REVSTAT – Statistical Journal
Volume 22, Number 4, October 2024, 455–477
https://doi.org/10.57805/revstat.v22i4.521

# Applications of Composite lognormal Distributions

Authors: Jiahang Lyu
– Department of Mathematics, University of Manchester, UK
jiahang.lyu@manchester.ac.uk

Saralees Nadarajah ✉
– Department of Mathematics, University of Manchester, UK
mbbsssn2@manchester.ac.uk

Received: December 2021 Revised: November 2022 Accepted: November 2022

Abstract:

- The use of a power law distribution to model upper tails is common in many areas, most notably in physics. In this paper, we consider two data sets published in the physics literature. We show that composite lognormal distributions can provide better fits than the power law distribution even when the former are applied to the full data (as described in the data section) and the latter is applied just to the upper tail of the data.

Keywords:

- *Kolmogorov–Smirnov statistic; probability plot; quantile plot.*

AMS Subject Classification:

- Primary 62E15.

✉ Corresponding author

## 1. INTRODUCTION

Heavy tails are common in many areas of the sciences and engineering. These are commonly modeled by a power law distribution applied to the upper tail of the data, ignoring the body of the data. Ignoring the body of the data implies loss of information and loss of the power of the model.

The use of the power law distribution to model heavy tails is most common in the physics literature. Two recent papers published in the literature applying the power law distribution to model heavy tails are Campolieti [8] and Balthrop and Quan [3]. Campolieti [8] modeled the top 100 richest net wealth data from Canada Business Magazine. Models of this kind can be used to describe the economy of a country, accurate models of fitting wealth data can give a better prediction of the financial condition of a country. Balthrop and Quan [3] modeled the U.S. cumulative coal production data. Coal productivity is a significantly important factor to the economy of a country. Energy or electricity production relies mostly on coal production. Hence a good model of coal production data is essential for predicting the price of coal.

The aim of this paper is to show that composite lognormal distributions (Nadarajah and Bakar [14, 15]) can be used to model the entirety of the data sets in Campolieti [8] and Balthrop and Quan [3]. In addition, we show that these distributions provide better fits than the power law distribution even when the former are applied to the full data (as described in the data section) and the latter is applied just to the upper tail of the data.

The use of composite lognormal distributions to model data is not new. Cooray and Ananda [9] were the first to suggest the use of composite lognormal distributions. But Nadarajah and Bakar [14, 15] were the first to write an R package (R Core Team [16]) to implement the use of composite lognormal distributions. More recent papers on composite distributions include Calderín-Ojeda [4, 5], Calderín-Ojeda and Kwok [7], Calderín-Ojeda [6], Aminzadeh and Deng [2], Kim *et al.* [11] and Mutali and Vernic [13]. The distributions in these papers have been used to model among others city sizes.

The contents of this paper are organised as follows. Some details of the composite lognormal and power law distributions are given in Section 2. The two data sets and their summary statistics are given in Section 3. The fits of the distributions to the data sets are discussed in Section 4. Finally, some conclusions are noted in Section 5.

## 2. METHODS

### 2.1. Composite lognormal distributions

In this section, we discuss the composite lognormal distribution in Nadarajah and Bakar [14, 15]. The composite lognormal distribution is made up by joining together two distinct distributions: one for the body and the other for the tail. The body is described by the lognormal distribution while the distribution for the tail can be arbitrary. The cumulative distribution function (cdf) of the composite lognormal distribution is

$$F(x) = \begin{cases} \dfrac{1}{1+\phi}\left[\Phi\left(\dfrac{\log\theta-\mu}{\sigma}\right)\right]^{-1}\Phi\left(\dfrac{\log x-\mu}{\sigma}\right), & \text{if } \ 0 < x \leq \theta, \\ \dfrac{1}{1+\phi} + \dfrac{\phi}{1+\phi}\dfrac{F_0(x)-F_0(\theta)}{1-F_0(\theta)}, & \text{if } \ \theta < x < \infty, \end{cases} \tag{2.1}$$

where $\phi > 0$, $\theta$ denotes the point at which the two distributions are joined together, $\Phi(\cdot)$ denotes the cdf of the standard normal distribution, $f_0$ denotes the probability density function (pdf) of the tail, and $F_0$ denotes the cdf of the tail. The following conditions ensure that $F(x)$ is continuous and differentiable at $\theta$:

$$\begin{gathered} \mu = \log\theta + \sigma^2 + \theta\sigma^2\frac{f_0'(\theta)}{f_0(\theta)}, \\ \phi = \left[\Phi\left(\frac{\log\theta-\mu}{\sigma}\right)\right]^{-1}\frac{1}{\theta\sigma}\psi\left(\frac{\log\theta-\mu}{\sigma}\right)\frac{1-F_0(\theta)}{f_0(\theta)}, \end{gathered} \tag{2.2}$$

where $\psi(\cdot)$ denotes the pdf of the standard normal distribution.

Different choices for $f_0$ and $F_0$ lead to different models for the composite lognormal distribution. In Section 4, we consider fourteen different models: the composite lognormal-Fréchet, composite lognormal-log logistic, composite lognormal-generalized Pareto, composite lognormal-Weibull, composite lognormal-inverse Weibull, composite lognormal-Pareto, composite lognormal-paralogistic, composite lognormal-inverse paralogistic, composite lognormal-Burr, composite lognormal-inverse Burr, composite lognormal-inverse Pareto, composite lognormal-inverse exponential, composite lognormal-exponential, composite lognormal-gamma, composite lognormal-inverse gamma, composite lognormal-transformed gamma and composite lognormal-inverse transformed gamma distributions. We fitted all of the distributions by the method of maximum likelihood. The best distribution was chosen according to the following information criteria:

- the Akaike Information Criterion (AIC) due to Akaike [1] defined by

  $$\mathrm{AIC} = 2k - 2\log\widehat{L},$$

  where $k$ denotes the number of parameters and $\widehat{L}$ denotes the maximized likelihood;
- the Bayesian Information Criterion (BIC) due to Schwarz [17] defined by

  $$\mathrm{BIC} = k\log n - 2\log\widehat{L},$$

  where $n$ denotes the number of data;
- the Hannan Quinn Criterion (HQC) due to Hannan and Quinn [10] defined by

  $$\mathrm{HQC} = -2\log\widehat{L} + 2k\log\log n.$$

The smaller the values of these criteria the better the fit. The goodness of fit of the distributions was assessed by the $p$-values of the Kolmogorov–Smirnov, Anderson Darling and Cramer von Mises statistics.

The fourteen different models considered include the following:

- the composite lognormal-inverse Burr distribution with

$$f_0(x)=\frac{\lambda_1\lambda_2\left(\frac{x}{\lambda_3}\right)^{\lambda_1\lambda_2}}{x\left[1+\left(\frac{x}{\lambda_3}\right)^{\lambda_2}\right]^{\lambda_1+1}}$$

  and

$$F_0(x)=\left[1+\left(\frac{x}{\lambda_3}\right)^{-\lambda_1}\right]^{-\lambda_2},$$

  where $\lambda_1$ and $\lambda_2$ are shape parameters while $\lambda_3$ is a scale parameter;

- the composite lognormal-generalised Pareto distribution with

$$f_0(x)=\frac{1}{\lambda_3}\left[1+\frac{\lambda_1(x-\lambda_2)}{\lambda_3}\right]^{-\frac{1}{\lambda_1}-1}$$

  and

$$F_0(x)=1-\left[1+\frac{\lambda_1(x-\lambda_2)}{\lambda_3}\right]^{-\frac{1}{\lambda_1}},$$

  where $\lambda_1$ is a shape parameter, $\lambda_2$ is a location parameter and $\lambda_3$ is a scale parameter;

- the composite lognormal-inverse paralogistic distribution with

$$f_0(x)=\frac{\lambda_1^2\left(\frac{x}{\lambda_2}\right)^{\lambda_1^2}}{x\left[1+\left(\frac{x}{\lambda_2}\right)^{\lambda_1}\right]^{\lambda_1+1}},$$

  where $\lambda_1$ is a shape parameter and $\lambda_2$ is a scale parameter.

## 2.2. Estimation

Suppose $x_1, x_2, ..., x_n$ is a random sample from (2.1). Let $\mathbf{\Lambda}$ denote the parameters specifying $f_0(\cdot)$ and $F_0(\cdot)$. The maximum likelihood estimates of $\theta$, $\sigma$ and $\mathbf{\Lambda}$, say $\widehat{\theta}$, $\widehat{\sigma}$ and $\widehat{\mathbf{\Lambda}}$, respectively, were obtained as follows:

**i**) Compute the likelihood function

$$L(\theta,\sigma,\mathbf{\Lambda})=\frac{\phi^{n-m}}{(1+\phi)^n[1-F_0(\theta)]^{n-m}}\left[\prod_{x_i\le\theta}\frac{\psi\left(\frac{\log x_i-\mu}{\sigma}\right)}{\Phi\left(\frac{\log\theta-\mu}{\sigma}\right)}\right]\left[\prod_{x_i>\theta}f_0(x_i)\right],$$

where

$$m=\sum_{i=1}^{n}I\{x_i\le\theta\}$$

and $I\{\cdot\}$ denotes the indicator function. $\mu$ and $\phi$ are given by (2.2). Hence, they are functions of $\theta$, $\sigma$ and $\mathbf{\Lambda}$.

**ii**) Take its log as

$$\begin{aligned}\log L(\theta,\sigma,\mathbf{\Lambda}) \;=\; & (n-m)\log\phi - n\log(1+\phi) \\ & - m\log\left[\Phi\left(\frac{\log\theta-\mu}{\sigma}\right)\right] + (m-n)\log[1-F_0(\theta)] \\ & + \sum_{x_i\le\theta}\log\psi\left(\frac{\log x_i-\mu}{\sigma}\right) \\ & + \sum_{x_i>\theta}\log f_0(x_i).\end{aligned}$$

**iii**) Set initial values for $\theta$, $\sigma$ and $\mathbf{\Lambda}$.

**iv**) Maximize the log-likelihood function to obtain

$$\widehat{\theta},\widehat{\sigma},\widehat{\mathbf{\Lambda}} = \text{argmax}_{\theta,\sigma,\mathbf{\Lambda}}\;\log L(\theta,\sigma,\mathbf{\Lambda}), \tag{2.3}$$

using the `optim` function in R.

**v**) Repeat steps iii) and iv) for a range of initial values to make sure that $\widehat{\theta}$, $\widehat{\sigma}$ and $\widehat{\mathbf{\Lambda}}$ are unique.

In Section 4, we compare the best of the composite lognormal distributions to the power law distribution given by the cdf:

$$F(x) = 1-\left(\frac{K}{x}\right)^{\alpha} \tag{2.4}$$

for $x>K$ and $\alpha>0$. For a given random sample $x_1,x_2,...,x_n$ from (2.4), the maximum likelihood estimates of $K$ and $\alpha$ are

$$\widehat{K} = \min(x_1,x_2,...,x_n)$$

and

$$\widehat{\alpha} = n\left[\sum_{i=1}^{n}\log\left(\frac{x_i}{\widehat{K}}\right)\right]^{-1},$$

respectively.

Campolieti [8] and Balthrop and Quan [3] fitted the power law distribution to the upper tail of the data. They used the following procedure to estimate the parameters:

**1**. Order the data as $x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$.

**2**. Let $\widetilde{K}=x_{(i)}$ and estimate $\alpha$ by

$$\widetilde{\alpha} = \left[\sum_{i=1}^{n} I\left\{x_i\ge\widetilde{K}\right\}\right]\left[\sum_{i=1}^{n}\log\left(\frac{x_i}{\widetilde{K}}\right)\right]^{-1}.$$

**3**. Compute the Kolmogorov–Smirnov statistic

$$\sup_{x\ge\widetilde{K}}\left|\widetilde{F}(x)-1+\left(\frac{\widetilde{K}}{x}\right)^{\widetilde{\alpha}}\right|,$$

where $\widetilde{F}(\cdot)$ denotes the empirical cdf of the data.

**4**. Repeat steps 2 and 3 for $i=1,2,...,n-1$.

**5**. Choose $\widetilde{K}$ and $\widetilde{\alpha}$ to correspond to the smallest value of the Kolmogorov–Smirnov statistic.

## 3. DATA SETS

The two data sets are described in Sections 3.1 and 3.2. We shall refer to the data as described in these sections as “full data” sets.

### 3.1. Canadian net wealth data

The data were collected from the rich 100 list “Canadian Business magazine”. The rich list is published every year on line by the magazine. However, due to the webpages being updated, we were able to get the data only for the years 2014–2018, 2012 and 2009. Due to changes in policy from one year to another, 2014 had 101 data points while for 2018 had 98 data points. The remaining years had 100 data points each.

**Table 1**: Summary statistics of net wealth data in billions of nominal Canadian Dollars (the deflated figures by the Consumer Price Index are given in the second row for each year).

| Year | Mean | Median | Standard deviation | Skewness | Kurtosis | Min | Max | CPI |
|---|---|---|---|---|---|---|---|---|
| 2018 | 3.44 | 2.11 | 4.61 | 6.02 | 47.39 | 1.07 | 41.14 | |
| | 0.0258 | 0.0158 | 0.0346 | 5.9245 | 43.4279 | 0.008 | 0.308 | 133.4 |
| 2017 | 3.07 | 2.03 | 4.25 | 6.47 | 53.48 | 0.875 | 39.13 | |
| | 0.0235 | 0.0155 | 0.03259 | 6.3692 | 49.41976 | 0.0067 | 0.3 | 130.4 |
| 2016 | 2.88 | 1.89 | 4.00 | 6.47 | 53.31 | 0.835 | 36.76 | |
| | 0.0224 | 0.0147 | 0.0311 | 6.375 | 49.2505 | 0.0065 | 0.2863 | 128.4 |
| 2015 | 2.56 | 1.80 | 3.34 | 6.40 | 52.47 | 0.782 | 30.738 | |
| | 0.0202 | 0.0142 | 0.0264 | 6.3065 | 48.4212 | 0.0062 | 0.2428 | 126.6 |
| 2014 | 2.29 | 1.46 | 2.91 | 5.91 | 46.35 | 0.721 | 26.075 | |
| | 0.0183 | 0.0116 | 0.0232 | 5.8207 | 42.438 | 0.0058 | 0.2083 | 125.2 |
| 2012 | 2.02 | 1.39 | 2.33 | 5.29 | 38.88 | 0.654 | 20.129 | |
| | 0.0166 | 0.0114 | 0.0191 | 5.2145 | 35.1037 | 0.0054 | 0.1654 | 121.7 |
| 2009 | 1.73 | 1.15 | 2.39 | 6.48 | 53.72 | 0.49 | 21.99 | |
| | 0.0151 | 0.01005 | 0.0208 | 6.3829 | 49.6533 | 0.0043 | 0.1922 | 114.4 |

Table 1 gives the summary statistics of the data in terms of nominal and real figures. The Consumer Price Index was taken from `https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000501`. Since the rich people are getting richer with time in terms of both nominal and real figures, the values of mean, median, maximum and minimum increase with year. The skewness is positive every year. The kurtosis is much greater than 3 every year, meaning that the data are heavy tailed.

## 3.2. Cumulative coal production data

The data were collected from the U.S. Energy Information Administration (EIA) website. Balthrop and Quan [3] used the cumulative yearly production data from 1983 to 2016. Due to the website being updated, we use the data from 2001 to 2018. The data contained a large number of zeros (over 800 data points were zero), and these were removed before fitting of the distributions.

**Table 2**: Summary statistics of coal data (unit: short tons).

| | |
|---|---|
| Mean | 4413710 |
| SD | 38518025 |
| Skewness | 27.62912 |
| Kurtosis | 954.948 |
| Min | 43 |
| Max | 1528026392 |
| Sample size | 4180 |

Table 2 shows that the skewness is positive. The kurtosis is once again much larger than 3, which indicates the data has a large heavy tail.

## 4. RESULTS AND DISCUSSION

In this section, we illustrate the flexibility of the composite lognormal distributions using the two real data sets. Fourteen of the composite lognormal distributions were fitted to both data sets in full. For comparison, the power law distribution is also fitted to the full data sets. The power law distribution is also fitted to the upper tail of the data sets.

In the discussion throughout Sections 4.1 and 4.2, "$p$-values" refer to $p$-values of the Kolmogorov–Smirnov statistic. But Tables 10 and 12 also report $p$-values of Anderson Darling and Cramer von Mises statistics. The conclusions based on these $p$-values are the same as those based on $p$-values of the Kolmogorov–Smirnov statistic.

In Section 4.3, we investigate finite sample performance of the maximum likelihood estimators of composite lognormal distributions to see if the conclusions reported in Sections 4.1 and 4.2 are reasonable.

## 4.1. Canadian net wealth data

Tables 3 to 9 give the best three distributions giving the smallest information criteria for each year. The power law distribution does not make the best three distributions for any of the years.

**Table 3**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2009.

| Model | $-2\log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −261.42 | −251.42 | −240.99 | −247.20 |
| Composite lognormal-inverse paralogistic | −251.60 | −243.60 | −235.79 | −240.44 |
| Composite lognormal-generalized Pareto | −209.98 | −199.98 | −189.56 | −195.76 |

**Table 4**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2012.

| Model | $-2\log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −231.72 | −221.72 | −211.30 | −217.50 |
| Composite lognormal-inverse paralogistic | −222.22 | −214.22 | −206.41 | −211.06 |
| Composite lognormal-generalized Pareto | −221.91 | −211.91 | −201.49 | −207.69 |

**Table 5**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2014.

| Model | $-2\log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −244.64 | −234.64 | −224.18 | −230.40 |
| Composite lognormal-inverse paralogistic | −237.97 | −229.97 | −222.13 | −226.80 |
| Composite lognormal-generalized Pareto | −228.74 | −218.74 | −208.28 | −214.51 |

**Table 6**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2015.

| Model | $-2\log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −211.45 | −201.45 | −191.03 | −197.23 |
| Composite lognormal-inverse paralogistic | −208.74 | −200.74 | −192.92 | −197.58 |
| Composite lognormal-generalized Pareto | −236.70 | −226.70 | −216.28 | −222.48 |

**Table 7**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2016.

| Model | $-2\log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −239.28 | −229.28 | −218.86 | −225.06 |
| Composite lognormal-inverse paralogistic | −229.61 | −221.61 | −213.80 | −218.45 |
| Composite lognormal-generalized Pareto | −234.94 | −224.94 | −214.52 | −220.72 |

**Table 8**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2017.

| Model | $-2 \log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −240.90 | −230.90 | −220.48 | −226.69 |
| Composite lognormal-inverse paralogistic | −240.07 | −232.07 | −224.25 | −228.91 |
| Composite lognormal-generalized Pareto | −215.20 | −205.20 | −194.78 | −200.98 |

**Table 9**: The three best composite lognormal distributions according to information criteria for Canadian net wealth data in 2018.

| Model | $-2 \log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-inverse Burr | −251.60 | −241.60 | −231.26 | −237.42 |
| Composite lognormal-inverse paralogistic | −244.83 | −236.83 | −229.08 | −233.69 |
| Composite lognormal-generalized Pareto | −233.40 | −223.40 | −213.06 | −219.22 |

Table 10 lists the $p$-values for the power law distribution and the very best composite lognormal distributions chosen as the ones having the smallest information criteria values.

**Table 10**: Fitted models and $p$-values for the Canadian net wealth data (the first row of $p$-values for each year is for the Kolmogorov–Smirnov statistic, the second row of $p$-values for each year is for the Anderson Darling statistic, the third row of $p$-values for each year is for the Cramer von Mises statistic).

| Year | $n$ | Composite model | | Power law fitted to full data | Power law fitted to upper tail | | |
|---|---|---|---|---|---|---|---|
| | | Best model | $p$-value | $p$-value | $p$-value | $\widetilde{K}$ | no of data > $\widetilde{K}$ |
| 2018 | 98 | Composite lognormal-inverse Burr | 0.973<br>0.970<br>0.974 | 0.060<br>0.055<br>0.058 | 0.99<br>0.98<br>0.99 | 2.77 | 37 |
| 2017 | 100 | Composite lognormal-inverse paralogistic | 0.853<br>0.855<br>0.855 | 0.011<br>0.010<br>0.008 | 0.994<br>0.994<br>0.995 | 2.96 | 27 |
| 2016 | 100 | Composite lognormal-inverse Burr | 0.9996<br>0.999<br>0.998 | 0.098<br>0.095<br>0.097 | 0.99<br>0.95<br>0.96 | 2.35 | 38 |
| 2015 | 100 | Composite lognormal-generalised Pareto | 0.844<br>0.840<br>0.851 | 0.020<br>0.030<br>0.035 | 0.98<br>0.99<br>0.96 | 1.96 | 48 |
| 2014 | 101 | Composite lognormal-inverse Burr | 0.921<br>0.935<br>0.922 | 0.063<br>0.065<br>0.068 | 0.92<br>0.93<br>0.90 | 1.85 | 42 |
| 2012 | 100 | Composite lognormal-inverse Burr | 0.996<br>0.999<br>0.995 | 0.065<br>0.061<br>0.062 | 0.9<br>0.9<br>0.91 | 1.48 | 46 |
| 2009 | 100 | Composite lognormal-inverse Burr | 0.998<br>0.995<br>0.999 | 0.020<br>0.025<br>0.022 | 0.959<br>0.966<br>0.954 | 1.17 | 48 |

The $p$-values for the very best composite lognormal distributions range from 0.8 to 0.99. The largest of these $p$-values is 0.9996 (2016) and the smallest is 0.844 (2015). The composite lognormal inverse Burr distribution gives the largest $p$-values for five of the seven years.

The $p$-values for the power law distribution are always less than 0.1 when applied to the full data. When applied to the tail (containing a fraction of the full data), the $p$-values are much closer to 1. But for four of the seven years the $p$-values for the very best composite lognormal distributions are still greater. For the year 2018, the $p$-value for the very best composite lognormal distribution is slightly smaller (0.973 compared to 0.99), but the power law tail models only 37 of the 98 observations. For the year 2017, the $p$-value for the very best composite lognormal distribution is again slightly smaller (0.853 compared to 0.994), but the power law tail models only 27 of the 100 observations. For the year 2015, the $p$-value for the very best composite lognormal distribution is again slightly smaller (0.844 compared to 0.98), but the power law tail models only 48 of the 100 observations.

The probability and quantile plots comparing the fits of the power law distribution and the very best composite lognormal distributions are shown in Figures 1 to 7.

Both the quantile and probability plots confirm that the composite lognormal distributions provide better fits than the power law distribution. Nearly all of the plotted points in the probability plots lie close to the 45 degree line for the composite lognormal distributions. The quantile plots show that the composite lognormal distributions provide good fits to the data except for a few extremely large observations. The power law distribution fitted to the full data gives poor fits. The power law distribution fitted to the tail gives much better fits but still not good as the composite lognormal distributions.

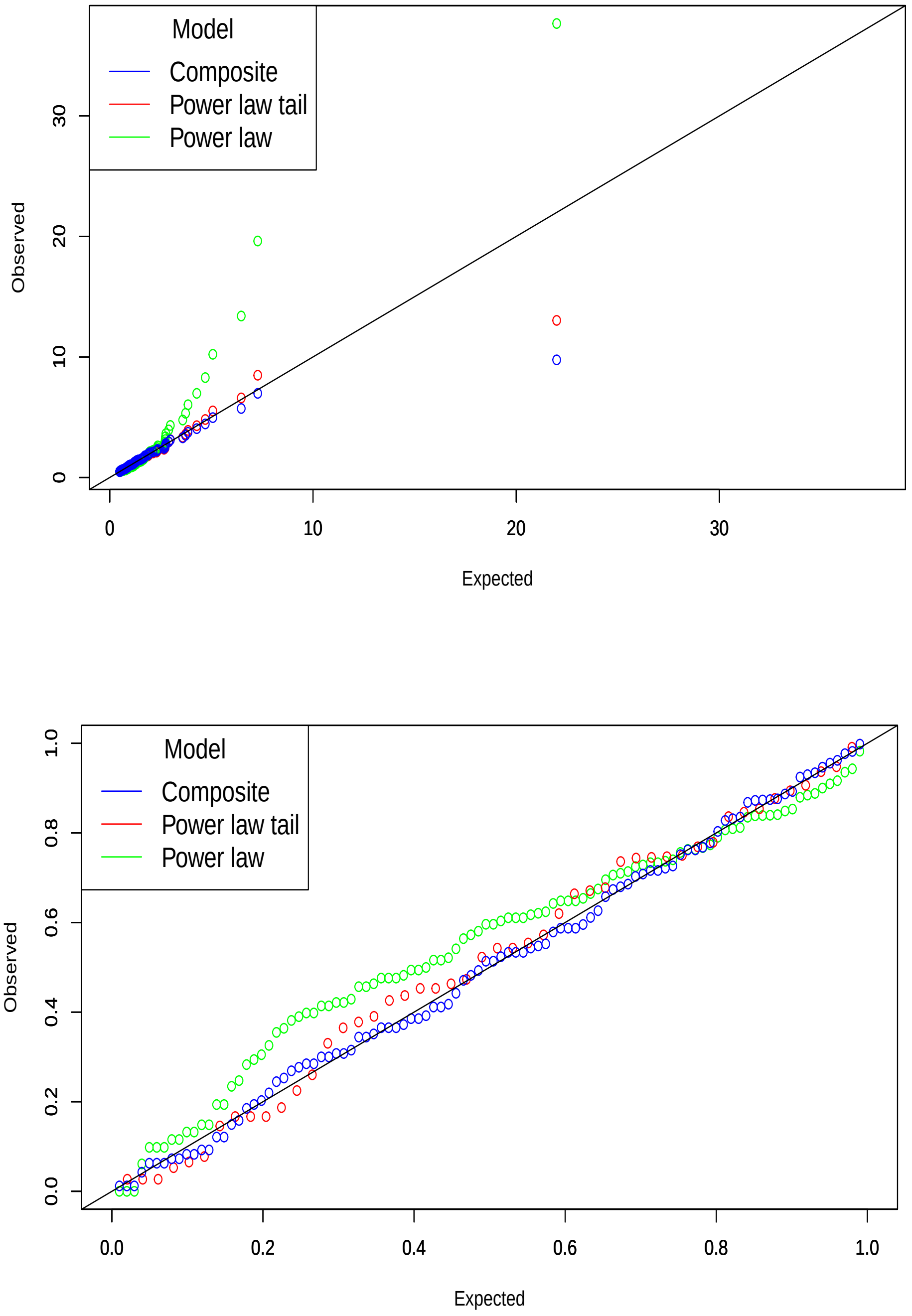


**Figure 1**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse Burr and power law distributions for Canadian net wealth data in 2009.

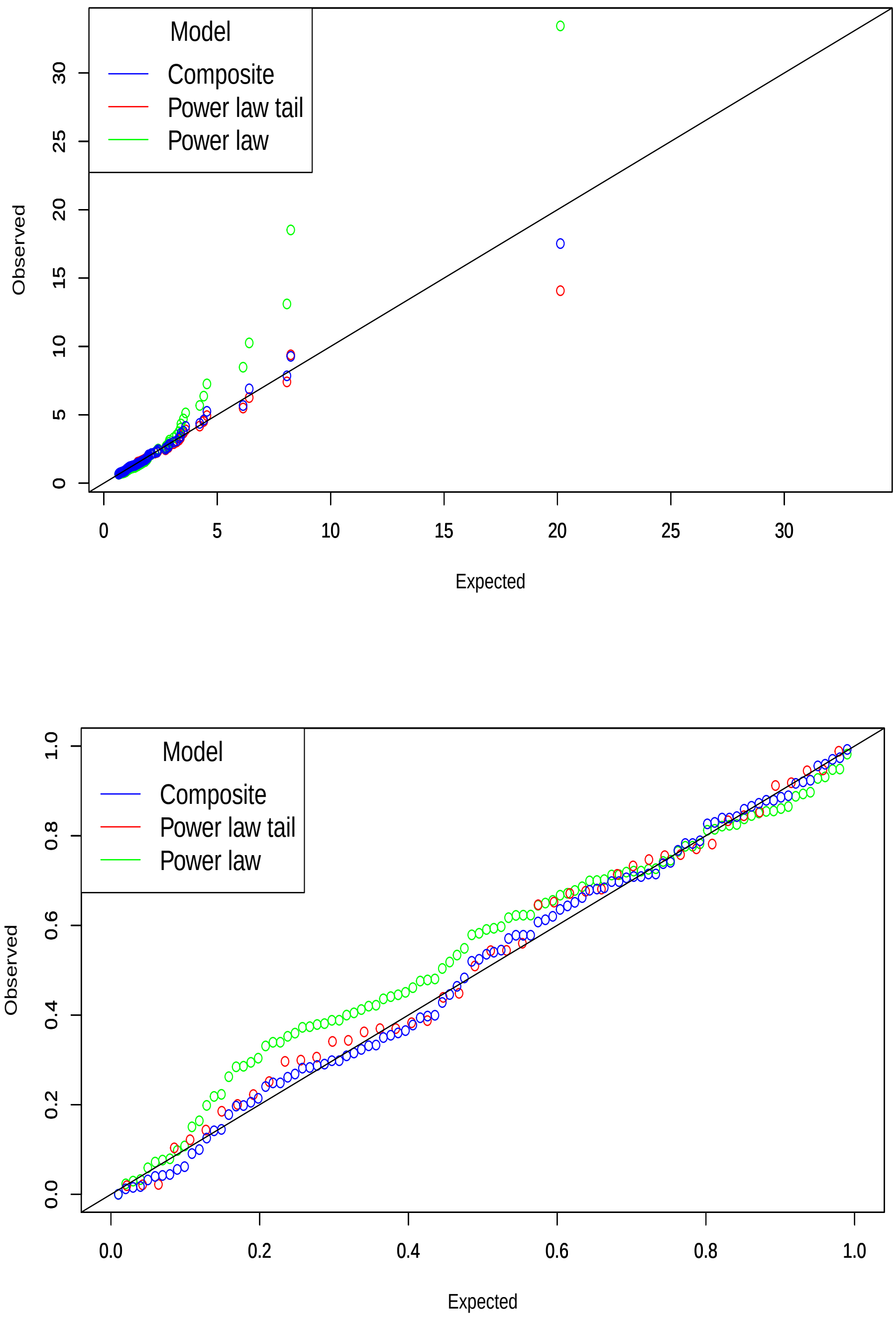


**Figure 2**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse Burr and power law distributions for Canadian net wealth data in 2012.

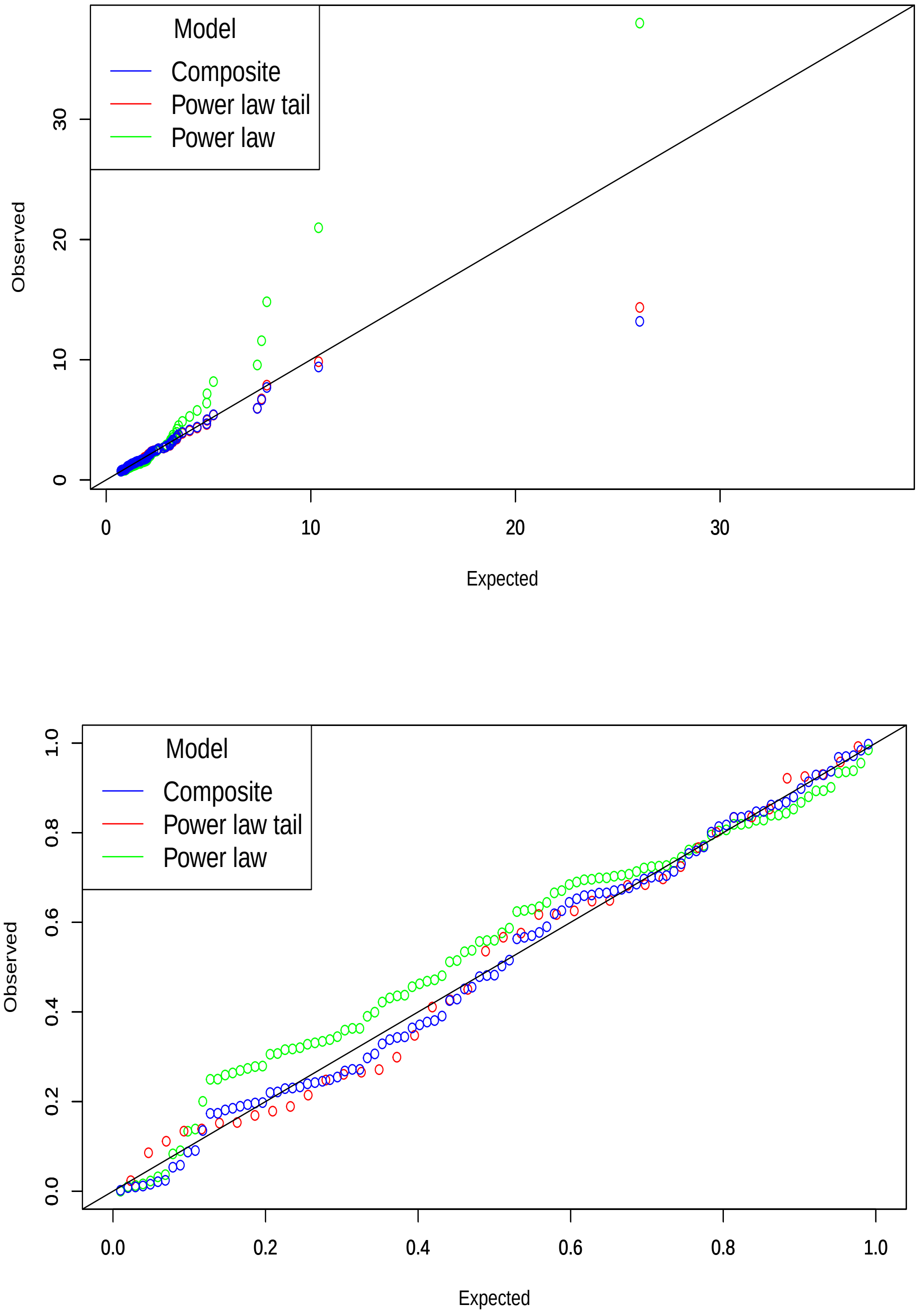
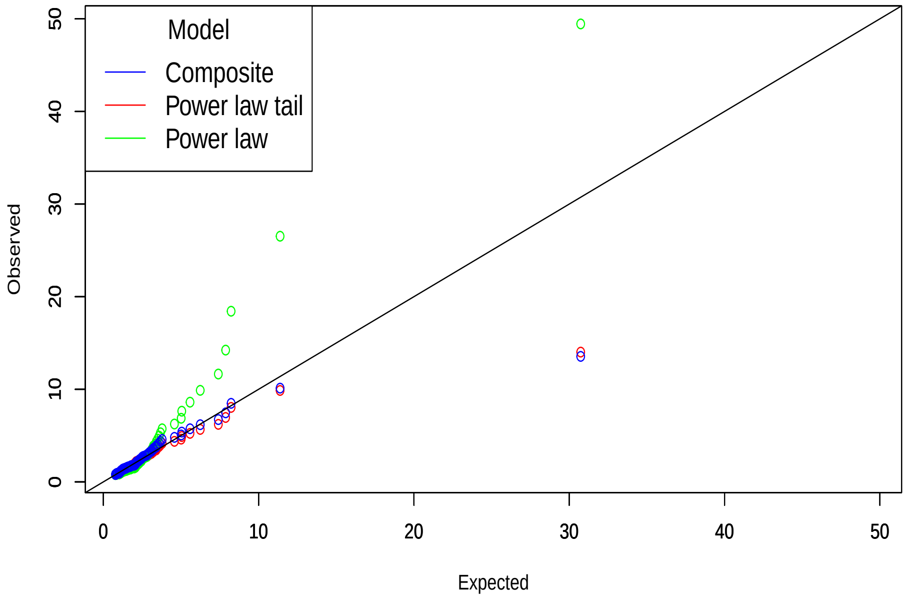


**Figure 3**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse Burr and power law distributions for Canadian net wealth data in 2014.



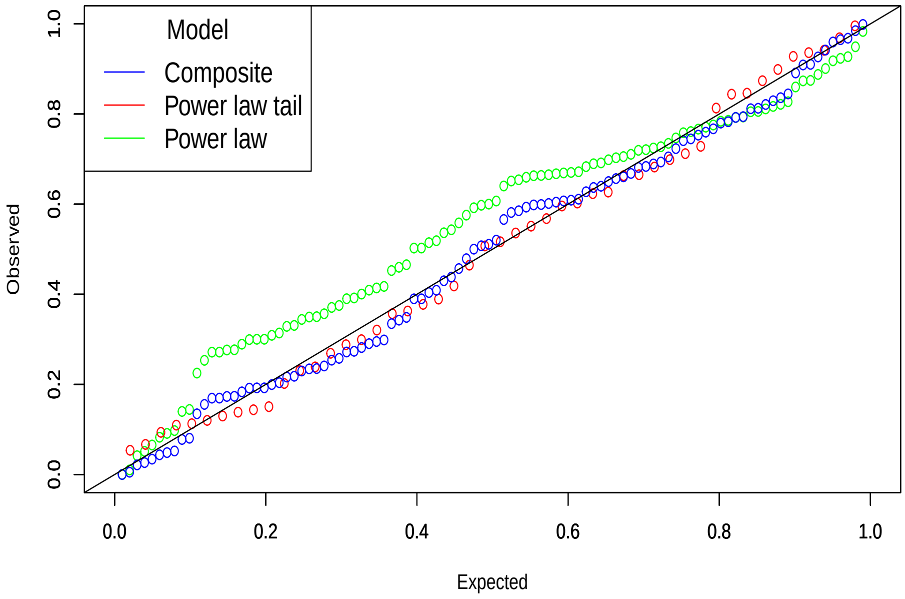


**Figure 4**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-generalised Pareto and power law distributions for Canadian net wealth data in 2015.

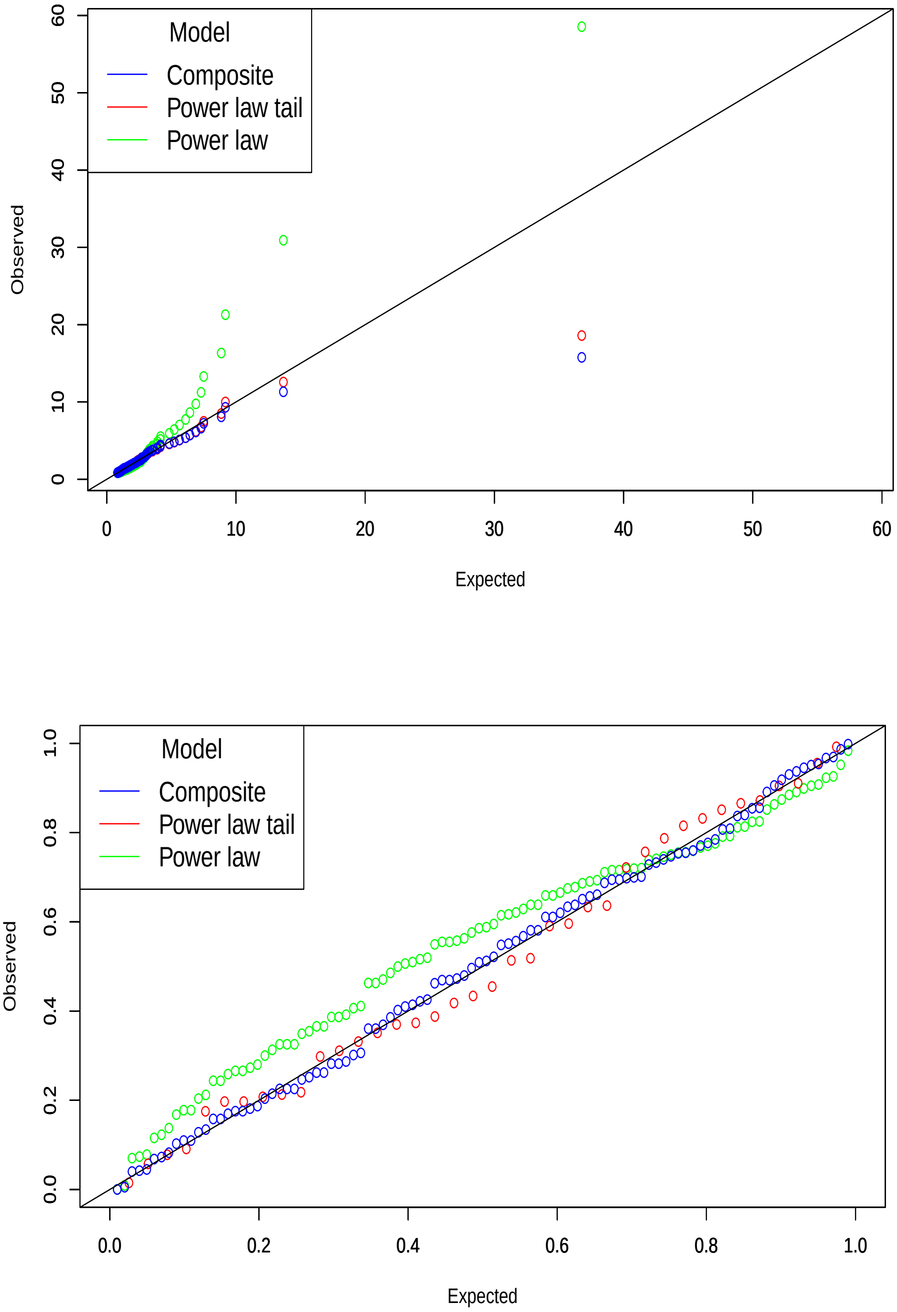


**Figure 5**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse Burr and power law distributions for Canadian net wealth data in 2016.

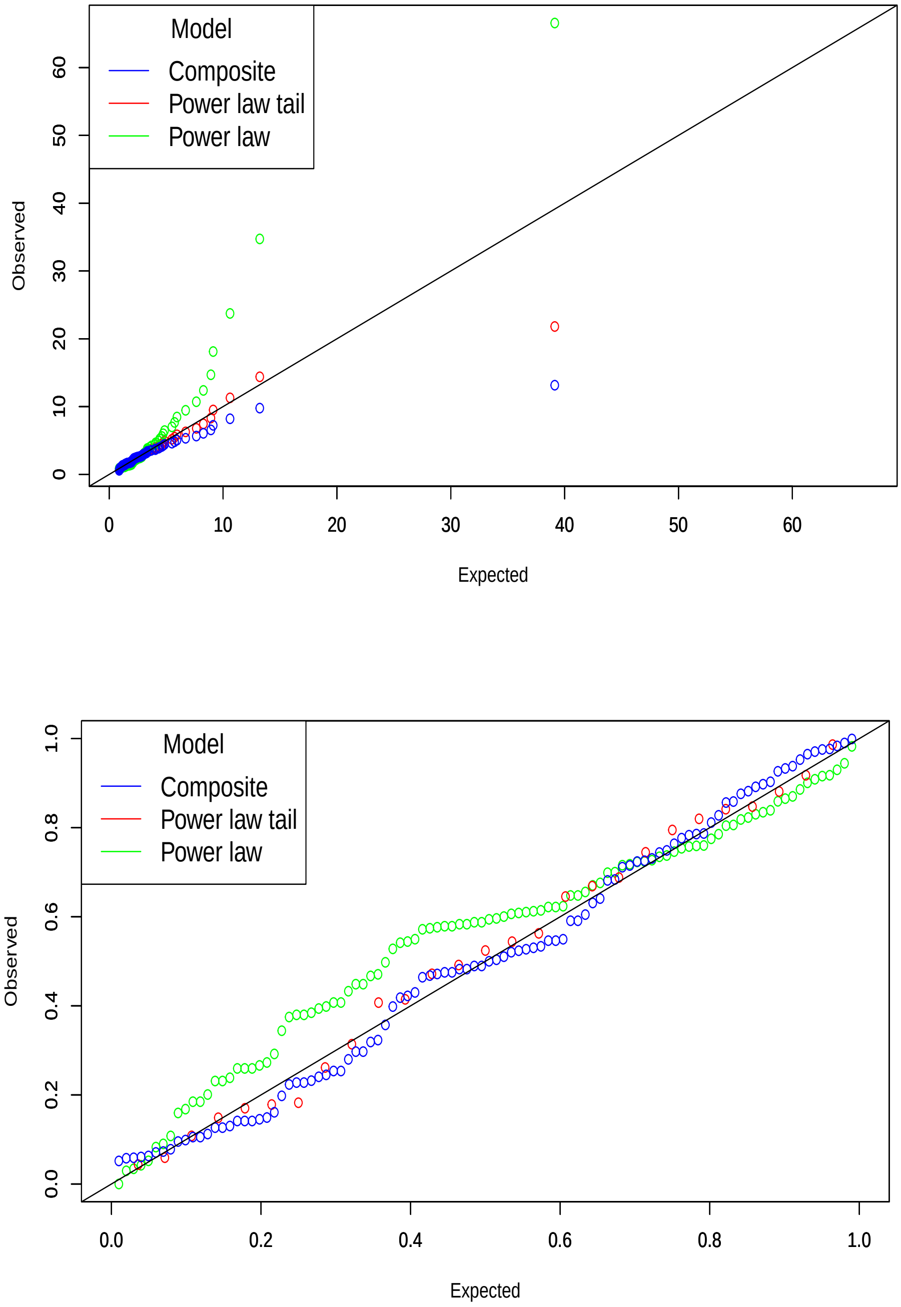


**Figure 6**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse paralogistic and power law distributions for Canadian net wealth data in 2017.

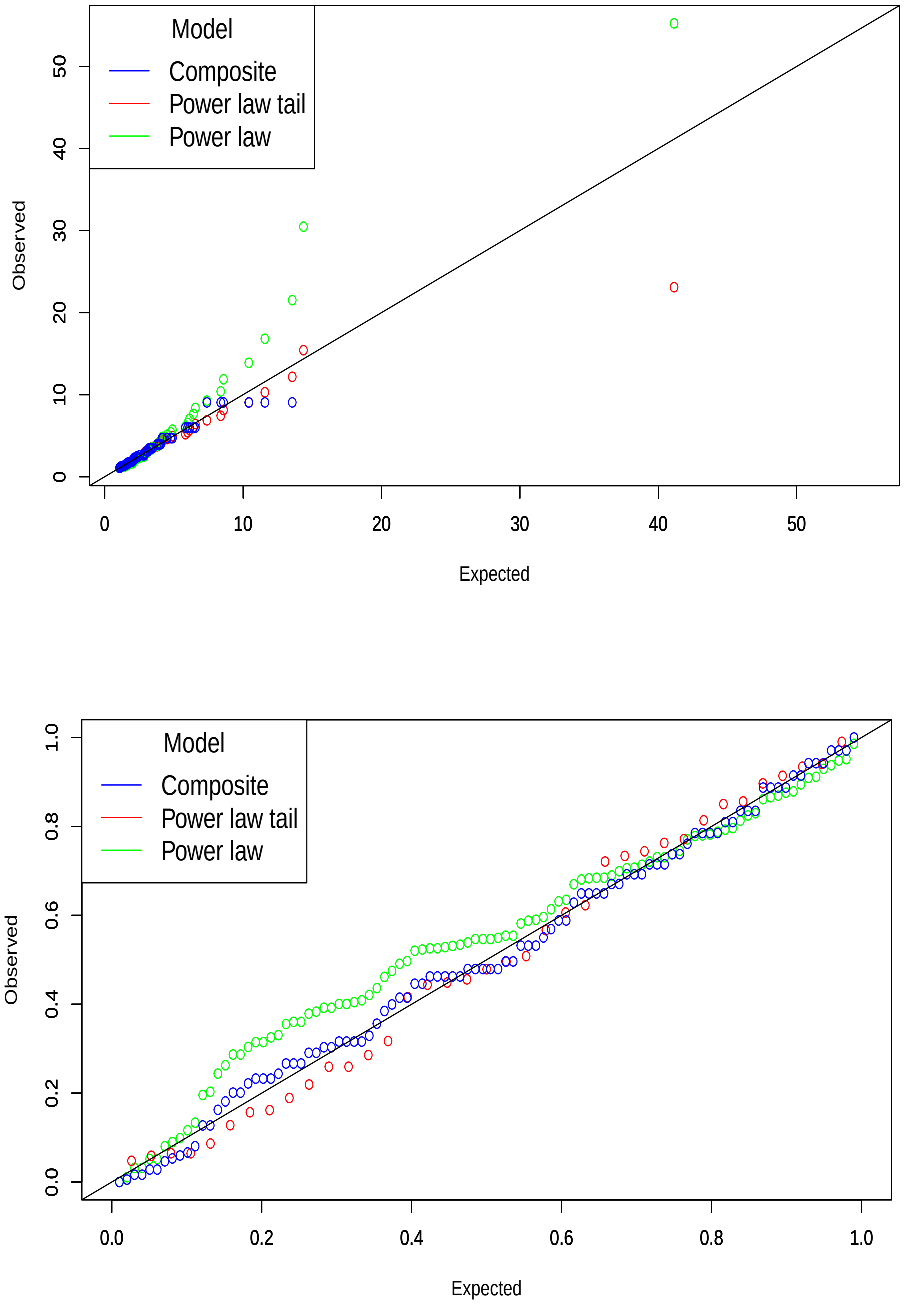


**Figure 7**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-inverse Burr and power law distributions for Canadian net wealth data in 2018.

## 4.2. Cumulative coal production data

Table 11 gives the best five distributions giving the smallest information criteria. The power law distribution once again does not make the best five distributions. The $p$-values of the best five distributions given in Table 12 range from 0.94 to 0.97. The largest $p$-value of 0.9723 is given by the composite lognormal-generalised Pareto distribution. The smallest $p$-value of 0.9407 is given by the composite lognormal-Pareto distribution. The composite lognormal-generalised Pareto distribution also gives the smallest information criteria values.

The fit of the power law distribution to the full data ($n = 4180$) gave a $p$-value $<$ $2.210^{-16}$. The fit of the power law distribution to the tail containing 1095 observations of the full data gave $\widetilde{K} = 1027417$ and $p$-value $= 0.108$. All of the $p$-values in Table 12 are much greater than 0.108.

**Table 11**: The five best composite lognormal distributions according to information criteria for cumulative coal production data.

| Model | $-2 \log L$ | AIC | BIC | HQC |
|---|---|---|---|---|
| Composite lognormal-generalised Pareto | −9464.96 | −9454.96 | −9429.60 | −9445.99 |
| Composite lognormal-log logistic | −9134.55 | −9126.55 | −9107.54 | −9119.83 |
| Composite lognormal-inverse paralogistic | −9270.34 | −9262.34 | −9243.33 | −9255.61 |
| Composite lognormal-paralogistic | −9219.03 | −9211.03 | −9192.01 | −9204.30 |
| Composite lognormal-Pareto | −9410.33 | −9402.33 | −9383.32 | −9395.60 |

**Table 12**: $p$-values for the five best composite lognormal distributions for cumulative coal production data.

| Model | $p$-values | | |
|---|---|---|---|
| | Kolmogorov–Smirnov | Anderson Darling | Cramer von Mises |
| Composite lognormal-generalised Pareto | 0.972 | 0.971 | 0.974 |
| Composite lognormal-log logistic | 0.965 | 0.964 | 0.966 |
| Composite lognormal-inverse paralogistic | 0.963 | 0.960 | 0.959 |
| Composite lognormal-paralogistic | 0.944 | 0.950 | 0.948 |
| Composite lognormal-Pareto | 0.941 | 0.940 | 0.938 |

The probability and quantile plots comparing the fits of the power law and composite lognormal-generalised Pareto distributions are shown in Figure 8.

The probability plot shows that the composite lognormal-generalised Pareto distribution provides a near perfect fit. The quantile plot shows that the composite lognormal-generalised Pareto distribution provides a good fit except for some extremely large observations. Neither of the two power law models provide as good a fit as the composite lognormal-generalised Pareto distribution.

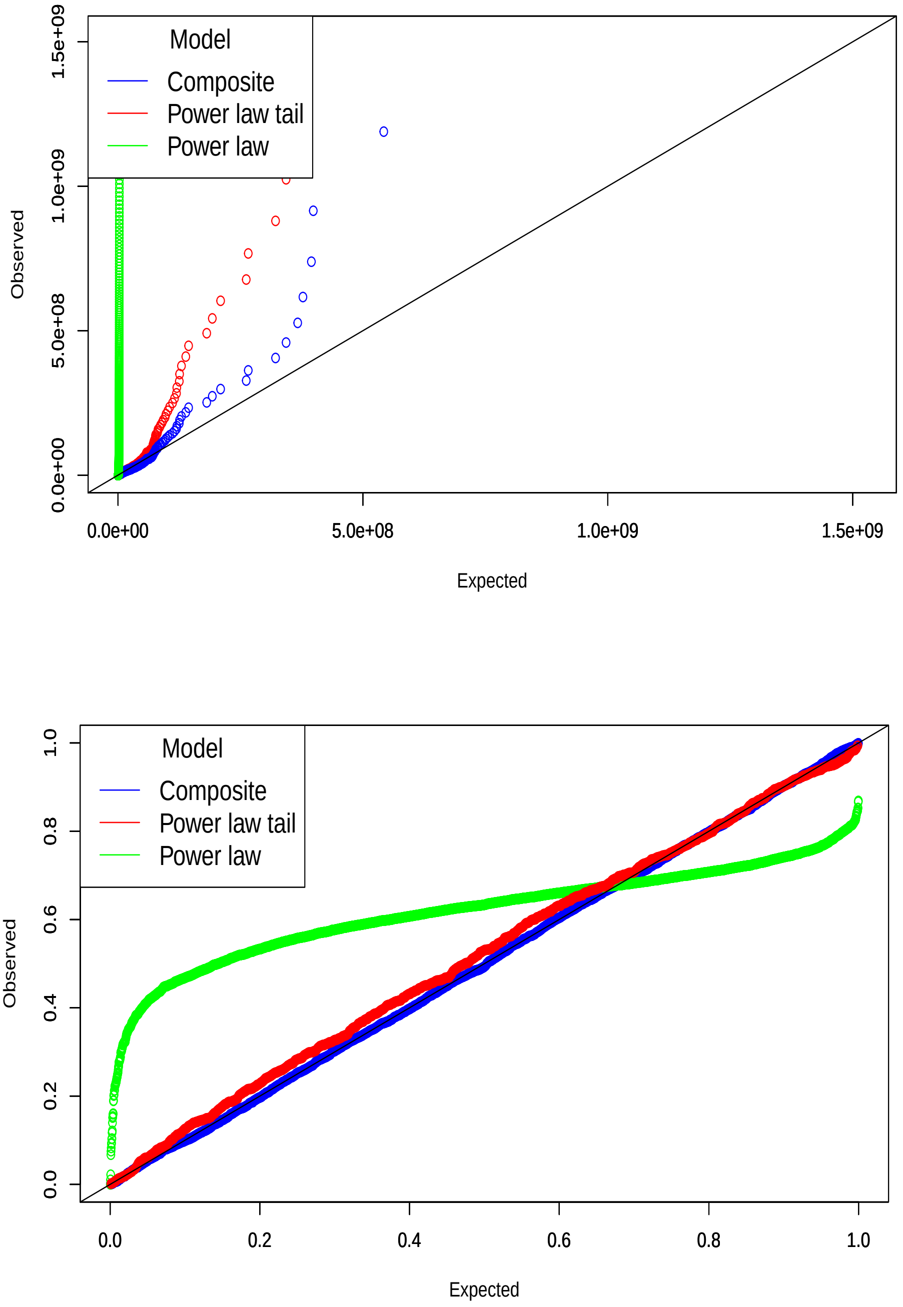


**Figure 8**: Quantile (left) and probability (right) plots for the fits of the composite lognormal-generalised Pareto and power law distributions for cumulative coal production data.

### 4.3. A simulation study

In this section, we assess the performance of the maximum likelihood estimates given by (2.3) with respect to sample size $n$. The assessment of the performance of the maximum likelihood estimates of $(\theta, \sigma, \mathbf{\Lambda})$ is based on a simulation study:

1. Generate ten thousand samples of size $n$ from the composite lognormal distribution by inverting
$$F(x) = u_k$$
for $k = 1, 2, ..., n$, where $u_1, u_2, ..., u_n$ is a random sample from uniform$(0, 1)$ and $F$ is given by (2.1);
2. Compute the maximum likelihood estimates for the ten thousand samples in step 1, say $\left(\widehat{\theta}_i, \widehat{\sigma}_i, \widehat{\mathbf{\Lambda}}_i\right)$ for $i = 1, 2, ..., 10000$.
3. Compute the biases and mean squared errors given by
$$\widehat{\text{bias}}_e(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{e}_i - e\right)$$
and
$$\widehat{\text{MSE}}_e(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{e}_i - e\right)^2$$
for $e = \theta, \sigma, \mathbf{\Lambda}$.

We repeated these steps for $n = 10, 11, ..., 500$ with $\theta = 1$, $\sigma = 1$ and $\mathbf{\Lambda}$ corresponding to the composite lognormal-inverse paralogistic distribution; so, computing $\widehat{\text{bias}}_\theta(n)$, $\widehat{\text{bias}}_\sigma(n)$, $\widehat{\text{bias}}_{\lambda_1}(n)$, $\widehat{\text{bias}}_{\lambda_2}(n)$, $\widehat{\text{MSE}}_\theta(n)$, $\widehat{\text{MSE}}_\sigma(n)$, $\widehat{\text{MSE}}_{\lambda_1}(n)$ and $\widehat{\text{MSE}}_{\lambda_2}(n)$ for $n = 10, 11, ..., 500$.

Figures 9 and 10 show how the biases and the mean squared errors vary with respect to $n$. The red line corresponds to the biases being zero. The following observations can be made:

1. The magnitude of the biases of the estimators generally decrease to zero;
2. The mean squared errors of the estimators generally decrease to zero;
3. The biases are generally negative for $\lambda_1$ and $\lambda_2$;
4. The biases appear largest in magnitude for $\lambda_1$ and $\lambda_2$;
5. The mean squared errors appear largest for $\theta$ and $\sigma$.

The results of the simulation study show that: the accuracy of the estimators of $\theta$, $\sigma$, $\lambda_1$ and $\lambda_2$ as measured by bias is reasonable for all $n \geq 300$; the accuracy of the estimators of $\theta$, $\sigma$, $\lambda_1$ and $\lambda_2$ as measured by mean squared error is reasonable for all $n \geq 300$. The sample size used in Section 4.2 is much greater than 300 but the sample sizes in Section 4.1 are not greater than 300. Hence, the conclusions in Section 4.2 should be reasonable but the conclusions in Section 4.1 should be treated conservatively.

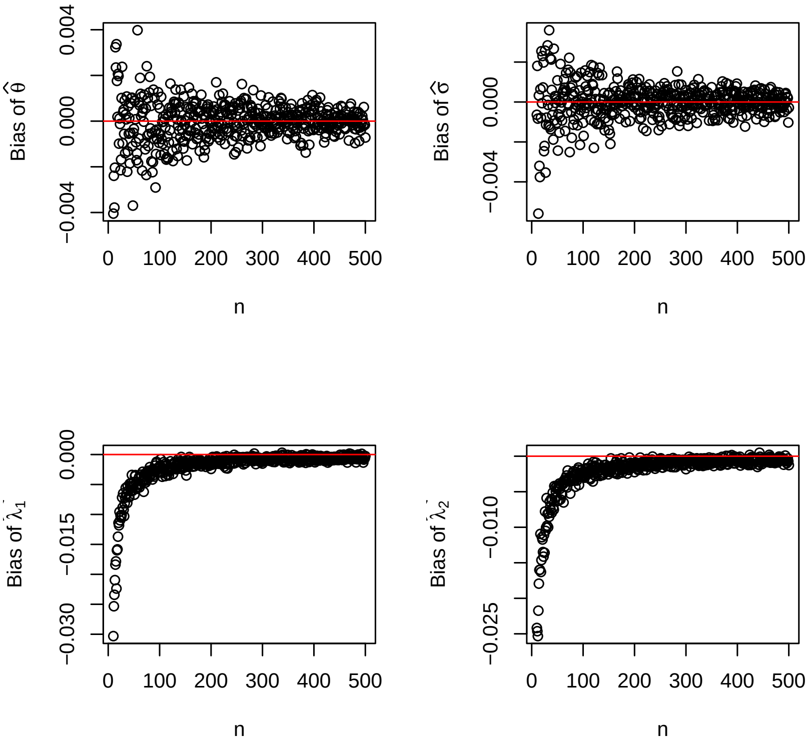


**Figure 9**: $\widehat{\text{bias}}_\theta(n)$, $\widehat{\text{bias}}_\sigma(n)$, $\widehat{\text{bias}}_{\lambda_1}(n)$ and $\widehat{\text{bias}}_{\lambda_2}(n)$ versus $n$.

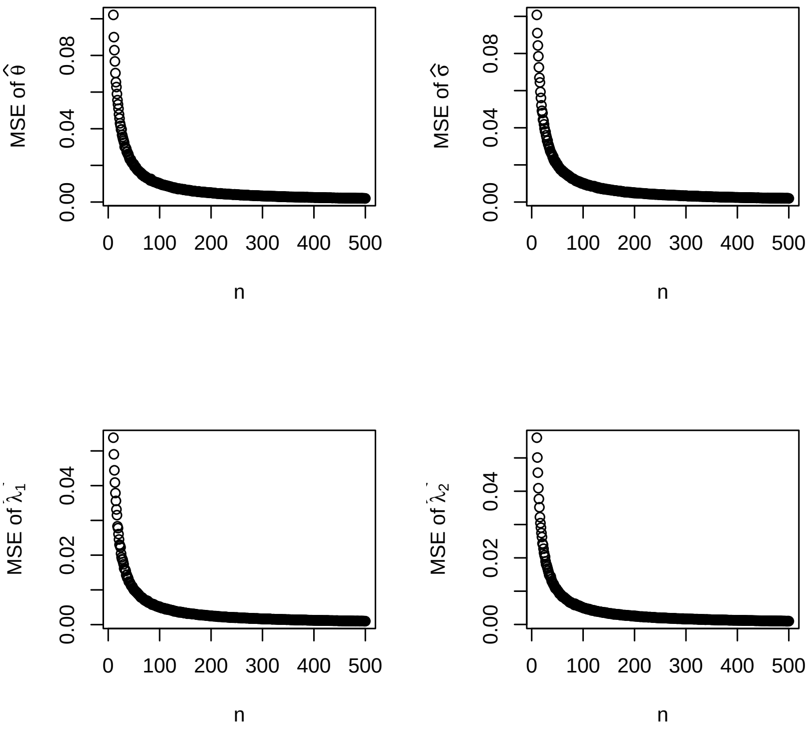


**Figure 10**: $\widehat{\text{MSE}}_\theta(n)$, $\widehat{\text{MSE}}_\sigma(n)$, $\widehat{\text{MSE}}_{\lambda_1}(n)$ and $\widehat{\text{MSE}}_{\lambda_2}(n)$ versus $n$.

We have presented results only for $\theta = 1$, $\sigma = 1$ and a particular composite lognormal distribution. But the results were similar for other choices for $\theta$ and $\sigma$ and other composite lognormal distributions.

## 5. CONCLUSIONS

In this paper, we have illustrated the power of composite lognormal distributions for two real data sets recently published in the physics literature. These data sets (in full or in part) have been previously modeled by the power law distribution. All of the composite lognormal distributions provide much better fits than the power law distribution when both were fitted to the full data sets. For the first data set, several of the composite lognormal distributions (composite lognormal-inverse Burr, composite lognormal-inverse paralogistic and composite lognormal-generalised Pareto distributions) provide better fits than the power law distribution even when the former were fitted to the full data and the latter was fitted only to the upper tail. For the second data set, all of the composite lognormal distributions provide much better fits than the power law distribution even when the former were fitted to the full data and the latter was fitted only to the upper tail. The goodness of fit was assessed by probability plots, quantile plots and $p$-values of the Kolmogorov–Smirnov, Anderson Darling and Cramer von Mises statistics. Software for fitting composite lognormal distributions is freely available from Nadarajah and Bakar [14].

Finally, we like to point out that the use of the power law distribution to model the two real data sets was motivated by a theoretical framework. Lee *et al.* [12] describe the rationale for the composite lognormal distributions in (2.1) as “the lognormal distribution models a large portion of the data well, but quickly fades away to zero. Thus it fits poorly a portion of the tail. On the other hand, $F_0$ fits the tail portion well, but fits the other portion poorly. By combining two distributions with one fitting the portion below a given threshold and the other fitting the portion larger than the threshold, the composite distribution (2.1) was proposed”. But to the best of our knowledge there is no theoretical motivation yet for the composite lognormal distributions. Finding a theoretical motivation for the composite lognormal distributions is a possible future work.

## ACKNOWLEDGMENTS

The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

## REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716–723.

[2] Aminzadeh, M.S. and Deng, M. (2019). Bayesian predictive modeling for inverse gamma-Pareto composite distribution, *Communications in Statistics – Theory and Methods*, **48**, 1938–1954.

[3] Balthrop, A. and Quan, S. (2019). The power-law distribution of cumulative coal production, *Physica A: Statistical Mechanics and Its Applications*, **530**, 121573.

[4] Calderín-Ojeda, E. (2015). On the composite Weibull – Burr model to describe claim data, *Communications in Statistics: Case Studies, Data Analysis and Applications*, **1**, 59–69.

[5] Calderín-Ojeda, E. (2016). The distribution of all French communes: A composite parametric approach, *Physica A – Statistical Mechanics and Its Applications*, **450**, 385–394.

[6] Calderín-Ojeda, E. (2018). A note on parameter estimation in the composite Weibull–Pareto distribution, *Risks*, **6**, doi: 10.3390/risks6010011

[7] Calderín-Ojeda, E. and Kwok, C.F. (2016). Modeling claims data with composite Stoppa models, *Scandinavian Actuarial Journal*, **9**, 817–836.

[8] Campolieti, M. (2018). Heavy-tailed distributions and the distribution of wealth: Evidence from rich lists in Canada, 1999–2017, *Physica A: Statistical Mechanics and Its Applications*, **503**, 263–272.

[9] Cooray, K. and Ananda, M.M.A. (2005). Modeling actuarial data with a composite lognormal-Pareto model, *Scandinavian Actuarial Journal*, 321–334.

[10] Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society, B*, **41**, 190–195.

[11] Kim, Y.; Kim, H.; Lee, G. and Min, K.-H. (2019). A modified hybrid gamma and generalized Pareto distribution for precipitation data, *Asia-Pacific Journal of Atmospheric Sciences*, **55**, 609–616.

[12] Lee, C.; Famoye, F. and Alzaatreh, A.Y. (2013). Methods for generating families of univariate continuous distributions in the recent decades, *Wiley Interdisciplinary Reviews: Computational Statistics*, **5**, 219–238.

[13] Mutali, S. and Vernic, R. (2020). On the composite lognormal – Pareto distribution with uncertain threshold, *Communications in Statistics – Simulation and Computation*, doi: 10.1080/03610918.2020.1743860

[14] Nadarajah, S. and Bakar, S.A.A. (2013). CompLognormal: An R package for composite lognormal distributions, *R Journal*, **5**, 97–103.

[15] Nadarajah, S. and Bakar, S.A.A. (2014). New composite models for the Danish fire insurance data, *Scandinavian Actuarial Journal*, 180–187.

[16] R Core Team (2022). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[17] Schwarz, G.E. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.