A Novel Asymmetric Distribution with Power Tails

Amita Singh[1], J. René van Dorp[2], Thomas A. Mazzuchi[3]

The George Washington University, Washington D.C. 20052

Abstract. In this paper we propose a four-parameter *asymmetric* doubly-Pareto uniform (DPU) distribution with support $(-\infty, \infty)$ whose density and cumulative distribution functions are constructed by seamlessly concatenating the left and right Pareto tails with a uniform central part. Properties of the distribution are described and a maximum likelihood estimation (MLE) procedure for its parameters is obtained. Two illustrative examples of the MLE procedure are provided. The first example utilizes an i.i.d. sample of standardized log-differences of bi-monthly 30-year U.S. conventional mortgage interest rates (1971-2004). The second example deals with the height of 100 female Australian athletes.

Keywords: Heavy-tailed and skewed distribution, maximum likelihood.

1. Introduction

Distributions of financial returns usually involve special functions, e.g., the Bessel function (Eberlein and Keller, 1995, Barndorff-Nielsen, 1997) or the hypergeometric function (Knight et al., 2002, Heinrich, 2004). This causes some difficulties in practice even with the aid of modern computers. The main purpose of this paper is to "humanize" financial distributions by proposing a distribution expressed in terms of elementary functions which seems to adequately represent some financial stochastic phenomena and is more transparent in its structure.

[1]Doctoral Candidate, Engineering Management and Systems Engineering Department, The George Washington University, 1776 G Street, NW, Suite 110, Washington D.C., 20052, e-mail: amitas@gwu.edu

[2]Corresponding Author and Associate Professor, e-mail: dorpjr@gwu.edu

[3]Professor and Chair, Same address, e-mail: mazzu@gwu.edu

Further confirmation that distribution of financial returns are fat-tailed, peaked and skewed have appeared in the relevant literature (e.g., Levy and Duchin, 2004, Kotz and van Dorp , 2004, McFall Lamm, 2003, Kotz et al., 2001, Solomon and Levy, 2000). This fact has reinvigorated the search for continuous distributions of this nature. Among the proposed distributions are the asymmetric Laplace (AL) distribution (e.g., Kotz et al., 2001)

$$f(x|\theta,\kappa,\sigma) = \begin{cases} \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2} exp\left[-\sqrt{2}\,\frac{1}{\sigma\kappa}(\theta - x)\right], & \text{for } x < \theta \\ \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2} exp\left[-\sqrt{2}\,\frac{\kappa}{\sigma}(x - \theta)\right], & \text{for } x \geq \theta, \end{cases} \tag{1}$$

where $\kappa, \sigma > 0$ and $\theta \in \mathbb{R}$, and the Pearson-type IV distribution (e.g., Nagahara, 1999),

$$f(x|m,\nu,a,\lambda) = k\left[1 + \left(\frac{x-\lambda}{a}\right)^2\right]^{-m} exp\left[-\nu \arctan\left(\frac{x-\lambda}{a}\right)\right] \tag{2}$$

where $m > 1/2,\ m, \nu, a,$ and λ are real constants and the normalization constant

$$k = \frac{2^{2m-2}|\Gamma(m + i\nu/2)|^2}{\pi a \Gamma(2m-1)}$$

involves the complex gamma function $\Gamma(\,\cdot\,)$. Both the pdf and cdf evaluations of Pearson-type IV distribution require the use of numerical approximations.

Lévy (1925) introduced a heavy-tailed family of probability laws referred to as "Lévy stable" distributions. Nolan (2005) among others describes basic facts about univariate stable distributions in detail. Lévy stable pdf's have also recently been applied to the modeling of financial tail behavior (e.g., Bardou et al., 2002, Huang and Solomon, 2001, Matacz, 2000, Solomon and Levy, 2000 and Bouchaud et al., 1998, amongst others) due to their attractive power-tail property. Unfortunately, the Gaussian and Cauchy distributions are the only two members in this family with a closed form expression for their pdfs. Other members are defined only via their characteristic functions (e.g., Nolan, 2005) requiring advanced numerical techniques for their application.

The asymmetric doubly Pareto-uniform (DPU) distribution to be discussed in this paper has support $(-\infty, \infty)$ and exhibits power laws in both its tails similar to the Lévy stable and Pearson-type IV distributions. A distinct advantage of the DPU family is that its pdf and cdf can be described using only elementary functions. In addition, the classical measures of skewness and kurtosis of DPU distributions may take values in $(-\infty, \infty)$ and $[1.8, \infty)$, respectively. A DPU distribution should not be confused with the recently proposed four-parameter double Pareto-Lognormal (DPL) distribution of Reed and Jorgensen (2003) with support $[0, \infty)$.

Our primary motivating example for DPU distributions is drawn from the financial domain. Figure 1A-C compare the Maximum Likelihood (ML) fit of the normal (Gaussian), asymmetric Laplace (0), and Pearson-type IV (0) distributions utilizing standardized bi-monthly log-differences of 30-year conventional U.S. mortgage interest rates (1971-2004) obtained via the Auto-Regressive Conditional Heteroscedastic (ARCH) time series model devised by Engle (1982). Figure 1D plots the ML fit of a DPU density. An empirical kernel density function is also displayed in Figures 1A-D. The empirical density was generated using the Bartlett-Epanechikov kernel (e.g., Izenman, 1991) combined with the over smoothed bandwidth (e.g., Sheather, 2004). Apparently, for the empirical density function in Figure 1, the normal distribution is "not peaked" sufficiently while the AL distribution is "overly peaked". Presumably, a sample kurtosis of 7.02 for the data in Figure 1 should be attributed to its heavy tailed behavior rather than its peakedness. A visual analysis of Figure 1 shows that both the DPU and Pearson-type IV pdf's provide a "better" fit than the normal and AL distributions. A detailed fit analysis is presented in Section 4.

The remainder of this paper is organized as follows: in Section 2 we present the pdf and cdf of the DPU family of distributions and discuss some of its properties. A maximum likelihood procedure for estimating the DPU parameters is described in Section 3. Section 4 provides two examples of the ML procedure – a detailed analysis of the mortgage interest rates data presented in Figure 1 and a

comparison of the ML fit of 100 heights of Australian athletes to the skew generalized normal (SGN) fit recently obtained by Arellano-Valle et al., 2004. Section 5 contains some concluding remarks. Mathematical details of the ML procedure are deferred to the Appendix.

2. DPU distributions and their properties

The construction of the DPU pdf $f(x|\alpha,\beta,m,n)$ utilizes the method of the *modified* mixture technique applied for the generalization of the trapezoidal distribution in Kotz and Van Dorp (2004). This technique seamlessly concatenates densities with disjoint but connecting supports. Specifically,

$$f(x|\alpha,\beta,m,n)=\sum_{i=1}^{3}\pi_i f_{X_i}(x),$$

where the pdfs are

$$\begin{cases} f_{X_1}(x)=\frac{m(\beta-\alpha)^m}{(\beta-x)^{m+1}}, & \text{for } x<\alpha<\beta,\\ f_{X_2}(x)=\frac{1}{\beta-\alpha}, & \text{for } \alpha\le x\le\beta,\\ f_{X_3}(x)=\frac{n(\beta-\alpha)^n}{(x-\alpha)^{n+1}}, & \text{for } x>\beta>\alpha, \end{cases}$$

and the weights (mixing probabilities) are given by

$$\begin{cases} \pi_1=\frac{n}{m+mn+n},\\ \pi_2=\frac{mn}{m+mn+n},\\ \pi_3=\frac{m}{m+mn+n}, \end{cases} \tag{3}$$

where $m,n>0$, the pdfs f_{X_i}, $i=1,3$ are Pareto distributions (e.g., Arnold, 1983) and f_{X_2} is a uniform distribution on $[\alpha,\beta]$. Hence, for the pdf we obtain

$$f(x|\alpha,\beta,m,n)=\mathcal{K}(m,n)\times\begin{cases} \frac{(\beta-\alpha)^m}{(\beta-x)^{m+1}}, & x<\alpha,\\ \frac{1}{\beta-\alpha}, & \alpha\le x\le\beta,\\ \frac{(\beta-\alpha)^n}{(x-\alpha)^{n+1}}, & x>\beta, \end{cases} \tag{4}$$

where the parameters $\alpha<\beta$, $m>0$, $n>0$ and

$$\mathcal{K}(m,n) = \frac{mn}{m+mn+n}. \tag{5}$$

The values of the mixing probabilities (3) ensure the continuity of the pdf at the boundaries of the central stage $[\alpha, \beta]$. The pdf (4) will be referred to as the doubly Pareto-uniform (DPU) distribution in view of its two Pareto tails and the central uniform stage. From the mixing probabilities (3) it immediately follows that

$$m = \pi_2/\pi_1 \text{ and } n = \pi_2/\pi_3. \tag{6}$$

Hence, the power parameter in the left tail (right tail) equals the ratio of the probability mass in the central part divided by the probability mass in the left tail less than α (right tail larger than β). Setting

$$Y = \frac{X-\alpha}{\beta-\alpha} \tag{7}$$

one obtains the "standardized" DPU pdf of the random variable Y from the pdf (4) given by:

$$f(y|0,1,m,n) = \mathcal{K}(m,n) \times \begin{cases} 1/(1-y)^{m+1}, & y<0, \\ 1, & 0 \le y \le 1, \\ 1/y^{n+1}, & y>1. \end{cases} \tag{8}$$

Standardized DPU distributions have a uniform$[0,1]$ central stage (see Figure 2A). The parameters m and n will be referred to as the *tail shape parameters* of the DPU distribution and the parameters α and β the *mode location (centrality) parameters*.

Integrating the pdf (4) we arrive at the corresponding cdf:

$$F(x|\alpha,\beta,m,n) = \begin{cases} \frac{n}{m+mn+n}\left(\frac{\beta-\alpha}{\beta-x}\right)^m, & x<\alpha, \\ \frac{mn(x-\alpha)+n(\beta-\alpha)}{(m+mn+n)(\beta-\alpha)}, & \alpha \le x \le \beta, \\ 1-\frac{m}{m+mn+n}\left(\frac{\beta-\alpha}{x-\alpha}\right)^n, & x>\beta. \end{cases} \tag{9}$$

Figure 2A (Figure 2B) plots an example DPU pdf (4) (cdf (9)), with parameters $\alpha = 0$, $\beta = 1$, $m = 5$, $n = 15$. Note that the central stage in Figure 2A is symmetric around $\frac{1}{2}$ and observe a linear behavior in the central part in both Figures 2A and 2B. Figure 2C (Figure 2D) plots an

example DPU pdf (4) (cdf (9)), with parameters $\alpha = -1\frac{1}{8}$, $\beta = -\frac{7}{8}$, $m = 5$, $n = 15$ exhibiting a more peaked behavior than the distribution in Figures 2A and B. The two distributions in Figure 2 are asymmetric due to differing tail behaviors.

2.1. Inverse Cumulative Distribution Function

The inverse cdf is immediately obtained from (9) as:

$$F^{-1}(y|\alpha,\beta,m,n) = \begin{cases} \lambda_1(\beta-\alpha)+\alpha, & 0 \le y < \pi_1, \\ \lambda_2(\beta-\alpha)+\alpha, & \pi_1 \le y \le 1-\pi_3\,, \\ \lambda_3(\beta-\alpha)+\alpha, & 1-\pi_3 \le y \le 1. \end{cases} \tag{10}$$

where:

$$\begin{cases} \lambda_1 = 1 - \sqrt[m]{\frac{\pi_1}{y}} \le 0, & 0 \le y < \pi_1, \\ 0 \le \lambda_2 = \frac{y-\pi_1}{\pi_2} \le 1, & \pi_1 \le y \le 1-\pi_3, \\ \lambda_3 = \sqrt[n]{\frac{\pi_3}{1-y}} \ge 1, & 1-\pi_3 \le y \le 1. \end{cases}$$

In (10) the quantiles of a DPU distribution are written as a linear combination of the mode location parameters α and β. Sampling from a DPU distribution is straightforward, utilizing the inverse cumulative distribution function theorem and a pseudo-uniform random number generator (e.g., Banks et al., 2001).

2.2. Limiting Distributions

The mixture probabilities π_1, π_2, π_3 (3) are solely functions of the powers m and n and not functions of the location parameters α and β. Keeping $m > 0$ fixed and letting $n \to \infty$, one obtains $\pi_1 \to \frac{1}{1+m}$, $\pi_2 \to \frac{m}{1+m}$, $\pi_3 \to 0$, and

$$\lim_{n \to \infty} \frac{n(\beta-\alpha)^n}{(x-\alpha)^{n+1}} = \begin{cases} 0 & x > \beta \\ \infty & x = \beta. \end{cases}$$

Thus in this case, the pdf (4) converges to the pdf

$$f(x|\alpha,\beta,m)=\begin{cases}\frac{1}{1+m}\frac{m(\beta-\alpha)^m}{(\beta-x)^{m+1}}, & x<\alpha,\\ \frac{m}{1+m}\frac{1}{\beta-\alpha}, & \alpha\le x\le\beta,\\ 0 & x>\beta,\end{cases} \tag{11}$$

which we shall call the *Left-Pareto uniform* (LPU) distribution. Similarly keeping $n>0$ fixed and letting $m\to\infty$, the pdf (4) converges to the *Right-Pareto uniform* (RPU) distribution with the density

$$f(x|\alpha,\beta,n)=\begin{cases}0, & x<\alpha,\\ \frac{n}{1+n}\frac{1}{\beta-\alpha}, & \alpha\le x\le\beta,\\ \frac{1}{1+n}\frac{n(\beta-\alpha)^n}{(x-\alpha)^{n+1}} & x>\beta.\end{cases} \tag{12}$$

Letting $m\to\infty$, $n\to\infty$, the original pdf (4) converges to a uniform distribution on $[\alpha,\beta]$. Finally, analogous to the uniform distribution with support $[\alpha,\beta]$, the pdf $f(x|\alpha,\beta,m,n)$ (4) converges to a single point mass at α (or β) when $\alpha\uparrow\beta$.

2.3. Moments of the DPU distribution

The central moments for X follow immediately from those of Y (7), utilizing

$$E[(X-E[X])^k]=(\beta-\alpha)^kE[(Y-E[Y])^k]. \tag{13}$$

Hence, from (13) we may conclude that for the DPU distribution the statistical measures of variability and shape such as the coefficient of variation (CV), skewness ($\sqrt{\beta_1}$) and kurtosis (β_2), are solely functions of the tail shape parameters m and n and not of the mode location parameters α and β. After some straightforward manipulation, the k-th moment of Y (7) about zero follows for $m>k, n>k$ from (8) to be

$$\mu'_k=E[Y^k]=\mathcal{K}(m,n)\times\Big[\frac{1}{k+1}\frac{n+1}{n-k}+(-1)^k k!\Big\{\prod_{i=0}^{k}(m-i)\Big\}^{-1}\Big], \tag{14}$$

where $\mathcal{K}(m,n)$ is defined by (5). Setting $k=1$ and $k=2$ yields

$$E[Y]=\mathcal{K}(m,n)\times\Big[\frac{1}{2}\,\frac{n+1}{n-1}-\frac{1}{m(m-1)}\Big], \tag{15}$$

where $m, n > 1$ and

$$E[Y^2] = \mathcal{K}(m,n) \times \Big[\frac{1}{3}\frac{n+1}{n-2} + \frac{2}{m(m-1)(m-2)} \Big], \qquad (16)$$

where $m, n > 2$. Expressions (15) and (16) allow for straightforward evaluation of the variance $Var[Y] = E[Y^2] - E^2[Y]$ for $m, n > 2$. Note that, $E[Y] \to -\infty\,(\infty)$ when $m \downarrow 1$ $(n \downarrow 1)$ keeping $n > 1$ fixed ($m > 1$ fixed). Similarly, $Var[X] \to \infty$ when either $m \downarrow 2$ or $n \downarrow 2$ while keeping the other tail parameter fixed at a value larger than 2. In the special case that $m = n > 1$, the mean value $E[Y]$ reduces to $\frac{1}{2}$ and for $m = n > 2$

$$Var[Y] = \frac{1}{12}\frac{n^2+n+6}{(n-1)(n-2)} \to \frac{1}{12} \quad \text{as } n \to \infty.$$

(Recall that the uniform distribution $[0, 1]$ is the limiting distribution of the pdf (8) as $m \to \infty, n \to \infty$.)

2.4. Moment Ratio Diagram

Moment ratio plots, popularized for Pearson-type distributions by Elderton and Johnson (1969), provide a visual assessment of the skewness and the kurtosis in a particular family of asymmetric distributions. A moment ratio diagram is a plot with the skewness $\sqrt{\beta_1}$ on the abscissa and the kurtosis β_2 on the ordinate, with the convention that $\sqrt{\beta_1}$ retains the sign of μ_3 (e.g., Kotz and Johnson, 1985). Values of $\sqrt{\beta_1}$ and β_2 for DPU distributions can be calculated using the general expression for the moments around the origin $\mu'_k = E[Y^k],\ k = 1, \ldots, 4,\ (14)$ and their relationship with the central moments μ_k=$E[(Y - E[Y])^k]$, $k = 2, 3, 4$, (e.g., Stuart and Ord, 1994). Explicit forms of $\sqrt{\beta_1}$ and β_2 for DPU distributions result in somewhat cumbersome expressions and are omitted.

Figure 3 displays the moment ratio diagram coverage for the DPU family (4). For reference purposes, the moment ratio coverage of the AL family (1) as well as the locations of the Gaussian

and uniform distributions are included in Figure 3. Recall that the left and the right exponential distributions indicated in Figure 3 are limiting distributions of the AL family (1). From Figure 3 we observe the leptokurtic behavior of the DPU family since its moment ratio coverage is not bounded from above (exceeding the Gaussian value of 3). Figure 3 may be somewhat misleading as it seems to indicate that the attainable range for skewness for the DPU family is less than that for the AL family of distributions. However, from (14) it immediately follows that for the DPU family of distributions $\sqrt{\beta_1} \rightarrow -\infty$ ($\rightarrow \infty$) when letting $m \downarrow 3$ and keeping n fixed (letting $n \downarrow 3$ and keeping m fixed). Values of β_2 for the DPU distribution do not exist for $m \leq 4$ or $n \leq 4$ (and thus Figure 3 only covers tail shape parameter ranges $m > 4$ and $n > 4$).

3. Maximum Likelihood Estimation

For a random sample $\underline{X} = (X_1, \ldots, X_s)$ of size s from the distribution (4) the likelihood function is by definition:

$$L(\underline{X}|\alpha, \beta, m, n) = \{\mathcal{K}(m,n)\}^s \left\{\frac{1}{\beta - \alpha}\right\}^s \times \left\{\prod_{i=1}^{r_1} \frac{\beta - \alpha}{\beta - X_{(i)}}\right\}^{m+1} \left\{\prod_{j=r_2+1}^{s} \frac{\beta - \alpha}{X_{(j)} - \alpha}\right\}^{n+1} \tag{17}$$

where $X_{(1)} < X_{(2)} < \ldots < X_{(s)}$ are the order statistics of $\underline{X}$, and r_1 and r_2 are such that:

$$X_{(r_1)} \leq \alpha < X_{(r_1+1)}, X_{(r_2)} \leq \beta < X_{(r_2+1)}, 0 \leq r_1 \leq r_2 \leq s.$$

By convention $X_{(0)} = -\infty$, $X_{(s+1)} = +\infty$. Note that, for the case that $r_1 = r_2 = r$, we have the restriction

$$X_{(r)} \leq \alpha < \beta \leq X_{(r+1)}, \; r = 0, \ldots, s.$$

Scenarios with $r_1 = r_2$ correspond to a set of order statistics such that no observations have been observed in the center stage of the DPU distribution. We propose the following direct algorithm to

maximize the likelihood $L(\underline{X}|\alpha, \beta, m, n)$ (17) and estimate the ML estimates of the parameters α, β, m, and n:

k-th Iteration:

Step 0: Set $k = 1$, $\alpha_1 = X_{(i)}$, $\beta_1 = X_{(j)}, i < j$, $m_1 = \frac{j-i}{i}, n_1 = \frac{j-i}{s-j}$.

Step 1: Determine $n_{k+1}$by maximizing $L(\underline{X}|\alpha_k, \beta_k, m_k, n)$ over n.

Step 2: Determine $m_{k+1}$by maximizing $L(\underline{X}|\alpha_k, \beta_k, m, n_{k+1})$ over m.

Step 3: Determine $\alpha_{k+1}$by maximizing $L(\underline{X}|\alpha, \beta_k, m_{k+1}, n_{k+1})$ over α.

Step 4: Determine $\beta_{k+1}$by maximizing $L(\underline{X}|\alpha_{k+1}, \beta, m_{k+1}, n_{k+1})$ over β.

Step 5: If $|L(\underline{X}|\alpha_k, \beta_k, m_k, n_k) - L(\underline{X}|\alpha_{k+1}, \beta_{k+1}, m_{k+1}, n_{k+1})| < \epsilon$ $STOP$

Else $k = k + 1$ and Goto Step 1.

In Step 0 above, we initialize α_1 and β_1 to be the i-th and j-th order statistic respectively. The starting values for m_1 and n_1 follow from (6) and the stage probability estimates

$$\widehat{\pi}_1 = \frac{i}{s}, \widehat{\pi}_2 = \frac{j-i}{s}, \widehat{\pi}_3 = \frac{s-j}{s}$$

follow in turn from $\alpha_1 = X_{(i)}$, $\beta_1 = X_{(j)}$, $1 \leq i < j < s - 1$. The advantage of this initialization set-up is that we can easily search for ML estimates from different starting points. This is necessary since the likelihood $L(\underline{X}|\alpha, \beta, m, n)$ (17) may have multiple local maxima. Figure 4 plots an example of a likelihood profile as a function of α and m, which exemplifies the existence of local maxima for the hypothetical data set of size $s = 8$:

$$\underline{X} = (0.10, 0.25, 0.30, 0.40, 0.45, 0.60, 0.75, 0.80). \tag{18}$$

For the data set (18) we obtain two local maxima

$$\widehat{\alpha} = X_{(2)} = 0.25, \widehat{\beta} = X_{(8)} = 0.80, \widehat{m} = 5.281 \text{ and } \widehat{n} \to \infty,$$

with a likelihood of 6.558 and

$$\widehat{\alpha} = X_{(1)} = 0.10, \widehat{\beta} = X_{(8)} = 0.80, \widehat{m} \to \infty \text{ and } \widehat{n} \to \infty,$$

with a likelihood of 17.347. Hence, the resulting ML fit for the data set (18) is the second local maximum. Observe that it coincides with the ML fit for a uniform distribution. (This is achieved in the algorithm above by setting the tail parameters to an arbitrarily large value.) The first local solution coincides with the local maximum in Figure 4, while the second one continues to follow the "ridge" in Figure 4.

A software program with the ML algorithm above (requiring a set of order statistics in an ASCII text file as input) is available from the authors upon request. Mathematical details regarding the algorithm steps above are presented in the Appendix.

It is perhaps noteworthy to point out that ML estimates for the mode location parameters α and β do not have to be attained at an order statistic. In addition, the ML procedure behaves in a manner such that it will not remove the central stage of the DPU distribution. Removal of this central stage would indicate that ML fit of the DPU distribution has converged to a single point mass (which has a likelihood of 0 when more than one distinct observation has been observed). Moreover, we observe from Figure 4 that $L(\underline{X}|\alpha, \beta, m, n)$ (17) is non-differentiable in the mode location parameter α (and the parameter β). Hence, asymptotic normality of the ML estimators with a minimum variance equal to the Cramer-Rao lower bound may not be attained when $s \to \infty$.

4. Illustrative examples

We shall illustrate the ML procedure for the DPU distribution for two separate data sets. The first one utilizes the monthly 30-year U.S. mortgage interest rates for the period from 1971-2004 as depicted in Figure 5. The second example considers the height of 100 female athletes derived from the data set reported by Cook and Weisberg (1994) containing thirteen variables measured on 202

athletes (102 males and 100 females) at the Australian Institute of Sport. Various aspects of the Cook and Weisberg (1994) data set have been analyzed (e.g., Azzalini and Capitanio, 1999, Arnold and Beaver, 2000). Arellano-Valle et al. (2004) recently fitted a skew generalized normal (SGN) density to this data. This is compared to our DPU ML fit.

4.1. Fit analysis for the data in Figure 5

One-step log differences of the mortgage rate time series in Figure 5 have an auto-correlation with lag 1 of 0.413. Hence, we split the time series of interest rates into two separate time series of bi-monthly log differences to reduce auto-correlation. That is, we set

$$\begin{cases} a_k = \ Ln(i_{2k+2}) - Ln(i_{2k}) & k = 0, 1, 2, \dots \\ b_k = \ Ln(i_{2k+1}) - Ln(i_{2k-1}) & k = 1, 2, 3, \dots . \end{cases} \tag{19}$$

The auto-correlations with lag 1 of the time series a_k and b_k equal -0.0203 and 0.1404, respectively. A hypothesis test of zero auto-correlation up to lag 6 is accepted for both time series with a *p*-value of 62% and 31%, respectively. Next, we apply an $ARCH(1)$ model to remove heteroscedasticity from the time series a_k. An $ARCH(1)$ is defined by

$$a_k = \sigma_k \nu_k, \sigma_k^2 = \alpha_0 + \alpha_1 a_{k-1}^2, \tag{20}$$

where α_i , $i = 0, 1,$ are constants, a_k is serially uncorrelated and ν_k is a sequence of i.i.d. random variables with mean zero and variance 1 (Engle, 1982). The second time series b_k shall be used for validation purposes using the parameters fitted for (20).

From the $ARCH(1)$ analysis, we obtain the following equation for $\widehat{\sigma}_k^2$:

$$\widehat{\sigma}_k^2 \ = 0.00164 + 0.35380 a_{k-1}^2, \tag{21}$$

where the parameters $\underline{\alpha} = (\alpha_0, \alpha_1) \ \approx (0.00164, 0.35380)$ were estimated using the least squares method (cf. (20)). The values of $\widehat{\sigma}_k^2$ given in (21) suggest that the standardized time series

$$\nu_k = \frac{a_k}{\widehat{\sigma}_k},\ k = 0, \ldots, 200, \tag{22}$$

should be a realization from an i.i.d. time series (by design), which would allow us to use standard maximum likelihood procedures. We have

$$\overline{\nu} = \frac{1}{201}\sum_{k=0}^{200}\nu_k = 0.0018,\ \frac{1}{200}\sum_{k=0}^{200}(\nu_k - \overline{\nu})^2 = 1.0652.$$

An analysis of the auto-correlation function (ACF) and the partial-autocorrelation function (PACF) of ν_k indeed suggests that the time-series ν_k is homescedastic and uncorrelated. (Details of our specific analysis results are available from the authors upon request).

The ML estimates of the DPU distribution fitted to the data $\nu_k,\ k = 0, \ldots, 200,$ are $\widehat{\alpha} = -0.517, \widehat{\beta} = 0.918,\ \widehat{m} = 2.546, \widehat{n} = 3.852.$ The empirical pdf of the standardized bi-monthly log-differences ν_k is depicted in Figure 1 of Section 1 together with ML fitted normal (Figure 1A), asymmetric Laplace (Figure 1B), Pearson-type IV (Figure 1C) and DPU (Figure 1D) distributions. QQ plots for the ML fitted distributions in Figure 1 are provided in Figure 6 from which we observe a better fit for the DPU and Pearson-type IV distributions compared to the normal and asymmetric Laplace distributions.

The results of a formal fit analysis are presented in Table 1. The Chi-square statistic is calculated utilizing 15 bins ($15 \in [\sqrt{201}, 201/5]$ as suggested by Banks et al., 2001). The boundaries of the bins are selected such that the number of observations O_i, $i = 1, \ldots, 15$, in each Bin i equal 13 or 14, totaling 201 data points. Such a boundary selection procedure partitions the support of the range of observed data in a similar manner as the "equal probability method of constructing classes" (e.g., Stuart and Ord, 1994). The DPU distribution results in the largest p-value (24.6%) of the chi-squared hypothesis test (due to more moderate bin to bin deviations), the Pearson-type IV in the second largest (19.6%), while the asymmetric Laplace and normal distributions attain a p-value of

less than 3% (taking into account the number of parameters of each distribution to determine the degrees of freedom). The DPU distribution and Pearson-type IV distribution perform equally well in terms of the KS -Statistic, but the Pearson-type IV distribution attains a slightly higher log-likelihood.

When conducting an analogous fit analysis for validation utilizing the second time series b_k (19) (using the parameter estimates from the first time series a_k (19)) the Pearson-type IV distribution outperforms all other distributions. The results are depicted in Table 2. The degrees of freedom in Table 2 are the same for all four distributions since no parameters have been estimated during validation. Surprisingly, all four distributions perform remarkably better in Table 2 in the chi-squared and log-likelihood statistics when compared to the analysis in Table 1. Some deterioration of the K-S statistic is observed in terms for the normal and asymmetric Laplace distributions, while the Pearson-type IV and DPU distributions show a slight improvement. Summarizing, while the Pearson-type IV distribution provides the better fit for the full data set under consideration, the DPU distribution can be considered a strong second.

4.2. Fit analysis for heights of 100 female Australian athletes

Arellano-Valle et al. (2004) recently fitted a skew generalized normal (SGN) density

$$f(z|\mu,\sigma,\lambda_1,\lambda_2) = \frac{2}{\sigma}\phi(\frac{z-\mu}{\sigma})\Phi\left(\frac{\lambda_1(z-\mu)}{\sqrt{\sigma^2+\lambda_2(z-\mu)^2}}\right), \qquad (23)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively, $\sigma, \mu, \lambda_1 \in \mathbb{R}, \lambda_2 \geq 0,$ to heights of 100 female athletes reported in the Cook and Weisberg (1994) data set. Sample statistics for the data are $\overline{z} = 174.594,\ s^2 = 67.934,\ \sqrt{\beta_1} = -0.568$ and $\beta_2 = 4.321$. Arellano-Valle et al. (2004) arrived at the ML estimates $\widehat{\mu} = 170.320,\ \widehat{\sigma} = 85.518,\ \widehat{\lambda}_1 = 4.381$ and $\widehat{\lambda}_2 = 24.184$ and a log-likelihood of -347.239 for the pdf (23).

Employing the ML algorithm in Section 4 we obtain the ML estimates $\widehat{\alpha} = 171.4,\ \widehat{\beta} = 180.5,$ $\widehat{m} = 2.011,\ \widehat{n} = 2.75$ and a slightly lower log-likelihood of -349.6. Note that the sample mean $\overline{Z}$ is contained here within the ML fitted central stage $[\widehat{\alpha}, \widehat{\beta}]$. Figures 7A-B compare the maximum likelihood (ML) fit of the SGN and the DPU distributions. An empirical kernel density function is also displayed in Figures 7A-B. Note that the data set is both skewed and fat tailed. The density was generated using the Bartlett-Epanechikov kernel (e.g., Izenman, 1991) combined with the over smoothed bandwidth (e.g., Sheather, 2004).

Figure 8 visually compares the ML fit of the SGN and DPU distributions using QQ plots. Both seem to provide an adequate fit of the data. The Kolmogorov-Smirnov statistics takes a value of 0.04 for the SGN and a value of 0.05 for the DPU distributions. On the other hand, the chi-squared test utilizing 10 bins yields a *p*-value of 12.5% for the SGN distribution and a *p*-value of 25.3% for the DPU distribution. Following the same procedure as in Section 4.1, the boundaries of the bins are selected such that the number of observations O_i, $i = 1, \ldots, 10$, in each Bin i equal $9, 10$ or 11, totaling 100 data points. Summarizing, both the ML fitted SGN and DPU distributions seem to provide reasonable fits for the data under consideration.

5. Closing Remarks

In the authors' opinion, the DPU distribution fits the data in an adequate manner for both data sets under consideration. Further, the attractiveness of this distribution is that the density function and the cdf of the DPU family may be evaluated using only elementary functions. In contrast, the Pearson-type IV distribution and SGN distribution (which perhaps perform slightly better in terms of fit than the DPU distribution in the examples in Section 4.1 and Section 4.2, respectively) require more complicated numerical analysis techniques. In this paper we have developed (to the best of our knowledge) the first continuous univariate distribution with *power tails* and support $[-\infty, \infty]$

that is leptokurtic, allows for asymmetry, and an ease of evaluation of its pdf, cdf and inverse cdf similarly being enjoyed by other continuous univariate distributions, e.g., the asymmetric Laplace distribution. Moreover, the parameters of the DPU distribution allow for a transparent interpretation. Its mode centrality parameters α and β define the central stage of the distribution, while the power tail parameters m and n equal the ratio of the central stage probability divided by the tail probabilities, respectively. The distribution is non-smooth at the boundaries of its central stage which may be considered to only be a cosmetic defect, if smoothness of a density is desirable.

Acknowledgements

We are indebted to Samuel Kotz who has been very gracious in donating his time to provide comments and suggestions in the development of this paper. We are thankful to the Associate Editor, a referee and the Editor-in-Chief of the Communications and Statistics journal whose comments improved the contents and presentation of the earlier versions.

Appendix: The maximum likelihood algorithm

Below we discuss some mathematical details regarding the executions of Step 1 and 3 of the algorithm in Section 3. The development of Step 2 (Step 4) is analogous to Step 1 (Step 3). The argument is similar to the one recently devised by Kotz and Van Dorp (2004) in connection with the ML algorithm of the Two-Sided Power (TSP) distribution.

STEP 1: We shall separately consider the cases $\beta < X_{(s)}$ and $\beta = X_{(s)}$.
Case A: $\beta < X_{(s)}$; From $X_{(r_2)} \leq \beta < X_{(r_2+1)}$ it follows that $r_2 \leq s-1$. Considering the loglikelihood profile of (17) as a function of n one can write :

$$Log\{L(\underline{X}|n)\} = sLog\{\mathcal{K}(m,n)\} + (n+1)Log\left\{ \prod_{j=r_2+1}^{s} \frac{\beta-\alpha}{X_{(j)}-\alpha} \right\} + \mathcal{C} \tag{24}$$

where $\mathcal{K}(m,n)$ is defined by (5) and $\mathcal{C}$ is a constant. Setting the partial derivative with respect to n in (24) equal to zero, the following quadratic equation in the parameter n is obtained

$$(m+1)n^2 + mn + mLog^{-1}\left\{ \sqrt[s]{ \prod_{j=r_2+1}^{s} \frac{\beta-\alpha}{X_{(j)}-\alpha} } \right\} = 0, \tag{25}$$

where $Log^{-1}(\,\cdot\,) = 1/Log(\,\cdot\,)$. Solving (25) for n and noting that $n > 0$ it immediately follows that

$$n^{k+1} = -\frac{m}{2(m+1)} + \frac{1}{2(m+1)} \sqrt{ m^2 - 4(m+1)Log^{-1}\left\{ \sqrt[s]{ \prod_{j=r_2+1}^{s} \frac{\beta-\alpha}{X_{(j)}-\alpha} } \right\} } \,. \tag{26}$$

Case B: $\beta = X_{(s)}$; We have from $X_{(r_2)} \le \beta < X_{(r_2+1)}$ that $r_2 = s$. Considering the likelihood profile of (17) as a function of n we now have :

$$L(\underline{X}|n) \propto \{\mathcal{K}(m,n)\}^s, \tag{27}$$

where $\mathcal{K}(m,n)$ is defined by (5). Maximizing the logarithm of the RHS of (27) and taking its derivative with respect to n yields the positive value $\mathcal{K}(m,n)/n^2$. Hence, the RHS of (27) is a strictly increasing function in n and $n_k \to \infty$. Hence, maximum likelihood estimation of a DPU distribution (4) reduces to maximum likelihood estimation of a LPU distribution (11), which may be achieved by setting n equal to an arbitrarily fixed large value in the algorithm presented in Section 3.

STEP 3: Maximizing the log-likelihood profile of (17) as a function of α we have

$$Log\{L(\underline{X}|\alpha)\} \propto \frac{\xi}{n+1} Log(\beta-\alpha) - \sum_{j=r_2+1}^{s} Log(X_{(j)}-\alpha) \tag{28}$$

where

$$\xi = r_1(m+1) + (s-r_2)(n+1) - s \tag{29}$$

and

$$\begin{cases} X_{(r_1)} \leq \alpha < X_{(r_1+1)} < \beta, \\ 0 \leq r_1 < r_2, \end{cases}$$

or

$$\begin{cases} X_{(r_1)} \leq \alpha < \beta, \\ 0 \leq r_1 = r_2. \end{cases}$$

Evidently, the two separate cases $\xi \leq 0$ and $\xi > 0$ ought to be considered.

Case A: $\xi \leq 0$; From functions of the form $Log(\theta - \alpha)$ being strictly decreasing functions for $\alpha < \theta, \xi \leq 0$ and $n > 0$ it immediately follows noting the restrictions

$$\alpha < X_{(j)}, j = r_2 + 1, \ldots, s \text{ and } \alpha < \beta,$$

that (28) is a strictly increasing function for $\alpha \in [X_{(r_1)}, \mathcal{U}]$, where

$$\mathcal{U} = \begin{cases} X_{(r_1+1)} & r_1 \in \{0, \ldots, r_2 - 1\} \\ \beta & r_1 = r_2. \end{cases} \tag{30}$$

Hence, (28) attains is maximum at the upper bound of the range $[X_{(r_1)}, \mathcal{U}]$.

Case B: $\xi > 0$; Taking the derivative of (28) with respect to α and equating it to zero yields

$$\sum_{j=r_2+1}^{s} \frac{\beta - \alpha}{X_{(j)} - \alpha} = \frac{\xi}{n+1}. \tag{31}$$

Taking the derivative of the LHS of (31) with respect to α yields

$$\sum_{j=r_2+1}^{s} \frac{\beta - X_{(j)}}{(X_{(j)} - \alpha)^2}.$$

Hence, from $\beta \leq X_{(r_2+1)}$, $\beta < X_{(j)}$, $j \in \{r_2+1, \ldots, s\}$ it follows that the LHS of (31) is a strictly decreasing function in α over the range $[X_{(r_1)}, \mathcal{U}]$, where $\mathcal{U}$ is given by (30). We are now able to conclude from (31) and the definition of ξ (29), that when $\xi > 0$ and:

1. $\sum_{j=r_2+1}^{s} \frac{\beta - \mathcal{U}}{X_{(j)} - \mathcal{U}} > \frac{\xi}{n+1} \Rightarrow$ the log-likelihood (28) attains it maximum at $\mathcal{U}$.
2. $\sum_{j=r_2+1}^{s} \frac{\beta - X_{(r_1)}}{X_{(j)} - X_{(r_1)}} < \frac{\xi}{n+1} \Rightarrow$ the log-likelihood (28) attains it maximum at $X_{(r_1)}$.
3. $\sum_{j=r_2+1}^{s} \frac{\beta - \mathcal{U}}{X_{(j)} - \mathcal{U}} < \frac{\xi}{n+1} < \sum_{j=r_2+1}^{s} \frac{\beta - X_{(r_1)}}{X_{(j)} - X_{(r_1)}} \Rightarrow$ the log-likelihood (28) attains it maximum at a stationary point $\alpha^* \in [X_{(r_1)}, \mathcal{U}]$.

References

Arnold, B.C. (1983). *Pareto Distributions.* Fairland, MD: International Co-Operative Publishing House.

Arnold, B.C., Beaver, R.J. (2000). Hidden truncation models. Sankhyā A 62:23 – 35.

Arellano-Valle, R.B., Gomez, H.W., Quintana, F.A. (2004). A new class of skew-normal distributions. Communications in Statistics, Theory and Methods 33 (7):1465 – 1480.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal Distribution. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 61:579-602.

Banks, J., Carson, J.S., Nelson, B.L., Nicol, D.M. (2001). *Discrete-Event System Simulation* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Bardou, F., Bouchaud, J., Aspect, A., Cohen-Tannoudji, C. (2002). *Lévy Statistics and Laser Cooling: How Rare Events Bring Atoms to Rest.* London: Cambridge University Press.

Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modeling. Scandanavian Journal of Statistics 24:1 – 13.

Bouchaud, J.P., Sornette, D., Walter, C., Aguilar, J.P. (1998). Taming large events, optimal portfolio theory for strongly fluctuating assets. International Journal of Theoretical and Applied Finance 1 (1): 25 – 41.

Cook, R. D., Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York:Wiley.

Eberlein, E., Keller, U. (1995). Hyperbolic distributions in finance. Bernoulli 1:281 – 299.

Elderton, W.P., Johnson, N.L. (1969). *Systems of Frequency Curves*. London: Cambridge University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica 50:987 – 1007.

Heinrich, J. (2004). A guide to the Pearson Type IV distribution. Note 6820, Collider Detector at Fermilab, Fermilab, Batavia, Illinois.

Huang, Z.F., Solomon, S. (2001). Power, Lévy, Exponential and Gaussian Regimes in Autocatalytic Financial Systems. The European Physical Journal. B20:601 – 607.

Izenman, A.J. (1991). Recent developments in nonparametric density estimation. Journal of the American Statistical Association 86 (413):205 – 224.

Knight, J.L., Satchell, S. E., Yu, J. (2002). Estimation of the stochastic volatility model by the empirical characteristic function method. Australian and New Zealand Journal of Statistics 44 (3):319 – 335.

Kotz, S., Johnson , N.L. eds. (1985). Moment ratio diagrams. *Encyclopedia of Statistical Sciences* Volume 5. New York, NY: Wiley.

Kotz, S., Kozubowski, T.J., Podgórski, K. (2001). *The Laplace Distribution and Generalizations*. Boston, MA: Birkhäuser.

Kotz, S., van Dorp, J.R. (2004). *Beyond Beta,Other Continuous Distributions with Bounded Support and Applications*, Singapore: World Scientific Publishing Company.

Levy, H., Duchin, R. (2004). Asset Return Distribution and the Investment Horizon, explaining contradictions. The Journal of Portfolio management 30 (3):47 – 62.

Lévy, P. (1925). *Calcul des probabilitiés*, 2nd part, Chapter VI. Paris: Gauthier-Villars.

Matacz, A. (2000). Financial Modeling and Option Theory with the Truncated Lévy Process. International Journal of Theoretical and Applied Finance 3 (1):143.

McFall Lamm Jr., R. (2003). Asymmetric Returns and Optimal Hedge Fund Portfolios.The Journal of Alternative Investments 9-21.

Nagahara, Y. (1999). The PDF and CF of Pearson Type IV distributions and the ML estimation of the parameters. Statistics & Probability Letters 43:251 – 264.

Nolan, J.P. (2005). *Stable Distributions* (Chapter 1). Forthcoming book, June 2005. Birkhauser (First edition).

Reed, W. J., Jorgensen, M. (2003). The double Pareto-lognormal distribution - A new parametric model for size distribution. Communications in Statistics – Theory & Methods 33 (8):1733 – 1753.

Sheather, S.J. (2004). Density estimation. Statistical Science 19 (4):588 – 597.

Solomon, S., Levy, M. (2000). Market Ecology, Pareto Wealth Distribution and Leptokurtic Returns in Microscopic Simulation of the LLS Stock Market Model. Proceedings of Complex Behavior in Economics, Aix en Provence (Marseille), France, May 4-6, 2000.

Stuart, A., Ord, J.K. (1994). *Kendall's Advanced Theory of Statistics*. New York, NY: Wiley.

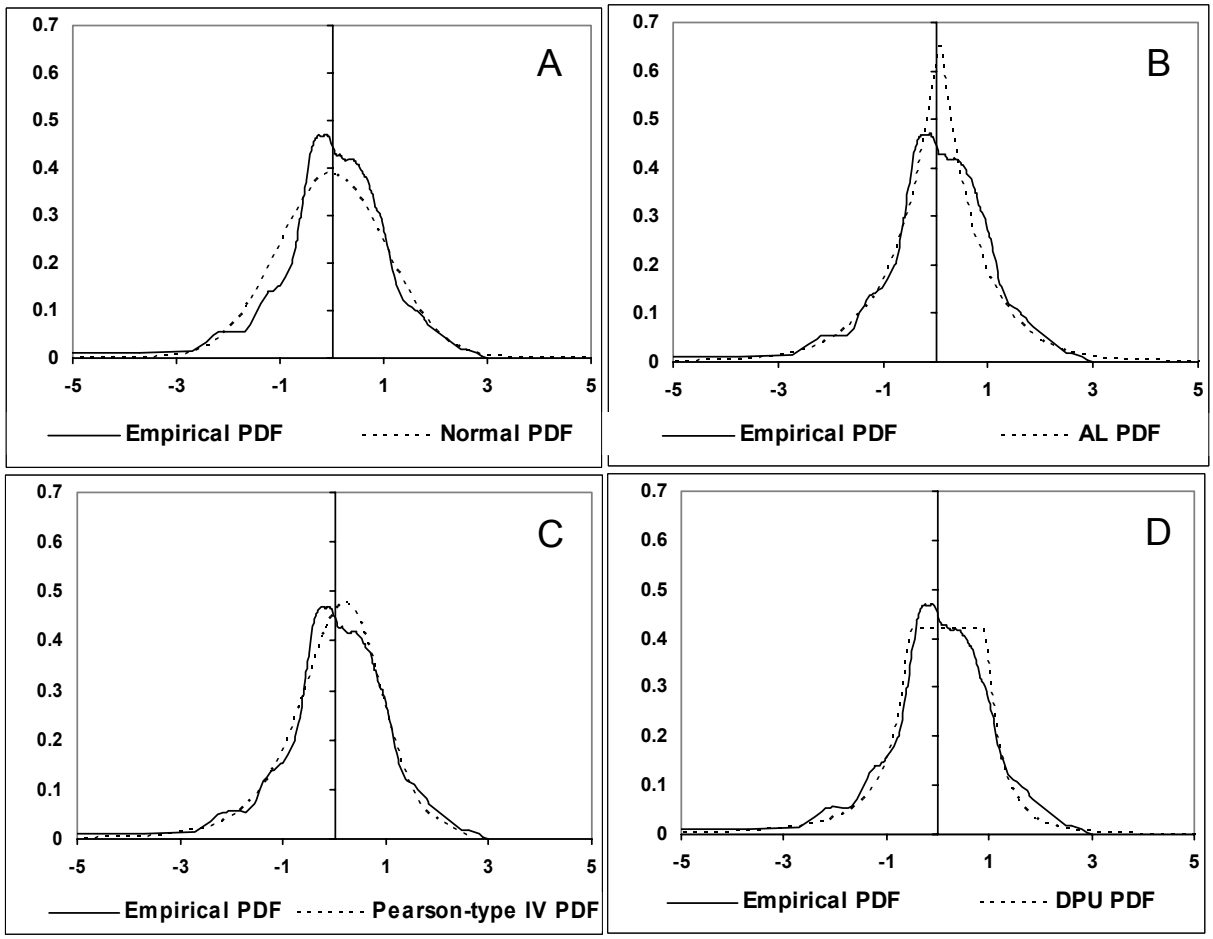

Figure 1. Empirical pdf's of standardized bi-monthly log-differences of 30-year conventional mortgage interest rates (1971-2004) together with the fitted (using the ML procedure) theoretical distributions.

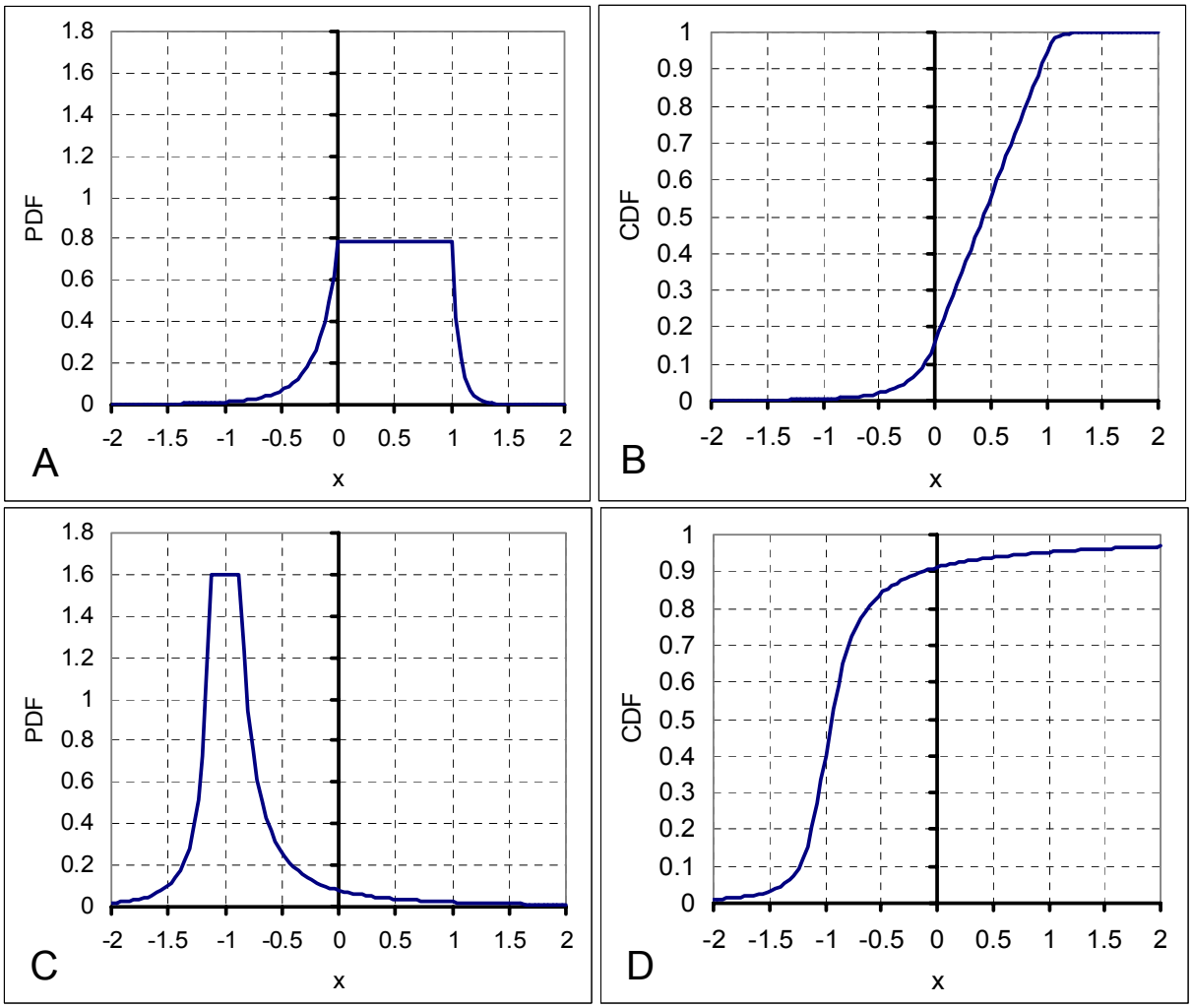

Figure 2. Cumulative distribution functions (9) (r.h. panels) and densities (4) (left-hand panels) of the DPU distribution. Panels A and B have parameters values $m = 5$, $n = 15$, $\alpha = 0$, $\beta = 1$; while panels C and D have $m = 2$, $n = 1$, $\alpha = -1\frac{1}{8}$, $\beta = -\frac{7}{8}$ (C-D).

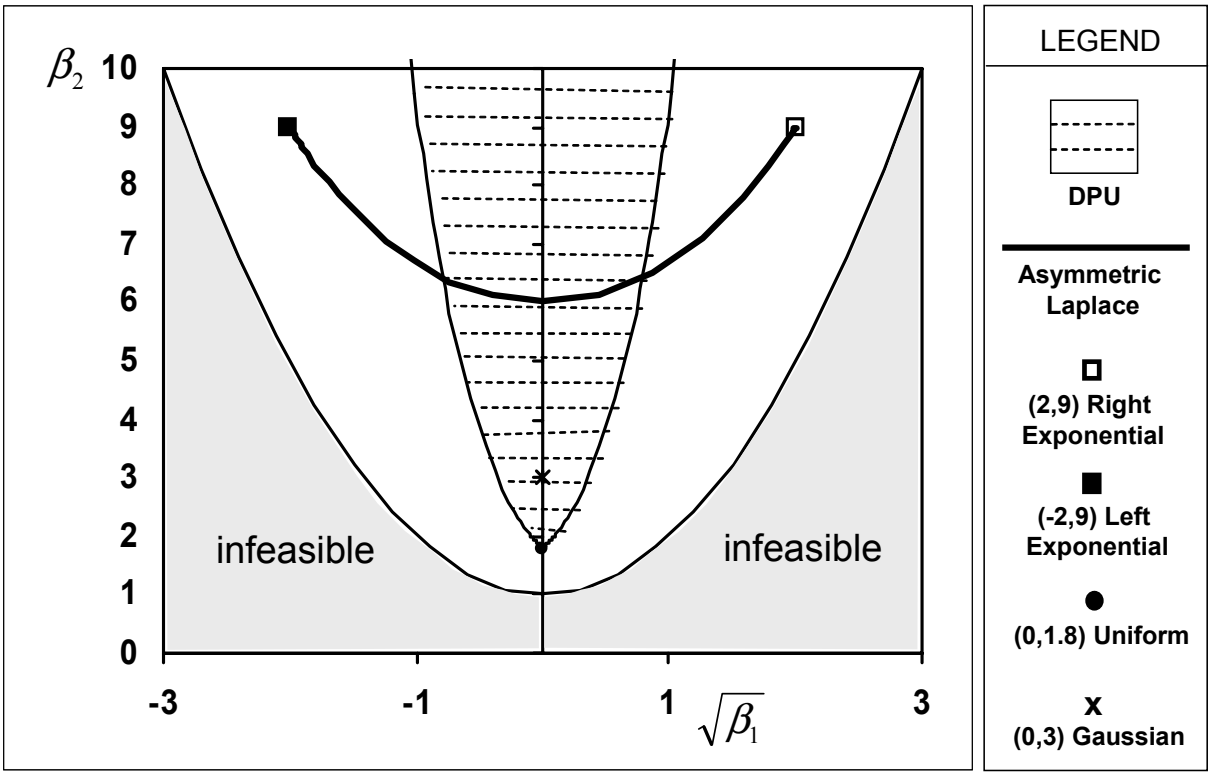

Figure 3. Moment Ratio Diagram $(\sqrt{\beta_1}, \beta_2)$ comparing the DPU family (4), the AL family (1), Gaussian family and uniform family of distributions.

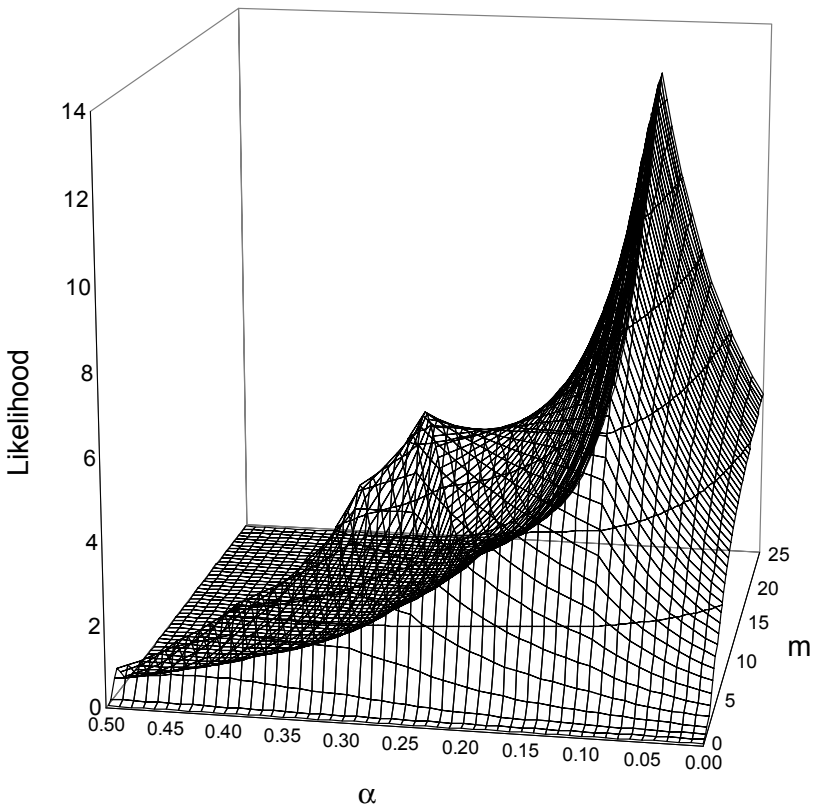

Figure 4. Example Likelihood Profiles of (17) as a function of the left tail shape parameter m and the left mode location parameter α for the hypothetical data set (18).

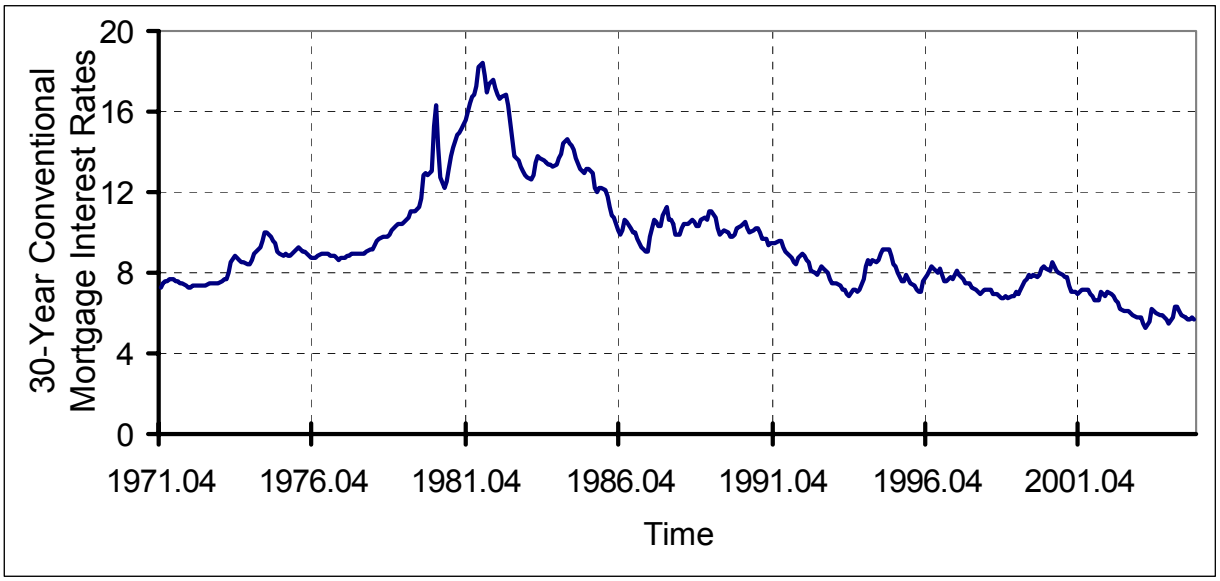

Figure 5. Time series of monthly 30-year U.S. mortgage interest rates from 1971 - 2004.

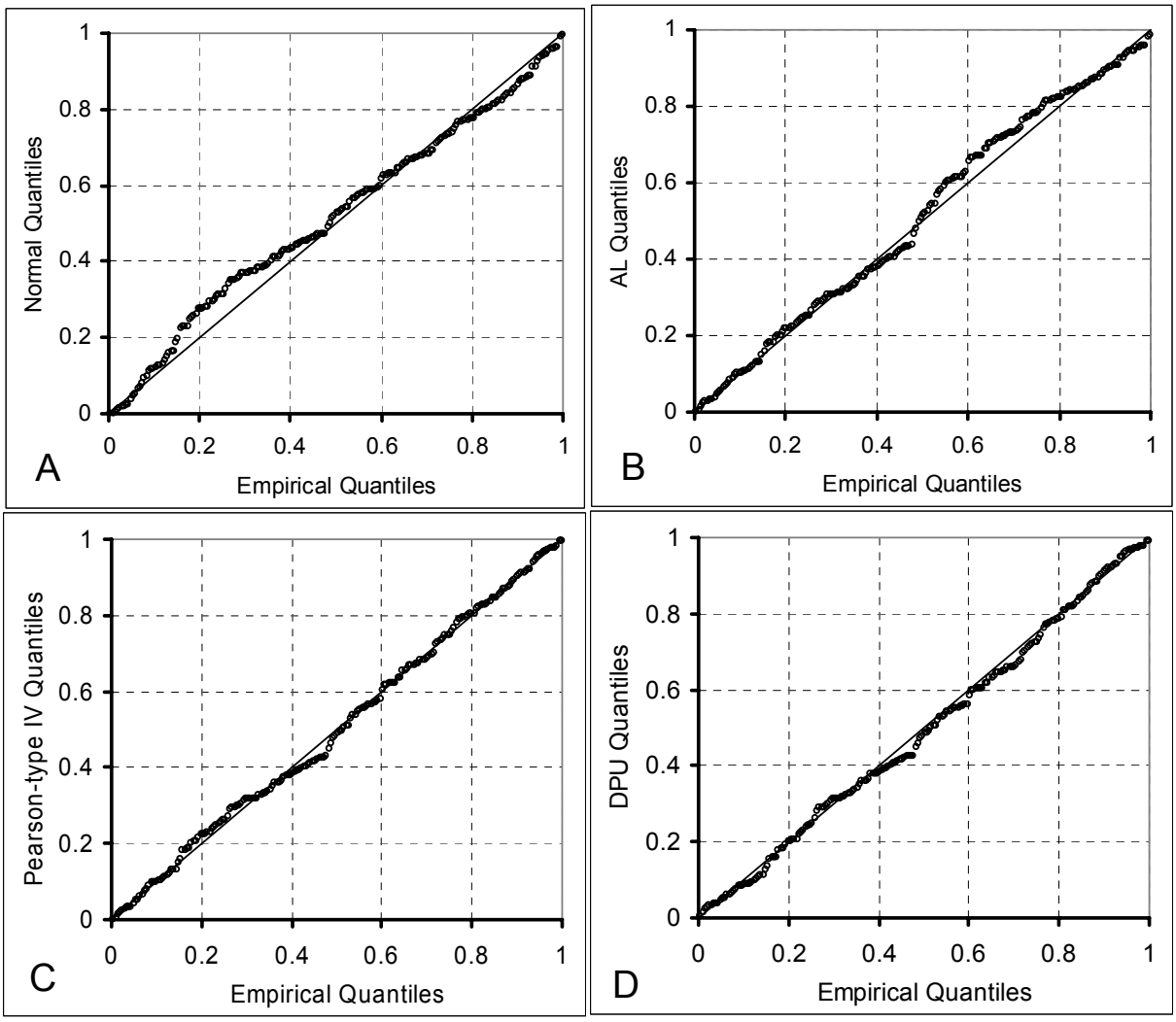

Figure 6. QQ plots for the ML fitted distributions in Figure 1. The fitted distributions are normal (Panel A), asymmetric Laplace (Panel B), Pearson-type IV (Panel C) and DPU (Panel D) distributions.

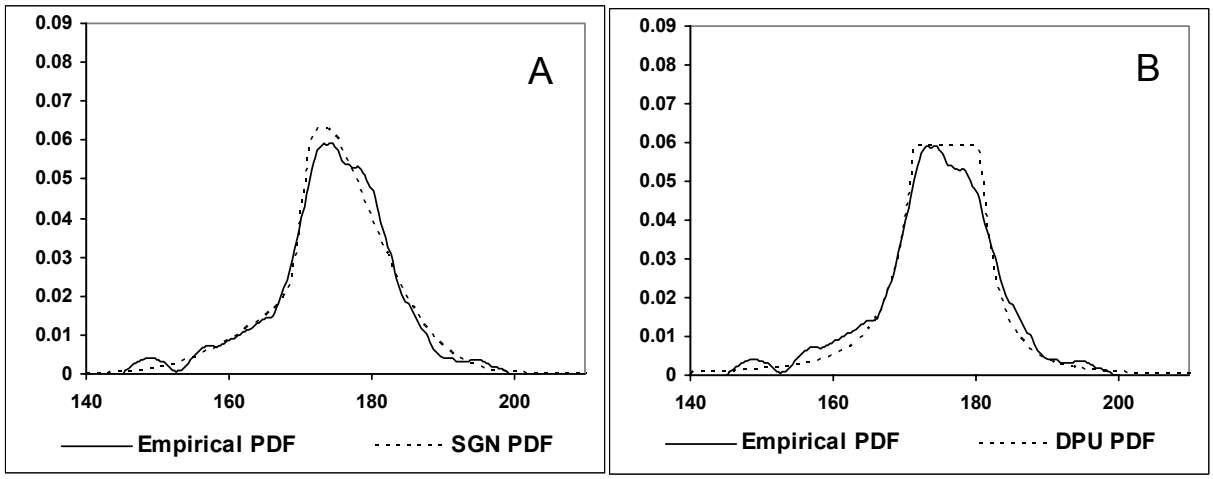

Figure 7. Empirical pdf's for the height of 100 female Australian athletes reported in the Cook and Weisberg (1994) data set, together with ML fitted SGN (Panel A) and DPU (Panel B) densities.

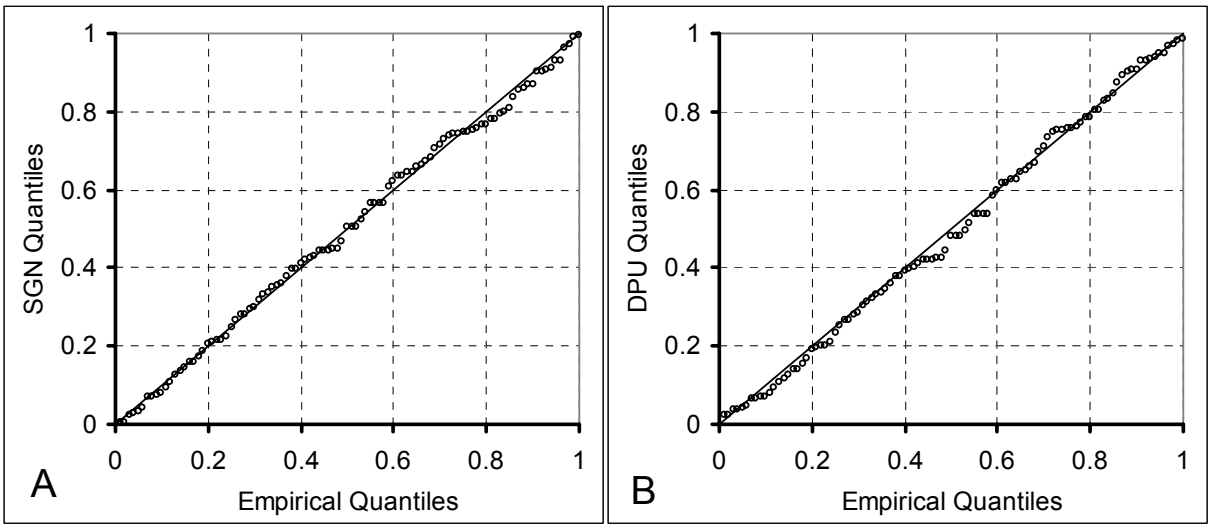

Figure 8. QQ plots for the ML fitted SGN (Panel A) and DPU (Panel B) distributions in Figure 7.

Table 1. Goodness of fit analysis utilizing the first time series a_k (19).

	Normal	AL	Pearson-type IV	DPU
Chi-Squared Statistic	23.98	21.51	13.52	12.62
Degrees of Freedom	12	11	10	10
P-value	2.0%	2.8%	19.6%	24.6%
K-S Statistic	0.080	0.059	0.050	0.050
Log-Likelihood	-291.0	-283.6	-278.0	-279.8

Table 2. Goodness of fit analysis utilizing the second time series b_k (19) for validation.

	Normal	AL	Pearson-type IV	DPU
Chi-Squared Statistic	18.26	13.32	6.90	11.42
Degrees of Freedom	14	14	14	14
P-value	19.5%	50.1%	93.8%	65.3%
K-S Statistic	0.082	0.068	0.033	0.040
Log-Likelihood	-284.7	-278.7	-275.1	-279.7