RISE | RICE INITIATIVE *for the* STUDY *of* ECONOMICS

RISE Working Paper 15-20

“The Volcano Distribution with an Application to StockMarket Returns”

by Michael Naaman and Robin Sickles

RICE
Department of Economics
Baker Hall, MS22
6100 Main Street, Houston, Texas 77005
https://economics.rice.edu

The Volcano Distribution with an Application to Stock Market Returns

Michael, Naaman
Christensen Associates

Robin Sickles
Rice University

August 1, 2015

Abstract

Power-law distributions have a wide range of applications including physics, economics, and biology. We derive a new family of densities with support on the real line which obeys a broken power law called the volcano distribution. An estimation procedure is also outlined.

The volcano density is very flexible as it can be unbounded or bimodal. It allows for an infinite mean or an undefined mean. The complexity of this distribution calls for a novel semiparametric estimation approach which we apply to stock market returns data.

1 Introduction

This paper seeks to extend the notion of the power law to a family of densities on the real line that allow for asymmetry. In statistical applications, densities on the real line tend to be symmetric about the mean or median. Asymmetry is reserved almost exclusively for densities with support on the positive or negative real numbers, but even these densities can be extended to symmetric densities on the real line. However, it is often the case that data exhibit fundamental asymmetry. It seems reasonable that stock market returns could have systematically different tail behavior for losses and gains. The distribution of wealth might behave quite differently for people with negative wealth as opposed to the very rich. We might even think that the distribution of wealth has infinite expectation which would have interesting policy implications for income inequality.

There are numerous other process in a wide variety of fields that have shown evidence of power laws. Gabaix (1999) gives many examples of processes that obey power laws in economics. One of which is that a reflected geometric Brownian motion can be described by a power law. In fact recalling that the Black-Scholes equation for a Eurpoean call option can be written as

$$\frac{dC}{dt} + \frac{\sigma^2 S^2}{2}\frac{d^2C}{dt^2} + rS\frac{dC}{ds} - rC = 0 \tag{1}$$

The solution to this PDE implies that the underlying distribution is lognormal, however, power-laws can also be a solution to (1). Ignoring the boundary conditions, an alternative solution is

$$C\left(S,t\right)=\frac{S^{-\zeta}}{\zeta}e^{-(\zeta+1)(r-\sigma^2/2)(T-t)} \tag{2}$$

Of course this solution doesn't correspond to the boundary conditions implied by a European call, but it does provide some intuition as to why power-laws might be relevant.

In the finance literature there is also some debate on whether stock market returns have infinite variance. This is sometimes discussed as a choice between Gaussian or stable distributions. While Taleb (2009) has written that this is an irrelevant issue, he does advocate using power laws to describe daily returns.

The problem with power laws is that they are unbounded at the origin, so they are restricted to a some domain which cannot include the origin.

In order to see the importance of power laws, consider a random variable, x, which follows a power law distributions if it is drawn from a probability distribution that has a density with the following property

$$p(x)\propto x^{-\alpha-1} \tag{3}$$

with $\alpha>0$ and the density will be scale invariant, which is a defining property of power law functions. In general it will not be possible to define a power law density over the entire real line. For example, if $\alpha=2$, the integral around 0 will diverge

$$\int_0^1 x^{-2}dx=+\infty \tag{4}$$

The solution to this problem is usually to restrict the integration over some $x_{\min}>0$ which will imply that x has a Pareto distribution. Axtell (2001) uses a discrete version of the Pareto distribution to determine that firm size obeys a power law. A weaker form of the power law distributions is to simply require that $p(x)$ be asymptotically scale invariant instead of scale invariant everywhere. Theses densities will have the form

$$p(x)\propto L(x)x^{\alpha-1} \tag{5}$$

$$\lim_{|x|\to+\infty}\frac{L(tx)}{L(x)}=1 \tag{6}$$

Where L is a slowly varying function which will preserve the asymptotic scale invariance. This means that the density will be approximately a power law for suitable large x. Grabchak and Samorodnitsky (2010) use this approach for a symmetric Pareto density smoothed with a Gaussian to allow for integration around 0. Of course this approach will only be scale invariant in the tails.

If we relax the assumption of global scale invariance, then the power law can be generalized so that there exists a mutually disjoint collection of sets,R_i, such that

$$p(x)\propto x^{\alpha_i-1} \tag{7}$$

For all $x \in R_I$. In general this will not be a globally scale invariant function because the power of x can vary; however, it will be scale invariant inside any of the sets that partition the real line. It is not immediately clear how this helps, but consider the following integrals

$$\int_0^1 x^{-2} dx = +\infty \tag{8}$$
$$\int_1^{+\infty} x^{-\frac{1}{2}} dx = +\infty \tag{9}$$
$$\int_0^1 x^{-\frac{1}{2}} dx = 2 \tag{10}$$
$$\int_1^{+\infty} x^{-2} dx = 1 \tag{11}$$
$$\tag{12}$$

The integrals in equations (8) and (9) diverge while the integrals equations (10) and (11) converge. This indicates that the minimum of the two functions can be integrated and we have the simple relation

$$\int_0^{+\infty} \frac{1}{x^2} \wedge \frac{1}{\sqrt{x}} dx = 3 \tag{13}$$

This leads to a density that represents a power law, but the power changes over the domain to ensure integrability. In application, power laws are often described as an effect of the tail behavior of a distribution, but it does not describe the entire distribution. This framework automatically allows for the tail behavior to be different than the middle part of the distribution, so the entire distribution can be described as a power law at least locally.

For example in stock market returns data, the volcano density is flexible enough to essentially allow for a different probability distribution when there are days with large gains or losses.

Bounded densities are used almost exclusively in application. Many proofs in mathematical statistics require the density to be bounded, but there are already unbounded densities lurking among us. The beta distribution with both parameters equal to .5 will be unbounded at 0 and 1 and it will still have nice properties such as finite mean and variance.

In this paper, a family of distributions will be introduced that is unbounded or bimodal. It can have fat-tails or heavy-tails. It can have a mean that is finite, infinite, or undefined. The density and its distribution function can be expressed analytically. Most importantly it is the natural family of random variables following a power law.

2 Model

Equation (13) can be extended to a family of densities with support on the whole real line This relationship can be exploited to define a density function. First we will restrict the density to have fixed scale, which determines the kink in the density, and location, which is where the density is unbounded. Then the location and scale parameters can allowed to vary afterwards.

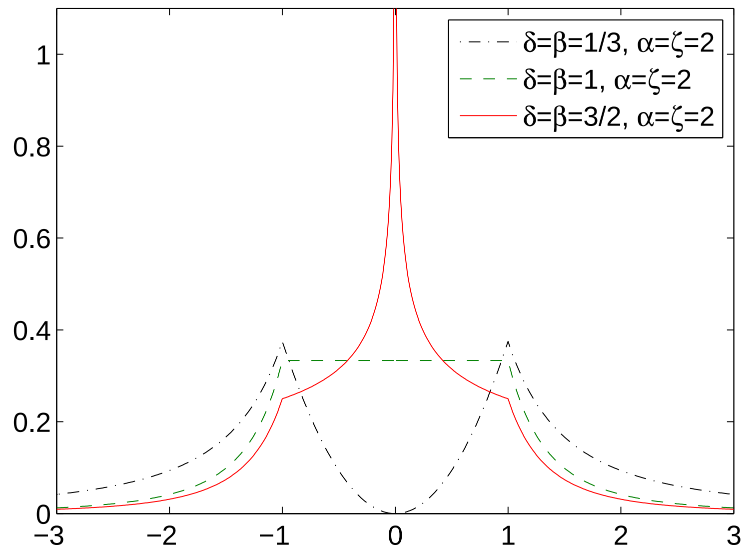

Figure 1: Volcano density for different parameters

$$v(x) = C\left\{\left[(-x)^{-\alpha-1}\wedge(-x)^{\frac{1}{\beta}-1}\right]H(-x)+\left[(x)^{-\zeta-1}\wedge(x)^{\frac{1}{\delta}-1}\right]H(x)\right\} \tag{14}$$

where $C = \left(\frac{1}{\alpha}+\beta+\delta+\frac{1}{\zeta}\right)^{-1}$ and $\min(\alpha,\beta,\delta,\zeta) > 0$. The density will be unbounded whenever $\min(\delta,\beta) > 1$; and it will be bimodal whenever $\min(\delta,\beta) < 1$. If $\min(\delta,\beta) = 1$, then the density will be flat on one or both sides of the 0. This fact is the impetus for the name of the distribution because as $\min(\delta,\beta)$ increases past unity the function has the appearance of an exploding volcano as can be seen in Figure 1.

In order to work with this density, it will be helpful to use the Heavyside step function. In some applications, the value of the function at the origin is defined differently or even left undefined, so to avoid confusion we define the Heavyside step function as

$$H(x) \equiv \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases}$$

which is the integrand of the Dirac delta function. Iverson brackets will also be used: for example,

$$[x > 0] = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

The continuity of the volcano density, except at the origin, justifies using the Iverson brackets interchangeably with the Heavyside step function.

This density is more difficult to work with than more standard densities that are bounded and unimodal. It could even be the case that there is a bounded density that is arbitrarily close to the volcano density. Any such bounded density would obviously be preferred, but there can be no such approximation as the following demonstrates.

Theorem 2.1. *There are no continuous or bounded functions that are equal to the standard volcano density almost everywhere.*

Proof. Suppose $g(x)$ is a continuous or bounded function that is equal to $v(x)$ a.e., furthermore, let $\delta = 2$, which means $v(x)$ is unbounded in a neighborhood of the origin. Clearly a continuous function must be bounded except maybe on a set of measure zero. Without loss of generality we can assume that the essential supremum of $g(x)$ is $M > 0$. Since $v(x)$ is decreasing whenever x is nonnegative, then $v(x) \geq M$ whenever $x \in \left[0, \frac{C^2}{M^2}\right]$. Since $g(x)$ is essentially bounded by M, it must be the case that

$$\int_0^{\frac{C^2}{M^2}} g - M dx \leq 0$$

However, if we integrate $v(x)$ over the same set of positive measure, we find

$$\int_0^{\frac{C^2}{M^2}} Cx^{-.5} - M dx = \left[2C\sqrt{\frac{C^2}{M^2}} - \frac{C^2}{M}\right] = \frac{C^2}{M} > 0$$

which contradicts the assumption that $g(x)$ is equal to $v(x)$ a.e. □

While this function is unbounded, it is still an integrable function so it has an absolutely continuous distribution function and we may conclude that the density represents an absolutely continuous random variable.

Lemma 2.2. *The standard volcano density can be rewritten as*

$$\begin{aligned} v(x) = C\left\{(-x)^{-\alpha-1}H\left(-x-1\right) + (-x)^{\frac{1}{\beta}-1}H(-x)H(x+1)\right\} + \\ C\left\{x^{\frac{1}{\delta}-1}H(x)H(1-x) + x^{-\zeta-1}H(x-1)\right\} \end{aligned} \quad (15)$$

Proof. Suppose that for nonnegative x, the inequality $x^{\frac{1}{\delta}-1} \leq x^{-\zeta-1}$ is satisfied. If we apply the natural logarithm to both sides of the inequality and rearrange, we have

$$\left(\frac{1}{\delta} + \zeta\right) \ln(x) \leq 0 \quad (16)$$

which is satisfied precisely when x is not greater than one. Similarly $x^{\frac{1}{\delta}-1} > x^{-\zeta-1}$ whenever $x > 1$. A similar argument will finish the result for negative x values.

□

Theorem 2.3. *The distribution function of the volcano density will be given by*

$$\begin{aligned} \Pr\left(x \leq t\right) = C\left\{\frac{\left[-\left(\mathrm{t} \wedge -1\right)\right]^{-\alpha}}{\alpha} + \beta\left[1 - \left[-\left(t \wedge 0\right)\right]^{\frac{1}{\beta}}\right] H\left(t+1\right)\right\} + \\ C\left\{\delta[t \wedge 1]^{\frac{1}{\delta}} H\left(t\right) + \frac{\left(1 - t^{-\zeta}\right)}{\zeta} H\left(t-1\right)\right\} \end{aligned} \quad (17)$$

Proof.

$$\int_{-\infty}^{t} v(x)\,dx = C\left\{\int_{-\infty}^{t\wedge -1} (-x)^{-\alpha-1}\, dx + H(t+1)\int_{-1}^{t\wedge 0} (-x)^{\frac{1}{\beta}-1} dx\right\} +$$

$$C\left\{H(t)\int_{0}^{t\wedge 1} (x)^{1/\delta-1}\, dx + \int_{1}^{t\vee 1} (x)^{-\zeta-1} dx\right\}$$

$$= C\left\{\frac{[-(\mathrm{t}\wedge -1)]^{-\alpha}}{\alpha} + \beta\left[1-[-(t\wedge 0)]^{\frac{1}{\beta}}\right] H(t+1)\right\} +$$

$$C\left\{\delta[t\wedge 1]^{\frac{1}{\delta}} H(t) + \frac{(1-t^{-\zeta}}{\zeta} H(t-1)\right\}$$

□

It is a relatively simple exercise to take the derivative of the distribution function to get back to the original density function. If location and scale parameters are added, the volcano density is given by

$$f(x) = \frac{C}{\sigma}\left\{\left(\frac{\mu-x}{\sigma}\right)^{-\alpha-1}[x\le\mu-\sigma] + \left(\frac{\mu-x}{\sigma}\right)^{\frac{1}{\beta}-1}[\mu-\sigma<x\le\mu]\right\} +$$

$$\frac{C}{\sigma}\left\{\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\delta}-1}[\mu<x\le\mu+\sigma] + \left(\frac{x-\mu}{\sigma}\right)^{-\zeta-1}[x>\mu+\sigma]\right\}$$

where f is now our volcano density with location parameter, μ, and scale parameter, σ. Of course this implies the standard properties of location scale families: namely, If $z \sim V(0,1,\alpha,\beta,\delta,\zeta)$, then $\sigma z + u \sim V(\mu,\sigma,\alpha,\beta,\delta,\zeta)$.

It is important to verify that this family of distributions is identified by which we mean for any $\theta \neq \theta_0$

$$v(x|\theta) \neq v(x|\theta_0)$$

where $\theta = (\alpha,\beta,\delta,\zeta,\mu,\sigma)$. The location parameter,μ, is identified by the spike or valley in the density and the scaling parameter, σ, changes the height and location of the kinks in the density. The other 4 parameters uniquely determine the shape of the density in each partition of the density.

Despite the complexity of this distribution the moments are easy to derive. It will be easier to compute the integrals if we integrate after standardizing by the location and scale parameters, $z = \left(\frac{x-\mu}{\sigma}\right)$, which leads to equation (18)

$$\begin{aligned} E\left[\left(\frac{x-\mu}{\sigma}\right)^{m}\right] &= C\left[\frac{\delta}{1+m\delta} + \frac{\beta(-1)^m}{1+m\beta} + \frac{1}{\zeta-m} + \frac{(-1)^m}{\alpha-m}\right] \\ &= CA_m \end{aligned} \tag{18}$$

Equation (18) is not based on central moments. In fact our location parameter is neither the mean nor the median nor the mode except under special circumstances, but it is the relevant measure of central tendency. However, in practice the location parameter is usually near the median.

The scale parameter will not be equal to the variance either, but it can still be interpreted as a measure of spread. As the variance increases the density becomes wider which is a property that the scale parameter also has, so even when the variance is infinite there is still a meaningful measure of the spread of the density. It also marks a fundamental change in the nature of the distribution as even in the fully symmetric case the density will have a change in slope at a location determined by the scale parameter.

One might consider using equation (18) to derive the moments of the volcano distribution. For example if the mean of the volcano distribution is finite, then we might consider a moment generating function of the type in equation (19).

$$\tilde{M}\left(t\right) = C \sum_{m=0}^{\lceil \alpha \wedge \zeta \rceil - 1} \frac{(t)^m}{m!} \left[\frac{\delta}{1+m\delta} + \frac{\beta(-1)^m}{1+m\beta} + \frac{1}{\zeta - m} + \frac{(-1)^m}{\alpha - m} \right] \tag{19}$$

Obviously equation (19) is insufficient as a moment generating function because it is possible to have an infinite mean for certain sets of parameters, but the moment generating can be considered as limit of (19).

Theorem 2.4. *The characteristic function of the volcano distribution is given by*

$$\begin{aligned} \phi\left(t\right) = e^{\mu i t} C \left[(\sigma i t)^{\alpha} \Gamma\left(-\alpha, \sigma i t\right) + (\alpha i t)^{-\frac{1}{\beta}} \gamma \left(\frac{1}{\beta}, \sigma i t \right) \right] + \\ e^{\mu i t} C \left[(-\sigma i t)^{-\frac{1}{\delta}} \gamma \left(\frac{1}{\delta}, -\sigma i t \right) + (-\sigma i t)^{\zeta} \Gamma\left(-\zeta, -\sigma i t\right) \right] \end{aligned} \tag{20}$$

Proof. Suppose we partition the sample space with $R_1 = [\frac{x-\mu}{\sigma} \leq -1]$,...,$R_4 = [\frac{x-\mu}{\sigma} \geq -1]$, then the law of total expectation can be applied

$$\phi\left(t\right) = \sum_{i=1}^{4} E\left[e^{itx} \mid R_i\right] P(R_i) = \sum_{i=1}^{4} \phi_i\left(t\right) P(R_i)$$

ϕ_1 and ϕ_4 will correspond to the characteristic functions of a Pareto distribution. It's also clear that ϕ_2 and ϕ_3 will have similar characteristic functions except the upper incomplete gamma function is replaced with the lower incomplete gamma function. □

The first thing to notice about this characteristic function is that it is split into partitions just like the density. There is a split at 0 and then the gamma function has been split into its lower and upper components. The characteristic function really shows the Pareto like nature of this distribution. The left and right tail are Pareto distributed. The two middle parts of the distribution are an inverted Pareto distributions that have the property of a Pareto density that is rotated up and scaled to the center to give the volcano like appearance. This is where one can see why this is the natural family for power law distributions as everything seems to just fit together nicely.

The moment generating function can easily by recovered from the characteristic function due to the identity, $\phi(t) = M(-it)$ which is given by

$$M(t) = e^{\mu t} C \left[(\sigma t)^{\alpha} \Gamma(-\alpha, \sigma t) + (\alpha t)^{-\frac{1}{\beta}} \gamma\left(\frac{1}{\beta}, \sigma t\right) \right] [t \geq 0]$$
$$+ e^{\mu t} C \left[(-\sigma t)^{-\frac{1}{\delta}} \gamma\left(\frac{1}{\delta}, -\sigma t\right) + (-\sigma t)^{\zeta} \Gamma(-\zeta, -\sigma t) \right] [t \leq 0] \qquad (21)$$

Unfortunately the function is somewhat erratic around the origin as its behavior is shown below.

$$\lim_{t \to 0^+} M(t) = C\left(\frac{1}{\alpha} + \beta\right)$$
$$\lim_{t \to 0^-} M(t) = C\left(\frac{1}{\zeta} + \delta\right)$$
$$M(0) = \lim_{t \downarrow 0} M(t) \;\; [t \geq 0] + \lim_{t \uparrow 0} M(t) \;\; [t \leq 0] = 1$$

The moment generating function will not be continuous at the origin and hence not differentiable. The moment generating function derived from (21) is consistent with (19) as can be seen from the power series below.

$$t^{-\frac{1}{\beta}} \gamma\left(\frac{1}{\beta}, t\right) = \sum_{m=0}^{+\infty} \frac{(t)^m}{m!} \frac{\beta(-1)^m}{1 + m\beta} \qquad (22)$$

However this issue will not be the case for the characteristic function. This is similar to the Pareto distribution which also has a discontinuity in the moment generating function, but a differentiable characteristic function.

For two middle parts of the distribution, the derivative of the characteristic function will have the following form

$$\lim_{x \to 0} \frac{d\left[x^{-s} \gamma(s, x)\right]}{dx} = \frac{-1}{s+1} \qquad (23)$$

For our purposes equation (23) would be used to evaluate ϕ_2 and ϕ_3. Since $\Gamma(s, x) = \Gamma(s) - \gamma(s, x)$, then it follows for the tails of the distribution

$$\frac{d\left[x^s \Gamma(-s, x)\right]}{dx} = s x^{s-1} \Gamma(-s) - \frac{d\left[x^s \gamma(-s, x)\right]}{dx} \qquad (24)$$

The first term of the sum is negligible as long as $s > 1$ which corresponds to the case of finite expectation; However if $s < 1$,then the first term dominates and diverges to infinity. The rate of convergence will be given by x^{s-1} because the gamma function will be finite for non-integer values which means that the smaller of α and ζ will determine whether the expectation is negative or positive infinity.

However this is not just a mathematical technicality; we took the sample mean of 100000 observations drawn from a $V(0, 1, 1, 1, 1, .5)$ and the minimum of a 1000 different sample means was on the order of 10^{17} implying that the sample mean really is diverging to positive infinity due to the heavier right tail. The only remaining case to consider is the symmetric case where $\alpha = \zeta \leq 1$.

Unfortunately the order of convergence is the same for both parameters and, at least in simulations, it appears that the sample mean converges to a distribution function of a non-degenerate random variable, so we interpret this behavior as undefined expectation. The results on the first moment are compiled in Equation 25.

$$E\left[x\right]=\begin{cases}\mu+\sigma CA_1 & min\left(\alpha,\zeta\right)>1\\ +\infty & min\left(\alpha,\zeta\right)\leq 1,\zeta<\alpha\\ -\infty & min\left(\alpha,\zeta\right)\leq 1,\zeta>\alpha\\ undefined & \zeta=\alpha\leq 1\end{cases} \tag{25}$$

Equation (19) can be used to compute the first two moments which can be used to find the variance which is computed after standardizing by the shadow parameters

$$Var\left(x\right)=\begin{cases}\sigma^2C\left(A_2-\ {A_1}^2\right) & min\left(\alpha,\zeta\right)>2\\ +\infty & min\left(\alpha,\zeta\right)\leq 2\end{cases} \tag{26}$$

3 Estimation

Estimating the parameters of this density is nontrivial, but the volcano density can be estimated by nonparametric methods. Since the density is not differentiable, it is hardly guaranteed that MLE approaches that rely on differentiation can be used. In order to estimate the parameters it is necessary to exploit specific features of the volcano density with nonparametric methods.

The first step will be to generating random samples from the volcano density which can be a bit tricky. Each partition has Pareto like properties, so a vector, X, of N random draws from a standard volcano distribution will have the form

$$X=\left(U_1^{-\frac{1}{\alpha}},U_2^{\beta},U_3^{\delta},U_4^{-\frac{1}{\zeta}}\right)$$

Where U_i is a vector of uniformly distributed random variable with the length of each vector given by the probability of being in a partition multiplied by the total number of draws. This way each random variable has its own distribution, but there will be a different number of observations in each partition. However, it may be the case that for a given N, one may not be able to evenly divide N for each partition. If this is the case, then we add an extra observation for that partition. Suppose for example that all partitions require a correction, then the probability of being in the first partition is

$$\frac{\lfloor N\frac{C}{\alpha}\rfloor+1}{N+4}=\frac{C}{\alpha}+\frac{\lceil N\frac{C}{\alpha}\rceil-\lfloor N\frac{C}{\alpha}\rfloor-4\frac{C}{\alpha}}{N+4}=\frac{C}{\alpha}+O\left(\frac{1}{N}\right)$$

This will produce a sample size with four extra observations, but these extra observations can be randomly removed.

Once random draws can be taken from the volcano distribution, it will be necessary to use a nonparametric kernel density to estimate the parameters in the model.

For any continuity point of the volcano density, the kernel density estimate will be consistent [1], so the only point of concern would be at the origin.

[1] This follows from Pagan and Ullah (1999), pg. 362.

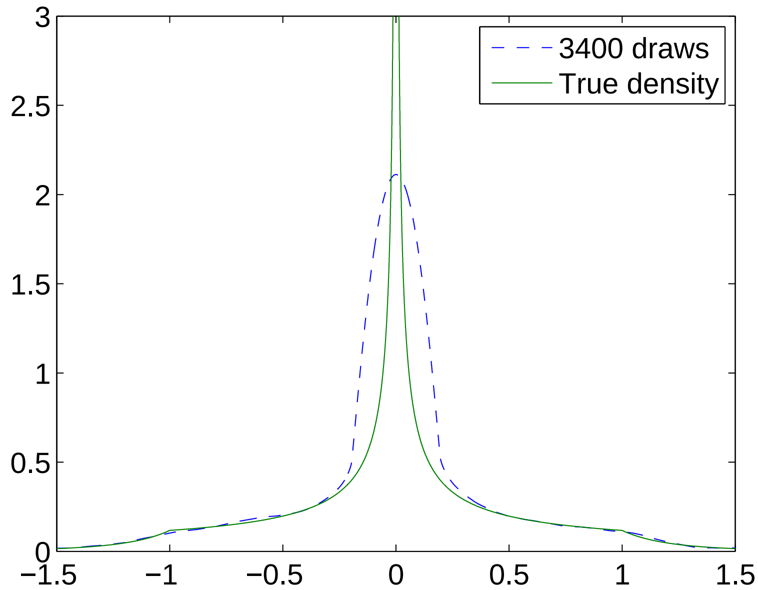

Figure 2: Kernel density estimate of volcano distribution with $\alpha = \beta = \delta = \zeta = 4$

Lemma 3.1. *Suppose K is the Bartlett kernel and $\hat{v}$ is the corresponding kernel density estimate for a standard volcano distribution with bandwidth, $h \propto n^{-\lambda}$ with $0 < \lambda \leq .5$, then the kernel density estimate is asymptotically unbiased and consistent at the origin.*

Proof. For the first part, the expectation of the kernel density estimate is

$$E\left[\hat{v}(0)\right] = \int_{-\infty}^{+\infty} K(\psi)v(h\psi)d\psi$$
$$= \frac{3Ch^{\frac{1}{\delta}-1}}{4}\int_0^1 \left(1-\psi^2\right)\psi^{\frac{1}{\delta}-1}d\psi + \frac{3Ch^{\frac{1}{\beta}-1}}{4}\int_{-1}^0 \left(1-\psi^2\right)\psi^{\frac{1}{\beta}-1}d\psi$$
$$= \frac{3Ch^{\frac{1}{\delta}-1}}{4}\left(\frac{2\delta^2}{1+2\delta}\right) + \frac{3Ch^{\frac{1}{\beta}-1}}{4}\left(\frac{2\beta^2}{1+2\beta}\right)$$

where $\psi = x/h$. Since $\lim_{n\to\infty} h = 0$ and either δ or β is greater than unity in the unbounded case, then we must have

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty} K(\psi)v(h\psi)d\psi = +\infty = v(0)$$

it will converge to 0 if both δ and β is less than unity, so the estimator is asymptotically unbiased. Even though the expectation may be infinite at the origin, it will be finite in all finite samples, so the variance may converge to 0. Since that data is i.i.d, the variance of the estimator is given by

$$V\left(\hat{v}(0)\right) = \frac{1}{nh}\int_{-\infty}^{+\infty} K(\psi)^2 v(h\psi)d\psi - \frac{\left(E\left[\hat{v}(0)\right]\right)^2}{n}$$

The second term will converge to zero because

$$\frac{\left(E\left[\hat{v}(0)\right]\right)^2}{n} \propto \frac{h^{2/\delta-2} \vee h^{2/\beta-2}}{n}$$
$$\propto n^{-2\lambda/\delta+2\lambda-1} \vee n^{-2/\lambda+2\alpha-1} \to 0$$

because $\lambda <= .5$. Similarly for the second term,

$$\frac{1}{nh}\int_{-\infty}^{+\infty} K(\psi)^2 v(h\psi)d\psi \propto \frac{h^{1/\delta-2} \vee h^{1/\beta-2}}{n}$$
$$\propto n^{-\lambda/\delta+2\lambda-1} \vee n^{-2/\lambda+2\lambda-1} \to 0$$

Since the estimator is asymptotically unbiased with variance going to 0, then consistency follows.

□

Using the Bartlet kernel and bandwidth equal to $n^{-1/5}$,the density estimate of the volcano distribution is shown in Figure 2[2]. The estimate seems to be a good fit except in a neighborhood of 0 which makes sense due to the spike at 0.

The volcano density will fail to have a continuous derivative at $x = \pm 1$ because of the kink in the density. The kink seems to pose little difficulty for the kernel density estimate, but it will still not fit well in a small neighborhood about $x = \pm 1$ even with a relatively large number of observations. This is not to say that estimation can't be done with fewer observations, but data from a volcano distribution may not exhibit the kinks at $\mu + \sigma$ and $\mu - \sigma$ on the horizontal axis in finite samples.

Figure 3 shows the nature of the convergence for the nonparametric estimate of the density around 0. It does seem to be the case that the peak of the estimate is approaching infinity and the neighborhood of poor fit is getting smaller, albeit very slowly.

Instead of using the optimally smoothed bandwidth of $n^{-1/5}$, we will use an under-smoothed kernel density estimate in order to reduce the finite sample bias of the kernel density estimate. Horowitz (2001) recommends using $h \propto n^{-1/3}$ in order to maximize the coverage probability of bootstrapped confidence intervals which we will be utilizing.

In Figure 4, one can see that the fit is improved around zero by using the under-smoothed estimate and it necessary to have a decent fit around 0 for estimation purposes. Even with an under-smoothed density estimate, it is noticeably different around the singularity even with 34000 observations.

Since it is not obvious that the regularity assumptions of maximum likelihood will hold, we will demonstrate the consistency of the estimators directly. We will demonstrate that the MLE estimators, conditional on the location ans scale parameters, are consistent but biased,so an analogue estimator is also presented that is consistent and conditionally unbiased which is our preferred approach.

[2]The choice of 3400 observations was made to ensure there would be an even number of observations in each of the four partitions, so there would be no need for the correction previously described

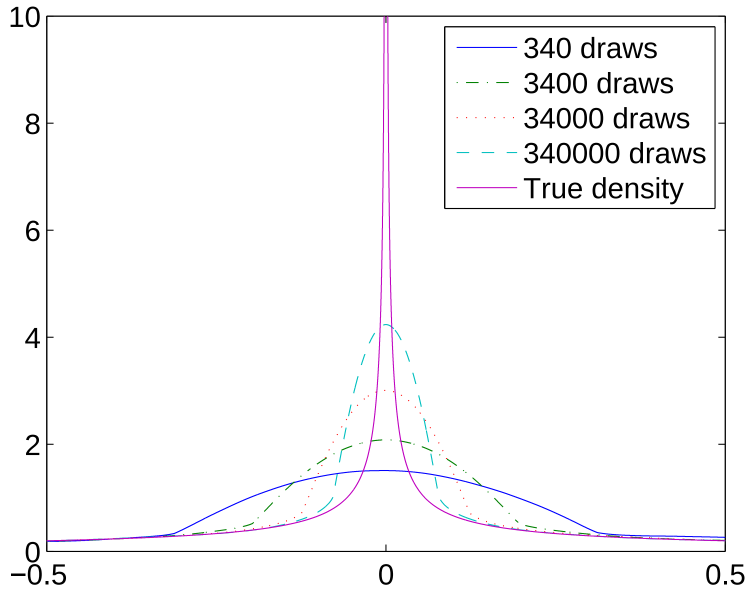

Figure 3: Kernel density estimate of volcano distribution with $\alpha = \beta = \delta = \zeta = 4$

Suppose consistent estimates of μ and σ were in hand, the data could be standardized to allow for the concentrated log-likelihood to be written as

$$L = \sum_i -(\alpha+1)\ln(-z_i)\,[z_i \leq -1] + \left(\frac{1}{\beta} - 1\right)\ln(-z_i)\,[-1 < z_i \leq 0] + \left(\frac{1}{\delta} - 1\right)\ln(z_i)\ \ [0 < z_i \leq 1] - (\zeta+1)\ln(z_i)\,[z_i > 1] + \ln C$$

This might seem like an incorrect application of the logarithm, but one must remember that the logarithm only applies in each of the four disjoint regions of the sample space. For example if the function is analytic then all of the cross terms are zero due to the partition, so the function will be applied on each partition. The first order conditions (FOC) can be rearranged to provide the following estimating equations.

$$A^2 = \frac{1}{n}\sum_i \ln(-z_i)\,[z_i \leq -1] = \frac{C}{\alpha^2} \tag{27}$$

$$B^2 = \frac{1}{n}\sum_i -\ln(-z_i)\,[-1 < z_i < 0] = C\beta^2 \tag{28}$$

$$D^2 = \frac{1}{n}\sum_i -\ln(z_i)\ \ [0 < z_i \leq 1] = C\delta^2 \tag{29}$$

$$Z^2 = \frac{1}{n}\sum_i \ln(z_i)\ \ [1 < z_i] = \frac{C}{\zeta^2} \tag{30}$$

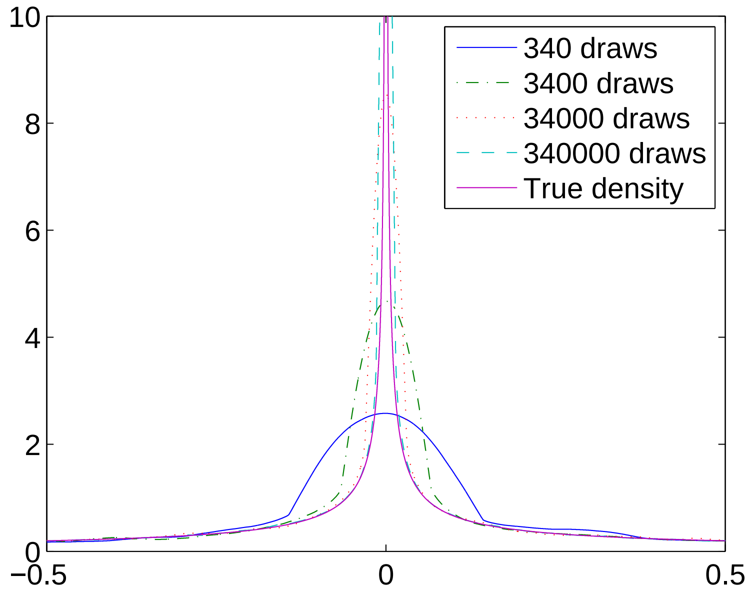

Figure 4: Kernel density estimate of volcano distribution with $\alpha = \beta = \delta = \zeta = 4$

The solution to the FOC is given by

$$\widetilde{\alpha}^{-1} = (A + B + D + Z)A \tag{31}$$
$$\widetilde{\beta} = (A + B + D + Z)B \tag{32}$$
$$\widetilde{\delta} = (A + B + D + Z)D \tag{33}$$
$$\widetilde{\zeta}^{-1} = (A + B + D + Z)Z \tag{34}$$

which is the maximum likelihood estimator conditional on the location and scale parameters.

Our preferred estimators conditional on the location and scale parameters are given by

$$\hat{\alpha}^{-1} = \frac{\frac{1}{n}\sum_i \ln(-z_i) \;\; [z_i \leq -1]}{\frac{1}{n}\sum_i [z_i \leq -1]} \tag{35}$$
$$\hat{\beta} = \frac{\frac{1}{n}\sum_i - \ln(-z_i) \;\; [-1 < z_i \leq 0]}{\frac{1}{n}\sum_i [-1 < z_i \leq 0]} \tag{36}$$
$$\hat{\delta} = \frac{\frac{1}{n}\sum_i - \ln(z_i) \;\; [0 < z_i \leq 1]}{\frac{1}{n}\sum_i [0 < z_i \leq 1]} \tag{37}$$
$$\hat{\zeta}^{-1} = \frac{\frac{1}{n}\sum_i \ln(z_i) \;\; [1 < z_i]}{\frac{1}{n}\sum_i [1 < z_i]} \tag{38}$$

Even though this is not conditional MLE,these estimators are the MLE conditional on being in any of the four regions. For example the estimator in equation (38) is the MLE of a Pareto distribution.

Lemma 3.2. *The conditional estimators defined by equations (35)-(38) are consistent and unbiased.*

Proof. It is enough to show $\hat{\zeta}^{-1}$ is consistent as the other parameters are similar. First note that

$$\Pr\left(z \le t | z > 1\right) = \frac{\Pr(z \le t, z > 1)}{\Pr(z > 1)} = \zeta \int_1^t z^{-\zeta-1} dz$$

So the expectation of the estimator is

$$E\left[\ln\left(z\right)|z > 1\right] = \zeta \int_{+\infty}^{1} \ln\left(z\right) z^{-\zeta-1}\ dz$$
$$= \zeta \left[-\ln\left(z\right)\frac{z^{-\zeta}}{\zeta}\right]_1^\infty + \zeta \int_1^\infty \frac{z^{-\zeta-1}}{\zeta} dz = \frac{1}{\zeta}$$

It is sufficient to verify that the variance of the estimate is finite, so

$$E\left[\left(\ln\left(z\right)\right)^2|z > 1\right] = \zeta \int_{+\infty}^{1} \left(\ln\left(z\right)\right)^2 z^{-\zeta-1}\ dz$$
$$= \zeta \left[-\left(\ln\left(z\right)\right)^2 \frac{z^{-\zeta}}{\zeta}\right]_1^\infty + 2\zeta \int_1^\infty \frac{z^{-\zeta-1}}{\zeta} \ln(z) dz = \frac{2}{\zeta^2}$$

so Chebyshev's inequality can be applied and we have

$$\hat{\zeta}^{-1} = \frac{\sum_i \ln\left(z_i\right)\ \ [1 < z_i]}{\sum_i [1 < z_i]} \to_p \zeta^{-1}$$

□

However it follows from Jensesn's inequality that the conditional MLE estimators will be biased. Of course even the conditional estimates are not feasible unless they remain consistent after the location and scale parameters are replaced with estimates.

In general unbiasedness will not be preserved. The four different regions are being estimated, so at least two of the estimates will have some finite sample bias caused by observations from the "wrong" region being included in the estimate; however, consistency will hold.

Theorem 3.3. *Suppose consistent estimates of μ and σ are available, then conditional estimators defined by equations (35)-(38) are consistent if, for each i, z_i is replaced by it's estimate $\hat{z}_i$.*

Proof. It will be helpful to rewrite the estimated standardizations as

$$\hat{z}_i = \frac{x_i - \hat{\mu}}{\hat{\sigma}} = \left(\frac{\sigma}{\hat{\sigma}}\right)\left(z_i - \frac{\mu - \hat{\mu}}{\sigma}\right)$$

It follows from the consistency of $\hat{\mu}$ and $\hat{\sigma}$ that $\hat{z}_i \to_p z_i$ Again we will demonstrate the consistency of $\hat{\zeta}^{-1}$ as the other estimators are similar. First let's prove the consistency of the denominator

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[\hat{z}_i \geq 1\right] - \frac{C}{\zeta}\right| \leq \left|\Pr\left(z \geq \frac{\hat{\sigma}}{\sigma} - \frac{\mu-\hat{\mu}}{\sigma}\right) - \Pr\left(z \geq 1\right)\right|$$
$$+\left|\frac{1}{n}\sum_{i=1}^{n}\left[z_i \geq \frac{\hat{\sigma}}{\sigma} - \frac{\mu-\hat{\mu}}{\sigma}\right] - \Pr\left(z \geq \frac{\hat{\sigma}}{\sigma} - \frac{\mu-\hat{\mu}}{\sigma}\right)\right|$$
$$\leq \left|\Pr\left(z \leq 1\right) - \Pr\left(z \leq \frac{\hat{\sigma}}{\sigma} - \frac{\mu-\hat{\mu}}{\sigma}\right)\right| + \sup_t\left|\frac{1}{n}\sum_{i=1}^{n}\left[z_i \leq t\right] - \Pr\left(z \leq t\right)\right|$$

Since the probability function is continuous, the first term becomes negligible. The second term after the first inequality is a measure of the distance between the empirical cdf and the actual cdf conditional on the estimates, but this may be bounded be the maximum distance which also becomes arbitrarily small.

Next we need to demonstrate the consistency of the numerator

$$\left|\frac{1}{n}\sum_{i=1}^{n}\ln\left(|\hat{z}_i|\right)\left[\hat{z}_i \geq 1\right] - \frac{C}{\zeta^2}\right| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(|\hat{z}_i|\right) - \ln\left(|z_i|\right)\right|\left[\hat{z}_i \geq 1\right] +$$
$$\frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(|z_i|\right)\right|\left|\left[\hat{z}_i \geq 1\right] - \left[z_i \geq 1\right]\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\ln\left(|z_i|\right)\left[z_i \geq 1\right] - \frac{C}{\zeta^2}\right|$$
$$\leq 2\left(\frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(|z_i|\right)\right|\right)\sup_t\left|\frac{1}{n}\sum_{i=1}^{n}\left[z_i \leq t\right] - \Pr\left(z \leq t\right)\right|$$
$$+\left|\frac{1}{n}\sum_{i=1}^{n}\ln\left(|z_i|\right)\left[z_i \geq 1\right] - \frac{C}{\zeta^2}\right| + \frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(|\hat{z}_i|\right) - \ln\left(|z_i|\right)\right|\left[\hat{z}_i \geq 1\right]$$

The first term is negligible because the first two moments of the of the natural logarithm of a volcano random variate is finite. As for the second term, consistency of the conditional estimator follows from the previous lemma. As for the third term, we may apply the continuous mapping theorem because the logarithm of the absolute value is discontinuous only at 0, a set of measure zero, so it follows $\left|\ln\left(|\hat{z}_i|\right) - \ln\left(|z_i|\right)\right| \to_p 0$ for each i which means the average also converges to zero.

□

Theorem 3.4. *The MLE estimators defined by equations (31)-(34) are consistent.*

Proof. It is enough to show $\hat{\zeta}^{-1}$ is consistent as the other parameters are similar.

$$Z^2 = \hat{\zeta}^{-1}\left(\frac{1}{n}\sum_i\left[\hat{z}_i > 1\right]\right)$$

Since Z^2 is simply the numerator of $\hat{\zeta}^{-1}$, then by the continuous mapping theorem it follows that

$$\sqrt{Z^2} \to_p \frac{\sqrt{C}}{\zeta}$$

The same argument holds for A, B, and D, so again by the continuous mapping theorem.

$$\hat{\zeta}^{-1} = (A + B + D + Z)\, Z \to_p \sqrt{C}\left(\frac{1}{\alpha} + \beta + \delta + \frac{1}{\zeta}\right)\frac{\sqrt{C}}{\zeta} = \frac{1}{\zeta}$$

By a similar arguments, the estimators defined by equations (31)-(34) are all consistent. □

All of these estimators are distributed normally even when the mean of the volcano distribution doesn't exist or is infinite. The asymptotic variance could be worked out in principle, but it will be much simpler to use the bootstrap for inference.

Since the location parameter and variance of the volcano distribution do not correspond to standard notions of mean and variance, the volcano distribution is of little use in empirical applications unless the location and scale parameters can be estimated.

If the density is unimodal, then the location parameter could be based on the mode of the estimated density.

$$\hat{\mu} = \text{argmax}\left[\ln \hat{f}(x)\right] \tag{39}$$

where $\hat{f}(x)$ is a nonparametric density evaluated at x and the domain of $\hat{f}$ is restricted to the observed data. The reason for such a restriction is because an arbitrary search grid may not be able to identify irrational parameters, but as the sample size grows there will be observed data points that are arbitrarily close to μ and $\mu \pm \sigma$. While it seems unlikely, equation (39) may return a set instead of a point in finite samples. In this case I would suggest taking the median of the observations in the set.

Theorem 3.5. *Suppose the volcano density is unimodal and the assumptions of lemma 3.1 are satisfied, then the estimator defined by equation (39) is consistent.*

Proof. it is clear that $\hat{\mu} = \text{argmax}\left[\ln \hat{f}(x)\right] = \text{argmax}\left[\hat{f}(x)\right]$, so we may work directly with the kernel density estimate. It follows from lemma 3.1 that $\hat{f}(\mu) \xrightarrow{p} f(\mu)$, but the following also holds

$$\hat{f}(\mu) \leq \hat{f}(\hat{\mu}) \leq f(\mu)$$

so it must also be the case that $\hat{f}(\hat{\mu}) \xrightarrow{p} f(\mu)$. It also follows from lemma 3.1 that the kernel density estimate is consistent for all x which means $\hat{\mu} \xrightarrow{p} z$. By assumption the volcano density is unimodal, so we must have $z = \mu$ □

While this is certainly a non-standard estimator, it is still found by maximizing an empirical log likelihood function.

In the bimodal case, the density will be 0 at the location parameter. It seems reasonable to use this property for estimation as well.

$$\hat{\mu} = \underset{|x|<C}{\operatorname{argmax}} \left[-\ln \hat{f}(x) \right] \tag{40}$$

where $C > 0$ determines the interval to search over. In the unimodal case, the domain was restricted to the observed data points, but in the bimodal case the domain must be restricted further to avoid the effect of the tails of the density.

Theorem 3.6. *Suppose the volcano density is bimodal and the assumptions of lemma 3.1 are satisfied and $C \leq \sigma$ then for the estimator defined by equation (40) is consistent.*

Proof. it is clear that $\hat{\mu} = \operatorname{argmax}\left[-\ln \hat{f}(x)\right] = \operatorname{argmin}\left[\hat{f}(x)\right]$, so we may work directly with the kernel density estimate. It follows from lemma 3.1 that $\hat{f}(\mu) \xrightarrow{p} 0$, but the following also holds as long as $C \leq \sigma$

$$0 \leq \hat{f}(\hat{\mu}) \leq \hat{f}(\mu)$$

so it must also be the case that $\hat{f}(\hat{\mu}) \xrightarrow{p} 0$. It also follows from lemma 3.1 that the kernel density estimate is consistent for all x which means $\hat{\mu} \xrightarrow{p} z$. If $C \leq \sigma$, then the kernel density will be unimodal over the relevant domain, so we must have $z = \mu$ □

Finally when the density is flat there will be an infinite number of modes. Previously we took advantage of the peakedness of the density at μ, so now we must take advantage of the flatness. Since the density is increasing at an increasing rate whenever $x < \mu - \sigma$ and decreasing at an increasing rate whenever $x > \mu + \sigma$, then the derivative of the density will reach its maximum at $x = \mu - \sigma$ and it's minimum at $x = \mu + \sigma$ which implies the following estimators.

$$\hat{\mu} = .5\left(\operatorname{argmax}\left[\hat{f}'(x)\right] + \operatorname{argmax}\left[-\hat{f}'(x)\right]\right) \tag{41}$$

The conistency of this estimator follows along similar lines as the other two estimators of the location parameter.

It may be obvious which one of the estimators in (39)-(41) to use by looking at a plot of the kernel density estimate. However, one may pretest the density for the number of modes, see Silverman (1986), and use this test to determine which of the three estimators to use. In the bimodal case, the domain of the minimization problem must be suitably close to the location parameter. If one were to use the approach of Silverman (1986), then C could be chosen to be one half of the range of the two modes.

The most difficult parameter to estimate is the scale parameter, σ, which represents the distance from the location parameter that corresponds to a change in the shape of the density. In the unimodal case, the points $\mu + \sigma$ and $\mu - \sigma$ correspond to the coordinates of largest circumscribed rectangle under the density.

From Figure 5 it is clear that for any $\epsilon > 0$, $|f(\mu + \sigma + \epsilon) - f(\mu + \sigma)| < \epsilon$ and $|f(\mu - \sigma - \epsilon) - f(\mu - \sigma)| < \epsilon$ which means that the points $\mu + \sigma$ and $\mu - \sigma$ correspond

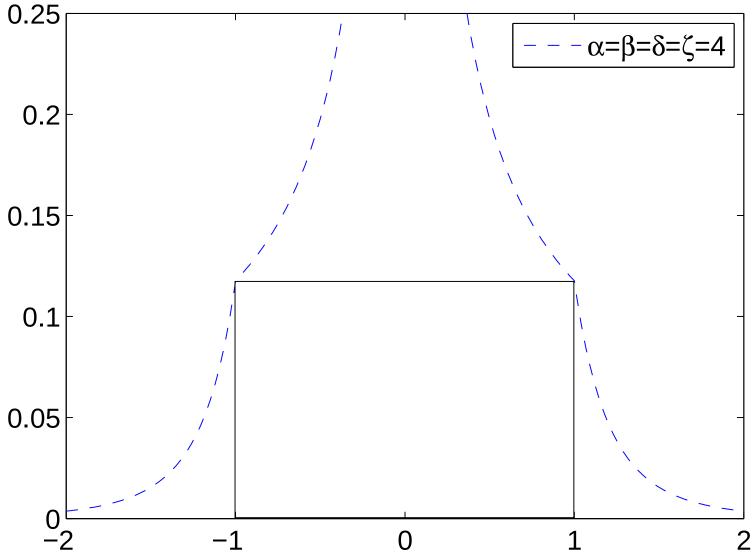

Figure 5: Largest circumscribed rectangle with parameters $\alpha = \beta = \delta = \zeta = 4$

to an area that is a local maximum. Since the volcano density is integrable, then any circumscribed rectangle with width converging to infinity must have a height that converges to zero faster than the width meaning the area of the rectangles converges to 0. It is obvious that the sequence of rectangles will have an area that monotonically converges to zero. A similar argument applies to rectangles converging to an infinite height and zero width and it follows that the area of the rectangle will be monotonically increasing until it reaches the points $\mu + \sigma$ and $\mu - \sigma$ on the horizontal axis and monotonically decreasing after those points showing that it is in fact a global maximum. In the bimodal case, the circumscribed rectangle argument doesn't apply, but a similar argument can be used to get to the same result.

It is not necessary to restrict ourselves to a single circumscribed rectangle. The same argument could be applied to the rectangle formed to the left and to the right of the location parameter. The volcano density is restricted so that the change in the shape of the density occurs at points that are equally far away from the location parameter, but it is not a necessary restriction. One could use the rectangle to the left of the location parameter to estimate the change in the density on the left side; and use the corresponding rectangle on the right side of the location parameter to estimate the change on the right side. In practice, I use the maximum of the two rectangles as my estimate for the scale parameter which is given in equation (44).

$$\hat{\sigma}_L = \hat{\mu} - \underset{x<\hat{\mu}}{\operatorname{argmax}} \left(\hat{f}(x) \left| x - \hat{\mu} \right| \right) \tag{42}$$

$$\hat{\sigma}_R = \underset{x>\hat{\mu}}{\operatorname{argmax}} \left(\hat{f}(x) \left| x - \hat{\mu} \right| \right) - \hat{\mu} \tag{43}$$

$$\hat{\sigma} = \hat{\sigma}_L \vee \hat{\sigma}_R \tag{44}$$

where we again restrict the domain to the observed data points.

Theorem 3.7. *Suppose $\hat{\mu}$ is a consistent estimator, then the estimator defined by equation (44) is consistent*

Proof. The first thing to notice is that if the expectation of a volcano random variable is finite then

$$|x-\mu|\, v(\mu,\sigma,\alpha,\beta,\delta,\zeta)=\sigma v\left(\mu,\sigma,\alpha-1,\frac{\beta}{1+\beta},\frac{\delta}{1+\delta},\zeta-1\right)$$

Since $1>\frac{\beta}{1+\beta}>0$, the transformed volcano density will be bimodal for all parameters and the two modes will be at $\mu\pm\sigma$ and even when the expectation is not finite the resulting improper density will still be bimodal which means

$$\operatorname{argmax}\left(f(x)\,|x-\mu|\right)=\{\mu+\sigma,\mu-\sigma\}$$

So it will be possible to back out σ if wither of the modes are identified. let's consider $\hat{\sigma}_R$

$$\begin{aligned}\left|\hat{f}(x)\,|x-\hat{\mu}|-\hat{\mu}-f(x)\,|x-\mu|+\mu\right| &\le \hat{f}\,||x-\hat{\mu}|-|x-\mu||+|\mu-\hat{\mu}|\\ &\quad+\left|\hat{f}-f\right||x-\mu|\\ &\le\left(\hat{f}+1\right)|\mu-\hat{\mu}|+\left|\hat{f}-f\right||x-\mu|\end{aligned}$$

As long as $x\neq\mu$, then both terms of the second inequality converge to zero in probability as both $\hat{\mu}$ and $\hat{f}$ are consistent estimators. If $x=\mu$, then it may be the case that the first term after the second inequality doesn't converge to 0 and $\hat{\sigma}_R=|\mu-\hat{\mu}|$. However, It follows that whenever $\mu<\hat{\mu}$, then the second inequality holds for all $x>\hat{\mu}$ which implies $|\hat{\sigma}_R-\sigma|=o_p(1)$. Since μ can only affect either $\hat{\sigma}_R$ or $\hat{\sigma}_L$, then

$$\begin{aligned}\Pr\left(|\hat{\sigma}-\sigma|\ge\epsilon\right)\le\Pr\left(|\hat{\sigma}_R-\sigma|\ge\epsilon\right)[\mu<\hat{\mu}]+\Pr\left(|\hat{\sigma}_L-\sigma|\ge\epsilon\right)[\mu>\hat{\mu}]\\ +\Pr\left(|\hat{\mu}-\mu|\ge\epsilon\right)\to 0\end{aligned}$$

so the maximum of $\hat{\sigma}_R$ and $\hat{\sigma}_L$ will be consistent. □

3.1 Simulated estimation

Since this approach is a little non-standard, it would be a good idea to simulate a data set and then see if the estimation procedure could identify the parameters correctly. In the following simulation, an under-smoothed estimate of the density with bandwidth equal to $n^{-1/3}$ as previously discussed.We simulated 340 observations from the standard volcano density with parameters $\alpha=\beta=\delta=\zeta=4$ and then estimated the parameters as previously outlined. Finally percentile bootstrapped confidence intervals were computed using 1000 replications and the results are compiled in Table 1.

Even though there are 340 observations, the confidence intervals for α and δ are very wide. But with this choice of parameters only about 6% of the data is used for the estimation

Table 1: Simulated estimation with 340 observations

Parameter	Estimate	Lower Bound of 95% CI	Upper Bound of 95% CI
μ	0.0014	-0.0045	0.0094
σ	0.0942	0.090	1.78
α	0.67	0.60	7.11
β	2.84	1.33	3.97
δ	1.69	1.32	3.82
ζ	0.70	0.63	4.85

Table 2: Simulated estimation with 21250 observations

Parameter	Estimate	Lower Bound of 95% CI	Upper Bound of 95% CI
μ	0.00011	-0.00019	0.00043
σ	1.00	0.81	1.03
α	4.00	3.34	4.33
β	4.09	2.91	4.20
δ	3.05	2.78	4.20
ζ	3.96	3.31	4.24

of α and ζ which translates to using roughly 10 observations for each estimate of α and ζ, which is very few observations for asymptotic results. Meanwhile there are roughly 160 observations used for each estimate of β and δ which is reflected in the shorter confidence intervals and both estimates are closer to their true values, but still fairly far away.

In fact with a simulation of 21250 observations, which is roughly the size of our stock market returns data, we can see from Table 2 that the estimates are now very close to their true values and the confidence intervals are fairly tight and they all contain the true parameter values. This indicates that this approach requires fairly large sample sizes.

3.2 Bivariate Extension

While a bit beyond the scope of this paper, the volcano distribution can be extended to the multivariate case. The generalization of the volcano distribution to a multivariate setting is far from trivial. The conditions for continuity and integrability are less obvious in the multivariate case. However, the following construction is made for the bivariate case with an eye towards estimation. Suppose we have independent random variables

$$x_i \sim V(\mu_i, \sigma_i, \alpha_i, \beta_i, \delta_i, \zeta_i)$$

where $i = \{1, 2\}$. There will be a corresponding distribution function

$$\begin{aligned}\Pr\left(x_1 \le s_1, x_2 \le s_2\right) &= \Pr\left(\tfrac{x_1-\mu_1}{\sigma_1} \le \tfrac{s_1-\mu_1}{\sigma_1}\right)\Pr\left(\tfrac{x_2-\mu_2}{\sigma_2} \le \tfrac{s_2-\mu_1}{s_2}\right)\\ &= \Pr(z_1 \le t_1)P(z_2 \le t_2)\end{aligned}$$

where $t_i = \frac{s_i-\mu_i}{\sigma_i}$ and $z_i = \frac{x_1-\mu_1}{\sigma_1}$. For each event corresponding to the two independent volcano distributions, there is an event and a pair of independent standard volcano random variables such that the probability of both events are the same. This is important because it

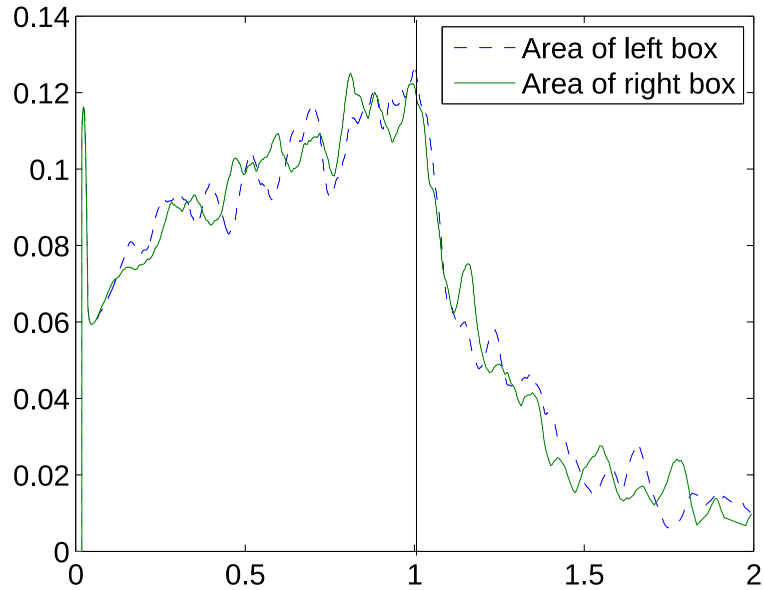

Figure 6: Estimation of σ using simulated data

will simplify the construction of the multivariate density. The normalization is fairly trivial once estimates of the marginal densities are in hand.

In the same spirit as the one dimensional case, it will be useful to define the bivariate volcano distribution through a transformation

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Assuming such a transformation is appropriate, the density for the bivariate case will be given by equation (45).

$$\begin{aligned} g\left(x_1, x_2\right) = \left(\frac{1}{1-\rho^2}\right) v\left(\frac{z_1-\rho z_2}{1-\rho^2}, \alpha_1, \beta_1, \delta_1, \zeta_1\right) * \\ v\left(\frac{z_2-\rho z_1}{1-\rho^2}, \alpha_2, \beta_2, \delta_2, \zeta_2\right) \end{aligned} \tag{45}$$

where $z_i = \frac{x_i - \mu_i}{\sigma_i}$. After one has estimated the marginal densities, this transformation follows easily. The information in the marginal densities can be used to estimate the dependency parameter in the following way

$$\hat{\rho} = \min_{-1 \leq \rho \leq 1} \sum_{i=1}^{n} \left(\widehat{g}\left(\widehat{x}_{1i}, \widehat{z}_{2i}\right) - \left(\frac{1}{1-\rho^2}\right) \prod_{j=1}^{2} v\left(\frac{\widehat{z}_{ji} - \rho \widehat{z}_{(3-j)i}}{1-\rho^2}, \widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\delta}_j, \widehat{\zeta}_j\right)\right)^2$$

where $\hat{g}\left(\widehat{z}_{1i}, \widehat{z}_{2i}\right)$ is a bivariate nonparametric kernel density estimator. In this simple case, a line search would be sufficient to solve the problem.

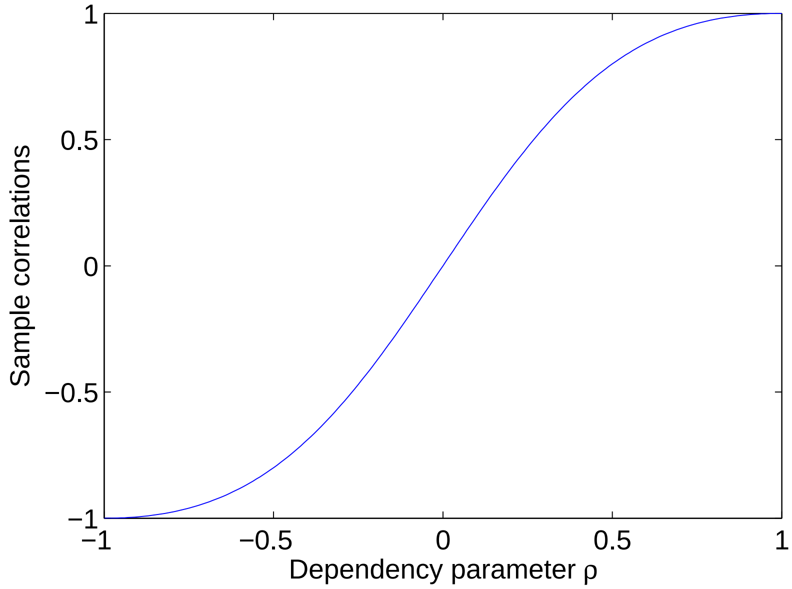

Figure 7: sample correlations $\alpha = \zeta = \frac{5}{2}$ and $\beta = \delta = 2$

The dependency parameter doesn't necessarily represent a correlation because the covariance may not exist. Even when the correlation does exist, the sample correlation will be a nonlinear function of the dependency parameter. The non-linear relationship between the sample correlation and the dependency parameter is demonstrated in Figure 7.

Each sample correlation is based on a single draw of 12000000 observations from a standard volcano density with unit variance. This allows the functional form to be isolated more easily. The sample correlations are roughly sinusoidal in nature and this general shape will remain the same for alternative parameters provided there are enough finite moments. This approach can be extended to higher dimensions, but the properties are not well known. This construction will preserve the box-like nature of the density, but it could be just as easily be generalized to a density in polar coordinates.

4 Application

4.1 Stock Market returns

It is well known that stock market returns data do not follow the normal distribution and they have heavy tails, but many finance models rely on finite variance assumptions. Schao et al. (2001) proposes a test statistic for finite variance in daily stock market returns, but it requires a parametric specification of the null hypothesis, finite variance, with finite fourth moments which is rather restrictive as there maybe finite second moments, but infinite third or fourth moments. One of the benefits of modeling heavy tailed data with the volcano distribution is that it is very easy to test whether or not a certain moment is finite because $min(\alpha, \zeta)$ determines the finiteness of any given moment. In order to test the assumption of finite variance, we use the daily data set from Schwert (1990). It has 22474 daily observations

Table 3: Stock market returns data

Parameter	Estimate	Lower bound of 95% CI	Upper bound of 95% CI
μ	0.00035	0.00023	0.00048
σ	0.02079	0.02076	0.02087
α	2.63	2.47	2.81
β	1.81	1.78	1.83
δ	1.75	1.73	1.78
ζ	2.42	2.25	2.64

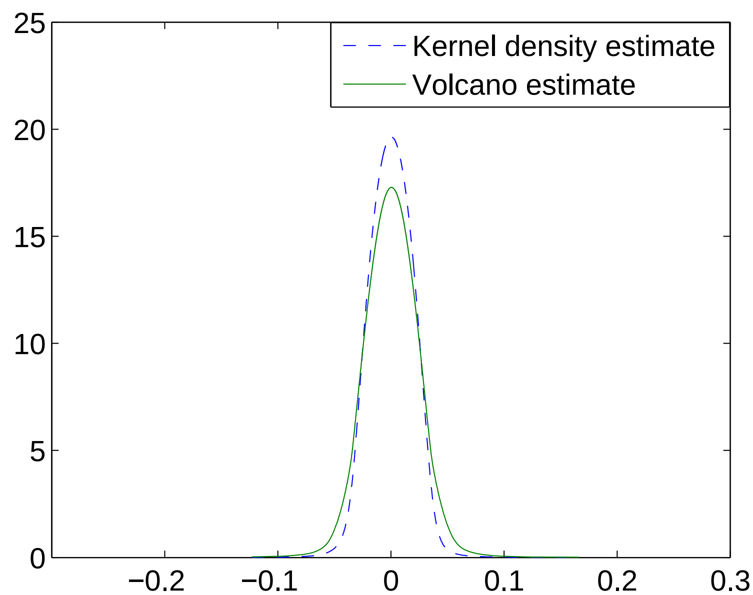

Figure 8: Fit for daily SP 500 stock market returns

of the S&P 500 from 1885 to 1962. The results of the estimation are presented in Table 3.

The location parameter is significantly larger than the sample mean which is 0.0003. Another implication is that the nature of stock returns is much different when the loss or gain is greater than 2% which makes intuitive sense on the daily level as this represents large swings in returns. The third and sixth rows provide information about the finiteness of the moments of stock market returns data. We reject infinite variance for the data because both $\hat{\alpha}$ and $\hat{\zeta}$ are statistically greater than two which supports the literature. Since $\hat{\alpha}$ and $\hat{\zeta}$ are also significantly less than 3, then we reject the hypothesis that the third moment is finite. This has important implications for models that rely on the existence of higher order moments in stock returns data. Furthermore we may also reject the hypothesis that the density function is square integrable. Finally the test of equality between α and ζ is rejected as is the test of $\beta = \delta$. This result is quite interesting because the rejection of symmetry implies that the way stock returns lose value is different than the way stock returns increase in value. The tails of the distribution are thicker for gains as opposed to losses.

In order to compare the fit of the volcano distribution, a random draw of 22474 was taken using the estimated parameters and kernel density estimates of the data and the random draw are presented in Figure 8. Both kernel density estimates are based on the same number of observations. The fit of seems reasonable, but the volcano distribution seems to underestimates peak of the density and overestimate the thickness of the tails of the stock market resturns data.

5 Conclusions

In this paper, a new family of probability distributions is introduced which represent a natural family for power law type distributions. However, there are many issues that remain unresolved for the volcano distribution. It would be interesting to find the solution of the FOC that corresponds to the global maximum of the log-likelihood function. We doubt that the method of moments estimator presented in this paper represents the global maximum.

A natural generalization would be to allow a left scaling factor and a right scaling factor. This would allow the bimodal volcano density to have varying heights. A more difficult generalization would be to describe the multivariate version of the density. It seems possible, but rather difficult especially the estimation of the scale parameter. The characteristic function could also be investigated further with uniqueness an open question.

This family of functions can also be applied in theoretical work because it is in some sense the largest unimodal, integrable function because careful choices of the parameters will result in an integrable function with tails declining at nearly the rate of $\frac{1}{x}$. It can be used as a bounding function in cases where the dominated convergence theorem is to be applied.

The volcano distribution is certainly an exotic distribution, but it is also very flexible. It can have varying symmetry, modality and boundedness. It can have an expectation that is infinity and with this can even be tested in a relatively straightforward way. It's greatest utility may even be as a tool for teaching because it challenges preconceived notions about continuity, boundedness, and integrability.

References

[1] Axtell, Robert L. (2001). U.S. Firms are Zipf distributed, *Science* **93** 1818–1820.

[2] Abramowitz, Milton and Stegun, Irene A. eds. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York.

[3] Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *Quarterly Journal of Economics.* **114** 739–767.

[4] Grabchak, M. and Samorodnitsky, G. (2010). Do financial returns have finite or infinite variance? A paradox and an Explanation. *Quantitative Finance.* **10** 883–893.

[5] Horowitz, J. (2001). "The Bootstrap" in *Handbook of Econometrics*, James J. Heckman and Edward Leamer (eds.). **5** 3160–3228.

[6] Kolmogorov, A.N. and Fomin, S.. (1975), *Introductory real analysis*. Dover Publications, New York. Translated from Russian by Richard A. Silverman.

[7] Lutmer, Erzo G.J. (2007). Selection, Growth, and the Size Distribution of Firms. *The Quarterly Journal of Economics*. **122** 1103–1144.

[8] Pagan, Adrian and Ullah, Aman (1999). *Nonparametric Econometrics*. Cambridge University Press, New York.

[9] Schwert, William G. (1990). Indexes of United States Stock Prices from 1802 to 1987. *The Journal of Business*. **63** 399–426.

[10] Shao, Q., Yu, H. and Yu, J. (2001). Do Stock Returns Follow a Finite Variance Distribution? *Annals of Economics and Finance* **2** 467–486.

[11] Talib, N.(2009). Finiteness of Variance is Irrelevant in the Practice of Quantitative Finance. *Complexity* **14** 66–76.