MAXIMUM LIKELIHOOD ESTIMATION OF STABLE PARETIAN DISTRIBUTIONS APPLIED TO INDEX AND OPTION DATA

STEPHEN KINSELLA AND FERGAL O’BRIEN

ABSTRACT. Ample evidence exists documenting the fat-tailed character of returns in financial markets (Rachev and Mittnik 2000). Several papers attempt to model these fat tailed distributions as power laws of the form $Pr(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}$, where $k$ is a positive integer measuring asset returns. $Pr(k)$ is the probability of actually observing $k$, $\alpha$ is the power law exponent, and $\zeta(\alpha)$ is the Riemann zeta function defined as $\sum_{k=1}^{\infty} k^{-\alpha}$. The method most employed in determining the power law exponent is graphical analysis of the log of the ranked data followed by regression. In this paper, we point out the flaws of this method of discovering power laws, and argue for a more direct method of discovery using maximum likelihood estimation over a bounded field. After deriving our model, we test it on two high frequency time series: AMEX-OIL, and European-style exercise options on FSTES 100. Relative to the standard OLS power law fitting procedure, we find a much more accurate fit to the data using the maximum likelihood method, estimating $\alpha$ exponents of approximately $3 - 5$, which we verify using Kolmogorov-Smirnov tests.

Date: February 4, 2009.

Version 1.0. JEL Classification: C10, C16, C19, G12. Keywords: Maximum Likelihood, Power Law, Asset Pricing.
ABSTRACT. Ample evidence exists documenting the fat-tailed character of returns in financial markets (Rachev and Mittnik [2000]). Several papers attempt to model these fat tailed distributions as power laws of the form $Pr(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}$, where $k$ is a positive integer measuring asset returns, $Pr(k)$ is the probability of actually observing $k$, $\alpha$ is the power law exponent, and $\zeta(\alpha)$ is the Riemann zeta function defined as $\sum_{k=1}^{\infty} k^{-\alpha}$. The method most employed in determining the power law exponent is graphical analysis of the log of the ranked data followed by regression. In this paper, we point out the flaws of this method of discovering power laws, and argue for a more direct method of discovery using maximum likelihood estimation over a bounded field. After deriving our model, we test it on two high frequency time series: AMEX-OIL, and European-style exercise options on FSTES 100. Relative to the standard OLS power law fitting procedure, we find a much more accurate fit to the data using the maximum likelihood method, estimating $\alpha$ exponents of approximately $3 - 5$, which we verify using Kolmogorov-Smirnov tests.

1. INTRODUCTION

If high frequency, long running financial time series exhibit stable Pareto, lognormal, and stretched exponential distributions, then it is important to estimate those distributions in as precise and consistent a manner as possible in order to price assets correctly.

This paper attempts an estimation process for two high frequency, long running financial time series, and provides a symbolic (and numerical) method by which the estimation may be carried out by practitioners via maximum likelihood, building on Rose and Smith (2000) and Clauset et al. (2007).

Given the abundance of evidence about lognormal and power-law distributions in financial economics, going back to Mandelbrot (1963), it is prudent to search for a scaling exponent when estimating returns data with suspected fat tails.

Where Pareto distributions are suspected, returns are said to follow ‘power laws’ of the form $Pr(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}$, where $k$ is a positive integer measuring the asset returns, $Pr(k)$ is the probability of actually observing $k$, $\alpha$ is the power-law exponent, and $\zeta(\alpha)$ is the Riemann zeta function defined as $\sum_{k=1}^{\infty} k^{-\alpha}$.

Because the power law takes the form $Pr(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}$, taking the log of both sides gives $\log Pr(k) = -\alpha \log k - \log \zeta(\alpha)$, which is a straight line, $\alpha$ obviously giving the slope.

In practice, a line is then fitted to the data by linear regression to determine the value of $\alpha$. This approach is flawed in two respects Clauset et al. (2007). First, power-law distributions are defined only over positive ranges, so all negative values must be discarded, because the characteristic equation of the Pareto distribution requires two positive parameters, $\alpha > 0$ and $\beta > 0$. Second, there is nothing in this procedure to remove the problem of being able to distinguish power law distributions from other distributions, such as the lognormal or stretched exponential. This article proposes a procedure to test for a power law by fitting the data in question to a power law distribution using maximum likelihood, then testing for goodness of fit using Kolmogorov-Smirnov tests.

A new simple maximum likelihood approach to power law estimation is important, because high frequency, long running time series in general do exhibit stable Pareto and lognormal distributions—recent studies (see Gabaix et al. (2003) Adler et al. (1998), and references
have found that the returns on many stocks, and stock indices, have a tail exponent $\alpha \approx 3$ while the distributions of trading volume, and the number of trades on financial markets obey power laws with exponents $\alpha \approx 1.5$ and $\alpha \approx 3.4$, respectively (Scherer et al. (2000), Danielsson et al. (2006), Theodossiou (1998), Duan (1994), and Silverberg and Verspagen (2003)). These estimates of the tail indices of $\alpha$ are robust to different types and sizes of financial markets, market trends, and are similar for different countries. Several recent studies have derived closed form likelihood expansions for the MLE approach in different settings, for example stochastic volatility modeling (A¨ıt-Sahalia, 2002, A¨ıt-Sahalia and Kimmel, 2007) and generalized linear mixed models McCulloch (1997).

Several frameworks have been proposed to model heavy-tailed phenomena, including stable distributions, Pareto distributions, multivariate $t$-distributions, mixtures of normals, power exponential distributions, stretched exponentials, ARCH processes, mixed diffusion jump processes, variance gamma, and normal-inverse gamma distributions (Rachev and Mittnik, 2000, Chapter 2), (Mittnik et al., 2000), (Phillips, 1994). However, the debate concerning the values of the tail indices for different heavy-tailed financial data and on appropriateness of their modeling based on the above distributions is still under way in the empirical literature. In particular, studies continue to find tail parameters of $\alpha \leq 2$ in different financial data sets, while also arguing that stable distributions are appropriate for their modeling (Clauset et al. 2007). Gabaix et al. (2003) propose a model which demonstrated power laws for stock returns, trading volume, and the number of trades, are explained by trading of large market participants—namely the largest mutual funds whose sizes have the tail exponent $\alpha \approx 1$.

DeVany and Walls (2004) presents evidence that stable distributions with tail indices $1 < \alpha < 2$ provide a good model for the distributions of profits in highly uncertain, adaptively contracted industries, like motion pictures.

This can be characterised as one of the first results in the economics of extreme events (DeVany and Walls, 2004). We also note that some studies also indicate the tail exponent is close to one or slightly less than one for such financial time series as the Bulgarian lev/US dollar exchange spot rates, and increments of the market time process for Deutsche Bank price records (see Rachev and Mittnik (2000), Scherer et al. (2000) and Silverberg and Verspagen (2003)) report the tail indices as considerably less than one for financial returns from technological innovations.

Our contribution in this paper is first to derive a symbolic and numerically tractable expression for the maximization of the likelihood of obtaining a reasonable estimate of $\alpha$, $\hat{\alpha}$, and test our estimation method on two data sets, comparing the results of the fits for $\alpha$ we obtain by simple OLS fitting. We perform this comparison using Kolmogorov-Smirnov tests. Results for this exercise are presented in table 5. We find a more accurate fit for these power-law distributed series using maximum likelihood, especially for at the money options and our Oil index data, with $\alpha (3, 5)$.

The rest of our paper is laid out as follows: section 2 describes the estimation method using a log likelihood form, section 4 describes our two data sets, and section 5 shows the findings of the application of the estimation methodology to the data. Section 6 concludes.

2. Estimation Method

The existence of unbiased estimators with minimum variance requires the sufficient statistics of the statistical model to be complete in the mathematical sense rather than the
statistical sense. In practice, statistical models often falter in this regard, so parameter estimators must be found by other means—usually ad-hoc methods.

Maximum likelihood Estimation (MLE) provides a way to generate parameter estimators which share some of the optimality properties of the standard OLS family of estimation techniques. To date, however, this estimation methodology has entailed computationally expense making it unsuitable for large scale financial analysis.


We define the likelihood function below, and illustrate its construction in a variety of settings, before exposing the method to the financial data in section 4.

2.1. Notation. Let \( X \) denote the variable(s) of interest that has (or is assumed to have) a probability density function (pdf): \( f(x; \theta) \), dependent upon a \((k \times 1)\) parameter \( \theta \in \Theta \subset \mathbb{R}^k \), whose true value \( \theta_0 \) is unknown.

We assume the functional form of \( f \) is known\(^2\).

Let \((X_1, \ldots, X_n)\) denote a random sample of size \( n \) drawn on \( X \), in practice a time series of arbitrary length. It is assumed that the pdf of the sample \( f_1, \ldots, n(x_1, \ldots, x_n; \theta) \) can be derived from the knowledge we have about \( f \), so the joint density we derive herein will depend upon the unknown parameter \( \theta \). A key point we need to stress is the likelihood function is mathematically equivalent to the joint distribution of the sample for the data types under study.

The real benefit of MLE to financial modelers in the age of vast computational resources is that instead of regarding the likelihood as a function of the \( X_i \)’s, the likelihood can be interpreted as a function of \( \theta \) defined over the parameter space \( \Theta \) for fixed values of each \( X_i = x_i \). The advantage of this approach is the recursive structure each \( x_i \) inherently has.

The simple likelihood for \( \theta \) is

\[
L(\theta|x_1, \ldots, x_n) \equiv f_{1, \ldots, n}(x_1, \ldots, x_n; \theta).
\]

(1)

It makes most sense to shorten the notation for the likelihood to just \( L(\theta) \), because the finer points of the definition are not used below.

Construction of the joint pdf may at first sight seem a daunting task. However, if the variables in \((X_1, \ldots, X_n)\) are mutually independent, then the joint pdf is given by the product of the marginals shown in equation 2

\[
f_{1, \ldots, n}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),
\]

(2)

\(^1\)We estimate simple returns rather than the standard log returns more commonly employed in the ‘power law’ literature. Coval and Shumway (2001) note that using logarithmic returns for options is problematic due to the fact that options held to maturity often generate returns of -100 %, and consequently the log transformation of any set of option returns over any finite holding period will be significantly lower than the raw net returns.

\(^2\)Or at least \( f \) can be guessed at, which makes sense for financial data of the type we are interested in.
which usually makes it easy to construct the joint pdf and hence the likelihood for \( \theta \). The literature on the necessary independence of explanatory variables and the controversy this assumption has created is summarised and criticized in Jeynes (1978), though his contribution has spawned a large literature which is beyond the scope of this paper.

It is important usually to distinguish between two forms of the likelihood for \( \theta \), namely, the likelihood function, and the observed likelihood function.

The likelihood function is defined as the likelihood for \( \theta \), given the random sample prior to observation; it is given by \( L(\theta|X_1, \ldots, X_n) \), and is a random variable.

Where there is no possibility of confusion, we will use ‘likelihood’ and ‘likelihood function’ interchangeably. The second form, the observed likelihood, is defined as the likelihood for \( \theta \) evaluated for a given sample of observed data, and it is not random. The examples developed below illustrate the construction of the likelihood, and its observed counterpart.

In simple cases, the likelihood function is easily constructed, since due to mutual independence, the joint distribution of the random sample is simply the product of the marginal distributions. In practical financial situations, however, mutual independence amongst the sampling variables does not occur, and so the derivation of the likelihood function requires more effort. Examples include financial time series models, defined as models pertaining to variables collected through time that depend on their past.

2.2. Example. Consider a random walk for time-dated returns \( X_{t,t-1} \) with drift model, following (Velupillai, 2005, pg. 65):

\[
X_t = \mu + X_{t-1} + U_t
\]

with initial condition \( X_0 = 0 \). The drift is given by the constant \( \mu \in \mathbb{R} \), while the disturbances \( U_t \) are assumed to be independently normally distributed with zero mean and common variance \( \sigma^2 \in \mathbb{R}^+ \); that is, \( U_t \sim N(0, \sigma^2) \), for all \( t = 1, \ldots, T \), and \( E[U_t, U_s] = 0 \) for all \( t \neq s \).

The objective is to construct the likelihood for parameter \( \theta = (\mu, \sigma^2) \). One (and the most common) approach is to use conditioning arguments. We begin by considering the joint distribution of the sample \( (X_1, \ldots, X_T) \). This cannot be written as the product of the marginals (by [1]) as \( X_t \) depends on \( X_{t-1}, \ldots, X_0 \), for all \( t = 1, \ldots, T \).

However, in light of this dependence, suppose instead that we decompose the joint distribution of the entire sample into the distribution of \( X_T \) conditional on all previous variables, multiplied by the joint distribution of all the conditioning variables:

\[
f_{1,\ldots,T}(x_1, \ldots, x_T; \theta) = f_{T|1,\ldots,T-1}(x_T|x_1, \ldots, x_{T-1}; \theta) \times f_{1,\ldots,T-1}(x_1, \ldots, x_{T-1}; \theta)
\]

where \( f_{T|1,\ldots,T-1} \) denotes the distribution of \( X_T \) conditional on \( X_1 = x_1, \ldots, X_{T-1} = x_{T-1} \), and \( f_{1,\ldots,T-1} \) denotes the joint distribution of \( (X_1, \ldots, X_{T-1}) \).

From the form of the standard random walk model, it should be clear that when fixing any \( X_t \), all previous \( X_s, (s < t) \) must also be fixed. Then it is easy to simplify the notation, as the conditional pdf on the right-hand side of (4) may be written as

\[
f_{T|1,\ldots,T-1}(x_T|x_1, \ldots, x_{T-1}; \theta) = f_{T|T-1}(x_T|x_{T-1}; \theta).
\]
Following the usual iid assumptions on the disturbances, it can be shown that

\[(6) \quad X_T | (X_{T-1} = x_{T-1}) \sim N(\mu + x_{T-1}, \sigma^2)\]

which makes it is easy to write down the conditional density given in (5). Consider now the joint distribution of \((X_1, \ldots, X_{T-1})\) on the right-hand side of equation 4. Here, again, the same idea is used to decompose the joint distribution of the remaining variables: the appropriate equations are (4) and (5) but with \(T\) replaced by \(T - 1\). By recursion,

\[(7) \quad f_{1,\ldots,T}(x_1, \ldots, x_T; \theta) = \prod_{t=1}^T f_t | t-1(x_t | x_{t-1}; \theta)\]

where each of the conditional densities in (7) is equivalent to (5) for \(t = 2, \ldots, T\), and \(f_{1|0}\) is the pdf of a \(N(\mu, \sigma^2)\) distribution because of the assumption \(X_0 = 0\). By and from (1), 7 is equivalent to the likelihood for \(\theta\).

2.3. Further derivation. Maximum likelihood parameter estimation is based on choosing values for \(\theta\) so as to maximise the likelihood function. That is, the MLE of \(\theta\), denoted \(\hat{\theta}\), is the solution to the maximized problem:

\[(8) \quad \hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta | X_1 = x_1, \ldots, X_n = x_n).\]

\(\hat{\theta}\) is the value of the argument of the likelihood, selected from anywhere in the parameter space that maximises the value of the likelihood after we have been given the sample. In other words, we seek the particular value of \(\theta\), namely, \(\hat{\theta}\), which makes it most likely to have observed the sample that we actually have. We may view the solution to 8 in two ways depending on whether the objective function is the likelihood function or the observed likelihood function. If the objective is the likelihood, then 8 defines the ML estimator \(\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)\); since this is a function of the random sample, \(\hat{\theta}\) is a random variable. If the objective is the observed likelihood, then 8 defines the ML estimate \(\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)\), where \(x_1, \ldots, x_n\) denotes observed data; in this case \(\hat{\theta}\) is a point estimate.

The solution to 8 is invariant to any monotonic increasing transformation of the objective. Since the natural logarithm is a monotonic transformation, it follows that

\[(9) \quad \hat{\theta} = \arg \max_{\theta \in \Theta} \log L(\theta),\]

which we will use from now on as the definition of the estimator (estimate). The natural logarithm of the likelihood, \(\log L(\theta)\), is called the log-likelihood function. A weaker definition of the MLE, but one that, in practice, is often equivalent to 8 is

\[(10) \quad \hat{\theta} = \arg \max_{\theta \in \Theta} \log L(\hat{\theta}).\]
where $\tilde{\Theta}$ denotes a finite, non-null set whose elements $\tilde{\theta}$ satisfy the conditions

\begin{equation}
\frac{\partial}{\partial \theta} \log L(\hat{\theta}) = 0
\end{equation}

and

\begin{equation}
\frac{\partial^2}{\partial \theta^2} \log(\hat{\theta}) \leq 0.
\end{equation}

The two parts of (12) express, respectively, the first- and second-order conditions familiar from basic calculus for determining local maxima of a function. Generally speaking, we determine MLE through computation of the (12), although the example below, finding MLE for the exponential parameter, relies on (9) alone. One further piece of notation is the so-called score (or 'efficient score' in some texts), defined as the gradient of the log-likelihood,

\begin{equation}
S(\theta) = \frac{\partial}{\partial \theta} \log L(\theta).
\end{equation}

For example, the first-order condition is simply $S(\tilde{\theta}) = 0$.

Let $X \sim \text{Exp}(\theta)$, where parameter $\theta \in \mathbb{R}_+$. Here is its pdf, assuming the domain of $f$ runs from $x$ to infinity, and $\theta > 0 : f = \frac{1}{\theta} e^{-x/\theta}$.

For a random sample of size $n$ drawn on $X$, the log-likelihood function is:

\begin{equation}
\log L\theta = \log \left[ \prod_{i=1}^{n} \frac{f}{.x \rightarrow x_i} \right]
\end{equation}

\begin{equation}
= -n\theta \log[\theta] + \sum_{i=1}^{n} x_i
\end{equation}

The score function is the gradient of the log-likelihood with respect to $\theta$:

\begin{equation}
\text{Score} = \frac{-n\theta + \sum_{i=1}^{n} x_i}{\theta^2}.
\end{equation}

Setting the score to zero and solving for $\theta$ corresponds to the first-order condition given in (12). We find:

\begin{equation}
\theta = \frac{\sum_{i=1}^{n} x_i}{n}.
\end{equation}

The unique solution corresponds to the sample mean, the first moment of the distribution. The nature of the solution is not yet clear; that is, does the sample mean correspond to a local minimum, local maximum, or saddle point of the log-likelihood? A check of the second-order condition, evaluated at the solution point gives

\begin{equation}
- \frac{n^3}{(\sum_{i=1}^{n} x_i)^2},
\end{equation}

which reveals that the Hessian is strictly negative at the sample mean and therefore the log-likelihood is maximized at the sample mean. Hence, the MLE of $\theta$ is

\begin{equation}
\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (X_i).
\end{equation}
Now let $X \sim N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$, with pdf $f(x; \mu, (\sigma^2))$:

$$f = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2}\right];$$

(19)

Assuming domain on $f$ of $-\infty \leftarrow x \rightarrow \infty$ and a real-valued mean, for a random sample of size $n$ drawn on $X$, the log-likelihood for parameter $\theta = (\mu, \sigma)$ is:

$$\log L(\theta) = \log \left[ \prod_{i=1}^{n} (f / x \rightarrow x_i) \right]$$

which gives

$$\frac{-1}{2\sigma^2} \left( n \left( \mu^2 + \sigma^2 \log[2\pi] + 2\sigma^2 \log[\sigma] \right) - 2\mu \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2 \right)$$

(21)

The score vector $S(\theta) = S(\mu, \sigma)$ is given by:

$$\left( \frac{-n\mu + \sum_{i=1}^{n} x_i}{\sigma^2}, \frac{n\mu^2 - n\sigma^2 - 2\mu \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2}{\sigma^3} \right)$$

(22)

Setting the score to zero and solving yields:

$$\left( \begin{array}{c} \sigma \rightarrow -\frac{(\sum_{i=1}^{n} x_i)^2}{\sqrt{n}} + \sum_{i=1}^{n} x_i^2 \sqrt{n} \\ \mu \rightarrow \frac{\sum_{i=1}^{n} x_i}{n} \end{array} \right), \quad \left( \begin{array}{c} \sigma \rightarrow \frac{-\left(\sum_{i=1}^{n} x_i \right)^2}{n} + \sum_{i=1}^{n} x_i^2 \sqrt{n} \\ \mu \rightarrow \frac{\sum_{i=1}^{n} x_i}{n} \end{array} \right)$$

(23)

Clearly, the negative-valued solution for $\sigma$ lies outside the parameter space and is therefore invalid; thus, the only permissible solution to the first-order conditions is:

$$\left( \begin{array}{c} \sigma \rightarrow \frac{-\left(\sum_{i=1}^{n} x_i \right)^2}{n} + \sum_{i=1}^{n} x_i^2 \sqrt{n} \\ \mu \rightarrow \frac{\sum_{i=1}^{n} x_i}{n} \end{array} \right)$$

(24)

Then

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma})$$

is the MLE of $\theta$, where $\hat{\mu}$ and $\hat{\sigma}$ are the formulae given as the solution values for $\theta$ (second-order conditions must be checked below). The functional form for $\hat{\sigma}$ may appear unfamiliar. However, if we utilize the following identity for the sum of squared deviations about the sample mean,

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i)$, then

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}.$$
Invoking MLE’s Invariance Property the MLE of $\sigma^2$ is

\[(\hat{\sigma})^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2\]

which is the 2nd sample central moment.

The second-order conditions may, for example, be checked by examining the eigenvalues of the Hessian matrix evaluated at $\hat{\theta}$:

\[\left( \frac{n^3}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2}, \frac{2n^3}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2} \right)\]

Given the identity for the sum of squared deviations, the eigenvalues of the Hessian are $-n\hat{\sigma}^{-2}$ and $-2n\hat{\sigma}^{-2}$, which clearly are negative. Thus, the Hessian is negative definite at $\hat{\theta}$ and therefore the log-likelihood is maximized at $\hat{\theta}$.

3. MLE for Pareto-distributed data

Let $X \sim \text{Pareto}(\alpha, \beta)$, where parameters $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_+$. Assuming $\alpha > 0, \beta > 0$ and allowing the domain of $f$ to fluctuate from $x \to \infty$, the pdf of $X$ is given by:

\[f = \alpha \beta^\alpha x^{-(\alpha+1)}\]

Since $X \geq \beta$, there exists dependence between the parameter and sample spaces. Given a random sample of size $n$ collected on $X$, the log-likelihood for $\theta = (\alpha, \beta)$ is:

\[(28) \quad (\log[\alpha] + \alpha \log[\beta]) - (1 + \alpha) \sum_{i=1}^{n} \log [x_i]\]

The score vector is given by:

\[(29) \quad \left( n \left( \frac{1}{\alpha} + \log[\beta] \right) - \sum_{i=1}^{n} \log [x_i], \frac{n\alpha}{\beta} \right)\]

If we attempt to solve the first-order conditions in the usual way, we cannot find a solution to the equations. However, if we focus on solving just the first of the first-order conditions, we find:

\[(30) \quad \alpha \to -\frac{n \log[\beta]}{n \log[\beta] - \sum_{i=1}^{n} \log [x_i]}\]

This time a solution is provided, albeit in terms of $\beta$; that is, $\hat{\alpha} = \hat{\alpha}(\beta)$. We now take this solution and substitute it back into the log-likelihood:

\[(31) \quad n \left( -1 + \log \left[ \frac{n}{-n \log[\beta] + \sum_{i=1}^{n} \log [x_i]} \right] \right) - \sum_{i=1}^{n} \log [x_i]\]

This function is known as the concentrated log-likelihood. It corresponds to $\log L(\hat{\alpha}(\beta), \beta)$. Since it no longer involves $\alpha$, we can maximise it with respect to $\beta$. Let $\hat{\beta}$ denote the solution to this optimization problem.

This solution can then be substituted back to recover $\hat{\alpha} = \hat{\alpha}(\hat{\beta})$; then $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ would be the MLE of $\theta$ by recursion.
In general, when the first-order conditions can be solved uniquely for some subset of parameters in $\theta$, then those solutions can be substituted back into the log-likelihood to yield the concentrated log-likelihood. The concentrated log-likelihood is then maximized with respect to the remaining parameters, usually using numerical techniques.

For my example, maximizing the concentrated log-likelihood using standard calculus will not work. This is because the parameter space depends on the sample space. However, by inspection, it is apparent that the concentrated log-likelihood is increasing in $\beta$. Therefore, we should select $\beta$ as large as possible. Now, since each $X_i \geq \beta$, we can choose $\beta$ no larger than the smallest observation. Hence, the MLE for $\beta$ is

$$\hat{\beta} = \min(X_1, X_2, \ldots, X_n)$$

which is the smallest order statistic. Replacing $\beta$ in $\hat{\alpha}(\beta)$ with $\hat{\beta}$ yields the MLE for $\alpha$,

$$\hat{\alpha} = n/\sum_{i=1}^{n} \log(\frac{X_i}{\min(X_1, X_2, \ldots, X_n)})$$

4. Data

We test our specification using two data sets, the LIFFE ESX index option data, and AMEX-OIL, a price-weighted index of the leading companies involved in the exploration, production, and development of petroleum.

4.1. LIFFE ESX. The LIFFE ESX index option data is end-of-day FTSE 100 European style exercise option data obtained from LIFFE through their on-line download service for the twelve-year period 01.01.1993—31.12.2004. Premiums are quoted in decimals and one point equals £10. The minimum tick size is £0.50. There are contracts expiring in March, June, September and December plus additional months such that the four nearest calendar months are always available for trading. The options expire on the third Friday of the month, and settlement is in cash on the first business day after the last trading day (which is the same as the exercise day). The interval between strike prices is either 50 or 100 index points, but the Exchange reserves the right to introduce tighter strike intervals (e.g., 25 points).

Figures 1 and 2 present data on calls and a puts in and around the money for the LIFFE ESX data set.

LIFFE prices the FTSE 100 stock index option as if it was an option on a matched maturity futures contract. To this end, LIFFE provides matched maturity futures price information with its option price data. However, since there are only four FTSE 100 futures contracts traded on LIFFE (March, June, September, and December) a true matched maturity futures price for a given option may not exist. The exchange, however, provides an implied futures price and since, in practice, only the near-dated futures contract trades heavily this is the contract used to calculate that implied price. This is achieved by adjusting the near-dated, or front-end, futures price using a cost of carry calculation. Information relevant to this calculation, most importantly, a consensus ex-ante assessment of the dividend yield on the index portfolio, is provided by market traders.

For each trading day in the period under investigation we have put and call options data for all available strikes and maturities. The data includes trade date, option type, expiry date, strike price, closing price, instrument settlement price, volume, volatility, and closing
Figure 1. Call option data for LIFFE-ESX. The top and bottom right panels show returns and absolute returns, respectively. The top left hand panel shows a leptokurtic distribution of returns relative to the 0-mean centered Gaussian, along with the extreme returns in 1999 and 2000 for this data giving fat (and long) tails to the distribution. The bottom panel reports the Autocorrelation function for 500 lags of the absolute returns, and computes a power law for the absolute returns $|x|$ by simple OLS fitting of the form $y(x) = a|x|^n$ of $y(x) = 0.99|x|^{-3.5}$, with $R^2 = 0.93$. We see from the bottom right panel that there are less persistent memory effects with options data, which makes intuitive sense. Source: LIFFE-ESX data and authors’ calculations.

For the LIFFE ESX contract the option price employed throughout the research is the settlement price. The settlement price in derivatives markets is the price exchanges use to compute daily gains and losses for open positions. To this end FTSE 100 ESX option positions are marked-to-market on a daily based on the previously mentioned implied futures price.

There are some limitations associated with using settlement prices:

(1) If there is no trade in an option during a trading day the settlement price is a theoretical price calculated by the exchange. Therefore any sample based on settlement prices will most likely include strikes that were not actually traded.

Note that the bid/offer data was not always accurately recorded, e.g., the bid/offer was regularly quoted as 0/0, and therefore the midpoint of the bid/offer could not be considered for use for the ESX options data.
Figure 2. Put option data for around the money returns. We see real and absolute returns in the top- and bottom-left panels, respectively, while the top right hand panels indicates the highly skewed nature of the data. Using the autocorrelation function and 500 lags, we can see the power law nature of the data in the bottom right hand panel, where absolute returns, a measure of volatility, are mapped onto their autocorrelation function (ACF). The ACF decays extremely slowly, and even after 500 lags is positive and significant. We estimate the equation \( ACF = x^{-\gamma} \) for lagged absolute returns \(|x|\), finding an equation of \( ACF = 1 \star |x|^{-0.88} \), with \( R^2 = 0.91 \). Source: LIFFE-ESX and authors’ calculations.

(2) Bliss and Panigirtzoglou (2004) note that as most option strikes trade infrequently and with great variations in time-of-last-trade, the market information used by the exchange when setting settlement prices is likely to be non-synchronous.

Crucially the settlement price reported may not be a price at which the option was actually traded at which has an implication for the calculation of price changes. This problem is to some extent mitigated by using the most liquid options available.

The focus is on the price changes for naked positions in put and call index options. The calculation of option price changes is not a straightforward task as the moneyness characteristic of options contracts changes according to movements in the underlying asset and there are a number of different expiries trading at any one time. Therefore, it is necessary to group options into “bins” based on type, moneyness measure, and time-to-maturity.
On every trade date time-to-maturity measures that are labelled first nearby and second nearby option contracts are identified. These contracts are among the most actively traded contracts on both exchanges. Following this the options data is further grouped according to what might be termed a moneyness measure. Option bins are classified according to X/S, the strike price divided by the underlying price, and range from 0.95 to 1.05 for puts and calls with intervals of 0.01. ITM calls will have moneyness measures less than 1.00, ATM calls will have moneyness measures close to 1.00 and OTM calls will have moneyness measures greater than 1.00. OTM puts will have moneyness measures less than 1.00, ATM puts will have moneyness measures close to 1.00 and ITM puts will have moneyness measures greater than 1.00. Using this measure of moneyness in conjunction with the time-to-maturity measure results in 20 call option bins and 20 put option bins for a total of 40 option bins.

Once it has been determined which bin a particular option lies in the option price change is calculated based on holding the option from the close of one trading day to the close of the next trading day. On days where a number of options fall in a particular bin one option is randomly selected and held as the single daily observation for that bin. On days where no option price change is observable, a missing value is recorded.

4.2. AMEX-OIL. AMEX-OIL is a price-weighted index of the leading companies involved in the exploration, production, and development of petroleum. It measures the performance of the oil industry through changes in the sum of the prices of component stocks. The index was developed with a base level of 125 as of August 27, 1984.

By way of illustration, Figure 4.2 below shows a comparison between two histograms. On the left, randomly-generated Gaussian data \( x \sim N(0, 1) \) and on the right the histogram of returns (open-close) to a high frequency (daily) data series AMEX-OIL from 26 Aug 1983–May 26, 2006, \( n = 5741 \). Compared to the typical Gaussian histogram, the AMEX-OIL data shows a skewness of 0.271, a quartile skewness of -0.0351, and a kurtosis of 12.073. We can clearly see the skewed nature of the data set.

A positive skewness indicates the distribution is skewed to the right, that is, compared to the left tail, the right tail is elongated. The kurtosis statistic reflects the peakedness of the centre compared to the normal distribution, so a value somewhere around three would be indicative of normality. Just finding a value above 12 should be convincing evidence that returns to this data series are non-Gaussian.

5. Analysis

Given the non normal nature of the two datasets presented in section 4, now we attempt an MLE estimation of the same data sets with 1 bin for the AMEX-OIL data, and four bins for the LIFFE-ESX data. We obtain the results shown in table 5.

---

4 First nearby options are identified by comparing an option’s trade date with its expiration date. Where an option’s trade date is in January and its expiration date in February, for example, it is identified for first nearby bins. First nearby options have, on average, 34 days to expiration, a minimum time to expiration of 15 days and a maximum time to expiration 51 days.

5 Like first nearby options, second nearby options are identified by comparing the option’s trade date with its expiration date. For example, if the trade date is in August and the expiration date in October then such an option is identified as second nearby. Second nearby options have, on average, 63 days to expiration, a minimum time to expiration of 44 days and a maximum time to expiration 81 days.
Figure 3. Data characteristics for AMEX-OIL. Data is daily recordings of AMEX-OIL from 1983-2006 with $n = 5,741$. Instead of the index itself, we show its returns in the upper left panel, showing increments between trading days. We see a large number of positive and negative movements, combined with some extreme movements in 1987 and 2001. We also see volatility clustering, or switches between calm and turbulent episodes. The top right panel shows the distribution of the series is kurtotic and leptokurtic relative to a Gaussian. Compared to the typical Gaussian histogram, the AMEX-OIL data shows a skewness of 0.271, a quartile skewness of -0.0351, and a kurtosis of 12.073. The lower left panel shows absolute returns, a well recognised measure of volatility. The power law nature of the data is shown in the bottom right hand panel, where absolute returns, a measure of volatility, are mapped onto their autocorrelation function (ACF). The ACF decays extremely slowly, and after 500 lags is positive and significant. Here we see both hyperbolic and exponential decay in the lag structure, we estimate the equation $ACF = |x|^{-\gamma}$ for lagged absolute returns $|x|$, finding an equation of $ACF = 0.55 * |x|^{-0.46}$, with $R^2 = 0.83$. Source: Yahoo! Finance and authors’ calculations.

Table 5 shows the results of estimation runs on the AMEX-OIL data, and all four LIFFE-ESX at the money bins, two calls, and two puts. We estimate $\alpha$ exponents for AMEX-OIL of 4.77, much higher than the $\alpha$ exponent reported by OLS regression to fit a model of the form $y(x) = a x^n$.

\[
\log(y(x)) = \log(x^{-\alpha}) \text{ of 3.89. The Kolmogorov Smirnov test is accepted for each run and for each data set we do in table 5 implying we have found a ‘true’ power law distribution in the data ranges we have been using.}
\]

Interpreting the option data is a little trickier: we do find evidence of a power-law distribution across the puts and calls of the LIFFE-ESX data set, but here the range is much greater, with \(3.73 < \alpha < 5.00\) for both calls and puts.

The variation in these put and call bins is extreme, as we can see from the top left hand panels of figures 1 and 2. Here we have to consider and take account of very extreme events in the distribution of returns across the various bins. These extreme events (in 1999 and 2001, for example), contribute much of the skewness and help produce the large values for \(\alpha\) we observe. Option data, however, is largely memoryless, as we can see from the computation of our ACF functions over 500 lags. Relative to the AMEX-OIL data, this data in fact carries little or no memory in fact, which argues against the long memory hypothesis championed by [Mandelbrot and Hudson (2004)](http://finance.yahoo.com/q), which may be a cause for concern. The time scale and frequency of the data being used do argue conclusively for estimation methods capable of including large deviations in likelihood functions with polynomial shape and high dimensionality, however.

Further work will investigate the structure of the autocorrelation function described in section 4 as well as exploring very out of the money option data for longer time frames.

<table>
<thead>
<tr>
<th>N</th>
<th>(&lt; r &gt;)</th>
<th>(\sigma)</th>
<th>Max</th>
<th>Min</th>
<th>(\hat{\alpha}) (MLE)</th>
<th>(\hat{\alpha}) (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEX-OIL</td>
<td>6,326</td>
<td>0.0005</td>
<td>0.0124</td>
<td>0.0915</td>
<td>-0.2021</td>
<td>4.7650</td>
</tr>
<tr>
<td>LIFEE Call (Bin 1)</td>
<td>2,554</td>
<td>-0.0089</td>
<td>0.2338</td>
<td>2.6522</td>
<td>-1.0000</td>
<td>4.6216</td>
</tr>
<tr>
<td>LIFEE Call (Bin 2)</td>
<td>2,504</td>
<td>-0.0019</td>
<td>0.1973</td>
<td>5.6197</td>
<td>-1.0000</td>
<td>5.0069</td>
</tr>
<tr>
<td>LIFEE Put (Bin 1)</td>
<td>2,555</td>
<td>-0.0086</td>
<td>0.2528</td>
<td>1.6190</td>
<td>-1.0000</td>
<td>3.7606</td>
</tr>
<tr>
<td>LIFEE Put (Bin 2)</td>
<td>2,477</td>
<td>0.0035</td>
<td>0.3090</td>
<td>12.7143</td>
<td>-1.0000</td>
<td>3.7364</td>
</tr>
</tbody>
</table>

Table 1. Results of MLE Method. Source: Authors’ calculations on AMEX-OIL and LIFFE-ESX using MLE and OLS.

---

6This D statistic is the minimum distance between two distributions, in our case a randomly-generated power law distribution \(X_p\), and our real world data, which we denote \(X\). The ‘D’ statistic is then given by \(D_{X_p} = \sup_x(F_{\text{AMEX-OIL, LIFFE}}(X) - F_{X_p}(X)).\)
6. Conclusion

When financial data are characterised by fat tailed distributions, it is important to estimate these distributions in an efficient manner to obtain an accurate price of the returns which make up the distribution. This paper derives a computationally efficient method of maximum likelihood estimation and applies the technique to two high frequency, long running financial datasets, one index, AMEX-OLIL, and one options data set, LIFFE-ESX, which we partition into two bins at around the money returns. We compare our method to the standard ordinary linear regression method in table 5 and judge the MLE method superior, using Kolmogorov-Smirnov statistics.

The method described in section 2 is recursive in structure, and works symbolically (following Brynjolfsson and Smith (2000)) as well as numerically (in the spirit of Clauset et al. (2007)). We show that the data we are using are good candidates for skewed distributions in section 4 and show the results of our estimation in section 4 also.

Further work will concentrate on extending the method to out of the money option pricing using a novel binning method developed by one of the authors, as well as increasing the time frequency of each data set to check for scaling exponents within each series. We will also extend the method to real time pricing of the autocorrelation function show for 500 lags in figures 1–4.

References


**DEPARTMENT OF ECONOMICS, KEMMY BUSINESS SCHOOL, UNIVERSITY OF LIMERICK, LIMERICK, IRELAND**  
*E-mail address: stephen.kinsella@ul.ie*

**DEPARTMENT OF ACCOUNTING AND FINANCE, KEMMY BUSINESS SCHOOL, UNIVERSITY OF LIMERICK, LIMERICK, IRELAND**  
*E-mail address: fergal.g.obrien@ul.ie*