



On parameter estimation in multi-parameter distributions

I.J.H. Visagie^{1,2}

¹*Statistics Department, University of Pretoria, South Africa*

²*DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS)*

Abstract Many-multi parameter distributions have limit cases containing fewer parameters. This paper demonstrates that, when fitting distributions to data realized from a distribution resembling one of these limit cases, the parameter estimates obtained vary wildly between estimators. Special attention is paid to the modelling of financial log-returns. Two classes of estimators are used in order to illustrate the behaviour of the parameter estimates; the maximum likelihood estimator and the empirical characteristic function estimator. This paper discusses numerical problems associated with the maximum likelihood estimator for certain distributions and proposes a solution using Fourier inversion. In addition to simulation results, parameter estimates are obtained by fitting the normal inverse Gaussian and Meixner distributions to smooth bootstrap samples from the log-returns of the Dow Jones Industrial Average index are included as examples.

Keywords Maximum likelihood estimator, empirical characteristic function estimator, Fourier inversion, normal inverse Gaussian distribution, Meixner distribution, log-returns.

AMS 2010 subject classifications 62F10, 62G09

DOI: 10.19139/soic.v6i3.583

1. Introduction and motivation

Multi-parameter distributions are used to model a variety of phenomena. When estimating the parameters of these distributions, the aim is often to accurately model the density function of the underlying distribution from which the data are realized (with little or no regard for the individual parameter estimates themselves). Consider, for instance, the use of a multi-parameter distribution in the modelling of financial log-returns in an option pricing model. The resulting option prices are functions of the density, but not of the individual parameter estimates.

Many multi-parameter distributions possess limit cases containing fewer parameters. The aim of this paper is to numerically demonstrate the instability of the resulting parameter estimates when a multi-parameter distribution is fitted to a distribution resembling one of these limit cases. In these cases, the instability in the parameter estimates can, at least partially, be explained by noting that a distribution containing redundant parameters is fitted to the data.

Two different estimators are used in order to illustrate the instability of the parameter estimates: the maximum likelihood estimator (*MLE*) and the empirical characteristic function estimator (*ECFE*). Many multi-parameter distributions have complicated densities. In some cases, the calculation of these densities lead to numerical problems. These problems and their effects on the calculation of the *MLE* are discussed in this paper. The *ECFE* is a popular alternative to the *MLE*, especially in cases where the characteristic function of the distribution has a form that is simpler to evaluate numerically than is the case for the density function.

The properties of the *MLE* are well-known and are available in many standard texts on asymptotic theory. For example, [4] discusses the strong consistency of *MLE*s together with the corresponding assumptions in Chapter

*Correspondence to: I.J.H. Visagie (Email: jaco.visagie@up.ac.za). Department of Statistics, University of Pretoria, South Africa.

4. The same text examines the asymptotic normality and asymptotic efficiency of *MLEs* in Chapters 18 and 20 respectively. Other excellent references on the properties of *MLEs* include Chapters 5 and 7 of [13] as well as Chapter 7 of [9].

The second estimation method, the *ECFE*, minimizes a distance measure between the characteristic function and its empirical counterpart. Minimum distance estimators, of which the *ECFE* is a special case, were introduced in a general setting in [15]. [11] proposed the use of the *ECFE* for estimating the parameters of the stable law. Thereafter, the asymptotic properties of this estimator for both the single and multi-parameter cases were considered in [8]. In the mentioned paper, the author showed that, under certain regularity conditions, the estimator is consistent and asymptotically normally distributed. The efficiency of the *ECFE* relative to the *MLE* is also considered in this paper. [5] presented results relating to the efficiency of the *ECFE*. A general exposition of the *ECFE* and related Fourier methods was presented in [6], while [7] considered the efficiency of empirical characteristic function procedures.

[16] provided an explanation of the use of the *ECFE* wherein the author favoured a practical exposition. The author pointed out that this estimation method may be preferred in cases where difficulties are encountered when applying likelihood based methods. The paper lists various assumptions implicit in the use of this estimation method and includes a number of examples. In addition to the Monte Carlo results presented, the author also provides a practical application, wherein the *ECFE* is used in order to fit a model to observed values of the Dow Jones Industrial Average index.

In this paper, the normal inverse Gaussian ($N \circ IG$) and Meixner distributions are used as specific examples in the context of financial modelling. These distributions are chosen since they are popular models for financial log-returns.

The remainder of the paper is structured as follows. In Section 2, we discuss the estimators used as well as the method used to obtain starting values for the numerical optimization procedures required in order to calculate the estimates. Five empirical examples are shown in Section 3; the examples relating to the $N \circ IG$ distribution and the Meixner distribution are discussed separately. Section 4 is the conclusion.

2. Estimation methods

Two different estimators, the maximum likelihood estimator (*MLE*) and the empirical characteristic function estimator (*ECFE*), are used in order to obtain the empirical results shown in Section 3.

2.1. The maximum likelihood estimator

The density functions of many multi-parameter distributions are known. In these cases it is, in principle, possible to calculate *MLEs*. Usually, no formulae are available for the *MLEs* and numerical optimization is used to maximize the likelihood function. However, as a result of the limited computational power of the computer packages used, the calculation of the density function (and likelihood function) may break down for certain parameter sets.

Often, distributions with complicated density (and likelihood) functions have much simpler characteristic functions. In these cases, the well-known Fourier inversion formula can be used to evaluate the density numerically. Denote by $f(x; \theta)$ and $\phi(t; \theta)$ the density and characteristic functions of some distribution with parameter set θ . The Fourier inversion formula can be expressed as:

$$f(x; \theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t; \theta) dt; \quad (1)$$

see [3]. In cases where the direct numerical calculation of the density function breaks down, (1) can be used to calculate the density function (and the likelihood function).

For distributions for which numerical problems are encountered when calculating the density function we distinguish between two methods of obtaining *MLEs*. In the first one we use an algorithm that sets the likelihood function to 0 for parameter sets for which the density cannot be calculated directly; we call this method *DMLE* (direct *MLE*). In effect, this method maximizes the likelihood function by varying the parameter estimates subject

to the restriction that the density function can be calculated directly (without the use of Fourier inversion). In the second method, (1) is used to evaluate the likelihood when direct calculation of the likelihood is not feasible. This method is referred to as *IMLE* (indirect *MLE*). Depending on the form of the characteristic function of the distribution considered, the calculation of the *IMLEs* may require substantial amounts of computer time.

2.2. The empirical characteristic function estimator

The *ECFE* obtains parameter estimates by minimizing a given distance measure between the theoretical characteristic function of the distribution and the empirical version of this function. The characteristic functions of the $N \circ IG$ and Meixner distributions are substantially simpler than the corresponding density functions, meaning that the *ECFE* is simpler to calculate for these distributions. Two variations of the *ECFE* are considered, the difference between the two estimators being the weight function used.

The empirical characteristic function of the dataset x_1, x_2, \dots, x_n is

$$\phi_n(u) = \frac{1}{n} \sum_{j=1}^n \exp(iux_j).$$

The *ECFE* of θ is defined as the minimizer over θ of the expression

$$D_n(\theta) = \int_{-\infty}^{\infty} |\phi_n(u) - \phi(u; \theta)|^2 w(u) du, \quad (2)$$

where w is some positive weight function satisfying

$$\int_{-\infty}^{\infty} w(u) du < \infty.$$

Two variations of the weight function are used:

$$w_1(u) = \exp(-|u|), \quad (3)$$

and

$$w_2(u) = \exp(-u^2). \quad (4)$$

Below, *ECFEA* and *ECFEN* respectively denote the estimators obtained using (3) and (4). As was the case for the *MLEs*, in general, no formulae exist for the *ECFEs* so numerical optimization is employed in order to calculate the estimates.

2.3. Starting values

The numerical optimization procedures used to obtain parameter estimates require starting values. Since optimal starting values are generally unknown in practice, it is often advisable to make use of a random procedure to choose starting values. In the numerical results shown below, starting values for the *MLEs* are obtained as follows. First, a range for the starting values of each of the parameters of the distribution is specified; this range is specified by the user and is chosen so as to include values deemed likely based on previous experience. Then, a possible starting value for each of the parameters is simulated independently from a uniform distribution. Next, the likelihood function is calculated for the realized parameter set. Finally, this process is repeated a large number of times and the parameter set associated with the largest value of the likelihood function is chosen as the set of starting values for the optimization procedure.

A similar approach is used in order to obtain starting values for the *ECFEs*. In this case, the distance defined in (2) is calculated and the parameter set resulting in the smallest distance is used as the starting values for the optimization algorithm.

The numerical results shown below were obtained by generating 10 000 random possible starting values for each optimization procedure used. The ranges chosen for the starting values are omitted for the sake of brevity.

In the case of the *IMLE*, the time required to calculate the likelihood function for all of the possible starting values is extreme; in some cases, the calculation of the *IMLE* required roughly 1 000 times as much computer time as was the case for the *DMLE*. In order to alleviate this computation burden, the likelihood function is not calculated for each possible set of starting values. Instead, the procedure used is as follows. The mean and the standard deviation of the dataset under consideration are calculated. The mean of the theoretical distribution for each of the parameter sets considered as possible starting values is also calculated. Thereafter, the likelihood function is only evaluated in cases where the mean of the theoretical distribution is within three standard deviations of the mean of the dataset. The remaining combinations of possible starting values are discarded. Similar techniques to reduce the time required to calculate the remaining estimates are deemed unnecessary.

The aim of this paper is to illustrate the instability of the parameter estimates obtained when using different estimators for multi-parameter distributions. In terms of the above-mentioned method, the algorithms corresponding to the different estimators may use different starting values. The author would like to emphasize that this is not the cause of the vast differences between the estimates obtained in the empirical results shown below. Given a dataset, the parameter estimates calculated using the different estimators vary wildly even when the procedures use the same starting values.

3. Empirical results

The empirical results shown below were achieved using the $N \circ IG$ and Meixner distributions. [1] presents a set of properties common to all financial returns. In his paper, the author argues for the use of distributions with at least four parameters as models for financial returns. Note that both the $N \circ IG$ and Meixner distribution each contain four parameters. Three examples relating to the $N \circ IG$ distribution and two relating to the Meixner distribution are presented. Some of the properties of these distributions are highlighted below before the examples are presented.

The empirical results relating to the first and second application of the $N \circ IG$ distribution were obtained as follows. 250 realizations (which roughly corresponds to the number of business days in a year) were simulated from a $N \circ IG$ distribution with known parameter set. The parameters were then estimated using each of the *DMLE*, *IMLE*, *ECFEA* and *ECFEN*. This process was repeated 1 000 times in each example. A similar approach is used in order to obtain the results presented in the fourth example, where the $N \circ IG$ distribution is replaced by the Meixner.

Unlike the examples mentioned above, the third and final examples are not concerned with data simulated from a known distribution. These examples are based on the observed log-returns of the Dow Jones Industrial Average (DJIA) index for the period starting on January 1, 2017 and ending on January 1, 2018. This index is comprised of a price-weighted average of 30 large publicly owned companies in the United States. A total of 250 log-returns were calculated from the prices for this index, retrieved from <http://finance.yahoo.com>.

Samples were drawn from the distribution underlying the calculated log-returns using the smooth bootstrap. (See [2].) The procedure used to sample from the underlying distribution is as follows. The bandwidth h used in the simulation was calculated using Silverman's rule of thumb;

$$h = \left(\frac{4}{3n} \right)^{1/5} \hat{\sigma},$$

where $\hat{\sigma}$ denotes the sample standard deviation and n denotes the sample size (250 in this case). The value of h was calculated to be 0.0015 for this dataset. Then, smooth bootstrap samples were achieved by resampling 250 values from the observed log-returns with replacement and adding an independent realizations from a $N(0, h^2)$ distribution to each sampled value. This process was repeated 1 000 times. The $N \circ IG$ and Meixner distributions were fitted to each of the resulting samples.

As mentioned above, both of the distributions considered are characterized by four parameters. For the sake of brevity, this paper discusses the estimates obtained for one of these parameters only in each example. In the case of the $N \circ IG$ distribution, the estimates of the parameter denoted by α are considered throughout. In the case of the Meixner distribution, the estimates achieved for the scale parameter α are considered in the simulated

example, while the results relating to the shape parameter δ are used in the analysis of the log-return data. Similar configurations were observed for the estimates of the remaining parameters. Boxplots are employed to showcase the resulting values of the chosen parameter estimates. Note that a cross is used in order to indicate an outlier in the boxplots presented throughout.

The numerical results achieved below were calculated using Matlab. The `fminsearch.m` subroutine in Matlab was used in order to calculate the presented parameter estimates. The mentioned subroutine uses the Nelder-Mead optimization algorithm. (See [10].)

3.1. The $N \circ IG$ distribution

A random variable $X : \Omega \rightarrow \mathbb{R}$ is said to follow a $N \circ IG$ distribution with parameter set $\theta = (\alpha, \beta, \mu, \delta)$ if it has density

$$f(x; \theta) = \frac{\alpha\delta}{\pi} \exp\left(\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)\right) \frac{K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)}{\sqrt{\delta^2 + (x - \mu)^2}}, \quad (5)$$

where $\alpha > 0$, $|\beta| < \alpha$, $\mu \in \mathbb{R}$, $\delta > 0$, and K_1 denotes the modified Bessel function of the third kind with index 1.

For the definition of this Bessel function; see [12]. The characteristic function of the $N \circ IG(\alpha, \beta, \mu, \delta)$ distribution is

$$\phi(t; \theta) = \exp\left(i\mu t - \delta\left(\sqrt{\alpha^2 - (\beta + it)^2} - \sqrt{\alpha^2 - \beta^2}\right)\right). \quad (6)$$

[14] introduced an alternative parameterization for the $N \circ IG$ distribution. Results similar to those shown below were obtained when using this parameterization.

When evaluating the density function of the $N \circ IG$ distribution for certain parameter sets, numerical difficulties are encountered because of the limited computation power available. The Bessel function $K_1(z) \downarrow 0$ as $z \uparrow \infty$. Since most computer packages used for numerical work round all sufficiently small positive numbers to 0, the Bessel function is rounded to 0 for sufficiently large arguments. If $\alpha \uparrow \infty$ and $\delta \uparrow \infty$, then the exponential and Bessel functions in (5) both have large arguments. If these arguments are large enough, the computer package used sets the value of the Bessel function to 0 and that of the exponential function to ∞ so the calculation of the density function breaks down. This problem can be remedied using characteristic function inversion. The $N \circ IG$ distribution is an example of a multi-parameter distributions with a special case containing fewer parameters.

Theorem 1

Let $X \sim N \circ IG(\alpha, \beta, \mu, \delta)$. If $\alpha \uparrow \infty$, $\delta \uparrow \infty$ and $\delta/\alpha \rightarrow \sigma^2$ while β is held constant, then the distribution of X converges to the $N(\mu + \beta\sigma^2, \sigma^2)$ distribution.

For a proof, see the appendix. The first example below simulates from a $N \circ IG$ with substantial skewness and excess kurtosis. The second example resembles the normal in that it exhibits little skewness and excess kurtosis. The two examples are contrasted in order to demonstrate the difference in the behavior of estimates when the true distribution resembles its limiting case and when it does not.

3.1.1. Example 1: The $N \circ IG(1, 0.5, 0, 1)$ distribution Figure 1 shows boxplots of the realized values of $\hat{\alpha}$ calculated using the four estimators discussed above. The boxplots suggest that the distribution of each of the estimators is positively skewed. The plots also clearly indicate that there are differences in the values obtained using the various estimators. Nevertheless, the median of each of the estimators is close to the true parameter value of 1. Furthermore, the largest estimate calculated using the *MLE* is less than 4, while the largest estimate associated with the *ECFE* is less than 14.

Below, the moments associated with the fitted distributions are compared to the moments of the underlying distribution. The first four standardized central moments (hereafter simply referred to as the moments) of the $N \circ IG$ distribution associated with each set of parameter estimates are calculated; let $\nu_j(\theta)$ denote the j^{th} moment of the distribution with parameter set θ . Table 1 shows the average and standard deviation (in brackets) of the moments associated with each estimator. The moments of the $N \circ IG(1, 0.5, 0, 1)$ distribution are included in the table (at the top of the table) for comparison.

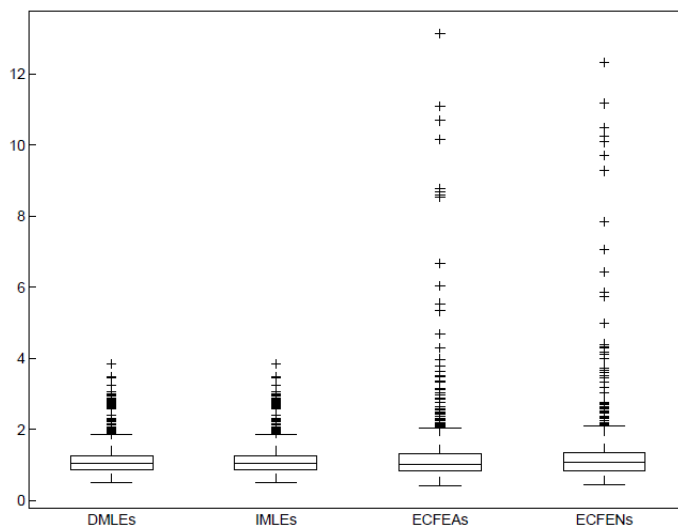


Figure 1. Boxplot of the estimates of α in Example 1.

	$\nu_1(\hat{\theta})$	$\nu_2(\hat{\theta})$	$\nu_3(\hat{\theta})$	$\nu_4(\hat{\theta})$
True values	0.577	1.540	1.612	9.928
<i>DMLE</i>	0.572 (0.078)	1.537 (0.277)	1.595 (0.458)	10.186 (3.351)
<i>IMLE</i>	0.572 (0.078)	1.537 (0.277)	1.595 (0.458)	10.185 (3.352)
<i>ECFEA</i>	0.576 (0.089)	1.581 (0.354)	1.639 (0.577)	10.738 (4.378)
<i>ECFEN</i>	0.575 (0.083)	1.562 (0.322)	1.619 (0.543)	10.474 (4.299)

Table 1: The moments associated with the various estimates obtained in Example 1.

Table 1 indicates that the moments associated with the fitted distributions closely resemble those of the true distribution. This observation, combined with the seemingly accurate estimates of α displayed in Figure 1, indicates that the estimation procedures perform satisfactorily in this case.

3.1.2. *Example 2: The $N \circ IG(50, 20, 2, 20)$ distribution* Figure 2 shows a boxplot of the *DMLEs* of α achieved by fitting the $N \circ IG$ distribution to the realized samples of the $N \circ IG(50, 20, 2, 20)$ distribution. The boxplots of the remaining estimators for α are shown in Figure 3.

In contrast to the *DMLE*, the estimators shown in Figure 3 clearly overestimate the true value of α (note that the scale on the vertical axis of the figure is 10^5). The large estimates of α can be explained by recalling that the $N \circ IG$ distribution approaches the normal distribution for large values of α and δ . In effect, we are fitting a (four-parameter) $N \circ IG$ distribution to data generated from a distribution resembling a (two-parameter) normal distribution.

Clearly the estimates calculated using the various estimators differ substantially. Moreover, it seems that no simple relations exist between the estimates. In order to illustrate the relationship between the estimates obtained, the correlations between the logarithms of the estimates are considered; logarithms are taken because the distributions of the estimates are positively skewed. The resulting correlation matrix is shown in Table 2. Note that the majority of the off-diagonal entries in this correlation matrix are close to 0.

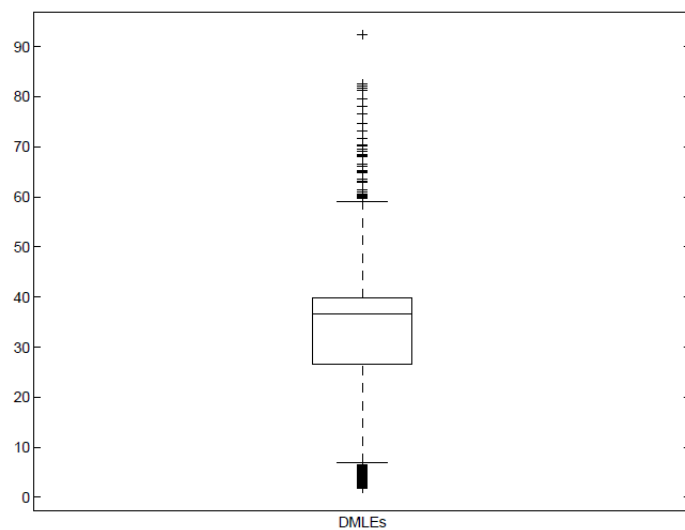


Figure 2. Boxplot of the $DMLEs$ obtained in Example 2.

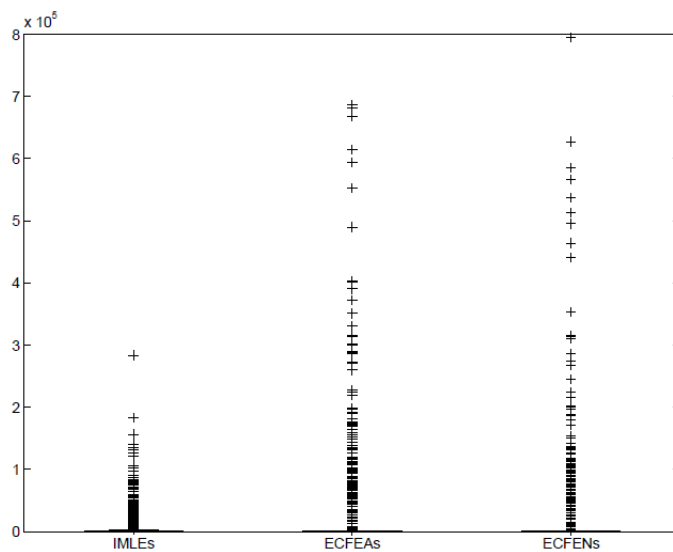


Figure 3. Boxplot of the remaining estimates of α in Example 2.

	$DMLE$	$IMLE$	$ECFEA$	$ECFEN$
$DMLE$	1.000	0.051	0.022	0.033
$IMLE$	0.051	1.000	0.107	0.035
$ECFEA$	0.022	0.107	1.000	0.047
$ECFEN$	0.033	0.035	0.047	1.000

Table 2: Correlation matrix of the logarithm of the estimates for α obtained in Example 2.

Keeping the vast differences in the parameter estimates, the associated density functions are more similar than one might expect. Let $\hat{\theta}_{DMLE}$ denote an estimate obtained using the $DMLE$, and let $\hat{\theta}_{IMLE}$ denote an estimate

obtained using the *IMLE*. For one of the datasets, the following estimates were obtained:

$$\hat{\theta}_{DMLE} = (5.031, 0.318, 10.634, 2.508),$$

$$\hat{\theta}_{IMLE} = (17624, -1356, 683, 8716).$$

Figure 4 shows the density functions associated with $\hat{\theta}_{DMLE}$ (dashed line) and $\hat{\theta}_{IMLE}$ (dotted line). The true density is also included in the figure (solid line) for reference. The density functions achieved are similar in spite of substantial differences in the estimates of the parameters.

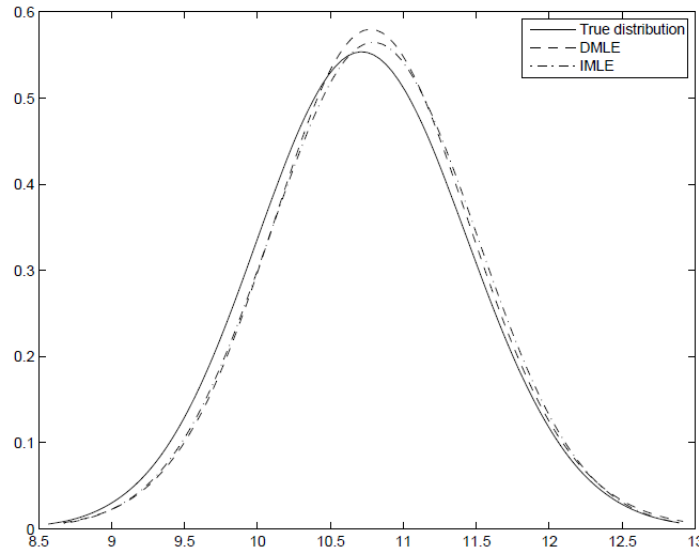


Figure 4. Comparison of estimated density functions.

In order to investigate the similarity between the fitted distributions that Figure 4 seems to imply, we turn to the moments associated with the fitted distributions. Table 3 shows the average and standard deviation (in brackets) of the moments associated with each estimator. As before, the true moments are included for reference.

	$\nu_1(\hat{\theta})$	$\nu_2(\hat{\theta})$	$\nu_3(\hat{\theta})$	$\nu_4(\hat{\theta})$
True values	10.729	0.520	0.040	3.005
<i>DMLE</i>	10.730 (0.045)	0.527 (0.079)	0.040 (0.140)	3.094 (0.197)
<i>IMLE</i>	10.730 (0.045)	0.518 (0.047)	0.043 (0.106)	3.048 (0.132)
<i>ECFEA</i>	10.728 (0.047)	0.522 (0.054)	0.036 (0.108)	3.056 (0.176)
<i>ECFEN</i>	10.728 (0.045)	0.520 (0.049)	0.020 (0.065)	3.026 (0.096)

Table 3: The moments associated with the various estimates obtained in Example 2.

The moments associated with each estimator closely resemble the moments of the true distribution. It is interesting to note that the closeness of the fitted densities to the true density (shown in Figure 4) and the close correspondence between the moments (illustrated in Table 3) imply that the estimators succeed in accurately modelling the shape of the true distributions. Nevertheless, the parameter estimates obtained by these estimators vary wildly.

3.1.3. *Example 3: The $N \circ IG$ distribution fitted to the log-returns of the DJIA* As a third example, the $N \circ IG$ distribution is fitted to samples obtained from the distribution of the log-returns of the DJIA. Figure 5 shows a boxplot of the *DMLEs* of α calculated using this distribution, while Figure 6 shows boxplots of the remaining estimates.

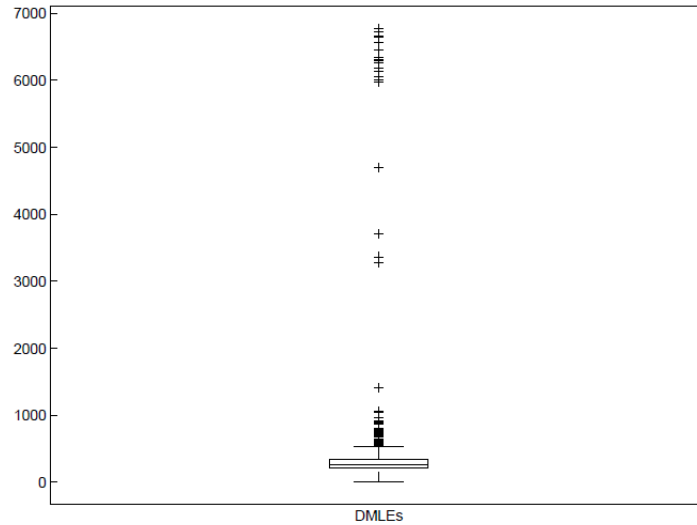


Figure 5. Boxplot of the *DMLEs* of α obtained in Example 3.

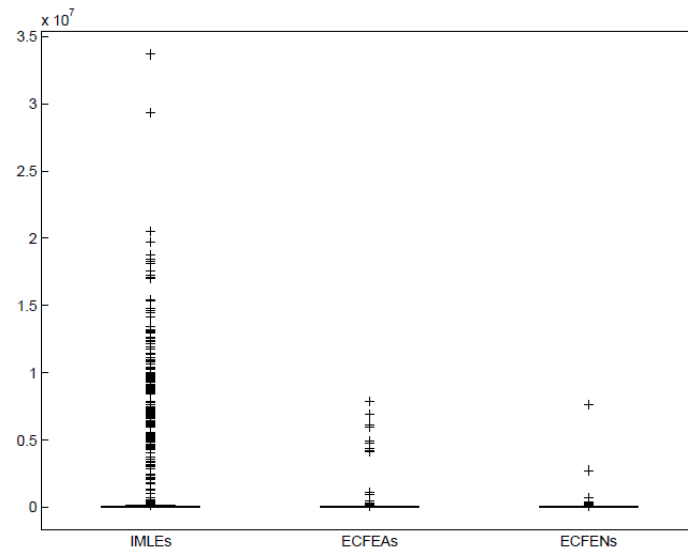


Figure 6. Boxplot of the remaining estimates for α obtained in Example 3.

As was the case in the previous example, it is clear that the estimates obtained by the various estimators differ substantially. It is also clear that extremely large estimates are often achieved. Following similar lines of inquiry as before, we are interested in determining whether or not the estimates associated with the different estimators are highly correlated. To this end, Table 4 contains the correlation matrix for the logarithms of the estimates calculated using the various estimators. The correlation between each pair of the different estimators is close to 0.

	<i>DMLE</i>	<i>IMLE</i>	<i>ECFEA</i>	<i>ECFEN</i>
<i>DMLE</i>	1.000	0.036	-0.000	0.031
<i>IMLE</i>	0.036	1.000	0.020	-0.007
<i>ECFEA</i>	-0.000	0.020	1.000	0.025
<i>ECFEN</i>	0.031	-0.007	0.025	1.000

Table 4: Correlation matrix of the logarithm of the estimates for α obtained in Example 3.

Next, we compare the moments of the fitted distributions to those of the data in order to determine the adequacy of the performance of the estimators. Table 5 shows the average and standard deviation (in brackets) of the moments associated with each estimator. Note that the entries relating to the first and second moments in Table 5 are multiplied by 10^3 and 10^4 respectively. This is done in order to ease comparisons between the results since these values are close to 0. The same convention is used in Table 8.

	$10^3 \times \nu_1(\hat{\theta})$	$10^4 \times \nu_2(\hat{\theta})$	$\nu_3(\hat{\theta})$	$\nu_4(\hat{\theta})$
Moments	0.871	0.174	-0.185	5.316
<i>DMLE</i>	0.844 (0.837)	4.689 (69.955)	0.121 (1.024)	8.202 (10.928)
<i>IMLE</i>	0.978 (0.326)	0.192 (0.025)	0.029 (0.223)	3.492 (1.128)
<i>ECFEA</i>	0.876 (0.268)	0.183 (0.027)	0.002 (0.028)	3.002 (0.010)
<i>ECFEN</i>	0.876 (0.267)	0.181 (0.026)	0.003 (0.032)	3.002 (0.019)

Table 5: The moments associated with the various estimates obtained in Example 3.

The results reported in Table 5 clearly indicate that the moments associated with the *DMLEs* do not perform as well as those associated with the remaining estimators. However, upon further investigation, it becomes clear that the *DMLE* algorithm is able to accurately model the underlying distribution in all but a few cases. In order to illustrate this, consider the variance implied by the various estimated parameter sets. The variance of the log-returns is 1.737×10^{-5} , while the mean of the variances implied by the obtained estimates is 4.168×10^{-3} . Nevertheless, the median of the implied variances is 2.009×10^{-5} , which closely corresponds to the variance of the log-returns. Figure 7 shows a boxplot of the variances implied by the *DMLEs*. The figure indicates that the large mean of the variances implied by *DMLEs* is a result of a few outliers and not a systematic overestimation. For a few samples, the *DMLE* algorithm was unable to converge to satisfactory results. This can, at least partially, be explained by the fact that the variance of the $N \circ IG$ distribution becomes small for large values of α , in which case direct evaluation of the likelihood function may not be possible.

The moments associated with the remaining parameter estimates correspond more closely to the moments of the true distribution. However, the remaining estimates slightly overestimate the skewness and the kurtosis of the log-returns. In this example, the implied moments achieved by the empirical characteristic function estimators are closer to those of the true distribution than is the case for the estimators based on the likelihood function.

3.2. The Meixner distribution

A random variable $X : \Omega \rightarrow \mathbb{R}$ is said to follow a Meixner distribution with parameter set $\theta = (\alpha, \beta, \mu, \delta)$ if it has density

$$f(x; \theta) = \frac{(2 \cos(\beta/2))^{2\delta}}{2\alpha\pi\Gamma(2\delta)} \exp\left(\frac{\beta(x - \mu)}{\alpha}\right) \left| \Gamma\left(\delta + \frac{i(x - \mu)}{\alpha}\right) \right|^2, \tag{7}$$

where $\alpha > 0$, $|\beta| < \pi$, $\mu \in \mathbb{R}$ and $\delta > 0$.

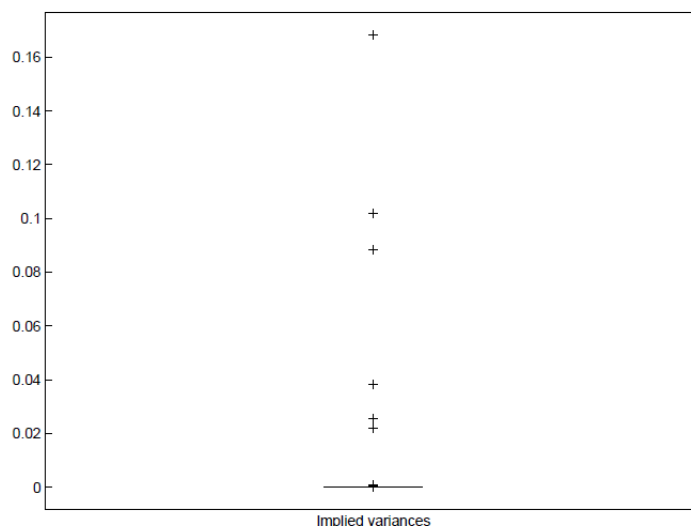


Figure 7. Boxplot of variances implied by the *DMLEs*.

The characteristic function of the *Meixner* $(\alpha, \beta, \mu, \delta)$ distribution is:

$$\phi(u; \theta) = \exp(i\mu u) \left(\frac{\cos(\beta/2)}{\cosh\left(\frac{\alpha u - i\beta}{2}\right)} \right)^{2\delta}.$$

Consider the density of the Meixner distribution given in (7). The functions making up this density function can lead to numerical difficulties: $\cos(\beta/2) \rightarrow 0$ if $\beta \rightarrow \pm\pi$, and $\Gamma(z) \uparrow \infty$ as $z \uparrow \infty$. As for the complex gamma function, if k is some constant and $z \uparrow \infty$, then $|\Gamma(z + ik)|^2 \uparrow \infty$. If, for example, the value of δ is excessively large, then both the gamma function and the complex gamma function take large values and the computer package used may set the value of both functions to ∞ . In this situation, the calculation of the density function breaks down.

Similar to the case for the $N \circ IG$ distribution, the Meixner distribution has the normal distribution as a limit case. The proof of the following theorem is deferred to the appendix.

Theorem 2

Let $X \sim \text{Meixner}(\alpha, \beta, \mu, \delta)$. If $\alpha \downarrow 0$, $\beta \rightarrow 0$ and $\delta \uparrow \infty$ so that $\alpha^2\delta/2 \rightarrow \sigma^2$ and $\beta/\alpha \rightarrow \gamma$, then the distribution of X converges to the $N(\mu + \gamma\sigma^2, \sigma^2)$ distribution.

3.2.1. Example 4: The Meixner $(0.1, -0.2, 0, 10)$ The parameter in this example are chosen such that the true distribution resembles the normal. Note, however, that the Meixner distribution presented in this example resembles the normal to a lesser degree than is the case for the $N \circ IG$ distribution of Example 2; the kurtosis of the distribution in the present example is 3.102 compared to the 3.005 of Example 2. As a result, the results presented below are less extreme than those of the mentioned example.

The Meixner distribution resembles the normal for small values of α . Since closeness to 0 is hard to judge on a figure, Figure 8 shows boxplots of $\hat{\alpha}^{-1}$. Note that the true value of $\alpha^{-1} = 10$. Several large outliers are omitted in the figure; the maximum values associated with the *IMLE* and *ECFE* respectively are 1 018 877 and 6 095.

The results presented in Figure 8 indicate that the value of α^{-1} is often overestimated by a substantial margin (meaning that the realized values of $\hat{\alpha}$ are often much closer to 0 than the true value). However, the estimators associated with the *ECFE* achieves medians which are close to the true value of α^{-1} , while the medians of both of the *MLEs* substantially overestimate this value. As before, the moments associated with the true and fitted distributions are compared in Table 6. The main entries in the table are the averages of the moments associated with the fitted distributions, while the associated standard deviations are reported in brackets.

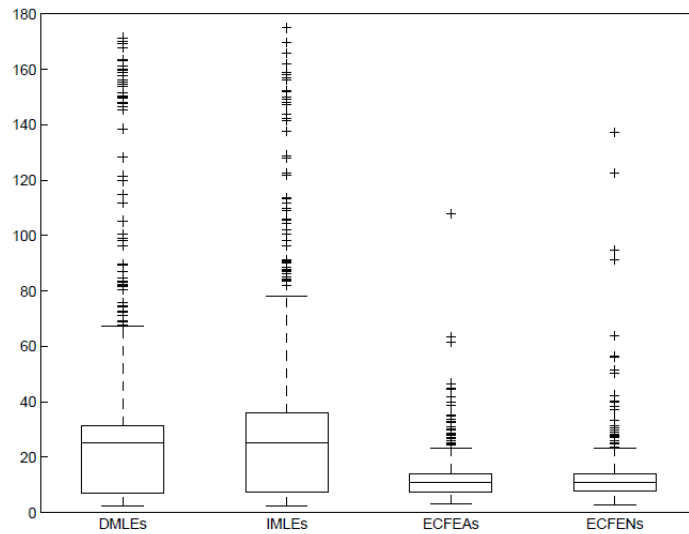


Figure 8. Boxplot of the inverse of the estimates of α in Example 4.

	$\nu_1(\hat{\theta})$	$\nu_2(\hat{\theta})$	$\nu_3(\hat{\theta})$	$\nu_4(\hat{\theta})$
True values	-0.100	0.051	-0.045	3.102
DMLE	-0.100 (0.015)	0.050 (0.005)	-0.040 (0.168)	3.184 (0.247)
IMLE	-0.100 (0.015)	0.050 (0.005)	-0.042 (0.167)	3.182 (0.248)
ECFEA	-0.100 (0.015)	0.050 (0.005)	-0.025 (0.233)	3.228 (0.188)
ECFEN	-0.100 (0.015)	0.050 (0.005)	-0.021 (0.174)	3.176 (0.153)

Table 6: The moments associated with the various estimates obtained in Example 4.

The averages and standard deviations associated with the first and second moments of the fitted distributions are identical. Nevertheless, the MLEs clearly outperform the ECFEs in their ability to accurately model the skewness and kurtosis of the underlying data. This observation is especially interesting in light of the fact that the boxplots presented in Figure 8 seemed to indicate that the ECFEs were better able to accurately estimate the true value of α^{-1} .

3.2.2. Example 5: The Meixner distribution fitted to the log-returns of the DJIA As a final example, the Meixner distribution is fitted to smooth bootstrap samples taken from the distribution underlying the log-returns of the DJIA. Boxplots of the estimates obtained for δ using the various estimators are shown in Figure 9. Once again, the distributions of the estimates are positively skewed.

Table 7 shows the correlation matrix for the logarithms of the estimates. As was the case when fitting the $N \circ IG$ distribution to these data, the off-diagonal elements of the correlation matrix are close to 0.

	DMLE	IMLE	ECFEA	ECFEN
DMLE	1.000	-0.022	-0.044	0.012
IMLE	-0.022	1.000	0.024	0.028
ECFEA	-0.044	0.024	1.000	0.026
ECFEN	0.012	0.028	0.026	1.000

Table 7: Correlation matrix of the logarithm of the estimates for δ obtained in Example 5.

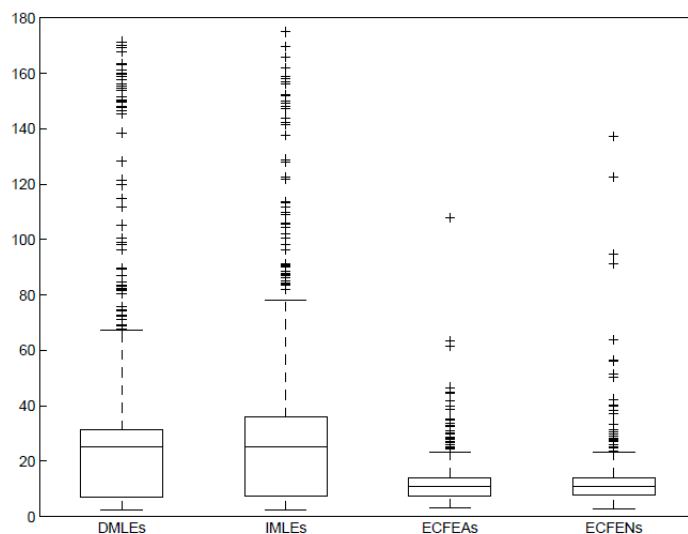
Figure 9. Boxplot of the estimates of δ in Example 5.

Table 8 shows the average and standard deviation (in brackets) of the moments associated with each estimator. The moments implied by each of the estimators closely match those of the observed log-returns. Interestingly, the *DMLEs* are best able to replicate the kurtosis of the log-returns. However, in line with the $N \circ IG$ distribution case, all of the estimators underestimate the kurtosis of the log-returns.

	$10^3 \times \nu_1(\hat{\theta})$	$10^4 \times \nu_2(\hat{\theta})$	$\nu_3(\hat{\theta})$	$\nu_4(\hat{\theta})$
Moments	0.871	0.174	-0.185	5.316
<i>DMLE</i>	0.856 (0.293)	0.193 (0.029)	0.075 (0.252)	4.056 (1.141)
<i>IMLE</i>	0.906 (0.304)	0.191 (0.026)	0.055 (0.208)	3.699 (1.021)
<i>ECFEA</i>	0.856 (0.284)	0.197 (0.061)	-0.007 (0.240)	3.148 (0.378)
<i>ECFEN</i>	0.856 (0.284)	0.187 (0.073)	0.004 (0.240)	3.142 (0.406)

Table 8: The moments associated with the various estimates obtained in Example 5.

4. Conclusions

This paper considers parameter estimation for multi-parameter distributions. It is demonstrated numerically that the parameter estimates obtained using various estimators vary wildly in cases where the data are realized from a distribution resembling a limit case of the distribution containing fewer parameters. The main contribution of the paper lies in the explanation of this phenomenon: the observed variability is, at least partially, due to the fact that a distribution containing redundant parameters is fitted to the data. In spite of the variability between estimates, in many cases the density functions achieved using various estimators are almost identical in spite of substantially different parameter estimates.

The maximum likelihood estimator (*MLE*) and the empirical characteristic function estimator (*ECFE*) are used to illustrate the mentioned behavior of the estimates. Many multi-parameter distributions have complicated density (and likelihood) functions. For some parameter sets the calculation of these densities may break down because of the limited computational power available to the modeler. Two methods for calculating *MLEs* are used. The first

discards parameter estimates for which the calculation of the likelihood function via the direct calculation of the density function breaks down while the second uses Fourier inversion to evaluate the likelihood in these cases. The *ECFE* obtains parameter estimates by minimizing a distance measure between the theoretical and empirical characteristic functions. Two variations of the *ECFE* are considered, the difference between the two being the weight functions used.

Specific empirical examples relating to the normal inverse Gaussian ($N \circ IG$) and Meixner distributions are included. It is shown that both of these distributions have the normal distribution as a limit case and the numerical problems associated with the calculation of the density functions of these distributions are explained.

Practical examples are obtained by simulating realizations from the $N \circ IG$ and Meixner distributions. Samples are also drawn from the distribution of the log-returns of the Dow Jones Industrial Average index using the smooth bootstrap. In spite of the substantially different parameter estimates obtained using the various estimators, the moments of the estimated distributions correspond closely to each other and to that of the distribution from which the data are realized in the vast majority of the cases considered. We conclude that, when multi-parameter distributions are fitted to data realized from a distribution resembling a limit case containing fewer parameters, the estimators are able to accurately model the characteristics of the underlying distribution, even when the estimates obtained differ substantially from the true parameters.

5. Acknowledgements

This research was done as part of the author's doctoral studies under the supervision of Prof. F. Lombard. The author would like to sincerely thank Prof. Lombard for his guidance.

The financial assistance of the National Research Foundation (NRF) and the DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS) is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF or the CoE-MaSS.

Appendix

Theorem 3

Let $X \sim N \circ IG(\alpha, \beta, \mu, \delta)$. If $\alpha \uparrow \infty$, $\delta \uparrow \infty$ and $\delta/\alpha \rightarrow \sigma^2$ while β is held constant, then the distribution of X converges to the $N(\mu + \beta\sigma^2, \sigma^2)$ distribution.

Proof

Consider the characteristic function of the $N \circ IG$ distribution with parameter set $\theta = (\alpha, \beta, \mu, \delta)$ where $\alpha \uparrow \infty$:

$$\begin{aligned} \phi(t; \theta) &= \exp\left(it\mu - \delta\left(\sqrt{\alpha^2 - (\beta + it)^2} - \sqrt{\alpha^2 - \beta^2}\right)\right) \\ &= \exp\left(it\mu - \alpha\delta\left(\sqrt{1 - \frac{(\beta + it)^2}{\alpha^2}} - \sqrt{1 - \frac{\beta^2}{\alpha^2}}\right)\right) \\ &= \exp\left(it\mu - \alpha\delta\left(-\frac{(\beta + it)^2}{2\alpha^2} + \frac{\beta^2}{2\alpha^2} + O(\alpha^{-4})\right)\right) \\ &= \exp\left(it\left(\mu + \beta\frac{\delta}{\alpha}\right) - \frac{t^2\delta}{2\alpha} + O(\alpha^{-3})\right). \end{aligned}$$

If $\delta = \alpha\sigma^2 + O(\alpha^\lambda)$ where $\lambda < 1$, then:

$$\begin{aligned}\phi(t; \theta) &= \exp\left(it(\mu + \beta(\sigma^2 + O(\alpha^{\lambda-1}))) - \frac{t^2}{2}(\sigma^2 + O(\alpha^{\lambda-1})) + O(\alpha^{-3})\right) \\ &\rightarrow \exp\left(it(\mu + \beta\sigma^2) - \frac{t^2}{2}\sigma^2\right),\end{aligned}$$

as $\alpha \uparrow \infty$. This is the characteristic function of the $N(\mu + \beta\sigma^2, \sigma^2)$ distribution. \square

Theorem 4

Let $X \sim \text{Meixner}(\alpha, \beta, \mu, \delta)$. If $\alpha \downarrow 0$, $\beta \rightarrow 0$ and $\delta \uparrow \infty$ so that $\alpha^2\delta/2 \rightarrow \sigma^2$ and $\beta/\alpha \rightarrow \gamma$, then the distribution of X converges to the $N(\mu + \gamma\sigma^2, \sigma^2)$ distribution.

Proof

Consider the characteristic function of the *Meixner* distribution with parameter set $\theta = (\alpha, \beta, \mu, \delta)$:

$$\begin{aligned}\phi(t, \theta) &= e^{i\mu t} \left(\frac{\cos\left(\frac{\beta}{2}\right)}{\cosh\left(\frac{\alpha t - i\beta}{2}\right)} \right)^{2\delta} \\ &= e^{i\mu t} \left(\frac{2 \cos\left(\frac{\beta}{2}\right)}{\exp\left(\frac{\alpha t - i\beta}{2}\right) + \exp\left(\frac{-\alpha t + i\beta}{2}\right)} \right)^{2\delta} \\ &= e^{i\mu t} \left(\frac{2 - \frac{\beta^2}{4} + O(\beta^4)}{2 - \frac{\beta(\beta + 2\alpha it)}{4} + \frac{\alpha^2 t^2 (8 - \beta^2)}{32} + O(\alpha^3) + O(\beta^3)} \right)^{2\delta},\end{aligned}\tag{8}$$

where the last equality above is obtained using Taylor expansions.

Taking logarithms in (8) and again using Taylor expansions we obtain:

$$\begin{aligned}\log(\phi(t, \theta)) &= i\mu t + 2\delta \log\left(2 - \frac{\beta^2}{4} + O(\beta^4)\right) \\ &\quad - 2\delta \log\left(2 - \frac{\beta(\beta + 2\alpha it)}{4} + \frac{\alpha^2 t^2 (8 - \beta^2)}{32} + O(\alpha^3) + O(\beta^3)\right) \\ &= i\mu t - \frac{\alpha^2 \delta t^2}{4} - \frac{32\alpha^2 \delta t^2}{(\beta^2 - 8)^2} - \frac{4\alpha \delta t (\alpha t + i\beta)}{\beta^2 - 8} + O(\alpha^3) + O(\beta^3).\end{aligned}\tag{9}$$

If we set

$$\delta = \frac{2\sigma^2}{\alpha^2} + O(\alpha^\varepsilon),$$

where $\varepsilon > -2$ and

$$\beta = \gamma\alpha + O(\alpha^\lambda),$$

where $\lambda > 1$, then after standard calculations (9) becomes as follows:

$$\begin{aligned}\log(\phi(t, \theta)) &= i\mu t - \frac{\sigma^2 t^2}{\left(\frac{\beta}{8} - 1\right)^2} + \frac{\sigma^2 t + \gamma\sigma^2 it}{1 - \frac{\beta^2}{8}} + O(\alpha^3) + O(\beta^3) \\ &\quad + O(\alpha^{\varepsilon+2}) + O(\alpha^{\lambda-1}) + O(\alpha^{\varepsilon+\lambda+1}).\end{aligned}\tag{10}$$

Letting $\alpha \downarrow 0$ and $\beta \rightarrow 0$ and simplifying, the characteristic exponent in (10) becomes:

$$\log(\phi(t, \theta)) = it(\mu + \gamma\sigma^2) - \frac{\sigma^2 t^2}{2},$$

which is the characteristic exponent of the $N(\mu + \gamma\sigma^2, \sigma^2)$ distribution. \square

REFERENCES

1. R. Cont, *Empirical properties of asses returns: stylized facts and statistical issues*, Quantitative Finance, 1:223–236. 2001.
2. A.C. Davison and D.V. Hinkley, *Bootstrap methods and their application*, Cambridge: Cambridge University Press. 1997.
3. W. Feller, *An Introduction to Probability Theory and Its Applications, volume 2*, London: John Wiley & Sons, Inc. 1971.
4. T.S. Ferguson, *A Course in Large Sample Theory.*, London: CRC Press. 1996.
5. A. Feuerverger and R.A. Mureika, *The empirical characteristic function and its applications*, The Annals of Statistics, 5:88–97. 1977.
6. A. Feuerverger and P. McDunnough, *On some Fourier methods for inference*, Journal of the American Statistical Association, 76:379–387. 1981.
7. A. Feuerverger and P. McDunnough, *On the efficiency of empirical characteristic function procedures*, Journal of the Royal Statistical Society: Series B, 43:20–27. 1981.
8. C.R. Heathcote, *The integrated squared error estimation of parameters*, Biometrika, 64:255–264. 1977.
9. E.L. Lehmann, *Elements of Large Sample Theory*, New York: Springer. 1999.
10. J.A. Nelder and R. Mead, *A simplex method for function minimization*, The Computer Journal, 7:308–313. 1965.
11. A.S. Paulson, E.W. Holcomb and R.A. Leitch, *The estimation of the parameters of the stable laws*, Biometrika, 62:163–170. 1975.
12. W. Schoutens, *Lévy Processes in Finance: Pricing Financial Derivatives*, Chichester: John Wiley & Sons, Inc.
13. A.W. Van der Vaart, *Asymptotic Statistics*, Cambridge: Cambridge University Press. 1998.
14. J.H. Venter, P.J. de Jongh and G. Griebenow, *NIG-GARCH models based on open, close, high and low prices.*, South African Statistical Journal, 39:79–101. 2005.
15. J. Wolfowitz, *Estimation by the minimum distance method*, The Annals of the Institute of Statistical Mathematics, 5:9–23. 1953.
16. J. Yu, *Empirical characteristic function estimation and its applications*, Econometric Reviews, 23:93-123. 2004.