poisson and then dropped the matter [15].

The issue moves into the background again until the Pitman-Koopman-Darmois theorem appears. Ronald Fisher had just discovered the existence of sufficient statistics and this raised the question of when, why and how does a sufficient statistic exist. Pitman, Koopman, and Darmois simultaneously discovered the same effect. A statistic sufficient for a parameter only exists for members of the exponential family. Koopman, in particular, showed the exception of Poisson, now called the Cauchy distribution could not have a sufficient point statistic [16]. This is important because non-Bayesian point estimators would lose information if used.

Indeed, the minimum variance unbiased estimator for the Cauchy distribution purposefully discards information. Since the Cauchy distribution lacks a mean, finding the sample average is pointless. Like all distributions, it does have a median. Rothenberg, *et al.*, in [17] determined that the mean of the central twenty-four percent is the minimum variance unbiased estimator of the population median. The remaining seventy-six percent of the data is discarded. The Bayesian estimator uses all of the data, and the estimate cannot be worse, in the Wald sense, than Rothenberg's method.

Bayesian methods are the oldest probability interpretation beginning with the writing of Thomas Bayes [18]. Formal axiomatization does not occur until later when de Finetti creates the first axioms of probability [13]. Other Bayesian axiom systems are also developed, notably by Savage in [19] and Cox in [20]. The writings of Abraham Wald would produce foundational work in decision theory, for both the Bayesian and the Pearson-Neyman school of thought [13].

In addition to work on probability and statistics, basic work on probability distributions was begun by Pearson [21]. The work would lead others to look at the properties of random variables drawn from these distributions such as additivity and the impact of multiplication or division. Curtiss in [22] published a method to determine the distribution of variables where there is multiplication or division. This was slightly generalized by Gurland in [23]. For the joint normal case, this was expanded by Marsaglia [24] [25].

Solving for the ratio of two random variates permits a general solution to the problems presented here.

Finally, work was performed by Mann and Wald in [26], with subsequent work by John White in [27] and M.M. Rao in [28] on autogregressive processes. Although not Bayesian in nature, these methods have a Bayesian interpretation that will be discussed later.

The economic literature presumes normality and so does not map to the

mathematical or statistical literature.

The paper provides the method of Curtiss found in [22] to move the economic literature back to the mathematical literature and provides other related tools. It starts with the simple assumptions of Markowitz and adds various complications such as correlated errors, systemic disequilibrium, alternative auction types, budget and liquidity limitations, bond certainties, dividends and regression methods. The purpose of this is not to be exhaustive, but rather to show how to approach the presence of uncertainty. The list of economic and financial problems not included in the paper is extensive, but those will be the work of others.

Because economic theory rests heavily upon its mathematics, this paper argues that the theory needs a fundamentally different grounding. Because most distributions lack a first moment, nothing resembling mean-variance finance remains possible. This goes deep into the risk management, legal and regulatory environment. Even at the state level, the *Uniform Prudent Investor Act* is predicated on the validity of a set of mathematical structures that do not exist. A raft of fundamental ideas in models where capital is present will need to be reviewed and reconsidered as to their implications.

2. A General Method to Derive Distributions of Returns for Equity Assets Using Ratios

2.1. Introduction

This article deals with four types transformations. They are $\dfrac{p_{t+1}}{p_t}$,

$\log(p_{t+1}) - \log(p_t)$, $p_{t+1} = Rp_t + \epsilon_{t+1}$ and $\log(p_{t+1}) = R\log(p_t) + \epsilon_{t+1}$. One of the difficulties has been that economic models have been constructed in one transformation but tested in another. While the budget constraint of the Capital Asset Pricing Model and the Security Market Line Beta representation are linearly additive in errors, the logarithmic approximation is usually tested resulting in a test of a different multiplicative model.

As will be shown, these transformations have different results. The model in use and the test of the model should match if at all possible. A goal of this article is to provide a deeper understanding of the consequences of the tools used.

As a simplification, $\dfrac{p_{t+1}}{p_t} - 1$ will be studied as $\dfrac{p_{t+1}}{p_t}$. The subtraction 1 has no impact on the distribution other than to shift it and constantly writing -1 does nothing to improve clarity.

Of the two classes of securities; equities and debt; equities are unique in that the final value is unknown. The distribution of returns for equity securities depends upon the distribution of potential valuations at the time of sale and the distribution of potential valuations at the time of purchase. The reward is the ratio of future values to present values.

Definition 1. The *reward* for investing is defined as the future value divided

by the present value. The *return* is the *reward* minus one.

There are three mechanisms to derive a distribution of rewards, or its cousin returns. The first, and standard method, is to derive the ratio distribution directly. This method was first described by Curtiss [22]. The second would be to convert the data into polar coordinates and to indirectly solve the problem using the angle of the return rather than a direct attack on the slope. The third is to regress the future value from the present value.

2.2. Curtiss' Method

Curtiss allows a general solution for the cases where there are two random variables, *Y* and *X*, related through the relationship

$$Z = \frac{Y}{X}. \tag{3}$$

The joint probability distribution for *X* and *Y* is $f(x, y)$. If $D(z)$ is the cumulative density function for *z*, and $p(z)$ is its probability density function, then they must be related through some transformation of the joint density function of *X* and *Y*.

Because this is a ratio, zero could potentially cause difficulty. This difficulty is avoided by also noting that the $\Pr(x = 0) = 0$ since the measure of a countable point over a continuum is zero.[1] Although its probability is zero, this does not make it an impossible event, only an event that can be removed as any other similar pole could without causing unintended consequences. All impossible events have a probability of zero, but not all events with a probability of zero are impossible.

Nonetheless, it does require that any solution is partitioned into the sets above and below zero. As a consequence, although

$$D(z) = \Pr(Z \le z), \tag{4}$$

This must be written as

$$D(z) = \Pr(Y \ge zX \mid X < 0) + \Pr(Y \le zX \mid X > 0). \tag{5}$$

This is extended into functional form through the equation

$$D(z) = \int_{-\infty}^{0} \int_{zx}^{0} f(x, y) \,\mathrm{d}y\mathrm{d}x + \int_{0}^{\infty} \int_{0}^{zx} f(x, y) \,\mathrm{d}y\mathrm{d}x, \tag{6}$$

where the integrals are assumed to be Lebesgue-Stieltjes integrals.[2]

The probability density function for $d(z)$ is

$$p(z) = \frac{\mathrm{d}\big[D(z)\big]}{\mathrm{d}z}, \tag{7}$$

which can be expressed as

---

[1]One could also assign an aribtrary density at $x = 0$ because the impact on the total sum would only be $k\mathrm{d}x$ which would vanish at the limit.

[2]For most ordinary cases and all cases in this article, the integrals can be assumed to be Lebesgue integrals or in some special cases not used here, they can be assumed to be Riemann integrals. See Curtiss in [22] for a complete discussion.

$$p(z) = \int_{-\infty}^{\infty} |x| f(x, zx) \, \mathrm{d}x. \tag{8}$$

2.3. A Trigonometric Method

A second solution is to note the relationship between a slope and an angle. If the reward for investing is Z as above then relationship between Θ and Z is

$$\Theta = \tan^{-1}(Z), \tag{9}$$

while

$$\Omega = \sqrt{X^2 + Y^2}. \tag{10}$$

The transformation of the function requires including the Jacobian for the tranformation of variables. Transforming the problem into polar coordinates results in

$$D(z) = \int_{-\frac{\pi}{2}}^{\tan^{-1}(z)} \int_0^{\infty} f(\omega, \theta) \omega \mathrm{d}\omega \mathrm{d}\theta + \int_{\frac{\pi}{2}}^{\tan^{-1}(z)+\pi} \int_0^{\infty} f(\omega, \theta) \omega \mathrm{d}\omega \mathrm{d}\theta. \tag{11}$$

The equation presumes no limitation on liability. The equation would need to be adjusted for the limitation of liability. This form has two potential advantages, one pedagogic, the other computational.

Since the cumulative density function for the standard Cauchy distribution is $\pi^{-1} \tan^{-1}(\tilde{x})$, this implies that a solution will be an angular transformation of the Cauchy distribution. This can be observed in the limits of the integrals. Using the trigonometric method makes it easier to see the underlying relationship between returns and the Cauchy distribution. Secondarily, some functions are simpler to solve in polar coordinates.

2.4. Ratio Distributions Based on Equilibrium Prices

Wealth invested in a security at time t is the price times the quantity owned at time t, that is to say $w_t = p_t \times q_t, \forall t$. For equity securities, not exchanged for cash or bankrupt, then the split adjusted price is equivalent to assuming that $q_t = 1, \forall t$. In that circumstance, the return on cash flows becomes the return on prices. To derive the distribution, with respect to the equilibrium prices, it is assumed that

$$p_t = p_t^* + \epsilon_t, \forall t. \tag{12}$$

As only the distribution of errors from the equilibrium are of interest the distribution will be of

$$\epsilon_t = p_t - p_t^*, \forall t. \tag{13}$$

Using this transformation transfers the point of calculation from (p_t, p_{t+1}) to $(0,0)$. Because this translation moves the ratio of prices to $(0,0)$ it will be necessary to translate the final distribution by R, which could be defined as

$$R = \frac{p_{t+1}^*}{p_t^*}. \tag{14}$$

3. Distribution for Going Concerns in Equilibrium

For historical reasons and because the complications of real life will be added later, this is being constructed in a classic Markowitzian model with many buyers and many sellers. There are no transaction costs, either there is infinite liquidity, or there is no market maker. Transactions occur in a double auction where potential buyers compete for best bid and sellers compete for best bid. Markets are not systematically away from equilibrium.

Because this is a double auction in equilibrium, there is no winner's curse. The rational behavior is for each actor to bid their expectation. The sampling distribution of the limit book will be normally distributed due to the central limit theorem. If the Markowitzian assumption of price taking behavior is included, then the appraisal errors are being committed by the counter-party. Additionally, it will be assumed that errors are independent. This may not be the case for some types of transactions, such as program trading.

As with Capital Asset Pricing Model style problems, there is no presumed positivity of future price. The assumption of limited liability will be added in Subsection 3.3. The sample space is the real numbers for future values of p.

Since both the entry and exit order should have normally distributed order books, it follows that the likelihood function is the ratio of two normal distributions. This solution is well known in statistics as the Cauchy distribution. The Cauchy distribution has no mean, and consequently, under the basic initial assumptions of mean-variance models, it is impossible for mean-variance finance to exist.

Using the method described in Section 2.4 one quickly arrives at the Cauchy distribution under the assumptions of the Markowitz style models as shown in Section 3.1 for the distribution of returns around the equilibrium points.

3.1. Using Curtiss' Method to Arrive at the Distribution

Curtiss' method involves solving

$$p(r_t) = \int_{-\infty}^{\infty} \frac{|\varepsilon_t|}{2\pi\sigma_t\sigma_{t+1}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_t^2}{\sigma_t^2} + \frac{r_t^2\varepsilon_t^2}{\sigma_{t+1}^2}\right)\right]d\varepsilon_t \tag{15}$$

causes us to arrive at

$$p(r_t) = \frac{1}{\pi}\frac{\Gamma}{\Gamma^2 + r_t^2}, \Gamma = \frac{\sigma_{t+1}}{\sigma_t}. \tag{16}$$

This equation still needs to be shifted by R. This preserves the nice interpretation of Γ as a measure of price heteroskedasticity so that the final form is

$$p(r_t) = \frac{1}{\pi}\frac{\Gamma}{\Gamma^2 + (r_t - R)^2}. \tag{17}$$

3.2. The Trigonometric Method

To provide an example of the trigonometric solution, consider the simple case where $\sigma_t = \sigma_{t+1} = 1$. Noting the symmetry, only one of the two integrals needs

solved and it would be multiplied by two. Since in equilibrium the center of location is at (0,0), the cumulative density function is

$$D(r_t) = \frac{2}{2\pi} \int_{-\frac{\pi}{2}}^{\tan^{-1}(r_t)} \int_0^\infty \omega e^{-\frac{\omega^2}{2}} \mathrm{d}\omega \mathrm{d}\theta \qquad (18)$$

$$D(r_t) = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\tan^{-1}(r_t)} \left[\lim_{\omega \to \infty} e^{-\frac{\omega^2}{2}} - \lim_{\omega \to 0} e^{-\frac{\omega^2}{2}} \right] \mathrm{d}\theta \qquad (19)$$

$$D(r_t) = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\tan^{-1}(r_t)} \mathrm{d}\theta \qquad (20)$$

$$D(r_t) = \frac{1}{\pi} \tan^{-1}(r_t) + \frac{1}{2}. \qquad (21)$$

The density function is

$$p(r_t) = \frac{1}{\pi(1 + r_t^2)}, \qquad (22)$$

which is the Cauchy distribution. The advantage of this method is at least partially pedagogic in that the relationship to the Cauchy distribution is obvious.

A review of Equation (18) shows that it could be broken up into two different densities. The distribution of ω is

$$p(\omega) = \omega e^{-\frac{\omega^2}{2}}, \qquad (23)$$

while the distribution of of θ is

$$p(\theta) = \frac{1}{\pi}. \qquad (24)$$

Equation (23) is the Rayleigh distribution, which is a special case of the Weibull distribution or the χ distribution [29] [30]. Equation (24) is the uniform distribution. This is also a restatement of Gull's Lighthouse problem [31]. Both ω and θ are independent of each other.

The author believes that greater knowledge may develop by exploring the role that the properties of the polar coordinate distributions possess. The uniform distribution, like the Cauchy is perfectly imprecise, that is it has no variance unless bounded. The Weibull is an extreme value distribution and under the Rayleigh parameterization implies that large values are probable. As the Rayleigh distribution, it is the limiting distribu- tion of the vector sum with "random amplitudes and uniformly distributed phases" [30]. Through the χ distribution it is the distribution of the square root of the sum of squares of independent draws from a normal distribution.

A further reason to consider this linkage is the work performed by Mc-Cullaugh in [32] and Burdzy in [33] on the linkage between the Cauchy distribution and the complex plane, which is tightly linked to trigonometric solutions through Euler's formula:

$$e^{ix} = \cos(x) + i\sin(x). \qquad (25)$$

Although more metaphorical and intuitive than rigorous, it may serve to point

out the error at time *t* is *real* in the sense that it is about to happen while the second error is, or at least feels, more *imaginary* in that it projections are being made, possibly decades in advance. While this work is being done in $\mathbb{R}^2$ significant work has been done in $\mathbb{C}$ instead, beginning with work by Burdzy in [33].

3.3. Limitation of Liability

The limitation of liability truncates the Cauchy distribution. Its parameters become

$$p(r_t) = \left[\frac{\pi}{2} + \tan^{-1}\left(\frac{R}{\Gamma} \right) \right]^{-1} \frac{\Gamma}{\Gamma^2 + (r_t - R)^2} \qquad (26)$$

This could be solved, as above, by restricting values for p_t and p_{t+1} to greater than zero.

An alternative representation would be to note that in bankruptcy $q_{t+1} = 0$ and so it is not a ratio of prices but quantities. That is to say, the ratio distribution of prices is multiplied by the ratio distribution of quantities. Given that the firm neither goes bankrupt nor merges out of existence during the interval, $\Pr(q_{t+1} = q_t) = 1$.

The truncated Cauchy distribution is reasonable approximation of returns on equity, even without discussing the issues of liquidity or the budget constraint discussed in Section 11. This is extensively discussed in [11]. The annual returns for going concerns, less the impact of dividends, collected from the population of data at the Center for Research on Security Prices and excluding such firms as shell companies and closed-end funds for the years 1962-2013 are shown graphically in **Figure 1**. The best fit models from parameters implied by the data are provided as well.
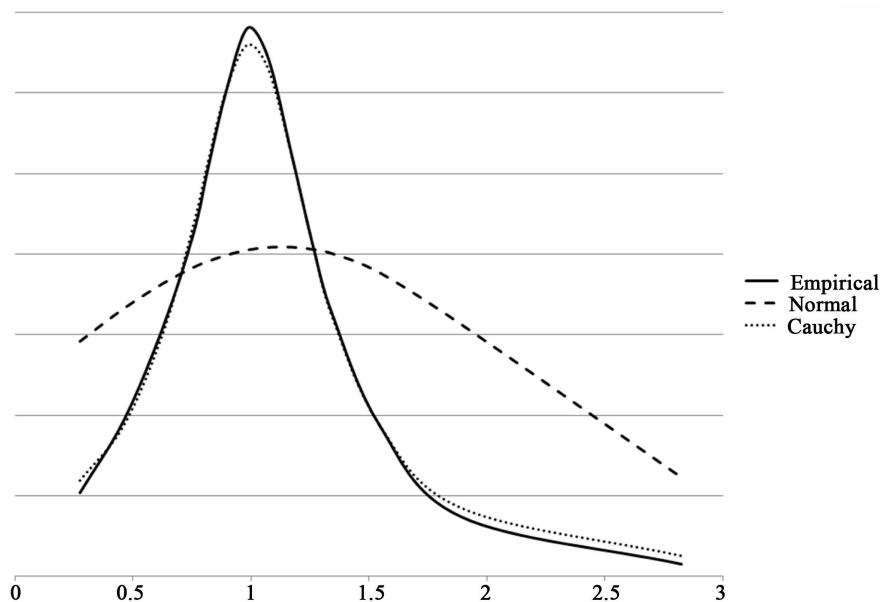

Figure 1. Empirical distribution versus implied Cauchy and Gaussian models from the data.

The three curves of 1 are disaggregated returns. This differs from the standard *market returns* shown throughout the literature, but rather are trade-by-trade returns across the period. All three curves are smoothed with spline fitting by Microsoft's Excel. The first curve is the empirical distribution of annual returns over the period. The second is the best fit Cauchy and Gaussian model of distributions. The Gaussian fits so poorly because the tails are so dense and long.

The Gaussian model has two conflicting issues, modeling the high density region and modeling the tails. Because of the squared impact of attempting to measure the variance the Gaussian model is flattened in the dense region, but still too thin at the extrema of the tails. Because the Cauchy model is a peaked distribution with long tails, there is no mathematical conflict in fitting a curve.

The fact of heavy tails has not been in dispute since at least Mandelbrot's article, but an explanation as to why it exists has not been made [2]. **Figure 1** is nothing more than a reminder of facts long settled but explained as anomalies. This is not an anomaly. This is how equity securities are supposed to work.

3.4. Observations on the Nature of the Cauchy Distribution

3.4.1. Lack of a Mean

Consider a generic variable, x_i, drawn from a Cauchy distribution, regardless if the form is a time series or some other data form, there is an understood relationship between the sampling distribution of the mean, $\overline{S}_n$, for a sample size of n and the density function of the raw data.

In order to shorten the exposition and without loss of generality, assume that x_i is drawn from the standard Cauchy distribution, that is

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}. \tag{27}$$

The sum of n Cauchy variates is

$$S_n = \sum_{i=1}^{n} x_i \tag{28}$$

and the sample mean is

$$\overline{S}_n = \frac{S_n}{n}. \tag{29}$$

The distribution of the sample mean can be found by using the characteristic function for the Cauchy distribution. The characteristic function is

$$\begin{aligned}
\phi(\tau) &= \int_{-\infty}^{\infty} e^{ix\tau} \frac{1}{\pi} \frac{1}{1+x^2} dx \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos\tau x + i\sin\tau x}{1+x^2} dx \\
&= \frac{2}{\pi} \int_{0}^{\infty} \frac{\cos\tau x}{1+x^2} dx = e^{-|\tau|}.
\end{aligned} \tag{30}$$

The characteristic function for the sum of n independent draws from the distribution described by Equation (27) is the product of the individual charac-

eristic functions resulting in a joint characteristic function of

$$\phi(\tau)^n. \tag{31}$$

Inverting the prior process yields

$$f(S_n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iS_n\tau - n|\tau|} d\tau = \frac{1}{\pi} \frac{n}{n^2 + S_n^2}. \tag{32}$$

Converting to the distribution of the sample mean yields

$$f(\bar{S}_n) = f(S_n)\frac{dS_n}{d\bar{S}_n} = \frac{1}{\pi} \frac{1}{1 + \bar{S}_n^2}. \tag{33}$$

The sampling distribution of the mean maps to the distribution of the raw data. The unexpected implication is that there is no difference in information about the center of location between looking at the sample mean of one thousand draws, two draws or simply looking at one draw of the raw data. This is equivalent to saying there is no gain in information by calculating the mean of any size over simply looking at one raw data point in determining the true center of location.

This is deeply problematic for the continued use of sample means or least-squares style methods in economics. The inference is purely random because the Cauchy distribution has no population mean and so the sample mean does not converge on a single point. The most that can be said about the center of location given the sample mean is that it is somewhere in the real numbers. As a consequence, when using mean-based methods, an inference regarding the true parameter β will be equivalent to having a sample size of one regardless of the true sample size. Such a finding is catastrophic for standard econometric methods. It can also be catastrophic to standard economic models involving capital.

3.4.2. Lack of Independence

The multivariate Cauchy distribution is a symmetric special case of the multivariate Student distribution. For the two dimensional case it is

$$\frac{1}{2\pi}\left[\frac{\Gamma}{\left(\Gamma^2 + \sum_{i=1}^{2}(r_i - R_i)^2\right)^{\frac{3}{2}}}\right]. \tag{34}$$

For the three dimensional case the density is

$$\frac{1}{\pi^2} \frac{\Gamma^2}{\sqrt{\Gamma^3}\left(\Gamma + \sum_{i=1}^{3}(r_i - R_i)^2\right)^2}. \tag{35}$$

Notice that at no point does a covariance matrix or some structure similar to a covariance matrix appear. Also note that the nth dimensional distribution is not just the n product of the univariate distributions. It is not obvious, but care must

be taken, that when solving as a student distribution that the traditional meaning of μ as a mean and σ as a standard deviation in Student's t distribution changes to μ as a mode and σ as a scale parameter. More particularly, σ is the half-width at half-maximum.

4. Assets Purchased in Single Auctions and Subject to the Winner's Curse

While stocks, in equilibrium, are not subject to a winner's curse, many assets are. This is because the high, or in some cases low, bidder purchases the asset. The difficulty is created as above, by each party bidding their expectation. While the sampling distribution of the mean is the Gaussian distribution, the sampling distribution of the extrema is the Gumbel distribution.

The determination of the ratio of two Gumbel distributions is complicated by the fact that there are multiple types of buyers. Professional buyers, collectors, and novice buyers may face different conditional distributions due to different information sets and different pricing goals.

An antique dealer in a network of dealers, or a construction contractor that regularly subcontracts, have significant information about the wholesale value while a novice actor generally can, at best, see retail pricing. Although the functional form for either will be the same, they would not be if the dealer did not purchase the asset at an auction. Rather than this auction based example, the dealers' returns would depend on the model of the dealer market.

The form of the Gumbel distribution for a single-sided, high-bid auction purchase is

$$p(\varepsilon_t) = \frac{1}{\sigma_t} e^{\left(-\frac{\varepsilon_t}{\sigma_t} - e^{-\frac{\varepsilon_t}{\sigma_t}}\right)}, \forall t. \tag{36}$$

Assuming the errors are independent and markets are in equilibrium, then

$$p(r_t) = \int_{-\infty}^{\infty} |\varepsilon_t| \frac{\exp\left(-\frac{r_t \epsilon_t}{\sigma_{t+1}} - e^{-\frac{r_t \epsilon_t}{\sigma_{t+1}}} - \frac{\epsilon_t}{\sigma_t} - e^{-\frac{\epsilon_t}{\sigma_t}}\right)}{\sigma_t \sigma_{t+1}} d\varepsilon_t, \tag{37}$$

remembering that the errors are centered around zero and not R. In a numerical solution, returns will need to be shifted by R.

The integral of the ratio of two Gumbel distributions is unknown. This creates two possible solutions. First, the method of histograms is available [34]. The second choice is to construct numerical approximations of the integral.

The second choice has the advantage that it is not strictly dependent on historical conditions. Still, this forces the construction of tables rather than allow for true parametric solutions. For $\sigma_t = \sigma_{t+1} = 1; R = 1$ the plot of the reward for investing is shown in **Figure 2**.[3]

---

[3]The author would prefer to name it the Maxham Distribution after his wife's maiden name since this does not appear to be a named distribution.

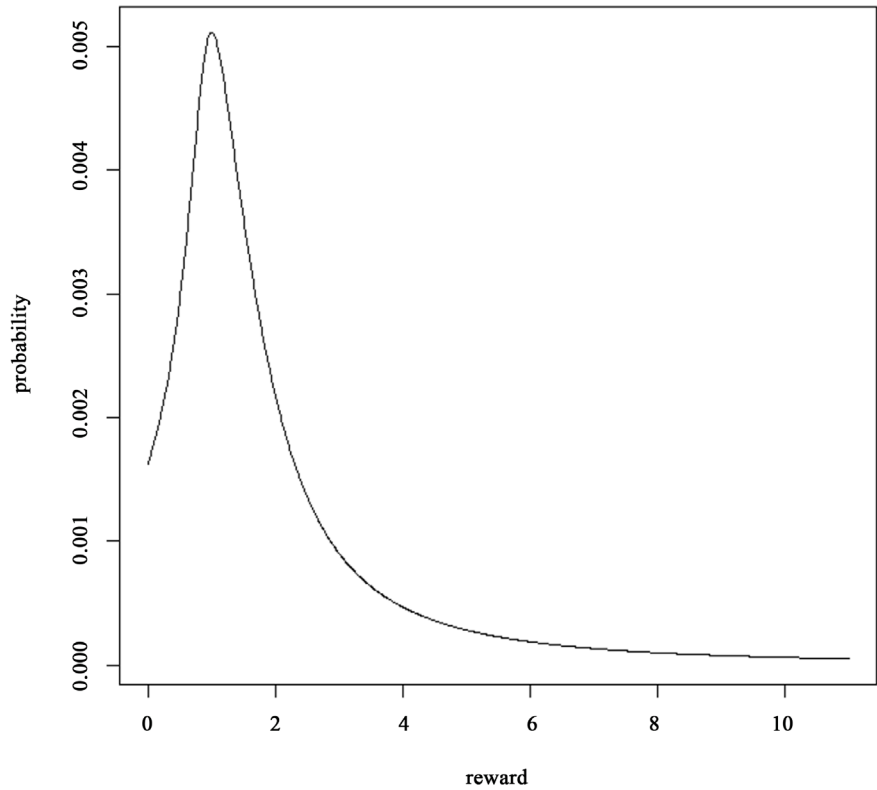**Figure 2.** Reward for investing in asset with winner's curse and selling with winner's curse.

5. Distribution of Accounting Ratios and Equity Returns with Dependent Errors

Accounting ratios and returns on equity securities without independent errors have a distribution that is very close to the Cauchy distribution. In this case, it is presumed that there is support on the entire real line. Adjustments for truncation would be necessary. The assumption is that an equilibrium exists. In this case, the joint distribution of errors is the bivariate normal centered on (0,0), but ε_{t+1} is correlated with ε_t through a correlation coefficient, ρ.

The distinction with the derivation of the Cauchy distribution is that for the Cauchy distribution, the covariance matrix for the normal distribution is

$$\Sigma = \begin{bmatrix} \sigma_{t,t} & 0 \\ 0 & \sigma_{t+1,t+1} \end{bmatrix}, \tag{38}$$

while in this case

$$\Sigma = \begin{bmatrix} \sigma_{t,t} & \sigma_{t,t+1} \\ \sigma_{t,t+1} & \sigma_{t+1,t+1} \end{bmatrix}. \tag{39}$$

Using Curtiss' method, the unshifted distribution of returns is

$$p(r_t) = \frac{\sigma_t \sigma_{t+1} \sqrt{1 - \dfrac{\sigma_{t,t+1}^2}{\sigma_t^2 \sigma_{t+1}^2}}}{\pi \left(r_t^2 \sigma_t^2 - 2 r_t \sigma_{t,t+1} + \sigma_{t+1}^2 \right)}. \tag{40}$$

If the above equation was multiplied by

$$\frac{\dfrac{1}{\sigma_t^2}}{\dfrac{1}{\sigma_t^2}}, \tag{41}$$

then the relationship between the Cauchy distribution becomes a bit more obvious by setting

$$\Gamma = \frac{\sigma_{t+1}}{\sigma_t}. \tag{42}$$

In that case, the distribution becomes

$$p(r_t) = \frac{\Gamma\sqrt{1-\rho^2}}{\pi\left(r_t^2 - 2r_t\sigma_{t,t+1}/\sigma_t^2 + \Gamma^2\right)}. \tag{43}$$

As with the Cauchy distribution, the return needs to be shifted by R so that the final formula is

$$p(r_t) = \frac{\Gamma\sqrt{1-\rho^2}}{\pi\left((r_t-R)^2 - 2(r_t-R)\sigma_{t,t+1}/\sigma_t^2 + \Gamma^2\right)}, \tag{44}$$

or with fewer parameters to estimate

$$p(r_t) = \frac{\sigma_t\sigma_{t+1}\sqrt{1 - \dfrac{\sigma_{t,t+1}^2}{\sigma_t^2\sigma_{t+1}^2}}}{\pi\left((r_t-R)^2\sigma_t^2 - 2(r_t-R)\sigma_{t,t+1} + \sigma_{t+1}^2\right)}. \tag{45}$$

Although the assumption of independence of errors is reasonable, program trading where few actors could be involved in moving large portfolios that are evaluated jointly, it is quite possible that in the case of quick turnover that the appraisal errors are highly associated with both the purchase and selling of a block of assets.

6. The Distribution of Returns with Markets in Disequilibrium

It is difficult to discuss markets in disequilibrium in economics. Events such as the financial crash of 2008 are problematic for equilibrium models. If the market is away from equilibrium at time t, then rather than considering

$$p_t = p_t^* + \epsilon_t \tag{46}$$

the model should be

$$p_t = p_t^* + \mu_t + \epsilon_t. \tag{47}$$

If $\mu > 0$ then returns should be smaller and, conversely, where $\mu < 0$ then returns should be larger. The question is "what is μ_t?" If μ_t is thought of as part of the error term, then it is a systematic bias. If it is thought as a systematic shift in the equilibrium, then the present value of cash flows does not equal the price where there is no pressure for change. In neither case, μ vanishes from

the numerator on the assumption that over time, prices must reflect future cash flows.

The two solutions imply different distribution rules.

6.1. Placing the Shift in the Error Term

If the market is believed to be away from the implied curve in equilibrium then it cannot be assumed that errors are centered on zero. If it is assumed that assets will be sold when market prices are no longer sticky and away from zero, then the mean error will be located at $(\mu_t, 0)$. If it also presumed that

$$\Gamma = \frac{\sigma_{t+1}}{\sigma_t}, \tag{48}$$

then it is possible to state everything in reference to time t. Using Curtiss' method

$$p(r_t) = \frac{\Gamma e^{-\frac{\mu_t^2}{\sigma_t}} \left(\sqrt{\pi} \mu_t e^{\frac{\Gamma^2 \mu_t^2}{(r_t - R)^2 + \Gamma^2 \sigma_t}} \operatorname{erf} \left(\frac{\mu_t}{\sigma_t \sqrt{\frac{(r_t - R)^2}{\Gamma^2} + \sigma_t}} \right) \right)}{4\pi \sigma_t \left((r_t - R)^2 + \Gamma^2 \sigma_t \right) \sqrt{\frac{(r_t - R)^2 + \Gamma^2 \sigma_t}{\Gamma^2 \sigma_t^2}}} \tag{49}$$

$$+ \frac{\Gamma e^{-\frac{\mu_t^2}{\sigma_t}} \left(\sqrt{\pi} \mu_t e^{\frac{\Gamma^2 \mu_t^2}{(r_t - R)^2 + \Gamma^2 \sigma_t}} + \sigma_t \sqrt{\frac{(r_t - R)^2 + \Gamma^2 \sigma_t}{\Gamma^2 \sigma_t^2}} \right)}{4\pi \sigma_t \left((r_t - R)^2 + \Gamma^2 \sigma_t \right) \sqrt{\frac{(r_t - R)^2 + \Gamma^2 \sigma_t}{\Gamma^2 \sigma_t^2}}}.$$

Buried inside this extensive expression is the Cauchy distribution. It is important to remember that if $\mu_t < 0$ then returns will be shifted to the right from equilibrium returns and conversely for $\mu_t > 0$.

Marsaglia's Extension of Curtiss' Method for the Normal Case

Marsaglia in [24] extended Curtiss' work by considering the more general case of

$$\frac{a + x}{b + y}, \tag{50}$$

where x and y are independent standard normal random variates, while a and b are constants. He extended this work to cover the cumulative density function for computational ease [25]. If $a = 0$, then this is equivalent to Equation (49) with several additional simplifying assumptions.

In the simplified case of Marsaglia [24] [25], he finds that the distribution is a convex combination of the Cauchy distribution and a distribution that can either be unimodal or bimodal and is the product of a Cauchy distribution and

other terms extracted from the normal distributions and constants. The condensed general solution to Marsaglia's problem is:

$$p(z) = \frac{\exp\left[-\frac{1}{2}\left(a^2 + b^2\right)\right]}{\pi\left(1 + z^2\right)}\left[1 + q\exp\left(\frac{1}{2}q^2\right)\int_0^q \exp\left(-\frac{1}{2}x^2\right)dx\right], q = \frac{b + az}{\sqrt{1 + z^2}} \quad (51)$$

For purposes of this article,

$$a = 0, b = \mu_t = \frac{\overline{p}_t - p_t^*}{\sigma_t}, \text{ and } z = r_t. \quad (52)$$

This splits the equation into two parts. The first part,

$$\frac{\exp\left[-\frac{1}{2}\mu_t^2\right]}{\pi\left(1 + r_t^2\right)}, \quad (53)$$

is the product of the Cauchy distribution of returns with an the non-normalized Gaussian kernel. The second portion,

$$1 + \frac{q}{\exp\left(-\frac{1}{2}q^2\right)}\int_0^q \exp\left(-\frac{1}{2}x^2\right)dx, \quad (54)$$

is one plus the elasticity of q with respect to the cumulative density function of q. As the available prices are systematically far from the equilibrium the distribution becomes inelastic. This stiffness of outcomes is generated by the fact that as μ_t goes to zero, the distribution goes to the Cauchy distribution. Conversely, as μ_t becomes far from the equilibrium price, such as when $\mu_t > 4$ then the role of the Cauchy distribution gets close to zero. The other distribution has all of its moments.

As a mixture, there are no moments, but the effect of the Cauchy distribution becomes small. This narrowing implies that as bubbles go on a poor outcome becomes increasingly certain while after a bear market a good outcome becomes increasingly certain.

If prices being far from an equilibrium can be thought of as type I or type II errors, then Equation (49) could be thought of as the statistical basis for value investing. Benjamin Graham's and David Dodd's "margin of safety" required of all investors would have a parallel "margin of excitement" for speculative investors, adapting the parlance of Graham's book, *The Intelligent Investor* [35].

This parallels the conversion of the cumulative density of returns into a Bernoulli distribution. Consider the problem of the

$$\Pr\left(r_t \geq j \mid p_t \leq k\right). \quad (55)$$

In this equation, j is a sufficient return to meet planned goals, while k is some limit price. Either the goal is met or it is not met. The variance maximizes when the probability of either case is fifty percent. A portfolio would carry a binomial distribution.

As p_t became large, the probability would become small and the variability

of outcomes would decline. Conversely, as p_t became small, the probability would become large and the variability of outcomes would decline. The instinct of the economist would be to ask "what is preventing the security from being at its equilibrium price" and the simple answer is "that is precisely what a type I or type II error would be in this case."

6.2. A Shift Moves the Equilibrium Away from the Present Value of Cash Flows

In this case,

$$R_t = \frac{\overset{*}{p}_{t+1}}{\overset{*}{p}_t + \mu_t}. \tag{56}$$

It follows that the errors would be a truncated Cauchy distribution, but shifted for the price shift.

6.3. Convex Combination

It is also quite possible that μ_t is split between a systematic bias and an equilibrium shift. How this function works in the real world is an empirical question to test under Bayesian model selection. If the shifting value were split into two components, μ_α and μ_β, where μ_α is the equilibrium shift and μ_β is the shift from bias, then the return distribution becomes

$$
p(r_t) = \frac{\Gamma e^{-\frac{\mu_\beta^2}{\sigma_t}}\left(\sqrt{\pi}\mu_\beta e^{\frac{\Gamma^2\mu_\beta^2}{(r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t}}\operatorname{erf}\left(\frac{\mu_\beta}{\sigma_t\sqrt{\frac{(r_t-R-\mu_\alpha)^2}{\Gamma^2}+\sigma_t}}\right)\right)}{4\pi\sigma_t\left((r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t\right)\sqrt{\frac{(r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t}{\Gamma^2\sigma_t^2}}}
$$

$$
+\frac{\Gamma e^{-\frac{\mu_\beta^2}{\sigma_t}}\left(\sqrt{\pi}\mu_\beta e^{\frac{\Gamma^2\mu_\beta^2}{(r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t}}+\sigma_t\sqrt{\frac{(r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t}{\Gamma^2\sigma_t^2}}\right)}{4\pi\sigma_t\left((r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t\right)\sqrt{\frac{(r_t-R-\mu_\alpha)^2+\Gamma^2\sigma_t}{\Gamma^2\sigma_t^2}}}. \tag{57}
$$

7. The Distribution of Mergers for Cash

Cash-for-stock mergers are a special limiting case. Whereas stock-for-stock provides the selling shareholders a contingent claim, cash for stock provides perfect liquidity. As cash is expensive, a cash purchase should have additional properties.

As this process is being thought of in a Bayesian framework, any estimate of return must include the possibility that the firm will be purchased for cash by another party. Given a cash merger will occur, it will have a defined future value

once the merger is announced. Prior to announcement a likelihood function is necessary to estimate a solution for pricing and returns. If it is assumed that the rates of return are efficiently priced once perfect liquidation is certain, then traditional economic tools should solve this problem. Going back to Equation (14) note that the future value is now fixed, given a cash-for-stock merger will happen with certainty.

Of course this is not certain, so Bayes theorem provides a solution. If

$$\phi\left(\mathbf{X}|\boldsymbol{\theta};M_c\right)\pi\left(\boldsymbol{\theta};M_c\right),\tag{58}$$

where M_c denotes a cash for stock merger, $\phi(\ )$ is the likelihood of observing data, given a set of parameters and the certainty of a cash for stock merger, times the probability of a cash for stock merger, then solving Equation (58) over the parameter space, $\boldsymbol{\theta}\in\Theta$, provides a value proportionate to the posterior probability for each possible value of the parameters. Once normalized and integrated over the parameters, the resulting predictive density is the best information about the distribution of future returns for a given security.

The concern here is the reward from investing from Equation (14). In that equation there are two prices, p_t and p_{t+1}. For a going concern, this is not an issue, because the firm will exist at time $t+1$. If a firm is merged out of existence for cash before time $t+1$ then what rule can be created for the future value?

In fact, there are two possible rules, either equally good for specific purposes, and less good for other purposes. The first rule would be to look forward in time starting at time t and ending at time $t+1$, partitioning the set into short, meaningful intervals. Then a probability of a merger completion would be calculated for each subinterval in the partitioned set. A likelihood function would also be created for each date.

The alternative rule would be to ask the probability of a cash-for-stock merger on or before the date denoted as $t+1$. In that case, r_t represents the reward received for having invested funds at time t. It does ignore reinvestment, which would be a separate question.

The question of a cash-for-stock merger has two important properties that can be illustrated in the derivation that is not done in prior derivations. The first is that a cash-for-stock merger has only one error in it, not two. The second, not mentioned above, is to separate the role of the investor and the economist or finance professional.

If it were known, with certainty, that a cash-for-stock merger was to occur, then the only question is what will the stock sell for in that merger.

The equation for the realized reward is

$$r_t = \frac{w_{t+1}}{p_t} = \frac{1}{e^{-\log(r_t)}},\tag{59}$$

where w is future wealth as there is no price, and which can be normalized to unity.

The reason for noting the relationship between nominal reward and logari-

thmic rewards has to do with how errors are conceptualized. Are errors multiplicative as would be implied by the logarithmic form or are they additive as the raw form would imply?

The issue is both theoretical and empirical. If the distribution of returns appears as either what one would expect from additive errors, or what one would expect from multiplicative errors, then a fundamental relationship to reality can be discussed.

The derivation of the proof is built upon a key proof by Landon in [36] that appears throughout economics and physics, but often without a realization of where that proof may have come from.

7.1. Jaynes Generalization of Landon's Proof

E.T. Jaynes in [14] tome on probability theory, *Probability Theory: The Language of Science*, uses part of a subsection to point out the importance of Landon's proof and to generalize it. Landon in [36] was trying to solve the problem of noise in communication circuits. The proof was for a particular case, but the proof immediately caught the attention of both Harold Jeffreys and John Maynard Keynes.

Landon's proof has several important components for economics. For Jaynes, Jeffreys, and Keynes, the importance was in noting that "by minor changes in the wording" the proof, "can be interpreted either as *calculating* a probability distribution, or *estimating* a frequency distribution" [14]. For purposes of this paper, the linkage between Bayesian and null hypothesis methods in this proof is unimportant. What is important is the distinction between methods available when there is one error in the decision process versus two errors as in normal equity trading. Additionally, the proof by Landon in [36] allows a direct linkage between the subjective decisions of individual actors and the calculation of the distribution that exists in equilibrium

Because of the relationship between subjective personal realities and objective market relationships, there will be a slight notation change. For the personal calculated rewards of a subjective actor, the notation will be $_i r_t$ at time *t* while the equilibrium return will be r_t.

This cumbersome notation is not present in *Probability Theory: The Language of Science* [14] because Jaynes, a physicist, is not concerned with the relationship between subjective personal realities of the actors with market equilibrium as an economist would be concerned. Additionally, Jaynes uses Dirac notation for expectations, whereas here the expectation of some random variable *x* will be notated $E(x)$. Although the physics notation is shorter, it is uncommon to see it in economics articles.

Model Assumptions

It is assumed that the observed market return is r_t and for this derivation, there will be additive errors. As with Markowitz, it will be assumed that there are many actors, or alternatively, that over time there will be many actors. With

Landon's original form of proof, the better assumption would be the latter. Further, it will be assumed that the markets are in equilibrium. As a consequence, any error term's expectation is zero.

In equation

$$p_{t+1} = Rp_t + \varepsilon_{t+1} \tag{60}$$

the error term is ε_t. This represents the movement of the price around the equilibrium trade price through time. The concern here, however, is with personal errors. Market and personal errors will be denoted ϵ_t or $_i\epsilon_t$.

As such, it is assumed that

$$E\left(_i\epsilon_t\right) = 0. \tag{61}$$

Further, for a given level of risk, it is assumed that

$$p\left(_i r_t | \sigma_t\right) \tag{62}$$

is known to each actor, denoted $i \in I$, where the set I indexes the market actors, and where σ_t is scale parameter, in this case, a measure of risk.

Landon in [36] then posited the question of the impact of a very small shock or error, $_i\epsilon_t$ to the evaluation of $_i r_t$ by looking at the equation

$$_i r_t' = {_i r_t} + {_i\epsilon_t}, \tag{63}$$

where $_i\epsilon_t$ is very small relative to σ_t. Landon also assumed that $_i\epsilon_t$ had a probability of being observed equal to

$$q\left(_i\epsilon_t\right) \mathrm{d}_i\epsilon_t, \tag{64}$$

and that this probability was independent of $p\left(_i r_t | \sigma_t\right)$.

The objective distribution of r_t' maps to some subjective value $p\left(_i r_t'\right)$ and depends upon the probability of the specific error or shock by the actor(s) and the separate probability of r_t'. The probability for this is

$$f\left(r_t'\right) = \int p\left(_i r_t' - {_i\epsilon_t} | \sigma_t\right) q\left(_i\epsilon_t\right) \mathrm{d}_i\epsilon_t. \tag{65}$$

This can be estimated by the nearby value r_t by expansion about the nearby value. The distribution can be estimated by

$$\begin{aligned} f\left(r_t\right) = {}& p\left(_i r_t | \sigma_t\right) - \frac{\partial p\left(_i r_t | \sigma_t\right)}{\partial r_t} \int {_i\epsilon_t} q\left(_i\epsilon_t\right) \mathrm{d}_i\epsilon_t \\ & + \frac{1}{2} \frac{\partial^2 p\left(_i r_t | \sigma_t\right)}{\partial r_t^2} \int {_i\epsilon_t^2} q\left(_i\epsilon_t\right) \mathrm{d}_i\epsilon_t + \cdots, \end{aligned} \tag{66}$$

which can be notationally shortened to

$$f\left(r_t\right) = p\left(_i r_t | \sigma_t\right) - E\left(\epsilon_t\right) \frac{\partial p\left(_i r_t | \sigma_t\right)}{\partial r_t} + \frac{1}{2} E\left(\epsilon_t^2\right) \frac{\partial^2 p\left(_i r_t | \sigma_t\right)}{\partial r_t^2} + \cdots. \tag{67}$$

Noting that the expectation of ϵ_t is zero and that the expectation of $_i r_t^2$ increments to $\sigma_t^2 + E\left(_i\epsilon_t^2\right)$ and the invariance property noted above, this leads to the equation

$$f\left(r_t\right) = p\left(_i r_t | \sigma_t\right) + E\left(_i\epsilon_t^2\right) \frac{\partial^2 p\left(_i r_t | \sigma_t\right)}{\partial \sigma_t^2} \tag{68}$$

which becomes

$$\frac{\partial p\left({}_i r_t \middle| \sigma_t\right)}{\partial \sigma_t} = \frac{1}{2}\frac{\partial^2 p\left({}_i r_t \middle| \sigma_t\right)}{\partial_i r_t^2},$$

(69)

whose known solution is the normal distribution.

7.2. Implications

Landon's proof allows an expansion about an equilibrium value when there is only one error. For most of the ratio distributions covered in this paper, there does not exist a first moment as they are all transformations of the Cauchy distribution on at least the half-plane. As such, no expansion can be created. A key tool is lost to economics.

By construction, it appears that cash-for-stock should carry a log-normal distribution.

8. Including the Impact of Bankruptcy

Bankruptcy is the simplest of the cases in that if the shareholders are wiped out, then $q_{t+1} = 0$. If B' is the posterior probability of bankruptcy, then returns simply become $0 \times B' + (1 - B')R'$, where R' is the posterior probability of returns for all other states. Of course in Bayesian statistics, these are not point estimates; the formula represents a joint distribution of returns and bankruptcy risk.

If returns were understood as independent of bankruptcy risk, then it would be the simple cross product over the parameter space. If the center of location and the scale parameter are a function of bankruptcy probability, then the situation becomes less clear.

The challenge is created by an absence of a unique or clear mechanism to model bankruptcy in the literature. Indeed, the plethora of ways to model bankruptcy generates a wide range of possible joint distributions. The prediction of potential future returns depends upon how the relationships among parameters are conceptualized. Logically the center of location for returns should be a function of spread and bankruptcy risk. The center of location for spread should be a function of bankruptcy risk, and the probability of bankruptcy should be independent. This ignores the impact of skew, dividends payments and other types of data other than just returns. While there is only one way for variables to be independent, there are an infinite number of ways for them to be dependent. Including bankruptcy is an important and unsolved problem.

A general outline of the predictive distribution for this problem is solved instead. The definition of the predictive distribution is

Definition 2 (Bayesian Predictive Distribution) *If $\tilde{x}$ is a future value to be predicted from a matrix or vector of observed data* **X** *, where the likelihood function uses a vector of parameters $\boldsymbol{\theta}$, with $\boldsymbol{\theta} \in \Theta$, where Θ is the parameter space and χ is the sample space, then the predictive probability that $\tilde{x} = k, k \in \chi$ is*

$$\pi'(\tilde{x} = k | \mathbf{X}) = \int_{\theta \in \Theta} \phi(\tilde{x} = k | \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{X}) d\boldsymbol{\theta},$$

where π' is the predictive distribution, π is the posterior distribution, and ϕ is the likelihood function, for all $k \in \chi$.

Considering only a model with a single location, scale and bankruptcy parameter, the predicted distribution of returns at time $t+1$ would be

$$p(\tilde{r}_{t+1}) = \int_0^\infty \int_0^\infty \int_0^1 \frac{p(\mathbf{X} | R; \Gamma; B) p(R | \Gamma; B) p(\Gamma | B) p(B)}{\int_0^\infty \int_0^\infty \int_0^1 p(\mathbf{X} | R; \Gamma; B) p(R | \Gamma; B) p(\Gamma | B) p(B) dBd\Gamma dR} dBd\Gamma dR \quad (70)$$

9. The Distribution of Returns for Zero Coupon Bonds

The only assets with a covariance matrix would be among n-risky fixed cash flows, where they share bankruptcy or some other risk. Zero coupon bonds with simultaneous maturity would be the simplest form of this. Several possibilities exist, such as the logistic distribution, the normal distribution or the multinomial distribution would work, but a distribution with a covariance matrix would account for covariance among the factors relating to bankruptcy.

10. The Likelihood Function for Solutions Involving Regression

10.1. Introduction

Regression creates a set of special cases and does not provide a neat or clear recommendation. The difficulty for regression that differs from non-Bayesian methods is that non-Bayesian methods are only concerned with the sample after it has been transformed by a statistical method. Ordinary least squares is an example of such a method. Bayesian statistics are concerned with how the data is generated in the first place. In practice, this means that many possible competing models co-exist and one hypothesis needs to exist for each possible model.

Discussions here cannot be exhaustive but can cover the major cases. Certain observations about the rules for regression are important here. First, for time series of the form

$$x_{t+1} = \beta x_t + \epsilon_{t+1}, |\beta| < 1, \tag{71}$$

then the likelihood function is the normal distribution. Indeed, an assumption of normal returns implies Equation (71), which would imply that people anticipated losses in every period. On the other hand, for the equation

$$x_{t+1} = \beta x_t + \epsilon_{t+1}, |\beta| > 1, \tag{72}$$

the likelihood was shown by White in [27] to be the Cauchy distribution. The reason White's proof is not better known is due to the implications of White's proof. Mann and Wald were able to show that the maximum likelihood estimator for β was the ordinary least squares estimator for all values of β, but White showed that the sampling distribution was the Cauchy distribution, implying that no solution existed for the problem [26] [27].

Although Thiel's regression would still work as a substitute for an actual solution, this is shown in separate work to not be admissible [12].

This issue does not exist for Bayesian methods which do not depend upon point statistics to function.

A rather peculiar problem is created by

$$x_{t+1} = \beta x_t + \epsilon_{t+1}, |\beta| = 1. \tag{73}$$

In Bayesian methods the probability that $\beta = 1$ is zero. This is due to the fact that any countable subset of an uncountable set is of measure zero. Extending White's logic, it is argued that the likelihood function for the equation

$$x_{t+1} = \beta x_t + \epsilon_{t+1}, |\beta| \geq 1, \tag{74}$$

is the Cauchy distribution. This would not be true for a non-Bayesian solution. For that, the Dickey-Fuller test statistic would be operative [37].

10.2. White's Proof

White in [26] worked on this problem in the Likelihoodist school of thinking. As with Landon in [36], the form of his proof has a Bayesian interpretation allowing a simple conversion from Likelihoodist to Bayesian thinking.

10.2.1. Assumptions and Simplifications

White made some simplifying assumptions that would result in no loss of generality or which would simplify notation. In particular, the notation for the vector of observations is

$$x' = (x_1, x_2, \cdots, x_t). \tag{75}$$

The variance of the error is one, that is

$$\sigma^2 = 1. \tag{76}$$

Additionally, the initial value is zero, that is

$$x_0 = 0. \tag{77}$$

The summation operator is without subscript or superscript and sums from one to T where T is the last observation. Following Mann and Wald in [26] the maximum likelihood estimator is

$$\hat{\beta} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2}. \tag{78}$$

Several simplifying notations exist for matrix relationships In particular several $T \times T$ matrices are created. These are

$$P = \begin{bmatrix} 1+\beta^2 & -\beta & 0 & 0 & & & \\ -\beta & 1+\beta^2 & -\beta & 0 & & & \\ 0 & -\beta & 1+\beta^2 & -\beta & & & \\ & & & & \ddots & & \\ & & & & -\beta & 1+\beta^2 & -\beta \\ & & & & 0 & -\beta & 1 \end{bmatrix}, \tag{79}$$

$$A = -\frac{1}{2} \begin{bmatrix} 2\beta & -1 & 0 & & & \\ -1 & 2\beta & -1 & & & \\ 0 & -1 & 2\beta & & & \\ & & & \ddots & & \\ & & & -1 & 2\beta & -1 \\ & & & 0 & -1 & 0 \end{bmatrix}, \tag{80}$$

and

$$B = \begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 0 \end{bmatrix}. \tag{81}$$

The joint distribution of x' is

$$f(x') = \frac{\exp\left(-\frac{1}{2}x'Px\right)}{(2\pi)^{\frac{T}{2}}} \tag{82}$$

and

$$\hat{\beta} - \beta = \frac{x'Ax}{x'Bx}. \tag{83}$$

The standard methods of finding the distribution due to Cramer in [38] of $\hat{\beta} - \beta$ were shown by White in [27] to diverge.

10.2.2. Observations about White's Proof

White's proof determines that the *sampling distribution* of the slope estimate is the Cauchy distribution for this case. His proof was to show that the limiting distribution of Equation (83) is the same as that found in Equation (17). $\hat{\beta}$ is a form of sample mean and as seen above in Equation (33), this is a catastrophic failure.

Although this is catastrophic for non-Bayesian methods, it is not a problem for Bayesian methods. In particular, the form of the proof used by White has a Bayesian interpretation.

Because standard methods diverge, White uses the product of the square root of Fisher information and the likelihood function to gather information about the sampling distribution. White's test statistic reframed into Bayesian language is the posterior density function. A question arises from White's use of ordinary least squares here.

Ordinary least squares is a form of the sampling mean and as shown in Equation (33) would have the same sampling distribution of the set of all possible slopes. Although this provides no information at all about the location of the non-existent population mean, it does provide the likelihood of a set of slopes where $|R| > 1$. At least as important, the solution would work in a Bayesian setting for $R \geq 1$ as it would still be the same ratio distribution.

10.3. Other Than Independent, Uncorrelated Errors

Although White's proof works fine for independent, uncorrelated errors, the other likelihoods would apply in other cases, such as an absence of independent errors and methods such as instrumental variable regression would still be necessary in cases where that would be appropriate.

The likelihood depends on model assumptions.

10.4. Log-Log Models

For the case where the distribution of errors in raw form would be the Cauchy distribution, or where the likelihood function would be the Cauchy distribution, it is already known that the hyperbolic secant distribution is the distribution of a logarithmic based model. This is obvious in the relationship between the cumulative density functions of the Cauchy and hyperbolic secant model.

The cumulative density function for the Cauchy distribution is

$$\frac{1}{2} + \tan^{-1}\left(\frac{x-\mu}{\gamma}\right), \tag{84}$$

while the cumulative density function for the hyperbolic secant distribution is

$$\frac{2}{\pi}\tan^{-1}\left\{\exp\left[\frac{\pi}{2}\left(\frac{x-\mu}{\sigma}\right)\right]\right\}. \tag{85}$$

The link between the two distributions is that if $X \sim \mathbb{C}(0,1)$ and $Y = \frac{2}{\pi}\log|X|$, then $Y \sim \mathbb{HS}(0,1)$ [39].

Ordinary Least Squares

A peculiar issue exists resulting from this relationship. Consider a regression model like

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_0 + \epsilon, \tag{86}$$

where x_1 and x_2 are logarithmic transformations of data drawn from a Cauchy distribution. Because a logarithmic model was used the central limit theorem still holds and so a covariance matrix will exist for the errors if ordinary least squares is used, although the actual likelihood function would be

$$\frac{1}{2\gamma}\operatorname{sech}\left[\frac{\pi}{2}\left(\frac{y - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \hat{\beta}_0}{\gamma}\right)\right]. \tag{87}$$

The important element of Equation (87) is that adding dimensions does not add a covariance matrix. Neither the Cauchy distribution nor the hyperbolic secant distribution has a structure similar to a covariance matrix. This is also true for the vector version of this, should multiple equations be estimated. The scale parameter changes, of course, for each model, but variables do not covary.

This does not exclude all forms of comovement, just that comovement described by the definition of covariance. Because of this, while x_1 and x_2 do not covary, they are also not independent either. This begs the interpretation of the covariance matrix in ordinary least squares since it is an estimator for a set of parameters that do not exist.

10.5. Example-Zimbabwe

Consider the special case of Zimbabwe documented by Richardson in [40]. Zimbabwe had been called the "jewel of Africa", due to its extensive natural resources and its ability to feed its people. Zimbabwe had a difficult colonial history. In the 1890's the colonial government seized native lands and transferred them to white colonists for farming. The consequence of this was that 4500 hundred white families owned almost all commercial farmland, while 840,000 black farmers worked on communal farms. The Mugabe government decided to return "stolen" land to the black population, although almost none of the existing white farmers could trace their families to the original colonists as the land had subsequently changed hands [40].

The transfer happened without compensation and was transferred to individuals in small plots who lacked the infrastructure and skills to manage such a farm. The resulting economic collapse is still felt to this day.

The question would be how to model this. Fortunately, Bayesian theory provides a disciplined solution called Bayesian model selection. In Bayesian model selection, the model is considered a parameter.

There are several possible hypothesis prior to seeing the actual data. The first would be that Zimbabwe grew constantly throughout the entire period so that the model would be

$$Model_1 = \left\{ y_{t+1} = \beta_0 y_t + \alpha_0 + \epsilon_{t+1}, \beta_0 \geq 1, \forall t \right\}. \tag{88}$$

A second model would be that Zimbabwe grew, then collapsed and never recovered. That model would be

$$Model_2 = \begin{cases} y_{t+1} = \beta_1 y_t + \alpha_1 + \epsilon_{t+1}, \beta_1 \geq 1, \forall t < k \\ y_{t+1} = \beta_2 y_t + \alpha_2 + \epsilon_{t+1}, \beta_2 < 1, \forall t \geq k \end{cases}. \tag{89}$$

Obviously one could continue partitioning the set of data into greater and greater numbers of subsets. The first model has three parameters to estimate. They are $Model_1, \beta_0$ and α_0. The second model has six parameters to estimate, they are $Model_2, \beta_1, \beta_2, \alpha_1, \alpha_2$ and k. This would be the conceptual equivalent to a structural break in Frequentist statistics except that a probability distribution would be assigned to possible dates for the break.

Of course one could continue this process adding a third growth process. This would necessitate the a distribution for the set $\{k_0, k_1, k_1 > k_0\}$. This would need to be facilitated by a contingent distribution for k_1 given a value of k_0.

Although this would seem to create a likelihood of overfitting the model to the data, the Bayesian posterior density naturally penalizes increased model structure in a mathematically coherent manner. Coherence, in this case, implying that fair gambles could be placed on models by governments and organizations.

10.6. Example-Calculating Dividends

Dividends are a stream of payments. How these payments are estimated depends entirely on the question being answered. Asking "what is the anticipated dividend of CNA Financial Corporation" is quite a different question from

"what are the total dividends to be paid by all members of the Standard and Poors 500". It is not credible to believe they would be estimated in the same manner as differing information governs.

For the case of CNA Financial Corporation, the dividend policy appears at various times in the disclosure filings with the Securities and Exchange Commission. Because CAN's board of directors is subject to time inconsistency a peculiar Bayesian problem forms, one that cannot truly be solved in a non-Bayesian manner. In particular, the problem is that the solution should be *subgame imperfect*. This implies that any model solution must be invalid unless it considers alternative models in its construction and assigns them a positive probability of being used.

A second problem would exist for CNA Financial. There always exists a positive probability the firm will suspend or not declare a cash dividend. So any dividend model must contain two components. The first is an estimate of the dividend, given one is declared, times the probability of declaration; the second element is a dividend of zero times the complementary probability.

Finally, it is probably better to model dividends either as a dividend yield or as a function of accounting measures such as income or free cash flows as the alternative policy to the stated one of the board. While it is not possible to state a clear likelihood function without facing a specific problem, some general principles can be discussed.

First, the modeler has to determine in advance whether they are seeking a point estimate or a distribution of estimates. If a distribution is sought then the Bayesian predictive distribution should govern the process.

On the other hand, if a point estimate is required, a cost function should be applied to the predictive distribution and a minimization of cost sought. This cost function should either be the true cost function of being wrong for the subjective actor, or the cost function of the marginal actor. For the marginal actor, in most cases, where the distribution is truncated at zero, the all-or-nothing cost function should be used. The logic behind has to do with an understanding of the marginal actor and the center of location.

If it is assumed that the marginal actor defines pricing in the system, then the price by the marginal actor is without "error". That is to say; it sits at the center of location for the distribution function. For a truncated distribution, this is usually at the mode. The all-or-nothing cost function, when applied to a continuous distribution, will be minimized by choosing the modal value.

Finally, if it is assumed that the dividends are being declared in an expanding economy or one with a fiat currency subject to persisting inflation, then it is also reasonable to believe that the likelihood function will be one without a defined mean.

11. Including the Budget Constraint

The impact of the budget constraint has two obvious solutions. The first is to explicitly model the cost of liquidity in the capital markets. Work on this can be

found in Abbott's chapter on measures of discount for liquidity and marketability [41]. The second would be to observe that only the numerator of the reward is impacted as someone cannot have a reward for an asset that he or she does not own, ignoring short-sales. As such, there is a numerator only if a counter-party has sufficient funds and is willing to expend them at the desired price. This changes the solution from the probability of a return to the probability of a return given the budget constraint is met times the probability the budget constraint of the counter-party will be met.

11.1. Abbott's Formulation of Liquidity

Abbott in [41] observes that the cost of liquidity is directly related to the half-life of the order size in the dealer's account. From the buyer's point of view, the price is marked up by a liquidity cost to the value normally called the ask. The seller receives a marked down price normally called the bid. For a single buy, where q_t is the quantity for sale, then the markup factor would be

$$\exp(\lambda q_t), \tag{90}$$

while the mark-down factor for a sale would be

$$\exp(-\lambda q_t). \tag{91}$$

This implies that you could rewrite prices as

$$p_t = \exp(\lambda q_t) p_t' \tag{92}$$

and

$$p_{t+1} = \exp(-\lambda q_{t+1}) p_{t+1}'. \tag{93}$$

Reformulating definition 1, we arrive at

$$r_t = \frac{\exp(-\lambda_{t+1} q_{t+1}) p_{t+1}'}{\exp(\lambda_t q_t) p_t'}. \tag{94}$$

If $q_t = q_{t+1}$ and $\lambda_t q = \lambda_{t+1}$, then from Section 7 then λ follows a normal distribution.

Because most people are really concerned with their net return, the real issue presented by Abbott in [41] is considering the net effect after liquidity costs. This could be presented as

$$r_t' = \frac{p_{t+1}'}{p_t'}. \tag{95}$$

This allows us to escape the question regarding the distribution of the individual terms by using regression. His work considers the impact of volume on the bid-ask spread. The advantage of this method is that it allows a net predictive distribution for planning purposes by institutional investors.

Abbott in [41] extends this discussion to discuss illiquidity or market failure. As λ becomes large, the probability of no counter-party trade increases. Conversely, volume increases λ becomes small. This permits a discussion of a

net return, given a trade happens, multiplied by the probability of that trade happening, which of course is essential to Bayesian thinking.

Of course, if one does not assume that $q_t = q_{t+1}$, then an explicit modeling of q has to happen as well. While this is stationary, it is outside of the scope of this article to explore how to model this.

11.2. Survival Functions

There is a second way to model the budget constraint. If the market maker is willing to buy or sell an asset at a cost, then it is as if there were no budget constraint.

If there is no market maker willing to buy an asset, then a sell must be inside the budget of the possible counterparties. This budget constraint is unknown, so needs to be modeled as a stochastic budget constraint. Further, the budget constraint should be a moving value as the money supply and disposable income changes.

For an asset purchased at time t at a price p_t, the sale at time $t+1$ depends on the existence of sufficient funds for a return to exist. Logistic regression is the logical function here, with two provisos. The first is that when the price is equal to zero, then there is a one hundred percent chance of a trade happening, while the second condition is that the probability of a trade goes to zero as price tends toward infinity.

12. Discussion

Decision making under uncertainty requires an understanding of how that uncertainty is structured. This article provides a framework to approach chance and uncertainty in outcomes when a reward or growth is anticipated. Equation (60) can be thought of as covering far more than only stocks. It includes any event where consumption is voluntarily deferred in the present, with the belief that the result will be a greater outcome in the future.

Voluntary marriage, voluntary involvement in religion, childrearing, the growth of output, profits and other social phenomena are covered by this math. The math differs from the math for phenomena such as defensive wars, insurance or casino gambling, for which the existing economic methods were well suited. It provides a split between models with anticipated gains versus those with anticipated losses.

The differences between the math revolve around the center of location and the spread parameter. The well-behaved properties of the Gaussian distribution are not appropriate for this class of problems.

That the center of location for the truncated distribution is the mode and that no first moment exists should also bring mean-variance finance to an end. Although one could have expected this earlier with the theoretical and empirical papers of Mandelbrot, Fama, MacBeth, French, and Roll; this provides a mathematical reason to exclude mean-variance models from any further use. [2]-[7] [10] At least in its current form, there is no longer any reason to continue

to support the Fama-French model as an alternative hypothesis.

It may appear problematic that the expectation of returns cannot exist, this is not the case. Humans anticipate the future. Other normative methods, such as Bayesian decision theory are not impacted by this loss of a tool.

A rather simple solution to avoid the headaches involved in the change in math is to note an expected utility exists for models with concave utility functions or doubly bounded reward distributions. The math excludes any discussion of *risk-neutral* pricing. It also eliminates the study of risk-loving behavior unless an additional item is added to the utility function such as a utility for engaging in activity or excitement. Alternatively, these can be studied by modeling returns as sufficient returns, making it a Bernoulli trial, rather than continuous returns.

Adding constraints on modeling is not a bad thing. Understanding the distributions allows for an understanding of how humans have to approach problems. It is certain that risk-loving behavior exists and risk-neutral behavior may also exist. Since this is the case, humans are doing something to overcome this issue. The enormous amount of physical and human capital is a testament to this fact.

While this may seem like many tools have been lost, a gain is observed. The catalog of unexplained anomalies may shrink to nothing as most of them were predicated on finite variance or at least a mean. Issues such as heteroskedasticity in models are gone because these distributions are askedastic. Economists can now redirect their efforts and attentions to retesting old models to see how they fare under the math and toward building new ones.

Some issues are not yet understood now, however. An index, such as the Standard and Poors 500 is now a statistic, and it is a statistic with unknown mathematical properties. Historically, the properties did not matter because it was approximately a weighted mean in a system believed to be well covered by the properties of the central limit theorem. Those properties may now matter. Also, traded variants as unit investment trusts may have very different properties. At this time we do not know. The traded form is an observable real world contract with specified properties. Linkages may exist among the securities by the contract that may not exist without it.

A new concept of comovment is required as well in order to understand regression. The method of ordinary least squares implies that a series is convergent. These are divergent time series. Capital is a source and not a sink by its very nature. A way of understanding the Cauchy distribution is as the solution to a pendulum problem. If X maps onto Y and both are drawn from a Cauchy distribution, then one could view regression as a double pendulum problem. Solving $Y|X$ since both are diverging implies that for a given value of X which could have been viewed as a movement of a pendulum, Y is drawn from a second pendulum attached to the end of the first.

The most that could be said of an equation such as $y = \beta x + \alpha$ where x and y are drawn from a Cauchy distribution is that fifty percent of the time $y \geq \beta x + \alpha$

and half of the time $y < \beta x + \alpha$. While, of course, a density could be provided this allows items at any given point to diverge while remaining linked. The simple concept of the Capital Asset Pricing Model that if $\beta = 2$ then if asset x increases by 1% then it is to be expected that asset y will increase by 2% is too strong of a statement here. The interpretation should instead be that fifty-percent of the time it should be the case that asset y will increase two or more percent and half of the time, it will not.

Additionally, the problem of local volatility, that is a variance over an interval of time is probably meaningless. For the univariate Cauchy distribution, the scale parameter describes the half-width at half-maximum. That is the points

$$\Pr(\mu \pm \gamma) = \frac{1}{2}\Pr(\mu) \tag{96}$$

and

$$\frac{1}{\pi}\int_{\mu-\gamma}^{\mu+\gamma}\frac{\gamma}{\gamma^2 + (x-\mu)^2} = \frac{1}{2}. \tag{97}$$

Spectral analysis would be required to determine how long one swing of the spectrum of returns would be. For shorter time periods is not clear how well one could predict the local, realized variability.

As mentioned in Section 3.4.2 there is no covariance matrix. Although there is a linkage between the separate scale parameters to the joint scale parameter in the multivariate case, it is not clear that this survives truncating the distribution for the limitation of liability. Further, by having a single scale parameter, errors are spherical. The implication is that the errors in growing economies linked by trade are altered by that trade. Errors cannot be independent.

This discussion is necessarily incomplete. Individuals with capital in their models will likely be reporting unexpected side effects for quite some time. New ways of thinking, verified by empirical observation, will emerge and the core tools of microeconomics will come to the rescue of researchers.

13. Conclusions

It is time for a change. It has been for quite some time. Now that a method of constructing distributions is understood, it is time to ask the questions believed answered earlier. Markowitz asked how humans made the risk and reward trade off [1]. This must be asked again. To this, we should also ask how the banking system impacts that process.

Knowledge of the likelihood function allows individuals to build and test economic models in a rational manner and also allows economists to minimize the number of required assumptions. Likewise, other fields that work with data that grow exponentially are impacted, they also have the same tools available. This does create a pedagogic issue however for undergraduates.

For business students, it will be necessary to teach both Bayesian and non-Bayesian methods. Null hypothesis methods will still be vital in fields such as marketing and accounting and of occasional use to fields like economics or

finance. Conversely, there will be times where Bayesian methods will be of great value to marketing or accounting. It is time to build a book larger than Fisher's cookbook solution to problems. The twentieth century began with only Bayesian methods, and null hypothesis methods exploded in value and use. The twenty-first century will need to bring balance to the utilization of both systems of thinking.

The twenty-first century is a time where tremendous levels of information will become available. Students and practitioners should not be taught one method or another, but both with a clear theoretical grounding in when one method or another method should be used and the trade-off that is created by such.

Research should proceed on anything that impacts cash flows, risks to those cash flows and the operation of financial intermediation through liquidity and credit services. This implies that bankruptcy, merger and dividend risk for equity securities are meaningful contributions. Likewise, the linkage between accountancy and market decision-making is possibly far more significant than in previous times. Under an assumption of efficient returns, how accountancy impacts decision-making was less important. Now that importance is no longer understood. Research should also proceed into the marketability of assets.

It is quite possible that since initial public offerings would have to be at least as good as existing securities in the secondary market, that initial offerings set the pricing for the system, that is they are the marginally priced security. While the ninety-day. Treasury bill was the implicit marginal asset under mean-variance finance, it may not be. As it stands, we do not know. Hence, the study of marketability discounts and the study of private firms may be critical to studying market pricing. It may be bills, it may be new public securities or it may be private firms with limited marketability. It is time to look.

Implicit in the research agenda above but at least as important is the study of governmental stability, the nature of policies and taxation. It is reasonable to believe that policy impacts returns, but how this happens is less than clear. Without an assumption of market efficiency for returns, the nature of this question changes fundamentally.

The important result from the Capital Asset Pricing Model, where net returns above the risk-free rate drive the process is no longer supported by the math, but we do not know what is supported by the math. The most basic trade-off questions need to be researched.

The research agenda is vast, and the distinction between fantasy and reality will only be settled with time, data and well-structured inference.

Acknowledgements

References

[1] Markowitz, H. (1952) Portfolio Selection. *The Journal of Finance*, **7**, 77-91.

[2] Mandelbrot, B. (1963) The Variation of Certain Speculative Prices. *The Journal of Business*, **36**, 394-419. https://doi.org/10.1086/294632

[3] Fama, E.F. (1963) Mandelbrot and the Stable Paretian Hypothesis. *Journal of Business*, **36**, 420-429. https://doi.org/10.1086/294633

[4] Fama, E. (1965) The Behavior of Stock Market Prices. *Journal of Business*, **38**, 34-105. https://doi.org/10.1086/294743

[5] Fama, E.F. and Roll, R. (1968) Some Properties of Symmetric Stable Distributions. *Journal of the American Statistical Association*, **63**, 817-836.

[6] Fama, E.F. and Roll, R. (1971) Parameter Estimates for Symmetric Stable Distributions. *Journal of the American Statistical Association*, **66**, 331-338. https://doi.org/10.1080/01621459.1971.10482264

[7] Fama, E.F. and MacBeth, J.D. (1973) Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy*, **81**, 607-636. https://doi.org/10.1086/260061

[8] Fama, E.F. and French, K.R. (2008) Dissecting Anomalies. *The Journal of Finance*, **63**, 1653-1678. https://doi.org/10.1111/j.1540-6261.2008.01371.x

[9] Yilmaz, B.Z. (2010) Completion, Pricing and Calibration in a Levy Market Model. Master's Thesis, The Institute of Applied Mathematics of Middle East Technical University, Ankara.

[10] Fama, E.F. and French, K.R. (1992) The Cross-Section of Expected Stock. *The Journal of Finance*, **47**, 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

[11] Harris, D. (2015) A Population Test of Distribution Assumptions in Mean-Variance Models for the Years 1925-2013. Working Paper at SSRN, Amsterdam.

[12] Harris, D. (2016) Why Practitioners Should Use Bayesian Statistics. Working Paper at SSRN, Amsterdam.

[13] Parmigiani, G. and Inoue, L. (2009) Decision Theory: Principles and Approaches. Wiley Series in Probability and Statistics, Chichester, 155-171. https://doi.org/10.1002/9780470746684

[14] Jaynes, E.T. (2003) Probability Theory: The Language of Science. Cambridge University Press, Cambridge, 205-207. https://doi.org/10.1017/CBO9780511790423

[15] 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**, 375-380. https://doi.org/10.1093/biomet/61.2.375

[16] Koopman, B.O. (1936) On Distributions Admitting a Sufficient Statistic. *Transactions of the American Mathematical Society*, **39**, 399-409. https://doi.org/10.1090/S0002-9947-1936-1501854-3

[17] Rothenberg, T.J., Fisher, F.M. and Tilanus, C.B. (1964) A Note on Estimation from a Cauchy Sample. *Journal of the American Statistical Association*, **59**, 460-463. https://doi.org/10.1080/01621459.1964.10482170

[18] Bayes, T. (1764) An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions,* **53**, 370-418. https://doi.org/10.1098/rstl.1763.0053

[19] Savage, L.J. (1954) The Foundations of Statistics. John Wiley & Sons, New York.

[20] Cox, R.T. (1961) The Algebra of Probable Inference. Johns Hopkins University Press, Baltimore.

[21] Markowitz, H. and Usmen, N. (1996) The Likelihood of Various Stock Market Return Distributions, Part 1: Principles of Inference. *Journal of Risk and Uncertainty*,

13, 207-219. https://doi.org/10.1007/BF00056153

[22] Curtiss, J.H. (1941) On the Distribution of the Quotient of Two Chance Variables. *Annals of Mathematical Statistics*, **12**, 409-421. https://doi.org/10.1214/aoms/1177731679

[23] Gurland, J. (1948) Inversion Formulae for the Distribution of Ratios. *The Annals of Mathematical Statistics*, **19**, 228-237. https://doi.org/10.1214/aoms/1177730247

[24] Marsaglia, G. (1965) Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, **60**, 193-204. https://doi.org/10.1080/01621459.1965.10480783

[25] Marsaglia, G. (2006) Ratios of Normal Variables. *Journal of Statistical Software*, **16**, 1-10. https://doi.org/10.18637/jss.v016.i04

[26] Mann, H. and Wald, A. (1943) On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica*, **11**, 173-200. https://doi.org/10.2307/1905674

[27] White, J.S. (1958) The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *The Annals of Mathematical Statistics*, **29**, 1188-1197. https://doi.org/10.1214/aoms/1177706450

[28] Rao, M.M. (1961) Consistency and Limit Distributions of Estimators of Parameters in Explosive Stochastic Difference Equations. *The Annals of Mathematical Statistics*, 32, 195-218. https://doi.org/10.1214/aoms/1177705151

[29] NIST (2012) Nist/Sematech E-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/

[30] Beckmann, P. (1964) Rayleigh Distribution and Its Generalizations. *Radio Science Journal of Research*, **68D**, 92-932. https://doi.org/10.6028/jres.068D.092

[31] Gull, S.F. (1988) Bayesian Inductive Inference and Maximum Entropy. Kluwer Academic Publishers, Berlin. https://doi.org/10.1007/978-94-009-3049-0_4

[32] McCullaugh, P. (1992) Conditional Inference and Cauchy Models. *Biometrika*, **79**, 547-559.

[33] Burdzy, K. (1984) Excursions of Complex Brownian Motion. University of California, Berkeley.

[34] Berger, J.O. (1980) Statistical Decision Theory, Foundations, Concepts, and Methods. Springer Series in Statistics, New York. https://doi.org/10.1007/978-1-4757-1727-3

[35] Graham, B. (1973) The Intelligent Investor: A Book of Practical Counsel. 4th Edition, Harper and Row, New York, 18-22.

[36] Landon, V.D. (1941) The Distribution of Amplitude with Time in Fluctuation Noise. *Proceedings IRE*, **29**, 50-54. https://doi.org/10.1109/JRPROC.1941.233980

[37] Dickey, D.A. and Fuller, W.A. (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, **74**, 427-431.

[38] Cramer, H. (1946) Mathematical Methods in Statistics. Princeton University Press, Princeton.

[39] Ding, P. (2014) Three Occurrences of the Hyperbolic-Secant Distribution. *The American Statistician*, **68**, 32-35. https://doi.org/10.1080/00031305.2013.867902

[40] Richardson, C.J. (2005) How the Loss of Property Rights Caused Zimbabwe's Collapse. *Cato Institute Economic Development Bulletin*, **4**, 1-4.

[41] Abbott, A. (2009) Valuation Handbook: Measures of Discount for Lack of Marketability and Liquidity. Wiley Finance, Hoboken, 474-507.

Appendix: Cancer and Other Related Problems

While this article was intended for use with models of capital, this, in fact, applies to anything that would grow at an exponential rate in the absence of a boundary condition. To provide a second example, consider the implications different distributions would have on cancer growth rates.

Although cancer grows in three dimensions, as in Equation (35), the multivariate equivalents to most of the univariate issues has not been solved. Viewing a tumor in one dimenion does allow a discussion of how those differing equations would be interpreted. For example, if cancer growth rates were well modeled by Equation (26), then this would imply that the cells acted as if independent. On the other hand, if the growth rates were to behave as in Equation (37), it would imply that the cells with the highest, or lowest, depending on what was being modeled, growth rates determined the overall growth rate of the system. It would imply heterogeneous growth rates within the tumor.

Equation (45) would imply that the growth rates depended on some factor, possibly a signal or a switch, while Equation (49) would imply a system in disequilibrium rather than in homeostasis.

Additionally, regression of data in raw form would have a posterior density of the Cauchy distribution and the hyperbolic secant distribution in log-log form in the case of flat priors.

Because Bayesian model selection allows the testing of multiple distributions and models are not restricted to single dimension models as implied in this article, for a given medical problem, Bayesian model selection will provide information about the underlying nature of the relationships among cells.