DISS. ETH N° 20694

# New Directions in Statistical Distributions, Parametric Modeling and Portfolio Selection

*A dissertation submitted to*

ETH ZURICH

*for the degree of*

Doctor of Sciences

*presented by*

YOHAN CHALABI

MS in Physics EPF Lausanne
born November 13, 1980
citizen of Martisberg, Switzerland

*accepted on the recommendation of*

PD Dr. Diethelm Würtz
Prof. Dr. Matthias Troyer
Prof. Dr. David J. Scott

2012

# Abstract

Quantitative approaches to solving portfolio optimization problems require an understanding of several areas of computational finance. These areas include statistical distributions, parametric modeling and portfolio design methods. Statistical distributions are used to model time series data. Parametric models are important components of advanced risk measures. Optimization methods are used to maximize increasingly complex portfolio objective functions. This thesis presents innovations in these three areas.

The first part considers the use of the generalized lambda distribution (GLD) family as a flexible distribution with which to model financial data sets. The GLD can assume distributions with a large range of shapes. Analysts can therefore work with a single distribution to model almost any class of financial assets, rather than needing several. This becomes especially useful in the estimation of risk measures, where the choice of the distribution is crucial for accuracy. The part presents a new parameterization of the GLD, wherein the location and scale parameters are directly expressed as the median and interquartile range of the distribution. The two remaining parameters characterize the asymmetry and steepness of the distribution. Conditions are given for the existence of its moments, and for it to have the appropriate support. The tail behavior is also studied. The new parameterization brings a clearer interpretation of the parameters, whereby the distribution's asymmetry can be more readily distinguished from its steepness. This is in contrast to current parameterizations of the GLD, where the asymmetry and steepness are simultaneously described by a combination of the tail indices. Moreover, the new parameterization can be used to compare data sets in a convenient asymmetry and steepness shape plot.

The second part of this work is related to the estimation of parametric models in the presence of outliers. The maximum likelihood estimator is often used to find parameter values. However, it is highly sensitive to abnormal points. In this regard, the weighted trimmed likelihood estimator (WTLE) has been introduced as a robust alternative. The part improves the WTLE by adding a scheme for automatically computing the trimming parameter and weights.

The last part goes beyond the traditional portfolio optimization idea. A new portfolio

selection framework is introduced where the investor seeks the allocation that is as close as possible to his "ideal" portfolio. To build such a portfolio selection framework, the $\phi$-divergence measure from information theory is used. There are many advantages to using the $\phi$-divergence measure. First, the allocation is made such that it is in agreement with the historical data set. Second, the divergence measure is a convex function, which enables the use of fast optimization algorithms. Third, the objective value of the minimum portfolio divergence measure provides an indication distance from the ideal portfolio. A statistical test can therefore be constructed from the value of the objective function. Fourth, with adequate choices of both the target distribution and the divergence measure, the objective function of the $\phi$-portfolios reduces to the expected utility function.

# Résumé

Les approches quantitatives pour résoudre les problèmes d'optimisation de portefeuille exigent une compréhension dans plusieurs domaines en finance quantitative. Ces domaines comprennent les distributions statistiques paramétriques, l'estimation de modèles paramétriques, et la construction de portefeuilles. Les distributions statistiques sont utilisées pour modéliser les séries chronologiques. Les modèles paramétriques sont importants pour la construction de mesures de risque avancées. Les méthodes d'optimisation sont nécessaires pour maximiser des fonctions objectives de portefeuille de plus en plus complexes. Cette thèse présente des innovations dans ces trois domaines.

La première partie concerne l'utilisation de la famille de distribution généralisée de lambda (GLD) comme distribution suffisamment flexible pour modéliser différentes sortes de séries temporelles. La GLD inclut des distributions qui ont une large gamme de formes. Les analystes peuvent donc travailler avec une distribution unique, plutôt que de devoir en utiliser plusieures. Cela devient particulièrement utile pour l'estimation des mesures de risque, où le choix de la distribution est crucial pour obtenir des valeurs précises. Cette partie présente un nouveau paramétrage de la GLD, dans laquelle les paramètres de localisation et d'échelle sont directement exprimés par la médiane et l'écart interquartile de la distribution. Les deux autres paramètres de forme caractérisent la dissymétrie et l'aplatissement de la distribution. Les conditions requises pour l'existence de ces moments, ainsi que pour son support sont présentées. Le comportement de la queue est également étudié. Le nouveau paramétrage apporte une interprétation plus claire de ces paramètres, de sorte que la dissymétrie de la distribution peut être plus facile à distinguer de son aplatissement. Ceci est en contraste avec les paramétrages actuels de la GLD, où la dissymétrie et l'aplatissement sont simultanément décrits par la combinaison de deux indices de queues. En outre, le nouveau paramétrage peut être utilisé pour comparer un ensemble de données dans un graphique de dissymétrie et d'applatissement.

La deuxième partie de ce travail se rapporte à l'estimation de modèles paramétriques en présence de valeurs aberrantes. L'estimateur de la vraisemblance maximal est souvent utilisé pour trouver les valeurs des paramètres. Cependant, il est très sensible aux points aberrants. A cet égard, l'estimateur de la vraisemblance pondérée et tronquée (WTLE) a

été proposé comme alternative. La partie améliore le WTLE en présentant une approche pour calculer automatiquement le paramètre de troncation et de poids.

La dernière partie cherche à dépasser l'idée d'optimisation de portefeuille traditionnelle. Dans un nouveau cadre pour la sélection de portefeuille, un investisseur cherche l'allocation la plus proche possible d'un portefeuille "idéal". Pour cela, on utilisera la mesure de divergence $\phi$ de la théorie de l'information. L'utilisation des mesures de divergence $\phi$ confère plusieurs avantages. Tout d'abord, la répartition est faite de telle sorte qu'elle est en accord avec les données historiques. Deuxièmement, la mesure de divergence est une fonction convexe qui permet l'utilisation d'algorithmes d'optimisation rapides. Troisièmement, la valeur optimale des portefeuilles $\phi$ indique l'éloignement par rapport au portefeuille idéal. Un test statistique peut donc être construit à partir de la valeur de la fonction objective. Quatrièmement, avec des choix adéquats de la distribution cible et de la mesure de divergence, la fonction objective des portefeuilles $\phi$ est réduite à la fonction d'utilité espérée.

# Acknowledgments

This work would not have been possible without the support of my supervisor, Diethelm Würtz, whose help, stimulating discussions and friendship were of great support. His intellectual fervor was very inspiring.

I would also like to express my gratitude to numerous researchers, colleagues, and friends, for the comments they made while attending scientific conferences and, in particular, the Meielisalp Workshop and Summer School held in Meielisalp, Switzerland.

I am most grateful to my family and my wife for their personal support and patience.

This work was made possible by financial support from Finance Online GmbH and from the Swiss Federal Institute of Technology Zurich.

# Contents

**Part I:**
**Flexible Distribution Modeling with the Generalized Lambda Distribution**

1 Introduction . . . 3
2 The generalized lambda distribution . . . 5
3 A shape parameterization of the GLD . . . 13
3.1 An asymmetry-steepness parameterization . . . 14
3.2 Support . . . 18
3.3 Moments . . . 18
3.4 Tail behavior . . . 20
3.5 Distribution shape . . . 22
3.6 Special cases . . . 29
3.7 Limiting cases . . . 34
4 Parameter estimation . . . 41
4.1 Robust moments matching . . . 41
4.2 Histogram approaches . . . 42
4.3 Goodness-of-fit approaches . . . 44
4.4 Quantile matching . . . 45
4.5 Trimmed L-moments . . . 45
4.6 MLE and MPS approaches . . . 46
4.7 Empirical study . . . 47
5 Applications . . . 53
5.1 Shape plot representation . . . 53
5.2 Quantile based risk measures . . . 55
5.3 Apparent scaling law . . . 57
6 Conclusion . . . 63

**Part II:**

**Robust Estimation with the Weighted Trimmed Likelihood Estimator**

7 Introduction . . . . . 67
8 The weighted trimmed likelihood estimator . . . . . 69
9 The auto-WTLE algorithm . . . . . 73
10 Robust GARCH modeling . . . . . 77
10.1 Introduction . . . . . 77
10.2 WTL GARCH(p,q) . . . . . 78
10.3 Simulation study . . . . . 80
11 Conclusion . . . . . 93

**Part III:**

**Portfolio Optimization Based on Divergence Measures**

12 Introduction . . . . . 97
13 Divergence measures . . . . . 101
14 Phi-portfolios . . . . . 105
15 Dual phi-portfolios . . . . . 107
15.1 Concept . . . . . 107
15.2 Dual form with the GLD . . . . . 108
15.3 Primal-dual interior point implementation . . . . . 110
16 Minimum phi-portfolios . . . . . 115
17 Choice of divergence measures . . . . . 119
18 Statistical test . . . . . 129
19 Expected utility theory . . . . . 133
20 Conclusion . . . . . 135

**Bibliography**

# Part I

# Flexible Distribution Modeling with the Generalized Lambda Distribution

# 1 Introduction

It is well known that distributions of financial returns can exhibit heavy-tails, skewness, or both. This trait is often taken as stylized fact, and can be modeled by $\alpha$-stable distributions. The drawback to using these distributions is that they do not all have closed form for their densities and distribution functions. In this regard, the generalized lambda distribution (GLD) offers an alternative. The four-parameter GLD family is known for its high flexibility. It can create distributions with a large range of different shapes. It shares the heavy-tail and skewness properties of the $\alpha$-stable distribution. Excepting Corrado [2001] and Tarsitano [2004], there has been little application of the GLD to financial matters. However, as will be covered in this chapter, the shape range of the GLD family is so large that it can accommodate almost any financial time series. Analysts can therefore work with a single distribution, rather than needing several. This becomes especially useful in the estimation of risk measures, where the choice of the distribution is crucial for accuracy. Furthermore, the GLD is defined by a quantile function. Its parameters can be estimated even when its moments do not exist.

This thesis introduces a more intuitive parameterization of the GLD that expresses the location and scale parameters directly as the median and interquartile range of the distribution. The two remaining shape parameters respectively characterize the asymmetry and steepness of the distribution. This is in contrast to the standard parameterization, where the asymmetry and steepness are described by a combination of the two tail indices. In Chapter 2 we review the history of the generalized lambda distribution. Then, in Chapter 3, the new formulation is derived, along with the conditions of the various distribution shape regions and the moment conditions.

The estimation of standard GLD parameters is notoriously difficult. For the new parameterization, fitting the GLD to empirical data can be reduced to a two-step estimation problem, wherein the location and scale parameters are estimated by their robust sample estimators. This approach also works when the moments of the GLD do not exist. Chapter 4 describes an extensive empirical study that compares different GLD parameter estimation methods. The author's contribution is to introduce the maximum product of spacing estimator (MPS) for the GLD. Also presented, is a new robust moment matching

method where the estimation of the two shape parameters is independent of the location and scale parameters. The methods developed in this work allow easy fitting of the GLD to empirical data, such as financial returns. An extensive simulation study to compare different GLD parameter estimators is also performed.

Applications of the new GLD parameterization are presented in Chapter 5. First, the new parameterization is used to compare data sets in a convenient asymmetry and steepness shape plot. The use of the asymmetry and steepness shape plot is illustrated by comparing equities returns from the NASDAQ-100 stock index. Second, widely used financial risk measures, such as the value-at-risk and expected shortfall, are calculated for the GLD. Third, the shape plot is used to illustrated the apparent scaling law and self-similarity of high frequency returns.

Later, in Part III the GLD will be used in the context of portfolio optimization based on $\phi$-divergence measures.

# 2 The generalized lambda distribution

The GLD is an extension of Tukey's lambda distribution [Hastings et al., 1947; Tukey, 1960, 1962]. Tukey's lambda distribution is symmetric and is defined by its quantile function. It was introduced to efficiently redistribute uniform random variates into approximations of other distributions [Tukey, 1960; Van Dyke, 1961]. Soon after the introduction of Tukey's lambda distribution, Hogben [1963], Shapiro et al. [1968], Shapiro and Wilk [1965], and Filliben [1975] introduced the non-symmetric case in sampling studies. Over the years, Tukey's lambda distribution became a versatile distribution, with location, scale, and shape parameters, that could accommodate a large range of distribution forms. It gained use as a data analysis tool and was no longer restricted to approximating other distributions. Figure 2.1 illustrates the four qualitatively different shapes of the GLD: unimodal, U-shaped, monotone and S-shaped. The GLD has been applied to numerous fields, including option pricing as a fast generator of financial prices [Corrado, 2001], and in fitting income data [Tarsitano, 2004]. It has also been used in meteorology [Ozturk and Dale, 1982], in studying the fatigue lifetime prediction of materials [Bigerelle et al., 2006], in simulations of queue systems [Dengiz, 1988], in corrosion studies [Najjar et al., 2003], in independent component analysis [Karvanen et al., 2000], and in statistical process control [Pal, 2004; Negiz and Çinar, 1997; Fournier et al., 2006].

The parameterization of the extended Tukey's lambda distributions has a long history of development. Contemporary parameterizations are those of Ramberg and Schmeiser [1974] and Freimer et al. [1988]. Ramberg and Schmeiser [1974] generalized the Tukey lambda distribution to four parameters called the RS parameterization:

$$Q_{\text{RS}}(u) = \lambda_1 + \frac{1}{\lambda_2}\left[u^{\lambda_3} - (1-u)^{\lambda_4}\right], \tag{2.1}$$

where $Q_{\text{RS}} = F_{\text{RS}}^{-1}$ is the quantile function for probabilities $u$, $\lambda_1$ is the location parameter, $\lambda_2$ is the scale parameter, and $\lambda_3$ and $\lambda_4$ are the shape parameters. In order to have a valid distribution function where the probability density function $f$ is positive for all $x$

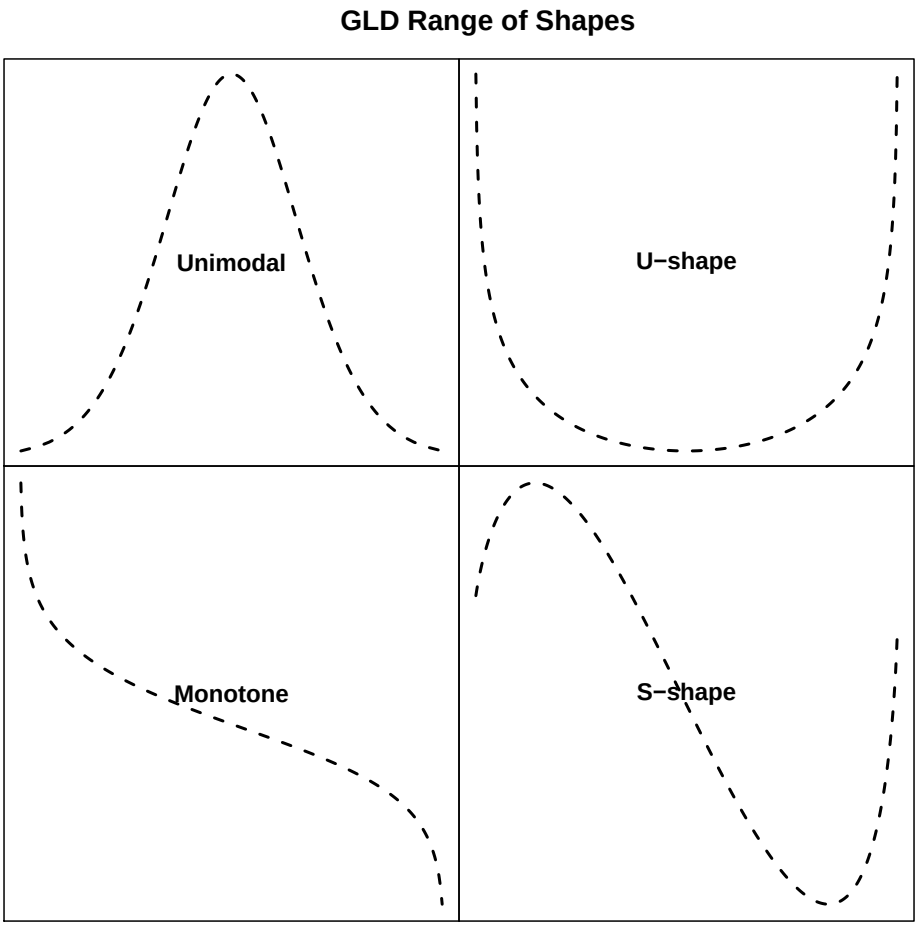


Figure 2.1: The GLD has four basic shapes: unimodal, U-shape, monotone, and S-shape.

| Region | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $Q(0)$ | $Q(1)$ |
|---|---|---|---|---|---|
| | | $\leq -1$ | $\geq 1$ | | |
| 1 | $< 0$ | $\begin{cases} -1 < \lambda_3 < 0 \text{ and } \lambda_4 > 1 \\ \frac{(1-\lambda_3)^{1-\lambda_3}(\lambda_4-1)^{\lambda_4-1}}{(\lambda_4-\lambda_3)^{\lambda_4-\lambda_3}} = \frac{-\lambda_3}{\lambda_4} \end{cases}$ | | $-\infty$ | $\lambda_1 + (1/\lambda_2)$ |
| | | $\geq 1$ | $\leq -1$ | | |
| 2 | $< 0$ | $\begin{cases} \lambda_3 > 1 \quad \wedge \quad -1 < \lambda_4 < 0 \\ \frac{(1-\lambda_4)^{1-\lambda_4}(\lambda_3-1)^{\lambda_3-1}}{(\lambda_3-\lambda_4)^{\lambda_3-\lambda_4}} = \frac{-\lambda_4}{\lambda_3} \end{cases}$ | | $\lambda_1 - (1/\lambda_2)$ | $\infty$ |
| | | $> 0$ | $> 0$ | $\lambda_1 - (1/\lambda_2)$ | $\lambda_1 + (1/\lambda_2)$ |
| 3 | $> 0$ | $= 0$ | $> 0$ | $\lambda_1$ | $\lambda_1 + (1/\lambda_2)$ |
| | | $> 0$ | $= 0$ | $\lambda_1 - (1/\lambda_2)$ | $\lambda_1$ |
| | | $< 0$ | $< 0$ | $-\infty$ | $\infty$ |
| 4 | $< 0$ | $= 0$ | $< 0$ | $\lambda_1$ | $\infty$ |
| | | $< 0$ | $= 0$ | $-\infty$ | $\lambda_1$ |

Table 2.1: Support regions of the GLD and conditions on the parameters given by the RS parameterization to define a valid distribution function (Karian and Dudewicz [2000]). The support regions are displayed in Fig. 2.2. Note that there are no conditions on $\lambda_1$ to obtain a valid distribution.

and integrates to one over the allowed domain, i.e.,

$$f(x) \geq 0 \quad \text{and} \quad \int_{Q(0)}^{Q(1)} f(x)\,\mathrm{d}x = 1,$$

the RS parameterization has complex constraints on the parameters and support regions as summarized in Table 2.1 and Fig. 2.2.

Later, Freimer et al. [1988] introduced a new parameterization called FKML to circumvent the constraints on the RS parameter values. It is expressed as

$$Q_{\text{FKML}}(u) = \lambda_1 + \frac{1}{\lambda_2}\left[\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4}\right]. \tag{2.2}$$

As in the previous parameterization, $\lambda_1$ and $\lambda_2$ are the location and scale parameters, and $\lambda_3$ and $\lambda_4$ are the tail index parameters. The advantage over the previous parameterization is that the only constraint on the parameters is that $\lambda_2$ must be positive. Figure 2.3 displays the support regions of the GLD in the FKML parameterization.

Estimation of the GLD parameters for empirical data is notoriously difficult because of interdependencies between the parameters, rapid changes in the distribution forms and supports. In particular, the support of the distribution is dependent upon the parameter

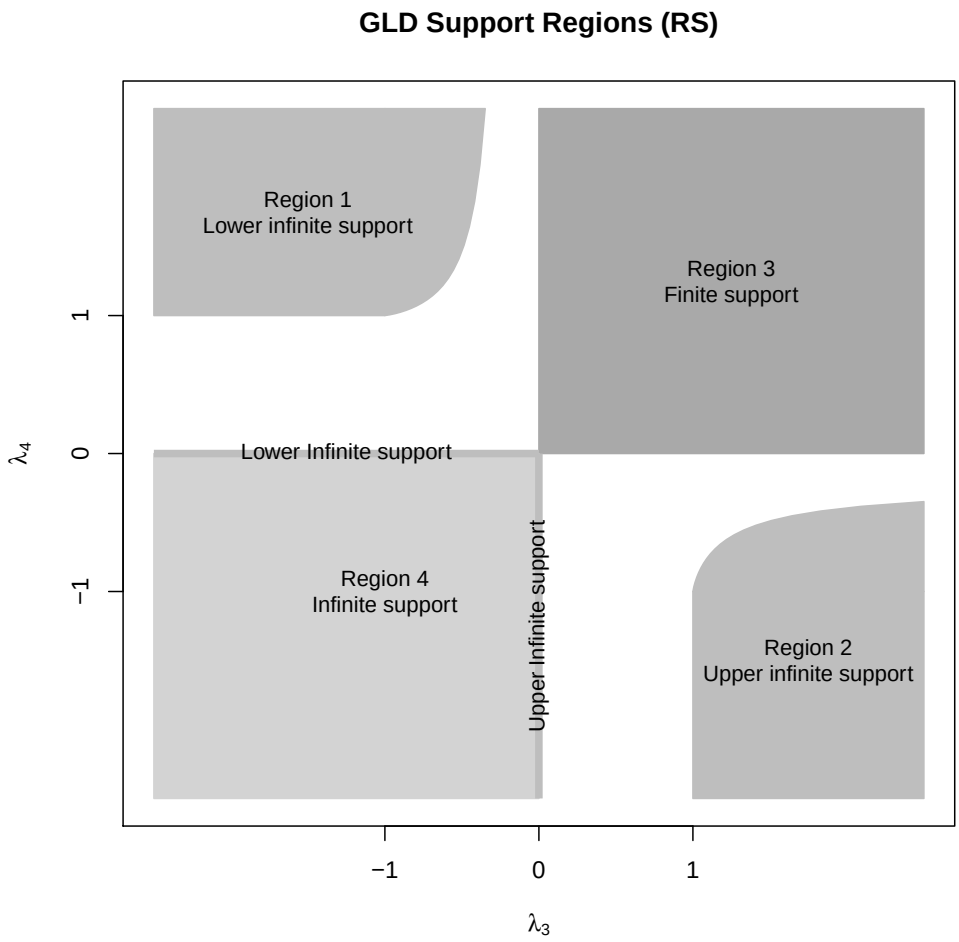


Figure 2.2: Support regions of the GLD in the RS parameterization that produce valid statistical distributions as described in Table 2.1.

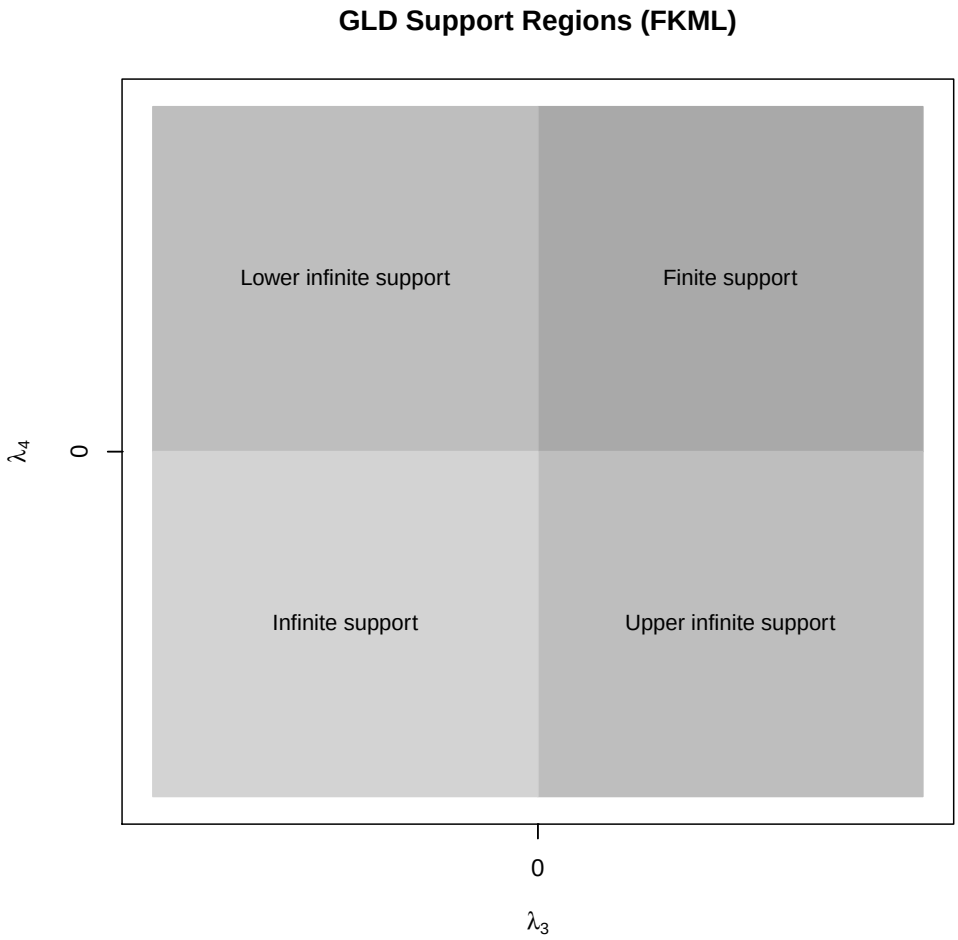


Figure 2.3: Support regions of the GLD in the FKML parameterization.

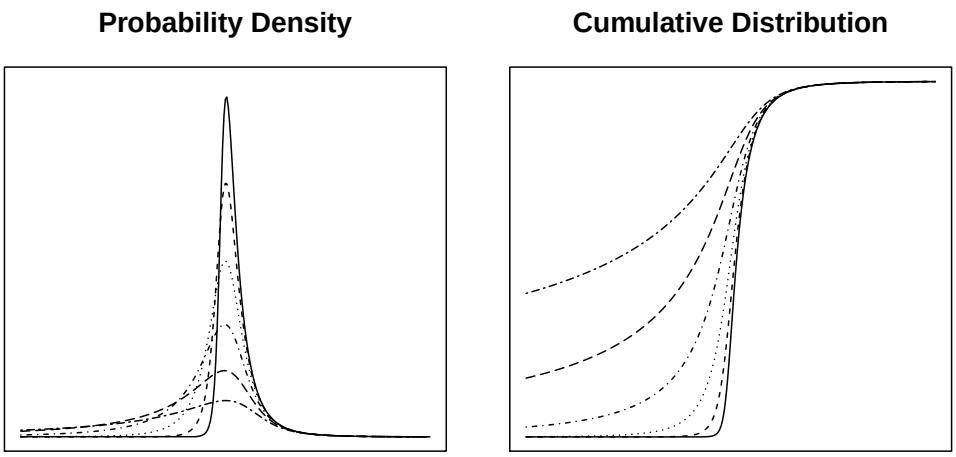


Figure 2.4: Probability density and cumulative distribution functions for the GLD in the RS parameterization. The location and scale parameters are, $\lambda_1 = 0$ and $\lambda_2 = -1$, respectively. The right tail index is fixed at $\lambda_4 = -1/4$ and the left tail index, $\lambda_3$, varies in the range $\{-1/16, -1/8, -1/4, -1/2, -1, -2\}$.

values, and can vary from the entire real line to a semi-infinite, or finite, interval. In the last decade, several papers have been published to discuss two different parameter estimation philosophies. On one side of the discourse are the direct estimation methods, such as least-squares estimation with order statistics [Ozturk and Dale, 1985] and with percentiles [Karian and Dudewicz, 1999, 2000; Fournier et al., 2007; King and MacGillivray, 2007; Karian and Dudewicz, 2003]; the methods of moments [Ozturk and Dale, 1982; Gilchrist, 2000], L-moments [Gilchrist, 2000; Karvanen and Nuutinen, 2008], and trimmed L-moments [Asquith, 2007]; and the goodness-of-fit method with histograms [Su, 2005] and with maximum likelihood estimation [Su, 2007]. On the other side, stochastic methods have been introduced with various estimators such as goodness-of-fit [Lakhany and Mausser, 2000] or the starship method [King and MacGillivray, 1999]. Moreover, Shore [2007] studied the $L^2$-norm estimator that minimizes the density function distance and the use of nonlinear regression applied to a sample of exact quantile values.

As noted by Gilchrist [2000], one of the criticisms of the GLD is that its skewness is expressed in terms of both tail indices $\lambda_3$ and $\lambda_4$ in Eqs. (2.1) and (2.2). In one approach addressing this concern, a five-parameter GLD was introduced by Joiner and Rosenblatt [1971], which, expressed in the FKML parameterization, can be written as,

$$Q_{\text{JR}}(u) = \lambda_1 + \frac{1}{2\lambda_2}\left[(1-\lambda_5)\frac{u^{\lambda_3}-1}{\lambda_3} - (1+\lambda_5)\frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right].$$

It has a location parameter $\lambda_1$, a scale parameter $\lambda_2$, and an asymmetry parameter, $\lambda_5$, which weights each side of the distribution and the two tail indices, $\lambda_3$ and $\lambda_4$. The conditions on the parameters are $\lambda_2 > 0$ and $-1 < \lambda_5 < 1$. However, the additional parameter can make the estimation of the parameter values even more difficult.

# 3 A shape parameterization of the GLD

In this chapter, it is shown how the four-parameter GLD of Eq. (2.2) can be transformed in terms of an asymmetry and steepness parameter without adding a new variable. Moreover, the location and scale parameters are reformulated in terms of quantile statistics. The median is used as the location parameter, and the interquartile range is used as the scale parameter.

The new parameterization brings a clearer interpretation of the parameters, whereby the asymmetry of the distribution can be more readily distinguished from the steepness of the distribution. This is in contrast to the RS and FKML parameterizations, where the asymmetry is described by a combination of the tail indices. The new formulation allows for relaxed constraints upon the support regions, and for the existence of moments for the GLD.

Another advantage of the new parameterization is that the estimation of the parameters for an empirical data set can be reduced to a two-step estimation problem, because the location and scale parameters can be directly estimated by their robust sample estimators. Note that sample quantiles can always be estimated, even when moments of the distribution do not exist. The two shape parameters can then be estimated by one of the usual estimation methods, such as the maximum likelihood estimator.

The remainder of this chapter is organized as follows. In Section 3.1, the new asymmetry-steepness parameterization of the GLD is introduced. In Section 3.2, the support of the distribution is derived in terms of its parameters, and the different support regions are illustrated in a shape plot. Then, in Section 3.3, conditions are given for the existence of its moments, and a shape plot representation is made. In Section 3.4, the tail behavior of the GLD in the new parameterization is studied. Based on the support and moment existence conditions, Section 3.5 studies the different shape regions of the distribution. Section 3.6 lists sets of parameter values that make the GLD a good approximation for several well-known distributions. Limiting cases of the parameter sets are presented in Section 3.7.

## 3.1 An asymmetry-steepness parameterization

In the following section, the new parameterization is expressed in the FKML form. Consider the GLD quantile function:

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2} S(u|\lambda_3, \lambda_4), \tag{3.1}$$

where

$$S(u|\lambda_3, \lambda_4) = \begin{cases} \ln(u) - \ln(1-u) & \text{if } \lambda_3 = 0,\ \lambda_4 = 0, \\ \ln(u) - \frac{1}{\lambda_4}\left[(1-u)^{\lambda_4} - 1\right] & \text{if } \lambda_3 = 0,\ \lambda_4 \neq 0, \\ \frac{1}{\lambda_3}\left(u^{\lambda_3} - 1\right) - \ln(1-u) & \text{if } \lambda_3 \neq 0,\ \lambda_4 = 0, \\ \frac{1}{\lambda_3}\left(u^{\lambda_3} - 1\right) - \frac{1}{\lambda_4}\left[(1-u)^{\lambda_4} - 1\right] & \text{otherwise,} \end{cases} \tag{3.2}$$

for $0 < u < 1$. $Q$ is the quantile function for probabilities $u$; $\lambda_1$ and $\lambda_2$ are the location and scale parameters; and $\lambda_3$ and $\lambda_4$ are the shape parameters jointly related to the strengths of the lower and upper tails. In the limiting case $u = 0$:

$$S(0|\lambda_3, \lambda_4) = \begin{cases} -\frac{1}{\lambda_3} & \text{if } \lambda_3 > 0, \\ -\infty & \text{otherwise.} \end{cases}$$

In the limiting case $u = 1$:

$$S(1|\lambda_3, \lambda_4) = \begin{cases} \frac{1}{\lambda_4} & \text{if } \lambda_4 > 0, \\ \infty & \text{otherwise.} \end{cases}$$

The median, $\tilde{\mu}$, and the interquartile range, $\tilde{\sigma}$, can now be used to represent the location and scale parameters. These are defined by

$$\tilde{\mu} = Q(1/2), \tag{3.3a}$$

$$\tilde{\sigma} = Q(3/4) - Q(1/4). \tag{3.3b}$$

The parameters $\lambda_1$ and $\lambda_2$ in Eq. (3.2) can therefore be expressed in terms of the median

and interquartile range as

$$\lambda_1 = \tilde{\mu} - \frac{1}{\lambda_2} S(\frac{1}{2}|\lambda_3, \lambda_4),$$
$$\lambda_2 = \frac{1}{\tilde{\sigma}} \left[ S(\frac{3}{4}|\lambda_3, \lambda_4) - S(\frac{1}{4}|\lambda_3, \lambda_4) \right].$$

As mentioned in the introduction, one of the criticisms of the GLD is that the asymmetry and steepness of the distribution are both dependent upon both of the tail indices, $\lambda_3$ and $\lambda_4$. The main idea in this chapter is to use distinct shape parameters for the asymmetry and steepness. First, it is clear, from the definition of the GLD in Eq. (3.1), that when the tail indices are equal, the distribution is symmetric. Increasing one tail index then produces an asymmetric distribution. Second, the steepness of the distribution is related to the size of both tail indices: increasing both tail indices results in a distribution with thinner tails. Now, formulate an asymmetry parameter, $\chi$, proportional to the difference between the two tail indices, and a steepness parameter, $\xi$, proportional to the sum of both tail indices. The remaining step is to map the unbounded interval of $(\lambda_3 - \lambda_4)$ to the interval $(-1, 1)$, and $(\lambda_3 + \lambda_4)$ to the interval $(0, 1)$. To achieve this, we use the transformation

$$y = \frac{x}{\sqrt{1+x^2}} \quad \leftrightarrow \quad x = \frac{y}{\sqrt{1-y^2}} \quad \text{where } y \in (-1, 1) \text{ and } x \in (-\infty, \infty).$$

The asymmetry parameter, $\chi$, and the steepness parameter, $\xi$, can then be expressed as

$$\chi = \frac{\lambda_3 - \lambda_4}{\sqrt{1 + (\lambda_3 - \lambda_4)^2}}, \tag{3.4a}$$

$$\xi = \frac{1}{2} - \frac{\lambda_3 + \lambda_4}{2\sqrt{1 + (\lambda_3 + \lambda_4)^2}}, \tag{3.4b}$$

where the respective domains of the shape parameters are $\chi \in (-1, 1)$ and $\xi \in (0, 1)$. When $\chi$ is equal to 0, the distribution is symmetric. When $\chi$ is positive (negative), the distribution is positively (negatively) skewed. Moreover, the GLD becomes steeper when $\xi$ increases. The parameterization of $\chi$ and $\xi$ in Eq. (3.4), yields a system of two equations for the tail indices $\lambda_3$ and $\lambda_4$:

$$\lambda_3 - \lambda_4 = \frac{\chi}{\sqrt{1-\chi^2}},$$

$$\lambda_3 + \lambda_4 = \frac{\frac{1}{2} - \xi}{\sqrt{\xi(1-\xi)}}.$$

This gives

$$\lambda_3 = \alpha + \beta,$$

$$\lambda_4 = \alpha - \beta,$$

where

$$\alpha = \frac{1}{2}\frac{\frac{1}{2} - \xi}{\sqrt{\xi(1-\xi)}}, \tag{3.5a}$$

$$\beta = \frac{1}{2}\frac{\chi}{\sqrt{1-\chi^2}}. \tag{3.5b}$$

The $S$ function of Eq. (3.2) can now be formulated in terms of the shape parameters $\chi$ and $\xi$;

$$S(u|\chi,\xi) = \begin{cases} \ln(u) - \ln(1-u) & \text{if } \chi = 0,\ \xi = \frac{1}{2}, \\ \ln(u) - \frac{1}{2\alpha}\left[(1-u)^{2\alpha} - 1\right] & \text{if } \chi \neq 0,\ \xi = \frac{1}{2}(1+\chi), \\ \frac{1}{2\beta}(u^{2\beta} - 1) - \ln(1-u) & \text{if } \chi \neq 0,\ \xi = \frac{1}{2}(1-\chi), \\ \frac{1}{\alpha+\beta}\left(u^{\alpha+\beta} - 1\right) - \frac{1}{\alpha-\beta}\left[(1-u)^{\alpha-\beta} - 1\right] & \text{otherwise,} \end{cases} \tag{3.6}$$

where $\alpha$ and $\beta$ are defined in Eq. (3.5), and $0 < u < 1$. When $u = 0$;

$$S(0|\chi,\xi) = \begin{cases} -\frac{1}{\alpha+\beta} & \text{if } \xi < \frac{1}{2}(1+\chi), \\ -\infty & \text{otherwise,} \end{cases}$$

and when $u = 1$;

$$S(1|\chi,\xi) = \begin{cases} \frac{1}{\alpha-\beta} & \text{if } \xi < \frac{1}{2}(1-\chi), \\ \infty & \text{otherwise.} \end{cases}$$

Given the definitions of $\tilde{\mu}$, $\tilde{\sigma}$, $\chi$, $\xi$, and $S$ in Eqs. (3.3), (3.4) and (3.6), the quantile function of the GLD becomes

$$\boxed{Q_{\text{CSW}}(u|\tilde{\mu},\tilde{\sigma},\chi,\xi) = \tilde{\mu} + \tilde{\sigma}\frac{S(u|\chi,\xi) - S(\frac{1}{2}|\chi,\xi)}{S(\frac{3}{4}|\chi,\xi) - S(\frac{1}{4}|\chi,\xi)}.} \tag{3.7}$$

Hereinafter the subscript CSW shall be used to denote the new parameterization.

Since the quantile function, $Q_{\text{CSW}}$, is continuous, it immediately follows by definition that the cumulative distribution function is $F_{\text{CSW}}[Q_{\text{CSW}}(u)] = u$ for all probabilities $u \in [0, 1]$. The probability density function, $f(x) = F'(x)$, and the quantile density function, $q(u) = Q'(u)$, are then related by

$$f\left[Q(u)\right]\ q(u) = 1. \tag{3.8}$$

The literature often refers to $f\left[Q(u)\right]$ as the density quantile function, $fQ(u)$.

The probability density function of the GLD can be calculated from the quantile density function. In particular, the quantile density function can be derived from the definition of the quantile function in Eq. (3.7). This gives

$$\boxed{q_{CSW}(u|\tilde{\sigma}, \chi, \xi) = \frac{\tilde{\sigma}}{S(\frac{3}{4}|\chi, \xi) - S(\frac{1}{4}|\chi, \xi)} \frac{\mathrm{d}}{\mathrm{d}\,u} S(u|\chi, \xi),} \tag{3.9}$$

where

$$\frac{\mathrm{d}}{\mathrm{d}\,u} S(u|\chi, \xi) = u^{\alpha+\beta-1} + (1-u)^{\alpha-\beta-1}\,,$$

with $\alpha$ and $\beta$ defined in Eq. (3.5).

It is interesting to note that the limiting sets of shape parameters; $\{\chi \to -1,\ \xi \to 0\}$ and $\{\chi \to 1,\ \xi \to 0\}$, produce valid distributions. When $\chi \to -1$ and $\xi \to 0$, the quantile function is

$$\lim_{\substack{\chi \to -1 \\ \xi \to 0}} Q_{CSW}(u) = \tilde{\mu} + \tilde{\sigma} \frac{\ln(u) + \ln(2)}{\ln(3)},$$

and the quantile density function is

$$\lim_{\substack{\chi \to -1 \\ \xi \to 0}} q_{CSW}(u) = \frac{\tilde{\sigma}}{\ln(3)} \frac{1}{u}.$$

When, instead, $\chi \to 1$ and $\xi \to 0$,

$$\lim_{\substack{\chi \to 1 \\ \xi \to 0}} Q_{CSW}(u) = \tilde{\mu} - \tilde{\sigma} \frac{\ln(1-u) + \ln(2)}{\ln(3)}, \tag{3.10}$$

and

$$\lim_{\substack{\chi \to 1 \\ \xi \to 0}} q_{CSW}(u) = \frac{\tilde{\sigma}}{\ln(3)} \frac{1}{1-u}.$$

Other sets of limiting shape parameters do not, however, yield valid distributions. Details

of the calculations for all limiting cases are shown in Section 3.7.

## 3.2 Support

The GLD can accommodate a wide range of distribution shapes and supports. In this section, the conditions imposed upon the shape parameters for the different support regions are calculated. The support of the GLD can be derived from the extreme values of $S$ in Eq. (3.6). When $u = 0$;

$$S(0|\chi,\xi) = \begin{cases} -\frac{2\sqrt{\xi(1-\xi)(1-\chi^2)}}{(\frac{1}{2}-\xi)\sqrt{1-\chi^2}+\chi\sqrt{\xi(1-\xi)}} & \text{if } \xi < \frac{1}{2}(1+\chi), \\ -\infty & \text{otherwise,} \end{cases} \tag{3.11}$$

and when $u = 1$;

$$S(1|\chi,\xi) = \begin{cases} \frac{2\sqrt{\xi(1-\xi)(1-\chi^2)}}{(\frac{1}{2}-\xi)\sqrt{1-\chi^2}-\chi\sqrt{\xi(1-\xi)}} & \text{if } \xi < \frac{1}{2}(1-\chi), \\ \infty & \text{otherwise.} \end{cases} \tag{3.12}$$

The GLD thus has unbounded infinite support, $(-\infty,\infty)$, when $\frac{1}{2}(1-|\chi|) \leq \xi$; semi-infinite support bounded above, $(-\infty, Q_{\text{CSW}}(1)]$, when $\chi < 0$ and $\frac{1}{2}(1+\chi) \leq \xi < \frac{1}{2}(1-\chi)$; and semi-infinite support bounded below, $[Q_{\text{CSW}}(0),\infty)$, when $0 < \chi$ and $\frac{1}{2}(1-\chi) \leq \xi < \frac{1}{2}(1+\chi)$. The distribution has finite support in the remaining region, $\xi \leq \frac{1}{2}(1-|\chi|)$.

As shown in Fig. 3.1, the support regions can be depicted using triangular regions in a shape plot with the asymmetry parameter $\chi$ versus steepness parameter $\xi$. This contrasts with the complex region supports of the RS parameterization displayed in Fig. 2.2. Of course, the region supports share the same intuitiveness as the FKML region supports in Fig. 2.3, since the shape-asymmetry parameterization is based on the FKML parameterization. The advantage of the CSW parameterization is that its shape parameters have finite domains of variation and the distribution family can therefore be represented by a single plot.

## 3.3 Moments

In this section, shape parameter ($\chi$ and $\xi$) dependent conditions for the existence of GLD moments are derived. As in Ramberg and Schmeiser [1974] and Freimer et al. [1988], conditions for the existence of moments can be determined by expanding the definition of the $k$th moment of the GLD into a binomial series where the binomial series can be

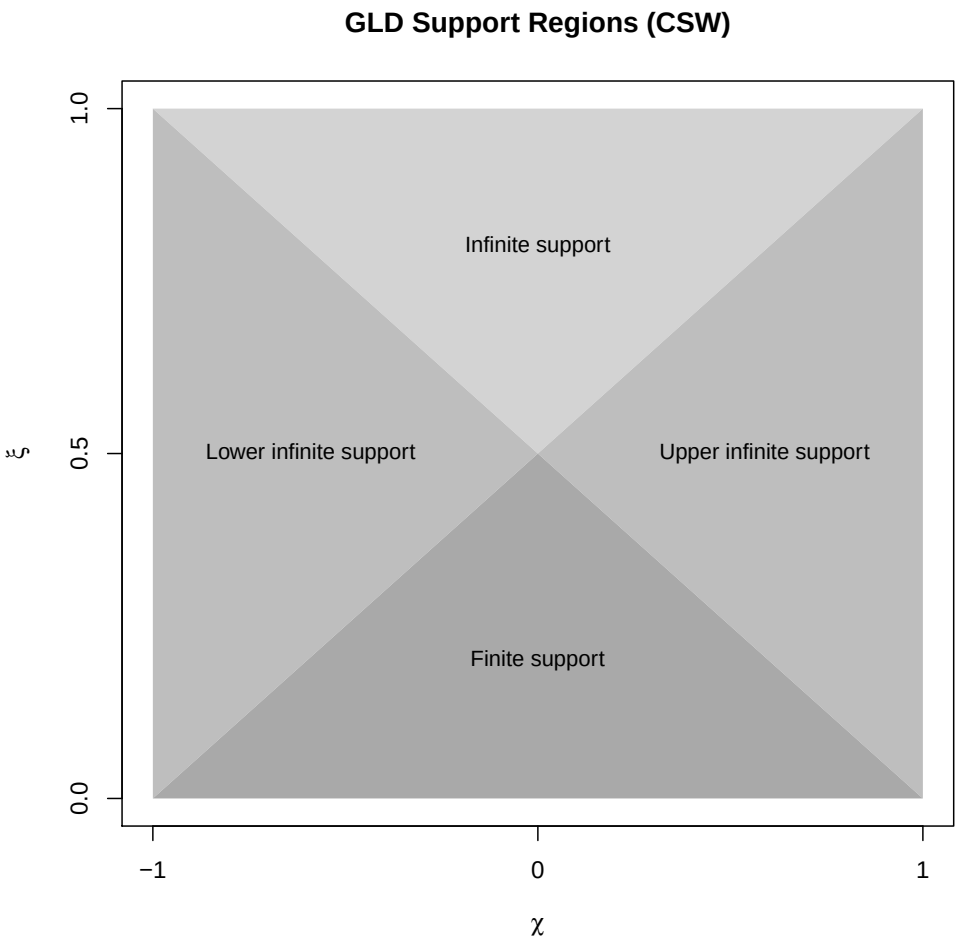


Figure 3.1: The four support regions of the GLD displayed in the shape plot with asymmetry parameter $\chi$ versus the steepness parameter $\xi$.

represented by the Beta function. The existence condition of the moment then follow the existence condition of the obtained Beta function. In the $\lambda$s representation of FKML in Eq. (2.2), the existence condition of the $k$th moment is $\min(\lambda_3, \lambda_4) > -1/k$. For the parameters $\alpha$ and $\beta$ defined in Eq. (3.5), this gives the existence condition

$$\min(\alpha + \beta, \alpha - \beta) > -1/k.$$

Some simple algebraic manipulation produces the condition of existence of the $k$th moment in terms of the shape parameters $\chi$ and $\xi$;

$$\xi < \frac{1}{2} - H\left(|\chi| - \sqrt{\frac{4}{4+k^2}}\right)\sqrt{\frac{1 - 2k|\beta| + k^2\beta^2}{4 - 8k|\beta| + k^2 + 4k^2\beta^2}},$$

where $H$ is a discontinuous function such that

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{otherwise,} \end{cases}$$

and $\beta$ is defined in Eq. (3.5b). Note that in the limiting cases of $\chi \to \pm 1$, the $k$th moment exists when and only when $\xi \to 0$. Figure 3.2 shows the condition lines of existence for the first four moments in the shape diagram. Any set of shape parameters $\chi$ and $\xi$ that is under the $k$th condition line defines a distribution with a finite $k$th moment.

## 3.4 Tail behavior

The tail behavior plays an important role in the modeling of stochastic processes. This is especially the case for financial returns that have been shown to have fat tail distributions. An important class of distributions are the regularly varying tail distribution such as the Student $t$, Fréchet, $\alpha$-stable and Pareto distributions that exhibit power law decays. They are suitable for modeling the large and small values typically observed in financial data set. Moreover, power law distributions become very handy in the derivation of the asymptotic distribution of estimators such as the regression estimators encountered in econometric models [Jessen and Mikosch, 2006].

We now restrict our attention to the shape parameter range for which the GLD has infinite right tail, $x \to \infty$; i.e., when $\xi > \frac{1}{2}(1 - \chi)$ as shown in Eqs. (3.11) and (3.12). From Eq. (3.7),

$$x = \tilde{\mu} + \tilde{\sigma}\frac{S(F(x)|\chi, \xi) - S(\frac{1}{2}|\chi, \xi)}{S(\frac{3}{4}|\chi, \xi) - S(\frac{1}{4}|\chi, \xi)},$$

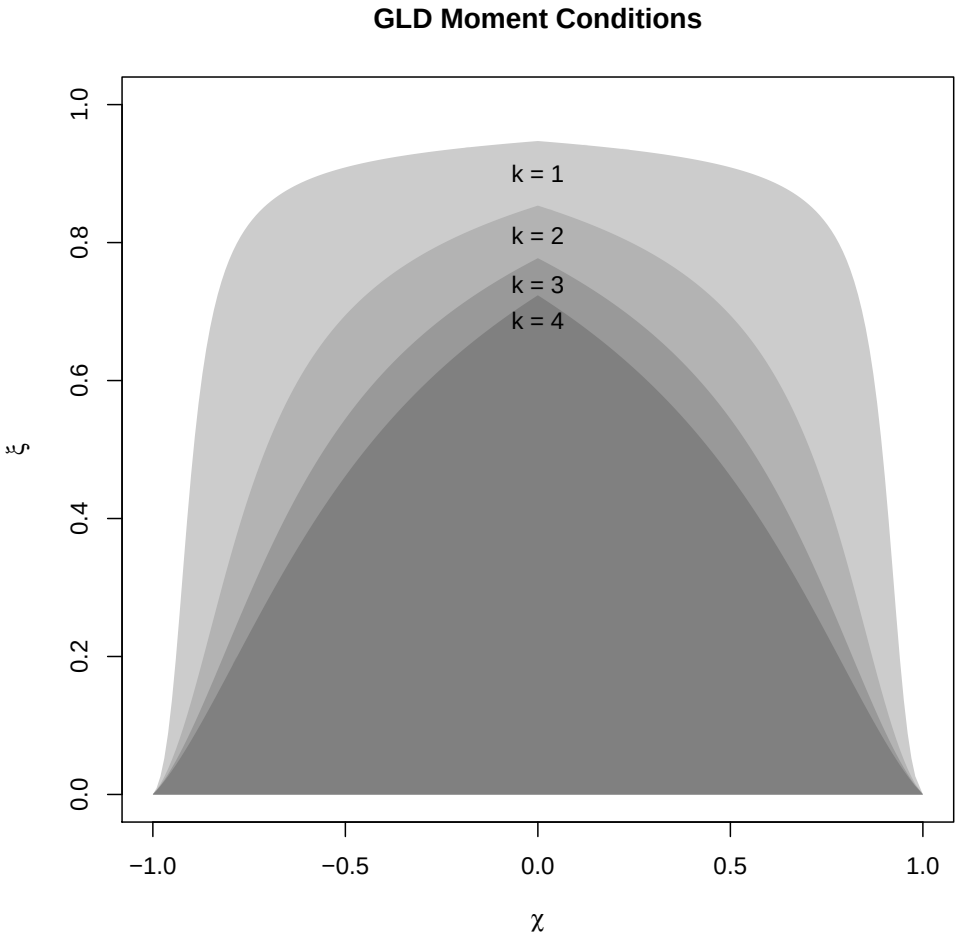


Figure 3.2: Shape conditions for the existence of moments $k$. Any set of parameters $\chi$ and $\xi$ that is under the $k$th line defines a distribution with finite $k$th moment.

where $S$ is defined in Eq. (3.6), $\alpha$ and $\beta$ are defined in Eq. (3.5). In region $\xi > \frac{1}{2}(1-\chi)$,

$$\begin{aligned}\lim_{x\to\infty} Q_{\mathrm{CSW}}(x) &\propto (1 - F_{\mathrm{CSW}}(x))^{\alpha-\beta} \\ &\propto \frac{\beta-\alpha}{\tilde{\sigma}} x,\end{aligned}$$

and hence

$$\frac{1 - F_{\mathrm{CSW}}(tx)}{1 - F_{\mathrm{CSW}}(x)} \sim t^{1/(\alpha-\beta)}.$$

The right tail of the GLD is thus regularly varying at $+\infty$ with index $-\frac{1}{\alpha-\beta}$ [see Embrechts et al., 1997, p. 129]. A similar argument shows that the left tail of the GLD is regularly varying at $-\infty$ with index $-\frac{1}{\alpha+\beta}$.

In the shape parameter region, $\xi > \frac{1}{2}(1+|\chi|)$, the left and right tail probabilities of the GLD are power laws in the asymptotic limit with index $-\frac{1}{\alpha+\beta}$ and $-\frac{1}{\alpha-\beta}$ respectively. Given the existence conditions of regularly varying distributions [Embrechts et al., 1997, Proposition A3.8], the $k$th moments of the GLD is therefore infinite when $k > \max(-\frac{1}{\alpha+\beta}, -\frac{1}{\alpha-\beta})$. This is in agreement with the existence conditions of moments presented in Section 3.3.

Moreover, from the monotone density theorem [Embrechts et al., 1997, Theorem A3.7], the density function is

$$f(x) \propto \frac{1}{\alpha-\beta}\left[\frac{\beta-\alpha}{\tilde{\sigma}} x\right]^{\frac{1}{\alpha-\beta}-1} \quad \text{as } x \to \infty,$$

and with the analogous result as $x \to -\infty$. Hence, The decay exponent for the left and right tail of the probability density function are:

$$\begin{aligned}\tau_{\mathrm{left}} &= \frac{1}{\alpha+\beta} - 1, \\ \tau_{\mathrm{right}} &= \frac{1}{\alpha-\beta} - 1,\end{aligned}$$

respectively. Since the left and right tail indices are related to the shape parameters $\chi$ and $\xi$, the parameter estimates can be used to form an estimate of the tail indices.

## 3.5 Distribution shape

The GLD exhibits four distribution shapes: unimodal, U-shape, monotone, and S-shape. In this section, conditions based on the derivatives of the quantile density $q$ in Eq. (3.9)

are derived for each distribution shape. Indeed, the density quantile function is defined as the multiplicative inverse of the quantile density function, as noted in Eq. (3.8). The first and second derivatives of $q$ are

$$\frac{d}{du}q(u|\tilde{\sigma},\chi,\xi) \propto (\alpha+\beta-1)u^{\alpha+\beta-2} - (\alpha-\beta-1)\,(1-u)^{\alpha-\beta-2}\,, \tag{3.13a}$$

$$\frac{d^2}{du^2}q(u|\tilde{\sigma},\chi,\xi) \propto (\alpha+\beta-1)(\alpha+\beta-2)u^{\alpha+\beta-3}+ (\alpha-\beta-1)(\alpha-\beta-2)\,(1-u)^{\alpha-\beta-3}\,, \tag{3.13b}$$

where $\alpha$ and $\beta$ are defined in Eq. (3.5). It is now possible to deduce the parameter conditions for each of the four shape regions of the GLD. Figure 3.3 summarizes the shape regions as described in the remaining part of this section.

**Unimodal** The GLD distribution is unimodal when its quantile density function is strictly convex; that is when $\frac{d^2}{du^2}q(u) > 0$ for all $u \in [0,1]$. Note that throughout this thesis, a unimodal distribution refers to a distribution with a single local maximum. From Eq. (3.13b), the GLD density function is then unimodal when

$$\alpha+\beta-2>0 \quad \text{and} \quad \alpha-\beta-2>0,$$

or when

$$\alpha+\beta-1<0 \quad \text{and} \quad \alpha-\beta-1<0.$$

After some tedious but straightforward calculations, the conditions are, in terms of the shape parameters $\chi$ and $\xi$,

$$0<\xi<\frac{1}{34}\left(17-4\sqrt{17}\right),$$
$$-2\sqrt{\frac{4-4\alpha+\alpha^2}{17-16\alpha+4\alpha^2}}<\chi<2\sqrt{\frac{4-4\alpha+\alpha^2}{17-16\alpha+4\alpha^2}},$$

or

$$\frac{1}{10}\left(5-2\sqrt{5}\right)<\xi<1,$$
$$-2\sqrt{\frac{1-2\alpha+\alpha^2}{5-8\alpha+4\alpha^2}}<\chi<2\sqrt{\frac{1-2\alpha+\alpha^2}{5-8\alpha+4\alpha^2}}.$$

Figure 3.4 illustrates the unimodal distribution shapes of the GLD.

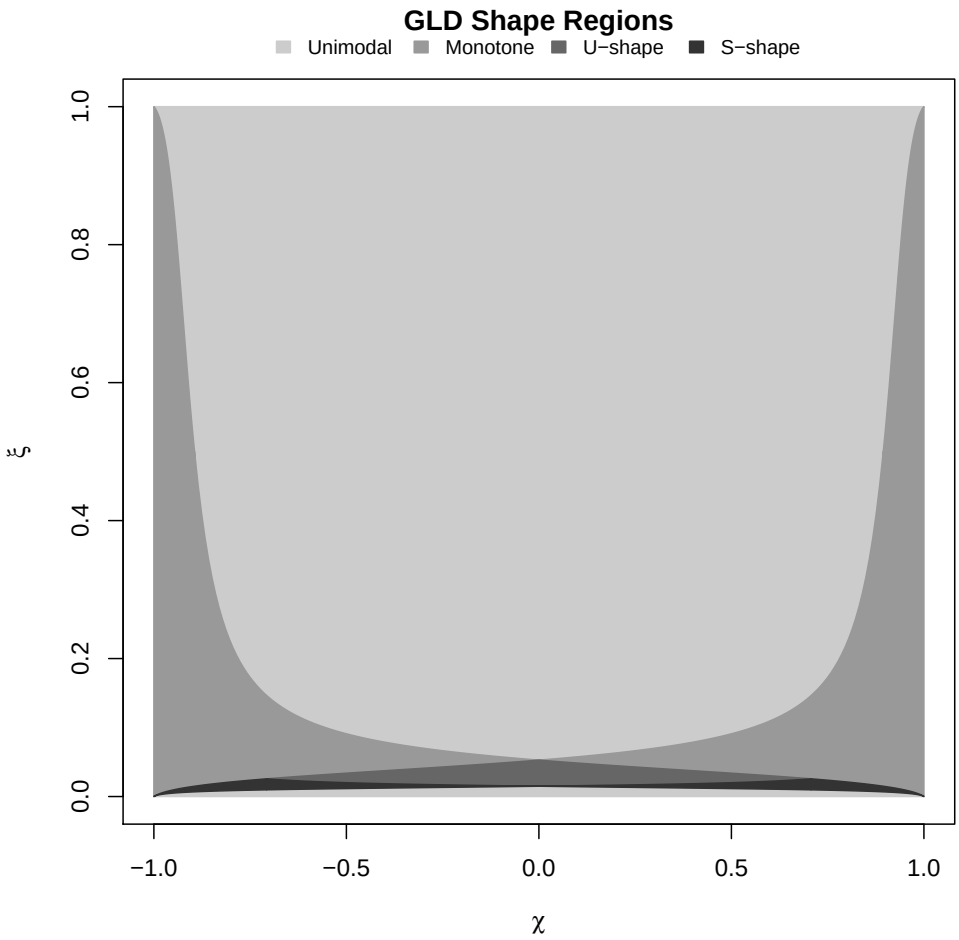


Figure 3.3: Parameter regions of the GLD shapes in the $(\chi, \xi)$-space.

**U-shape** Likewise, the GLD is U-shaped when the quantile density function is strictly concave, $\frac{d^2}{du^2}q(u) < 0$ for all $u \in [0,1]$. This is the case when

$$0 < \alpha + \beta - 1 < 1 \quad \text{and} \quad 0 < \alpha - \beta - 1 < 1,$$

giving the equivalent conditions

$$\frac{1}{20}\left(10 - 3\sqrt{10}\right) \leq \xi < \frac{1}{10}\left(5 - 2\sqrt{5}\right),$$
$$-2\sqrt{\frac{1 - 2\alpha + \alpha^2}{5 - 8\alpha + 4\alpha^2}} < \chi < 2\sqrt{\frac{1 - 2\alpha + \alpha^2}{5 - 8\alpha + 4\alpha^2}},$$

or

$$\frac{1}{34}\left(17 - 4\sqrt{17}\right) < \xi < \frac{1}{20}\left(10 - 3\sqrt{10}\right),$$
$$-2\sqrt{\frac{4 - 4\alpha + \alpha^2}{17 - 16\alpha + 4\alpha^2}} < \chi < 2\sqrt{\frac{4 - 4\alpha + \alpha^2}{17 - 16\alpha + 4\alpha^2}}.$$

Figure 3.4 illustrates U-shaped distribution shapes of the GLD.

**Monotone** The probability density function is monotone when its derivative is either positive or negative for all possible values in its support range. From Eq. (3.13a), this is the case when

$$\alpha + \beta - 1 > 0 \qquad \text{and} \qquad \alpha - \beta - 1 < 0,$$

or

$$\alpha + \beta - 1 < 0 \qquad \text{and} \qquad \alpha - \beta - 1 > 0.$$

In terms of the shape parameters, the monotone shape conditions are

$$0 < \chi \leq \frac{2}{\sqrt{5}},$$
$$\frac{1}{2} - \sqrt{\frac{1 + 2\beta + \beta^2}{5 + 8\beta + 4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{1 - 2\beta + \beta^2}{5 - 8\beta + 4\beta^2}},$$

or

$$\frac{2}{\sqrt{5}} < \chi < 1,$$

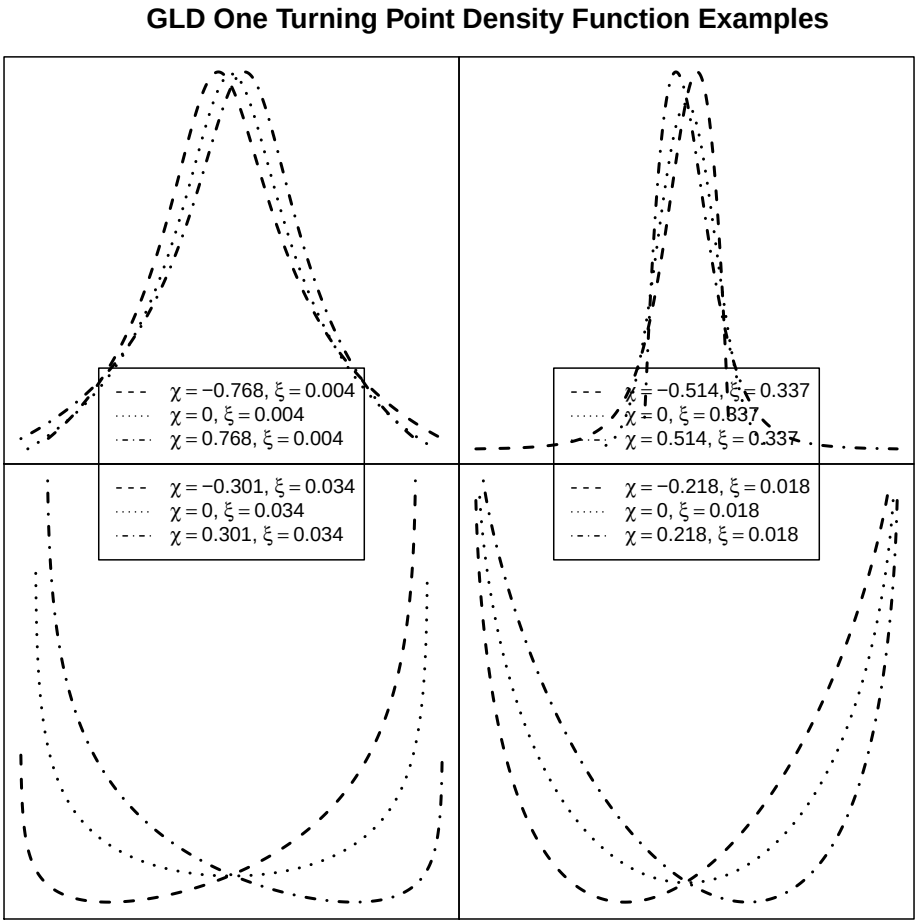


Figure 3.4: One turning point probability density function of the GLD with different sets of shape parameters.

$$\frac{1}{2} - \sqrt{\frac{1+2\beta+\beta^2}{5+8\beta+4\beta^2}} < \xi < \frac{1}{2} + \sqrt{\frac{1-2\beta+\beta^2}{5-8\beta+4\beta^2}},$$

or

$$-1 < \chi \leq -\frac{2}{\sqrt{5}},$$

$$\frac{1}{2} - \sqrt{\frac{1-2\beta+\beta^2}{5-8\beta+4\beta^2}} < \xi < \frac{1}{2} + \sqrt{\frac{1+2\beta+\beta^2}{5+8\beta+4\beta^2}},$$

or

$$-\frac{2}{\sqrt{5}} < \chi < 0,$$

$$\frac{1}{2} - \sqrt{\frac{1-2\beta+\beta^2}{5-8\beta+4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{1+2\beta+\beta^2}{5+8\beta+4\beta^2}}.$$

Figure 3.5 illustrates monotones shape of the GLD with different sets of shape parameters.

**S-shape** Since $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$, the remaining shape parameter regions are:

$$\alpha + \beta > 2 \quad \text{and} \quad 1 < \alpha - \beta < 2,$$

and

$$1 < \alpha + \beta < 2 \quad \text{and} \quad \alpha - \beta > 2.$$

Observe that the first derivative of the quantile density function at $u = 0$ has the same sign as the first derivative of the quantile density function at $u = 1$. This indicates that the gradients at the two end points of the probability density function tend toward the same direction. However, the gradient of the quantile density function has not the same sign for all possible values of the distribution. The GLD therefore has an S-shape density when

$$0 < \chi \leq \frac{1}{\sqrt{2}},$$

$$\frac{1}{2} - \sqrt{\frac{4+4\beta+\beta^2}{17+16\beta+4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{4-4\beta+\beta^2}{17-16\beta+4\beta^2}},$$

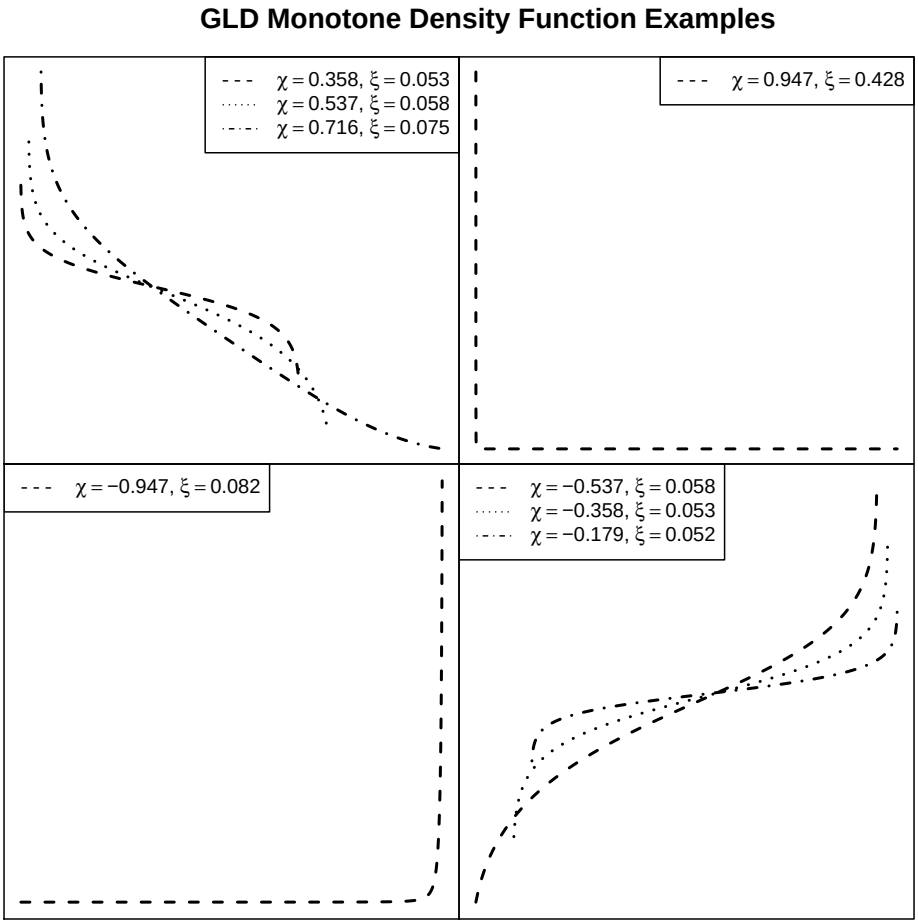


Figure 3.5: Monotone probability density function of the GLD with different sets of shape parameters.

or

$$\frac{1}{\sqrt{2}} < \chi < 1,$$

$$\frac{1}{2} - \sqrt{\frac{4+4\beta+\beta^2}{17+16\beta+4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{1+2\beta+\beta^2}{5+8\beta+4\beta^2}},$$

and

$$-1 < \chi \leq -\frac{1}{\sqrt{2}},$$

$$\frac{1}{2} - \sqrt{\frac{4-4\beta+\beta^2}{17-16\beta+4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{1-2\beta+\beta^2}{5-8\beta+4\beta^2}},$$

or

$$-\frac{1}{\sqrt{2}} < \chi < 0,$$

$$\frac{1}{2} - \sqrt{\frac{4-4\beta+\beta^2}{17-16\beta+4\beta^2}} < \xi < \frac{1}{2} - \sqrt{\frac{4+4\beta+\beta^2}{17+16\beta+4\beta^2}}.$$

Figure 3.6 illustrates S-shaped probability densities of the GLD.

## 3.6 Special cases

As seen in the previous section, the GLD can assume a wide range of distribution shapes. Here, values of the parameters in Eq. (3.7) are found such that $Q_{CSW}$ replicates several common distributions. The method used to accomplish this is as follows. Taking a set of equidistant probabilities, $p_i = \frac{i}{N+1}$, with $i = 1, \ldots, N$ and $N = 500$, the respective quantiles, $x_i$, and densities, $d_i = f(x_i)$, are calculated for each of the target distributions. The shape parameters, $\hat{\chi}$ and $\hat{\xi}$, are then fitted, by minimizing the maximum absolute quantile error (MQE):

$$\sup_{\forall i} |Q_{CSW}(p_i|\tilde{\mu}, \tilde{\sigma}, \hat{\chi}, \hat{\xi}) - x_i|.$$

The median and interquartile range of the target distributions are used for the location and scale parameters, $\tilde{\mu}$ and $\tilde{\sigma}$. Note that $N = 500$ was explicitly chosen in order to compare the results of this work with those of the previous studies of Gilchrist [2000], King and MacGillivray [2007] and Tarsitano [2010].

Table 3.1 lists the fitted shape parameters for some common distributions along with the

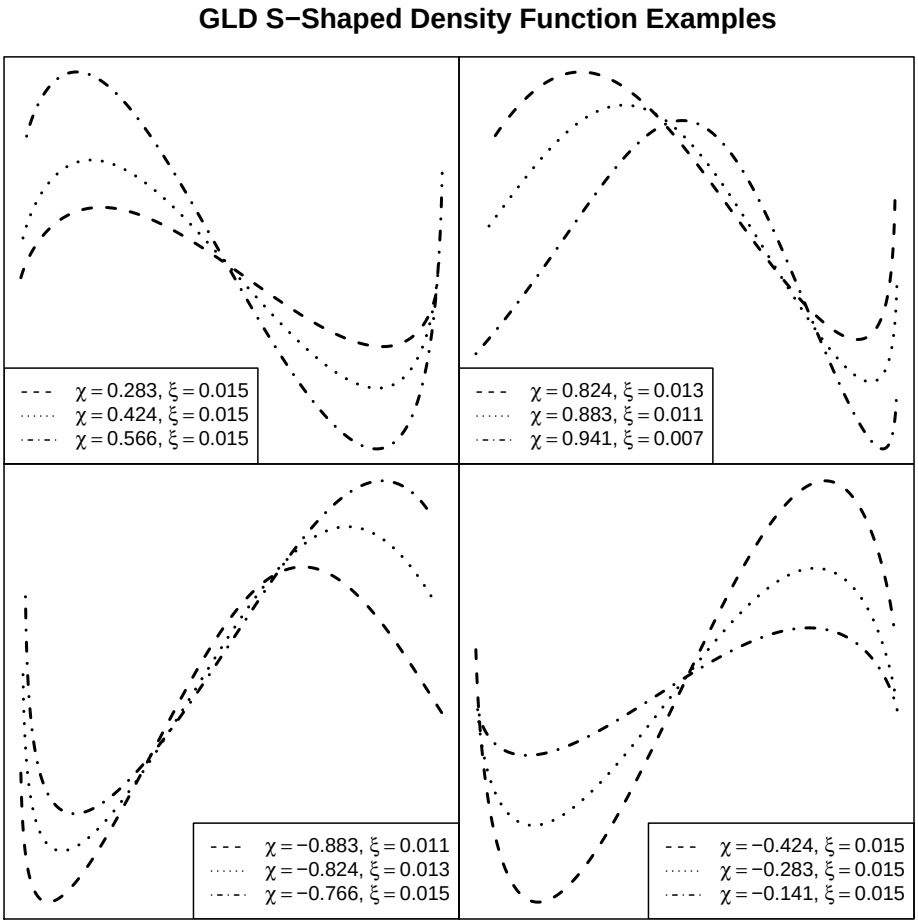


Figure 3.6: S-shape probability density function of the GLD with different sets of shape parameters.

| Distribution | Parameters | $\hat{\chi}$ | $\hat{\xi}$ | $sup\|\hat{Q}\|$ | $\sup\|\hat{F}\|$ | $\sup\|\hat{f}\|$ |
|---|---|---|---|---|---|---|
| Normal | $\mu = 0,\ \sigma = 1$ | 0.0000 | 0.3661 | 0.012 | 0.001 | 0.001 |
| Student's $t$ | $\nu = 1$ | 0.0000 | 0.9434 | 1.587 | 0.005 | 0.012 |
| | $\nu = 5$ | 0.0000 | 0.5778 | 0.069 | 0.003 | 0.004 |
| | $\nu = 10$ | 0.0000 | 0.4678 | 0.033 | 0.002 | 0.003 |
| Laplace | $\mu = 0,\ b = 1$ | 0.0000 | 0.6476 | 0.257 | 0.015 | 0.093 |
| Stable | $\alpha = 1.9,\ \beta = 0$ | 0.0000 | 0.5107 | 0.399 | 0.010 | 0.010 |
| | $\alpha = 1.9,\ \beta = 0.5$ | 0.0730 | 0.5307 | 0.584 | 0.014 | 0.013 |
| Gamma | $k = 4,\ \theta = 1$ | 0.4120 | 0.3000 | 0.120 | 0.008 | 0.012 |
| $\chi^2$ | $k = 3$ | 0.6671 | 0.1991 | 0.295 | 0.015 | 0.072 |
| | $k = 5$ | 0.5193 | 0.2644 | 0.269 | 0.011 | 0.017 |
| | $k = 10$ | 0.3641 | 0.3150 | 0.233 | 0.007 | 0.004 |
| Weibull | $k = 3,\ \lambda = 1$ | 0.0908 | 0.3035 | 0.007 | 0.003 | 0.016 |
| Log Normal | $\mu = 0,\ \sigma = 0.25$ (log scale) | 0.2844 | 0.3583 | 0.011 | 0.007 | 0.052 |
| Gumbel | $\alpha = 0.5,\ \beta = 2$ | -0.3813 | 0.3624 | 0.222 | 0.010 | 0.010 |
| Inv. Gaussian | $\mu = 1,\ \lambda = 3$ | 0.5687 | 0.2957 | 0.096 | 0.022 | 0.175 |
| | $\mu = 0.5,\ \lambda = 6$ | 0.3267 | 0.3425 | 0.008 | 0.008 | 0.125 |
| NIG | $\mu = 0,\ \delta = 1,\ \alpha = 2,\ \beta = 1$ | 0.2610 | 0.4975 | 0.124 | 0.014 | 0.029 |
| Hyperbolic | $\mu = 0,\ \delta = 1,\ \alpha = 2,\ \beta = 1$ | 0.2993 | 0.4398 | 0.198 | 0.021 | 0.030 |

Table 3.1: Shape parameters of the GLD to approximate common distributions.

maximum MQE appearing in the $\sup|\hat{Q}|$ column. Also reported are the maximum probability error, $\sup|\hat{F}|$, and the maximum density error, $\sup|\hat{f}|$. The maximum probability error (MPE) is defined as

$$\sup_{\forall i}|F_{CSW}(x_i|\tilde{\mu}, \tilde{\sigma}, \hat{\chi}, \hat{\xi}) - p_i|.$$

The maximum density error (MDE) is defined as

$$\sup_{\forall i}|f_{CSW}(x_i|\tilde{\mu}, \tilde{\sigma}, \hat{\chi}, \hat{\xi}) - d_i|.$$

An important distinction between this method and those of previous studies, is that here the appropriate values of the location and scale parameters, being identified with the median and interquartile range of the target distribution, are known. In previous works, the location and scale parameters had to be estimated alongside the tail parameters. In so doing, the fitted parameters can sometimes produce a GLD that does not well approximate the target distribution around its center. This problem arises from adjustment of the location and scale parameter values to improve a poor fit over the tails of the distribution. This is at the expense of having a worse fit around the center. Figure 3.7 demonstrates this issue with the Student $t$-distribution of two degrees of freedom. The left-hand-side plot displays the fitted GLD obtained using the exact values for $\tilde{\mu}$ and $\tilde{\sigma}$. The right-hand-side

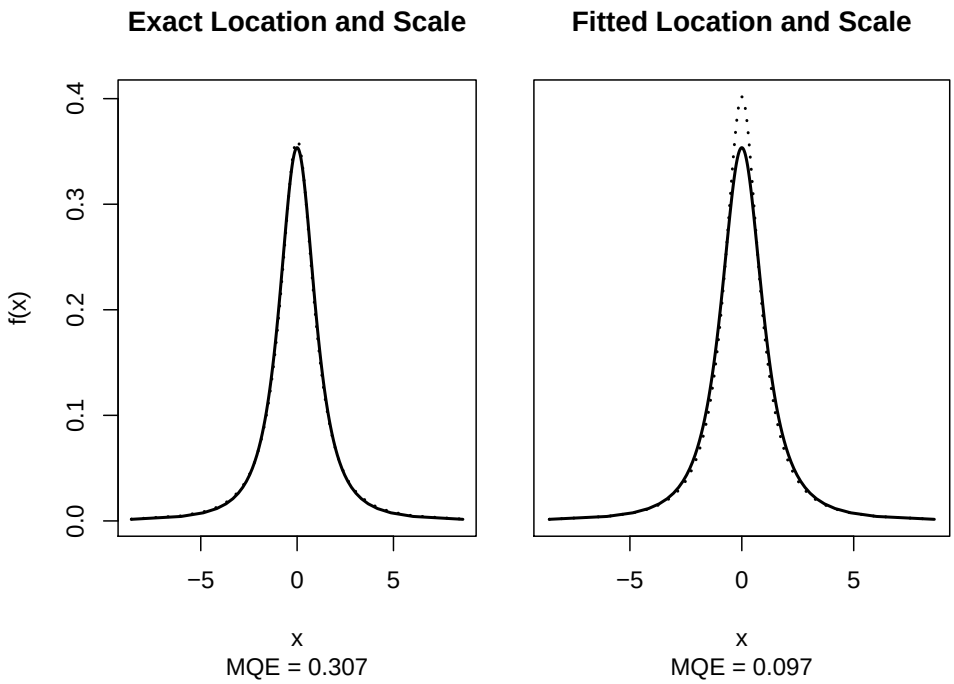


Figure 3.7: Approximation of the Student $t$-distribution (dotted lines) with two degrees of freedom by either fitting or using the true values for the location and scale parameters, $\tilde{\mu}$ and $\tilde{\sigma}$. In the left-hand-side figure, the fitted GLD where the true values for the median and interquartile range were used. In the right-hand-side figure, the location and scale parameters were included in the estimation. Note the maximum quantile error (MQE) for the left hand side figure, `0.307`, is larger than the one for the right hand side figure, `0.097`, although it has a better visual fit.

plot shows the fitted GLD obtained when including the location and scale parameters in the inexact estimation. Clearly, the center of the distribution is not well described by the latter solution. Using the known values for $\tilde{\mu}$ and $\tilde{\sigma}$ yields a greater MQE value (of `0.307`, in this case) than that of the alternate approach (for which, the MQE value was `0.097`).

It is interesting to note that the MPE could have been used in place of the MQE while fitting the shape parameters. This would have resulted in a smaller MPE value. However, in practice, GLDs fitted according to the MPE estimator do not well approximate the tails of the target distribution. This is especially the case with fat-tailed distributions. Figure 3.8 illustrates this tendency. The left-hand-side plot displays the fitted log-CDF obtained by minimizing the MQE for the Student $t$-distribution of two degrees of freedom. The right-hand-side plot is the fitted log-CDF obtained with the MPE estimator. Note that the true median and interquartile range were used in both cases. The parameter-value

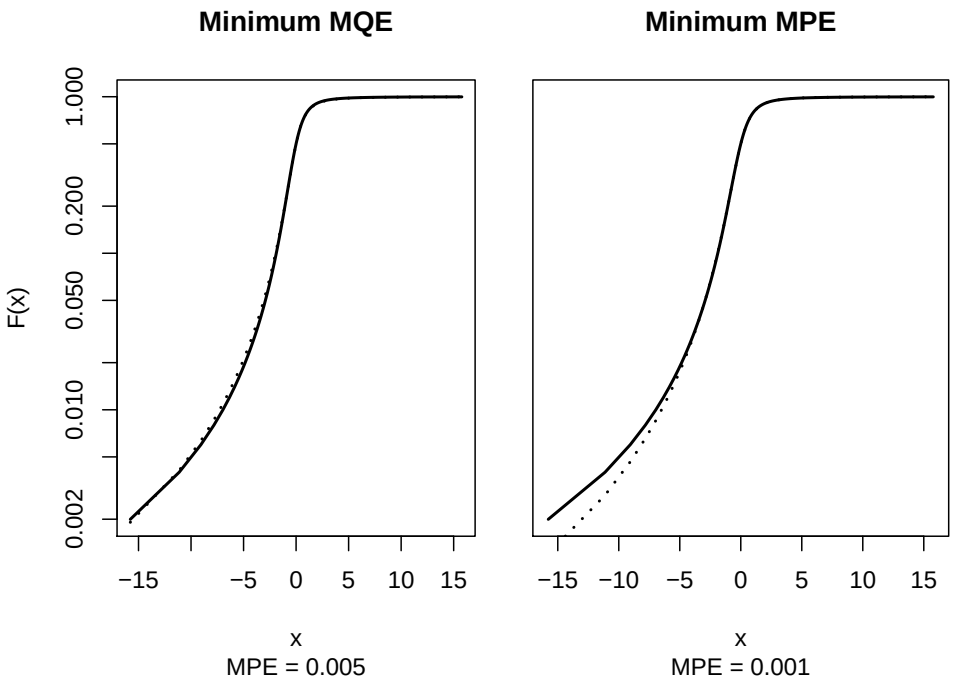


Figure 3.8: Comparison of the fitted GLD (dotted lines) to the Student $t$-distribution with two degrees of freedom by either using the MQE or the MPE estimator. The left-hand-side figure was obtained by minimizing the MQE and the right-hand-side plot with the MPE. The MPE for the left hand side plot, `0.005`, is larger than the one for the right hand side figure, `0.001`, although it has a better visual fit.

set obtained with the MQE has produced the better visual fit.

Another important consideration in our approach is to ensure that the fitted parameters yield a GLD with a support that includes all $x_i$. This is especially the case for the gamma, $\chi^2$ and Wald distributions. Indeed, it is possible to find parameter-value sets for which the MQE is small, but where the fitted distribution does not include all $x_i$. We therefore add constraints in the optimization routine to ensure that all $x_i$ are included in the fitted GLD support range.

These considerations motivated the decisions to use the known values for the location and scale parameters, to ensure that all points are included in the support of the fitted distribution, and to use the MQE estimator to fit the shape parameters.

The GLD includes the uniform, logistic, and exponential distributions as special cases. The corresponding GLD parameter values, obtained from Eqs. (3.6), (3.7) and (3.10), are summarized in Table 3.2. Note that the exponential distribution corresponds to the

| | $\tilde{\mu}$ | $\tilde{\sigma}$ | $\chi$ | $\xi$ |
|---|---|---|---|---|
| Uniform $(a, b)$ | $\frac{1}{2}(a+b)$ | $\frac{1}{2}(b-a)$ | 0 | $\frac{1}{2} - \frac{1}{\sqrt{5}}$ |
| | $\frac{1}{2}(a+b)$ | $\frac{1}{2}(b-a)$ | 0 | $\frac{1}{2} - \frac{2}{\sqrt{17}}$ |
| Logistic $(\mu, \beta)$ | $\mu$ | $\beta \ln(9)$ | 0 | $\frac{1}{2}$ |
| Exponential $(\lambda)$ | $\frac{1}{\lambda}\ln(2)$ | $\frac{1}{\lambda}\ln(3)$ | 1 | 0 |

Table 3.2: Special cases of the GLD in the CSW parameterization.

limiting case $\{\chi \to 1,\ \xi \to 0\}$, as mentioned at the end of Section 3.1.

## 3.7 Limiting cases

In this section, the quantile and quantile density functions, as given in Eqs. (3.7) and (3.9), are derived for special cases where the shape parameters tend to their limiting values. These cases are:

1. $\chi \to -1$ and $\xi \to 0$,
2. $\chi \to 1$ and $\xi \to 0$,
3. $\chi \to 1$ and $\xi \to 1$,
4. $\chi \to -1$ and $\xi \to 1$,
5. $\chi \to -1$,
6. $\chi \to 1$,
7. $\xi \to 1$,
8. $\xi \to 0$.

The derivation of these limiting cases is important for the implementation of the GLD, as well as for recovering all included special cases distributions of the GLD as seen in Section 3.6. The most important limiting cases are $\{\chi \to -1,\ \xi \to 0\}$ and $\{\chi \to 1,\ \xi \to 0\}$, because they define valid probability distributions. The other listed cases do not. The complete calculations follow. Readers whom are uninterested in these details may skip ahead to Chapter 4.

The combined limits $\chi \to -1$ and $\xi \to 0$ equate to the particular case where $\xi = \frac{1}{2}(1+\chi)$ and $S(u) = \ln(u) - \frac{1}{a}\left[(1-u)^a - 1\right]$ with $a = \frac{\frac{1}{2}-\xi}{\sqrt{\xi(1-\xi)}}$ as seen in Eq. (3.6). Hence,

$$\begin{aligned}\lim_{\substack{\chi\to-1\\ \xi\to 0}} S(u) &= \ln(u) - \lim_{a\to\infty}\frac{1}{a}\left[(1-u)^a - 1\right]\\ &= \ln(u).\end{aligned}$$

From Eq. (3.7), the quantile function is

$$\lim_{\substack{\chi\to-1\\ \xi\to 0}} Q_{CSW}(u) = \tilde{\mu} + \tilde{\sigma}\frac{\ln(u)+\ln(2)}{\ln(3)}.$$

The quantile density function, $q = Q'$, is then

$$\lim_{\substack{\chi\to-1\\ \xi\to 0}} q_{CSW}(u) = \frac{\tilde{\sigma}}{\ln(3)}\frac{1}{u}.$$

The validity conditions for a quantile density distribution, $q$, of probabilities, $u$, is that q(u) is finite and $q(u) \geq 0$ for $0 < u < 1$. Both of these conditions are satisfied by the derived quantile density function above, and hence the derived quantile function describes a valid probability distribution.

When $\chi \to 1$ and $\xi \to 0$, it follows that $S(u) = \frac{1}{b}(u^b - 1) - \ln(1-u)$ with $b = \frac{\chi}{\sqrt{1-\chi^2}}$. Hence,

$$\begin{aligned}\lim_{\substack{\chi\to 1\\ \xi\to 0}} S(u) &= \lim_{b\to\infty}\frac{1}{b}(u^b-1) - \ln(1-u)\\ &= -\ln(1-u).\end{aligned}$$

The quantile function becomes

$$\lim_{\substack{\chi\to 1\\ \xi\to 0}} Q_{CSW}(u) = \tilde{\mu} - \tilde{\sigma}\frac{\ln(1-u)+\ln(2)}{\ln(3)},$$

and the quantile density function is

$$\lim_{\substack{\chi\to 1\\ \xi\to 0}} q_{CSW}(u) = \frac{\tilde{\sigma}}{\ln(3)}\frac{1}{1-u}.$$

Again, it is straightforward to show that the derived quantile function describes a valid probability distribution.

When $\chi \to 1$ and $\xi \to 1$, it follows that $S(u) = \ln(u) - \frac{1}{a}\left[(1-u)^a - 1\right]$ where $a = \frac{\frac{1}{2}-\xi}{\sqrt{\xi(1-\xi)}}$. Hence,

$$\lim_{\substack{\chi\to 1\\ \xi\to 0}} Q_{CSW}(u) = \tilde{\mu} + \tilde{\sigma}\lim_{a\to-\infty}\frac{\ln(u) - \frac{1}{a}\left[(1-u)^a - 1\right] + \ln(2) + \frac{1}{a}\left[\left(\frac{1}{2}\right)^a - 1\right]}{\ln(3) - \frac{1}{a}\left[\left(\frac{1}{4}\right)^a - 1\right] + \frac{1}{a}\left[\left(\frac{3}{4}\right)^a - 1\right]}$$

$$
\begin{aligned}
&= \tilde{\mu} + \tilde{\sigma} \lim_{a \to -\infty} \frac{\left(\frac{1}{2}\right)^a - (1-u)^a}{\left(\frac{3}{4}\right)^a - \left(\frac{1}{4}\right)^a} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{a \to -\infty} \frac{2^a - (4-4u)^a}{3^a - 1} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{a \to -\infty} (4-4u)^a \\
&= \begin{cases} \tilde{\mu} & (0 < u < 3/4)\,, \\ \tilde{\mu} + \tilde{\sigma} & (u = 3/4)\,, \\ \infty & (3/4 < u < 1)\,. \end{cases}
\end{aligned}
$$

The resulting quantile function diverges for $3/4 < u < 1$, and therefore does not correspond to a valid probability distribution.

When $\chi \to -1$ and $\xi \to 1$, it follows that $S(u) = \frac{1}{b}(u^b - 1) - \ln(1-u)$ with $b = \frac{\chi}{\sqrt{1-\chi^2}}$. Hence,

$$
\begin{aligned}
\lim_{\substack{\chi \to 1 \\ \xi \to 0}} Q_{CSW}(u) &= \tilde{\mu} + \tilde{\sigma} \lim_{b \to -\infty} \frac{\frac{1}{b}\left(u^b - 1\right) - \ln(1-u) - \frac{1}{b}\left[\left(\frac{1}{2}\right)^b - 1\right]}{\frac{1}{b}\left[\left(\frac{3}{4}\right)^b - 1\right] - \frac{1}{b}\left[\left(\frac{1}{4}\right)^b - 1\right]} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{b \to -\infty} \frac{u^b - \left(\frac{1}{2}\right)^b}{\left(\frac{3}{4}\right)^b - \left(\frac{1}{4}\right)^b} \\
&= \tilde{\mu} - \tilde{\sigma} \lim_{b \to -\infty} (4u)^b \\
&= \begin{cases} -\infty & \text{if } 0 < u < 1/4, \\ \tilde{\mu} - \tilde{\sigma} & \text{if } u = 1/4, \\ \tilde{\mu} & \text{if } 1/4 < u < 1. \end{cases}
\end{aligned}
$$

The resulting quantile function diverges for $0 < u < 1/4$, and therefore does not correspond to a valid distribution.

For the remaining limiting cases, $\{\chi \to \pm 1,\ \xi \to 0,\ \xi \to 1\}$, the GLD quantile function becomes

$$
Q^*_{CSW}(u) = \tilde{\mu} + \tilde{\sigma} \frac{\frac{u^{\alpha+\beta}-1}{\alpha+\beta} - \frac{(1-u)^{\alpha-\beta}-1}{\alpha-\beta} - \frac{\left(\frac{1}{2}\right)^{\alpha+\beta}-1}{\alpha+\beta} + \frac{\left(\frac{1}{2}\right)^{\alpha-\beta}-1}{\alpha-\beta}}{\frac{\left(\frac{3}{4}\right)^{\alpha+\beta}-1}{\alpha+\beta} - \frac{\left(\frac{1}{4}\right)^{\alpha-\beta}-1}{\alpha-\beta} - \frac{\left(\frac{1}{4}\right)^{\alpha+\beta}-1}{\alpha+\beta} + \frac{\left(\frac{3}{4}\right)^{\alpha-\beta}-1}{\alpha-\beta}}, \tag{3.14}
$$

where $\alpha$ and $\beta$ are defined in Eq. (3.5). Note that the superscript $^*$ is used to distinguish the above expression from the full definition of the GLD quantile function in Eq. (3.7). When $\chi \to -1$ and $0 < \xi < 1$, it is necessary to consider the limit of $\tilde{Q}_{CSW}$ as $\beta \to -\infty$. From Eq. (3.14),

$$\begin{aligned}
\lim_{\chi \to -1} Q_{CSW}(u) &= \lim_{\beta \to -\infty} Q^*_{CSW}(u) \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\beta \to -\infty} \frac{u^\beta - \left(\frac{1}{2}\right)^\beta}{\left(\frac{3}{4}\right)^\beta - \left(\frac{1}{4}\right)^\beta} \\
&= \tilde{\mu} - \tilde{\sigma} \lim_{\beta \to -\infty} (4u)^\beta .
\end{aligned}$$

Hence, when $\chi \to -1$,

$$\lim_{\chi \to -1} Q_{CSW}(u) = \begin{cases} -\infty & \text{if } 0 < u < \frac{1}{4}, \\ \mu - \sigma & \text{if } u = \frac{1}{4}, \\ \mu & \text{if } \frac{1}{4} < u < 1. \end{cases}$$

The resulting quantile function is divergent for $0 < u < 1/4$, and therefore does not correspond to a valid probability distribution.

For $\chi \to 1$ and $0 < \xi < 1$, it follows that $\beta \to \infty$. From Eq. (3.14),

$$\begin{aligned}
\lim_{\chi \to 1} Q_{CSW}(u) &= \lim_{\beta \to \infty} Q^*_{CSW}(u) \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\beta \to \infty} \frac{-(1-u)^{-\beta} + \left(\frac{1}{2}\right)^{-\beta}}{\left(\frac{3}{4}\right)^{-\beta} - \left(\frac{1}{4}\right)^{-\beta}} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\beta \to \infty} \frac{1}{(4-4u)^\beta}.
\end{aligned}$$

Hence,

$$\lim_{\chi \to 1} Q_{CSW}(u) = \begin{cases} \tilde{\mu} & \text{if } 0 < u < 3/4, \\ \tilde{\mu} + \tilde{\sigma} & \text{if } u = 3/4, \\ \infty & \text{if } 3/4 < u < 1. \end{cases}$$

The resulting quantile function diverges when $3/4 < u < 1$, and therefore does not produce a valid probability distribution.

The conditions $\xi \to 1$ and $-1 < \chi < 1$ imply $\alpha \to -\infty$. From Eq. (3.14),

$$\begin{aligned}
\lim_{\xi\to 1} Q_{CSW}(u) &= \lim_{\alpha\to-\infty} Q^*_{CSW}(u) \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\alpha\to-\infty} \frac{u^{\alpha+\beta} - (1-u)^{\alpha-\beta} - 2^{-\alpha-\beta} + 2^{\beta-\alpha}}{\left(\frac{3}{4}\right)^{\alpha+\beta} - 4^{\beta-\alpha} - 4^{-\alpha-\beta} + \left(\frac{3}{4}\right)^{\alpha-\beta}} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\alpha\to-\infty} \frac{4^{\alpha+\beta}u^{\alpha+\beta} - 2^{\alpha+\beta} + 2^{\alpha+3\beta}}{3^{\alpha+\beta} - 16^{\beta} - 1 + 16^{b}3^{a-b}} \\
&\quad - \tilde{\sigma} \lim_{\alpha\to-\infty} \frac{(4-4u)^{\alpha-\beta} - 2^{\alpha-3\beta} + 2^{\alpha-\beta}}{16^{-\beta}3^{\alpha+\beta} - 1 - 16^{\beta} + 3^{\alpha-\beta}} \\
&= \tilde{\mu} - \frac{\tilde{\sigma}}{1+16^{\beta}} \left[ \lim_{\alpha\to-\infty} (4u)^{\alpha+\beta} - 16^{\beta} \lim_{\alpha\to-\infty} (4-4u)^{\alpha-\beta} \right].
\end{aligned}$$

Hence,

$$\lim_{\xi\to 1} Q_{CSW}(u) = \begin{cases} -\infty & \text{if } 0 < u < 1/4, \\ \tilde{\mu} - \tilde{\sigma}\left(1+16^{\beta}\right)^{-1} & \text{if } u = 1/4, \\ \tilde{\mu} & \text{if } 1/4 < u < 3/4, \\ \tilde{\mu} + 16^{\beta}\ \tilde{\sigma}\left(1+16^{\beta}\right)^{-1} & \text{if } u = 3/4, \\ \infty & \text{if } 3/4 < u < 1. \end{cases}$$

The resulting quantile function diverges for $1 < u < 1/4$ and also for $3/4 < u < 1$. It does not, therefore, correspond to a valid probability distribution.

When $\xi \to 0$ and $-1 < \chi < 1$, it follows that $\alpha \to \infty$. Equation (3.14) becomes

$$\begin{aligned}
\lim_{\xi\to 0} Q_{CSW}(u) &= \lim_{\alpha\to\infty} Q^*_{CSW}(u) \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\alpha\to\infty} \frac{u^{\alpha+\beta} - (1-u)^{\alpha-\beta} - 2^{-\alpha-\beta} + 2^{\beta-\alpha}}{\left(\frac{3}{4}\right)^{\alpha+\beta} - 4^{\beta-\alpha} - 4^{-\alpha-\beta} + \left(\frac{3}{4}\right)^{\alpha-\beta}} \\
&= \tilde{\mu} + \tilde{\sigma} \lim_{\alpha\to\infty} \frac{\left(\frac{4}{3}\right)^{\alpha+\beta} u^{\alpha+\beta}}{1 - 16^{\beta}3^{-\alpha-\beta} - 3^{-\alpha-\beta} + 9^{-\beta}16^{\beta}} \\
&\quad - \tilde{\sigma} \lim_{\alpha\to\infty} \frac{\left(\frac{4}{3}\right)^{\alpha-\beta} (1-u)^{\alpha-\beta}}{9^{\beta}16^{-\beta} - 3^{\beta-\alpha} - 16^{-\beta}3^{\beta-\alpha} + 1} \\
&= \tilde{\mu} + \frac{\tilde{\sigma}}{9^{\beta}+16^{\beta}} \left[ 9^{\beta} \lim_{\alpha\to\infty} \left(\frac{4u}{3}\right)^{\alpha+\beta} - 16^{\beta} \lim_{\alpha\to\infty} \left(\frac{4}{3} - \frac{4u}{3}\right)^{\alpha-\beta} \right],
\end{aligned}$$

and so

$$\lim_{\xi \to 0} Q_{CSW}(u) = \begin{cases} -\infty & \text{if } 0 < u < 1/4, \\ \tilde{\mu} - 16^{\beta}\ \tilde{\sigma}\left(9^{\beta} + 16^{\beta}\right)^{-1} & \text{if } u = 1/4, \\ \tilde{\mu} & \text{if } 1/4 < u < 3/4, \\ \tilde{\mu} + 9^{\beta}\ \tilde{\sigma}\left(9^{\beta} + 16^{\beta}\right)^{-1} & \text{if } u = 3/4, \\ \infty & \text{if } 3/4 < u < 1. \end{cases}$$

The resulting quantile function diverges for $1 < u < 1/4$ and also for $3/4 < u < 1$. Therefore, it does not correspond to a valid probability distribution.

# 4 Parameter estimation

Estimating parameter values for the generalized lambda distribution (GLD) is notoriously difficult. This is due to abrupt distribution shape changes that accompany variations of the parameter values in the different shape regions (Fig. 2.2). In this section, an extensive Monte-Carlo study of different GLD estimators is presented. This includes maximum log-likelihood, maximum product of spacing, goodness-of-fit, histogram binning, quantile sequence, and linear moments. Besides these estimators, we introduce the maximum product of spacing and a robust moment matching to estimate the shape parameters of the GLD.

## 4.1 Robust moments matching

A robust moment approach, for the estimation of the distribution parameters, is presented. The method is based on the median, interquartile range, Bowley's skewness and Moors' kurtosis. It yields estimators for the two shape parameters, which are independent of the location and scale parameters. A similar approach has been introduced by King and MacGillivray [2007].

The method suggested by Ramberg et al. [1979] for fitting GLD distributions, uses a method of moments wherein the first four moments are matched. Clearly, this approach can only be applied to regions within the first four moments exist. In addition, parameter estimates based upon sample moments are highly sensitive to outliers. This is especially true of third and fourth moment. To circumvent the estimation of sample moments, Karian and Dudewicz [1999] considered a quantile approach that estimates the parameters from four sample quantile statistics. In that approach, the four statistics depend upon the four parameters of the GLD, thereby leading to a system of four nonlinear equations to be solved.

In the CSW parameterization, because the location and scale parameters are, respectively, the median and interquartile range, they can be directly estimated by their sample estimators. What remains to be estimated, are two shape parameters; $\chi$ and $\xi$. This is achieved by using the robust skewness ratio $\tilde{s}$ of Bowley [1920], and the robust kurtosis

ratio $\tilde{\kappa}$ of Moors [1988]. These robust moments are defined as

$$\tilde{s} = \frac{\pi_{3/4} + \pi_{1/4} - 2\pi_{2/4}}{\pi_{3/4} - \pi_{1/4}},$$

$$\tilde{\kappa} = \frac{\pi_{7/8} - \pi_{5/8} + \pi_{3/8} - \pi_{1/8}}{\pi_{6/8} - \pi_{2/8}},$$

where $\pi_q$ indicates the $q$th quantile, and the tilde indicates the robust versions of the moments. For a detailed discussion of Bowley's skewness and Moors' kurtosis statistics, please refer to [Kim and White, 2004].

Recall that the quantile function of the GLD in the CSW parameterization is

$$Q_{\text{CSW}}(u|\tilde{\mu}, \tilde{\sigma}, \chi, \xi) = \tilde{\mu} + \tilde{\sigma}\frac{S(u|\chi, \xi) - S(\frac{1}{2}|\chi, \xi)}{S(\frac{3}{4}|\chi, \xi) - S(\frac{1}{4}|\chi, \xi)},$$

as seen in Eq. (3.7) where $S$ is defined in Eq. (3.6). The population robust skewness and robust kurtosis thus depend only upon the shape parameters. Explicitly,

$$\tilde{s} = \frac{S(3/4|\chi, \xi) + S(1/4|\chi, \xi) - 2S(2/4|\chi, \xi)}{S(3/4|\chi, \xi) - S(1/4|\chi, \xi)}, \tag{4.1a}$$

$$\tilde{\kappa} = \frac{S(7/8|\chi, \xi) - S(5/8|\chi, \xi) + S(3/8|\chi, \xi) - S(1/8|\chi, \xi)}{S(6/8|\chi, \xi) - S(2/8|\chi, \xi)}. \tag{4.1b}$$

Estimates of the shape parameters, $\chi$ and $\xi$, can then be calculated by solving the nonlinear system of equations in Eq. (4.1). The advantage of this method, wherein the GLD shape parameters are readily obtained from quantile-based estimators, is that it reduces the nonlinear system from four to two unknowns. As a secondary application, this simplified two-equation system can be efficiently solved by a lookup method in order to obtain good initial estimates of the shape parameters for use by other fitting methods that are sensitive to initial values and might otherwise get trapped in local minima.

## 4.2 Histogram approaches

The histogram approach is appealing and simple; the empirical data are binned in a histogram and the resulting probabilities, taken to be at the midpoints of the histogram bins, are fitted to the true GLD density. This approach was considered by Su [2005] for the GLD. To obtain the best estimates, it is vital to choose an appropriate number of bins for the histogram. Three different methods of estimating the number of bins were investigated. In the following discussions of these, $n$ denotes the sample size, and $b_n$

denotes the number of bins for that sample size.

**Sturges breaks** The approach of Sturges [1926] is currently the default method for calculating the number of histogram bins in the `R` Environment for Statistical Computing [R Core Team]. Since Sturges' formula computes bin sizes from the range of the data, it can perform quite poorly when the sample size is small in particular when, $n$ is less than 30. For this method

$$b_n = \lceil \log_2(n+1) \rceil.$$

The brackets represent the ceiling function, by which $\log_2(n+1)$ will be rounded up to the next integer value.

**Scott breaks** Scott's choice [see Scott, 1979] is derived from the normal distribution and relies upon an estimate of the standard error. It offers a more flexible approach than the fixed number of bins used by Sturges. In Scott's method, the number of bins is

$$b_n = \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil,$$

where

$$h = 3.49\hat{\sigma}n^{1/3}.$$

When all data are equal, $h = 0$, and $b_n = 1$.

**Freedman–Diaconis breaks** The Freedman–Diaconis choice [Freedman and Diaconis, 1981] is a robust selection method based upon the full and interquartile ranges of the data. The number of bins is

$$b_n = \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil,$$

where

$$h = \frac{\pi(0.75) - \pi(0.25)}{n^{1/3}}.$$

If the interquartile range is zero, then $h$ is instead set to the median absolute deviation.

The Sturges breaks do not use any information concerning the form of the underlying distribution, and so the Sturges histogram bins are often far from being optimal. The Scott approach is reliant upon the distribution being approximately normal, and so, for GLD applications, the histogram bins are not expected to be optimally chosen over the tails. Since it is a robust method based on the interquartile range, the Freedman–Diaconis

approach seems to be the most promising option for creating histograms from random variates drawn from the heavy-tailed GLD.

## 4.3 Goodness-of-fit approaches

To test the goodness-of-fit of a continuous cumulative distribution function, to a set of data, statistics based on the empirical distribution function can be used. Examples of these are the Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling statistics. These statistics are measures of the difference between the hypothetical GLD distribution and the empirical distribution.

To use these statistics as parameter estimators, they are minimized with respect to the unknown parameters of the distribution. It is known, from the generalized Pareto distribution, that these approaches are able to successfully estimate the parameters even when the maximum likelihood method fails. This is shown by Luceño [2006], who calls this approach the method of maximum goodness-of-fit. Definitions and computational forms of the three goodness-of-fit statistics named above are also given in the paper of Luceño [2006].

The following discussions consider a random sample of $n$ *iid* observations, $\{x_1, x_2, \ldots, x_n\}$, with the order statistics of the sample denoted by $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$, and write $F_i = F(x_{(i)})$ the corresponding probability of the event $x_{(1)}$.

**The Kolmogorov–Smirnov statistic** The Kolmogorov–Smirnov statistic measures the maximum distance between the empirical cumulative distribution function and the theoretical probabilities, $S_n$;

$$D_n = \sup_x |F_x - S_n(x)|,$$

where the sample estimator is

$$\widehat{D_n} = \frac{1}{2n} + \max_{1 \le i \le n} \left| F_i - \frac{1 - i/2}{n} \right|.$$

Minimizing this leads to the optimization of a discontinuous objective function.

**The Cramér–von Mises statistic** The Cramér–von Mises statistic uses mean-squared differences, thereby leading to a continuous objective function for parameter optimization:

$$W_n^2 = n \int_{\infty}^{\infty} [F_x - S_n(x)]^2 dF(x),$$

which reduces to

$$\widehat{W_n^2} = \frac{1}{12n} + \sum_{i=1}^{n} \left( F_i - \frac{1 - i/2}{n} \right)^2.$$

**The Anderson–Darling statistic** The Anderson–Darling statistic is a tail-weighted statistic. It gives more weight to the tails and less weight to the center of the distribution.

This makes the Anderson–Darling statistic an interesting candidate for estimating parameters of the heavy-tailed GLD. It is defined as

$$A_n^2 = n \int_{\infty}^{\infty} \frac{[F_x - S_n(x)]^2}{F(x)(1 - F(x))} dF(x),$$

where its sample estimator is

$$\widehat{A_n^2} = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\ln(F_i) + \ln(1 - F_{n+1-i}).$$

## 4.4 Quantile matching

As described by Su [2010], the quantile matching method consists of finding the parameter values that minimize the difference between the theoretical and sample quantiles. This approach is especially interesting for the GLD, since the distribution is defined by a quantile function.

Consider an indexed set of probabilities, $p$, as a sequence of values in the range $0 < p_i < 1$. The quantile matching estimator is defined For a cumulative probability distribution $F$ with a set of parameters $\theta$, the quantile matching estimator yield the parameter values

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{i}^{n} (F^{-1}(p_i|\theta) - Q(p_i))^2,$$

where $n$ is the cardinality of $p$, $F^{-1}(p|\theta)$ is the quantile function, and $Q$ the sample quantile function.

## 4.5 Trimmed L-moments

A drawback of using moment matching to fit a distribution, is that sample moment estimators are sensitive to outliers. To circumvent this issue, the L-moments have been defined as linear combinations of the ordered data values. The method of L-moments is to find the distribution parameter values that minimize the difference between the sample

L-moments and the theoretical ones. Elamir and Seheult [2003] introduced the trimmed L-moments, a robust extension of the L-moments.

The trimmed L-moments, $l_r(t_1, t_2)$, on the order statistics $X_{1:n}, \ldots, X_{n:n}$ of a random sample $\mathbf{X}$ of size $n$ with trimming parameters $t_1$ and $t_2$, are defined as

$$TL(r, t_1, t_2) = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E\left[r + t_1 - k,\, r + t_1 + t_2\right], \tag{4.2}$$

where

$$E\left[i,\, r\right] = \frac{r!}{(i-1)!(r-i)!} \int_0^1 Q(u) u^{i-1} (1-u)^{r-i}.$$

$Q$ is the quantile function of the distribution. Note that when $t_1 = t_2 = 0$, this reduces to the definition of the standard L-moment. The concept behind the trimmed L-moments is to extend the sample size by $t_1 + t_2$ and trim the $t_1$ smallest and $t_2$ largest order statistics.

Using the unbiased estimator of the expectation of the order statistics, $E$, of Downton [1966], Elamir and Seheult [2003] obtains the sample estimator $\hat{E}$:

$$\hat{E}(i, r) = \frac{1}{\binom{n}{r}} \sum_{t=1}^{n} \binom{t-1}{i-1} \binom{n-t}{r-i} X_{t:n},$$

which can be used to estimate the sample's $r$th trimmed L-moment in Eq. (4.2).

The derivation of the trimmed L-moments for the GLD is cumbersome. Asquith [2007] calculated them for the RS parameterization.

## 4.6 MLE and MPS approaches

The Kullback–Leibler (KL) divergence [Kullback, 1959], also known by the names information divergence, information gain, and relative entropy, measures the difference between two probability distributions: $P$ and $Q$. Typically, $P$ represents the empirical distribution, and $Q$ comes from a theoretical model. The KL divergence allows for the interpretation of many other measures in information theory. Two of them are the maximum log-likelihood estimator and the maximum product of spacing estimator.

**Maximum log-likelihood estimation** The maximum log-likelihood method was introduced by Fisher [1922]. For a more recent review, please refer to Aldrich [1997]. Consider a random sample of $n$ *iid* observations, $\{x_1, x_2, \ldots, x_n\}$ drawn from the probability distribution with parameter set $\theta$. Then the maximum value of the log-likelihood function, $\mathcal{L}$,

defined by

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta),$$

returns optimal estimates for the parameters $\theta$ of the probability density function $f$. The maximum of this expression can be found numerically using a non-linear optimizer that allows for linear constraints.

**Maximum product of spacing estimation** The maximum likelihood method may break down in certain circumstances. An example of this is when the support depends upon the parameters to be estimated [see Cheng and Amin, 1983]. Under such circumstances, the maximum product of spacing estimation (MPS), introduced separately by Cheng and Amin [1983]; Ranneby [1984], may be more successful. This method is derived from the probability integral transform, wherein the random variates of any continuous distribution can be transformed to variables of a uniform distribution. The parameter estimates are the values that make the spacings between the cumulative distribution functions of the random variates equal. This condition is equivalent to maximizing the geometric means of the probability spacings.

Consider the order statistics, $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$, of the observations $\{x_1, x_2, \ldots, x_n\}$, and compute their spacings or gaps at adjacent points,

$$D_i = F\left(x_{(i)}\right) - F\left(x_{(i-1)}\right),$$

where $F\left(x_{(0)}\right) = 0$ and $F\left(x_{(n+1)}\right) = 1$. The estimate of the maximum product of spacing becomes

$$\hat{\theta} = \arg\max_{\theta \in \boldsymbol{\Theta}} \sum_{i=1}^{n+1} \ln D(x_{(i)}).$$

The maximizing solution can be found numerically with a non-linear optimizer.

Note that the concept underlying the MPS method is similar to that of the starship estimation GLD estimation method introduced by King and MacGillivray [1999]. However, the starship method uses the Anderson–Darling test to find the parameters that make the spacings closest to a uniform distribution.

## 4.7 Empirical study

In this section, we compare the aforementioned estimators. In this study we specifically take: the Anderson–Darling statistic for the goodness-of-fit approach, the Friedman–

Diaconis breaks for the histogram approach, the L-moments estimator, the MLE, the MPS estimator, the quantile matching estimator, the robust moment matching and the trimmed L-moments estimator with $t_1 = 1$ and $t_2 = 1$. All methods have been implemented for the GLD in the statistical software environment `R` [R Core Team] and are available within the package **gldist**.

Figure 4.1 shows the fitted log-density of the GLD for the Google equity log-returns with the different methods. Overall, The estimates differ by small amounts in the lower tail. However, the upper tail fit of the histogram approach is substantially different from the other methods.

To compare the estimation methods, we generated one thousand time series of one thousand points each, for each of the set GLD parameter sets: $\{\tilde{\mu} = 0, \tilde{\sigma} = 1, \chi = 0, \xi = 0.35\}$. We perform both approaches, when the location and scale parameters, $(\chi, \xi)$, are estimated by their sample estimators, and when they are included in the estimator. Table 4.1 shows the median and interquartile range of the minimum discrimination information criterion (MDE). The MDE is a divergence between two distributions measure used in information statistics. It is part of the $\phi$-divergence measure, which will be studied in Part III. The MDE between two continuous distributions, $f$ and $g$, is defined as

$$D_\phi(f, g) = \int_{-\infty}^{\infty} \phi\left(\frac{f(x)}{g(x)}\right) g(x)\, dx,$$

where

$$\phi(x) = -\log(x) - x + 1.$$

A divergence measure is used to evaluate how closely the estimator can fit the distribution. This is in contrast to the traditional mean absolute deviation and mean absolute error of the parameter estimates, which focus on how closely the estimator can fit the parameters. The distinction is important because, although the fitted parameters might substantially differ from those used to generate the time series, their combination can produce a fitted distribution that closely resembles the generating distribution. Table 4.1 also shows how many of the estimation routines have successfully converged to a minimum, and the median computation times. From the results, the performance of the estimators in terms of the MDE is not affected by estimating the parameter values of the GLD using the two-step procedure: the location and scale parameter estimated by their sample estimates, and the two shape parameter values fitted by the estimator into consideration. However, the computation times of convergence are much shorter in the former case. This is expected, since the optimization routine only seeks solutions for the shape parameters, leading to a

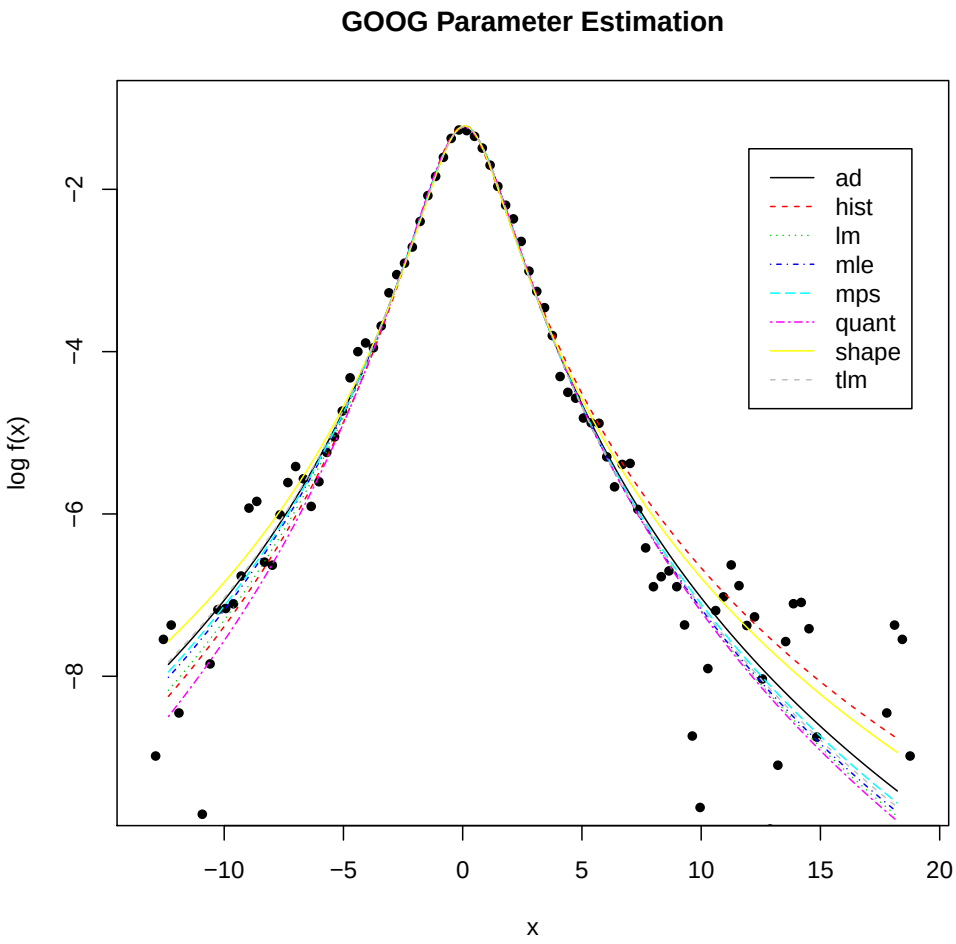


Figure 4.1: Parameter estimation for the Google equity. The full lines are drawn from the fitted distribution function and the points are taken from a kernel density estimate of the time series. The methods compared are: the Anderson–Darling goodness-of-fit method (ad), the histogram binning method (hist), the linear moments method (lm), the maximum likelihood estimator (mle), the maximum product of spacing estimator (mps), the quantile sequence method (quant), the robust moment matching method (shape), and the trimmed linear moments estimator (tlm) with $t_1 = t_2 = 1$.

simpler optimization path. This result can be useful when dealing with a large data set, in which case the computation time is commonly critical.

| | sample $(\tilde{\mu}, \tilde{\sigma})$ | | | optim $(\tilde{\mu}, \tilde{\sigma})$ | | |
|---|---|---|---|---|---|---|
| | MDE | conv. | time [s] | MDE | conv. | time [s] |
| mle | 0.0022 (2.50E-03) | 100% | 0.15 | 0.0019 (1.94E-03) | 100% | 0.43 |
| hist | 0.0046 (5.29E-03) | 100% | 0.01 | 0.0054 (5.85E-03) | 100% | 0.02 |
| prob | 0.0069 (7.49E-03) | 100% | 0.14 | 0.0048 (5.55E-03) | 100% | 0.34 |
| quant | 0.0030 (3.74E-03) | 100% | 0.02 | 0.0021 (2.50E-03) | 100% | 0.03 |
| shape | 0.0090 (1.44E-02) | 100% | 0.01 | 0.0090 (1.44E-02) | 100% | 0.01 |
| ad | 0.0035 (3.84E-03) | 100% | 0.14 | 0.0026 (2.50E-03) | 100% | 0.36 |
| mps | 0.0022 (2.21E-03) | 100% | 0.10 | 0.0018 (1.76E-03) | 100% | 0.19 |

(a) With no outliers

| | sample $(\tilde{\mu}, \tilde{\sigma})$ | | | optim $(\tilde{\mu}, \tilde{\sigma})$ | | |
|---|---|---|---|---|---|---|
| | MDE | conv. | time [s] | MDE | conv. | time [s] |
| mle | 0.0082 (6.64E-03) | 100% | 0.16 | 0.0073 (5.70E-03) | 100% | 0.37 |
| hist | 0.0070 (1.10E-02) | 100% | 0.01 | 0.0071 (6.40E-03) | 100% | 0.02 |
| prob | 0.0069 (7.14E-03) | 100% | 0.14 | 0.0051 (5.27E-03) | 100% | 0.35 |
| quant | 0.0047 (4.77E-03) | 100% | 0.01 | 0.0045 (5.18E-03) | 100% | 0.03 |
| shape | 0.0086 (1.33E-02) | 100% | 0.01 | 0.0086 (1.33E-02) | 100% | 0.01 |
| ad | 0.0055 (4.88E-03) | 96% | 0.20 | 0.0042 (3.20E-03) | 91% | 0.40 |
| mps | 0.0101 (7.25E-03) | 100% | 0.09 | 0.0088 (6.34E-03) | 100% | 0.19 |

(b) With 1% of outliers with scale d = 3

| | sample $(\tilde{\mu}, \tilde{\sigma})$ | | | optim $(\tilde{\mu}, \tilde{\sigma})$ | | |
|---|---|---|---|---|---|---|
| | MDE | conv. | time [s] | MDE | conv. | time [s] |
| mle | 0.1391 (1.41E-02) | 100% | 0.14 | 0.1300 (1.38E-02) | 100% | 0.48 |
| hist | 0.1968 (8.52E-03) | 100% | 0.01 | 0.8170 (2.27E-01) | 100% | 0.04 |
| prob | 0.0585 (1.99E-02) | 100% | 0.14 | 0.0578 (1.92E-02) | 100% | 0.40 |
| quant | 0.1259 (8.93E-03) | 100% | 0.02 | 0.2061 (5.00E-02) | 100% | 0.03 |
| shape | 0.0400 (2.36E-02) | 100% | 0.01 | 0.0400 (2.36E-02) | 100% | 0.01 |
| ad | 0.1174 (1.78E-02) | 100% | 0.12 | 0.1155 (1.77E-02) | 100% | 0.34 |
| mps | 0.1418 (1.39E-02) | 100% | 0.12 | 0.1329 (1.36E-02) | 100% | 0.23 |

(c) With 10% of outliers with scale d = 10

Table 4.1: Simulation study of the estimators. Median and interquartile of the minimum discrimination information criterion (MDE) is reported. The convergence success rates of the optimization routines, as well as the median computation times are reported. The simulated samples are of length 1000. The number of Monte-Carlo replication is 1000. The set of GLD parameters used to simulate the data are, $\{\tilde{\mu} = 0, \tilde{\sigma} = 1, \chi = 0, \xi = 0.35\}$

# 5 Applications

## 5.1 Shape plot representation

A nice consequence of having a closed domain of variation for the shape parameters $\chi$ and $\xi$, is that they can be represented by a shape plot, such as. Figure 5.1 illustrates the different shapes associated with location parameter $\tilde{\mu} = 0$ and scale $\tilde{\sigma} = 1$. The shape plot affords a simple interpretation. The $x$-axis measures the asymmetry, and the $y$-axis expresses the heaviness of the tails. The shape plot is thus ideal for comparing the fitted shape parameters of a data set between different time series. This section illustrates the use of the shape plot with shape parameters fitted to equities from the NASDAQ-100 index.

The NASDAQ-100 index comprises 100 of the largest US domestic and international non-financial companies listed on the NASDAQ stock market. The index reflects the share prices of companies across major industry groups, including computer hardware and software, telecommunications, retail and wholesale trade, and biotechnology. The listed equity returns are expected to exhibit a wide range of distribution shapes. The GLD is therefore a good candidate for modeling their various distributions. The NASDAQ-100 financial index series financial index series used in this study has records from 2000–01–03 to 2011–12–31 [1]. These are listed in Table 5.1. The log-returns of the adjusted closing prices were used. Note that the survivor bias [Elton et al., 1996] has not been considered in this experiment, because the objective is merely to illustrate the use of the GLD with real time series data.

First, the location and scale parameters were estimated from their sample estimators. The maximum likelihood estimator is then used to fit the shape parameters, $\chi$ and $\xi$. Figure 5.2 shows the fitted shape parameters. It is interesting to note that the fitted parameters are to the symmetric vertical line at $\chi = 0$. However, the fitted shape parameters have values that are well above those that best describe the standard normal distribution; these are represented by a triangle in the shape plot. The GLDs fitted to the NASDAQ data have "fatter" tails than does the normal distribution. This is one of the

1. Data downloaded from `finance.yahoo.com`

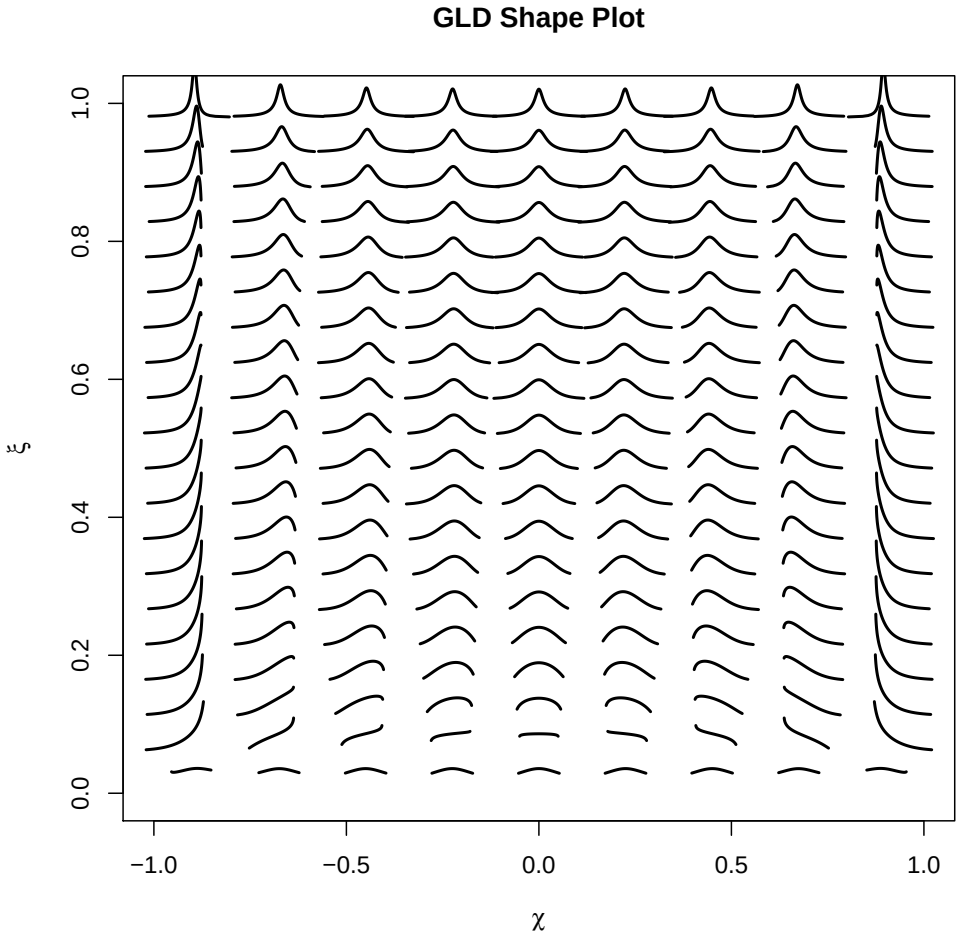


Figure 5.1: This figure illustrates the different probability density shapes for various steepness and asymmetry parameters. The location and scale of the distribution are, respectively, $\tilde{\mu} = 0$ and $\tilde{\sigma} = 1$.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| AKAM | CHRW | DLTR | FISV | KLAC | MSFT | ORCL | RIMM | SYMC | XLNX |
| AMAT | CMCSA | EBAY | FLEX | LIFE | MU | ORLY | ROST | TEVA | XRAY |
| CA | COST | ERTS | FLIR | LLTC | MXIM | PAYX | SBUX | URBN | YHOO |
| CELG | CSCO | ESRX | HSIC | LRCX | MYL | PCAR | SIAL | VOD | |
| CEPH | CTSH | EXPD | INFY | MAT | NTAP | PCLN | SNDK | VRSN | |
| CERN | CTXS | FAST | INTC | MCHP | NVDA | QCOM | SPLS | VRTX | |
| CHKP | DELL | FFIV | INTU | MICC | NWSA | QGEN | SRCL | WFMI | |

Table 5.1: NASDAQ Symbols. The 66 components of the NASDAQ-100 index that have records from 2001–01–03 to 2011–12–31.

so-called stylized facts of financial returns.

This example is important because it illustrates how the GLD can model time series with tails that are fatter than those of the normal distribution. The ability of the GLD to model time series with different types of tails could be used in an assets selection process. For example, assets could be classified according to their kurtosis order, which is along the vertical access of the GLD shape plot.

## 5.2 Quantile based risk measures

Two risk measures commonly used in portfolio optimization are the value-at-risk (VaR) and the expected shortfall risk (ES). On one hand, the $\mathrm{VaR}_\gamma$ is the maximum loss forecast that may happen with probability $\gamma \in [0,1]$ given a holding period. On the other hand, the $\mathrm{ES}_\gamma$ is the averaged $\mathrm{VaR}_\gamma$ in the confidence interval $[0,\gamma]$. These risk measures are thus related to the quantiles of the observed equity price sample distribution. Since the quantile function takes a simple algebraic form, these two risk measures are easily calculated:

$$\mathrm{VaR}_\gamma = Q_{\mathrm{CSW}}(\gamma|\tilde{\mu},\tilde{\sigma},\chi,\xi),$$

and

$$\begin{aligned}
\mathrm{ES}_\alpha &= \frac{1}{\gamma}\int_0^\gamma \mathrm{VaR}_\nu d\nu \\
&= \frac{1}{\gamma}\int_0^\gamma Q_{\mathrm{CSW}}(\nu|\tilde{\mu},\tilde{\sigma},\chi,\xi)d\nu \\
&= \begin{cases}
\gamma(B+\tilde{\mu}+A\ln\gamma)+(A-A\gamma)\ln(1-\gamma) & \text{if } \chi=0,\ \xi=\frac{1}{2}, \\
\gamma\tilde{\mu}+A\gamma(\frac{1}{2\alpha}+\ln\gamma-1)+A\frac{(1-\gamma)^{1+2\alpha}-1}{2\alpha+4\alpha^2}+B\gamma & \text{if } \chi\neq 0,\ \xi=\frac{1}{2}(1+\chi), \\
\frac{A\gamma\left(-1+4\beta^2+\gamma^{2\beta}\right)}{2\beta(1+2\beta)}+\gamma(B+\tilde{\mu})+(A-A\gamma)\ln(1-\gamma) & \text{if } \chi\neq 0,\ \xi=\frac{1}{2}(1-\chi), \\
\dfrac{A\left((1-\gamma)^{1+\alpha}-(1-\gamma)^{\beta}\right)(1-\gamma)^{-\beta}}{(\alpha-\beta)(1+\alpha-\beta)}+B\gamma & \\
+\dfrac{A\gamma}{\alpha-\beta}-\dfrac{A\gamma}{\alpha+\beta}+\dfrac{A\gamma^{1+\alpha+\beta}}{(\alpha+\beta)(1+\alpha+\beta)}+\gamma\tilde{\mu} & \text{otherwise,}
\end{cases}
\end{aligned}$$

where $Q_{\mathrm{CSW}}$ is defined in Eq. (3.7), $\alpha$ and $\beta$ are defined in Eq. (3.5), and

$$A = \tilde{\sigma}\left[S(3/4|\chi,\xi)-S(1/4|\chi,\xi)\right]^{-1},$$

$$B = -\left[S(3/4|\chi,\xi)-S(1/4|\chi,\xi)\right]^{-1}.$$

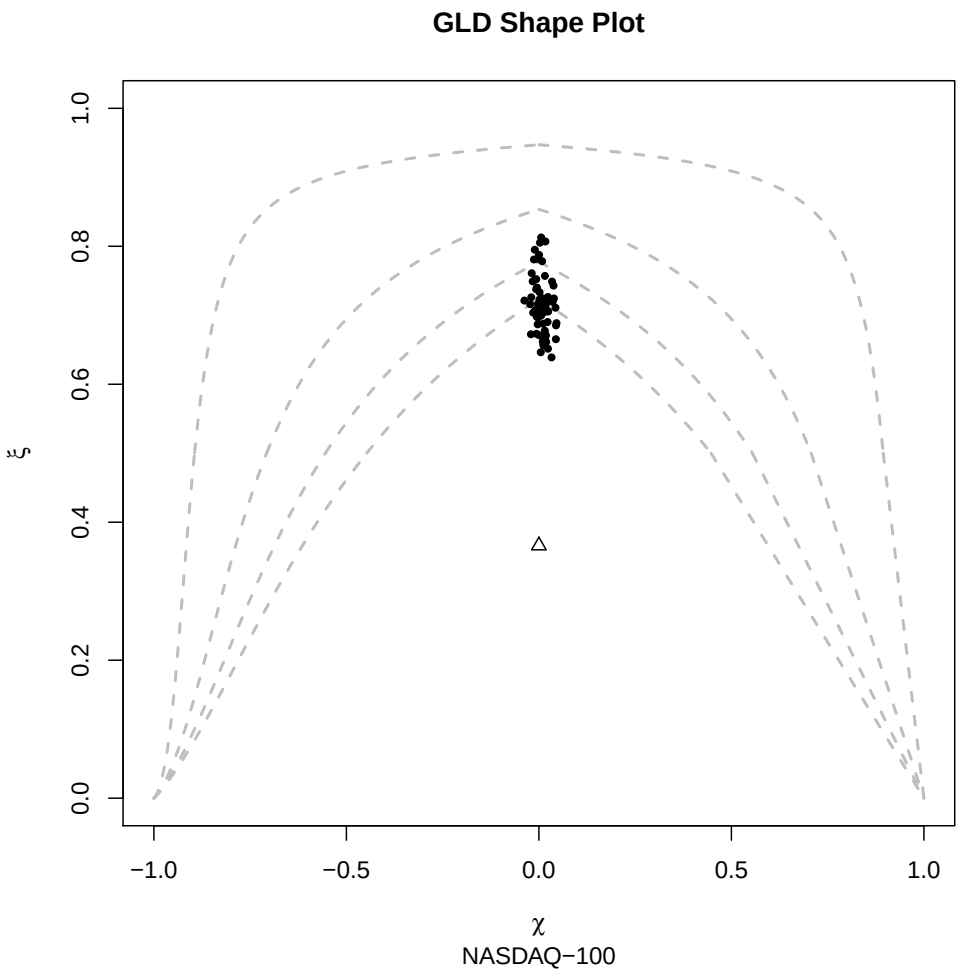


Figure 5.2: Shape plot of the fitted shape parameters of the 66 components of the NASDAQ-100 Index (Table 5.1). The dashed lines represent the existence conditions of the moments of the GLD starting from the existence of the first moment from the top to the existence of higher moments. The triangle symbol corresponds to the GLD shape parameters that best approximate the standard normal distribution.

## 5.3 Apparent scaling law

This section follows the examples of Barndorff-Nielsen and Prause [2001], wherein the normal inverse Gaussian distribution (NIG) has been used to demonstrate the existence of an apparent scaling law for financial returns. By repeating that exercise with the GLD, the ability of the GLD to model both financial returns and the NIG will be demonstrated.

The first so-called scaling law was discovered by Mandelbrot [1963] with cotton prices. Mandelbrot has shown that the averaged daily squared log-returns of cotton prices are self similar. That is, the aggregated returns have the same distribution as the original returns up to a scaling factor. Another well known scaling law, reported by Mueller et al. [1990], is the scaling law of the empirical volatility in high-frequency exchange rate returns. Schnidrig and Würtz [1995] observed the same scaling law in the US dollar and Deutschmark (USD/DEM) exchange rates. Herein, the volatility is defined as the average of absolute logarithm price changes. The volatility over a duration time, $\tau$, is

$$v_\tau = \langle |x_t - x_{t-\tau}| \rangle$$

where $x_t = \ln \frac{P_t}{P_{t-1}}$ are the logarithm returns and $P_t$ are the prices at time $t$. There is a scaling law

$$v_\tau = \left(\frac{\tau}{T}\right)^H v_T,$$

where $T$ is an arbitrary time interval and $H$ is the scaling law index factor. In the last decade, several new scaling laws have been reported. The interested reader may wish to refer to [Glattfelder et al., 2011] for a review of known scaling laws.

Barndorff-Nielsen and Prause [2001] have shown that the apparent scaling of the volatility of high-frequency data is largely due to the semi-heavy-tailedness of financial distributions. The scaling parameter, $H$, can be estimated by a linear regression on the logarithm of the scaling relation in the scaling law.

Figure 5.3 displays the scaling laws of random variates generated by the GLD with location parameter $\tilde{\mu} = 0$, scale parameter $\tilde{\sigma} = 5 \times 10^{-4}$, and shape parameter $\chi = 1$. It is clear that varying the skewness parameter $\xi$ of the GLD changes the scaling constant (i.e., the regression coefficient in the log-log transform). This is a good illustration of how a semi-heavy fat-tailed distribution can reproduce a scaling law, and of how the GLD parameters can be modified to obtain a desired scaling constant.

Another typical effect of the temporal aggregation of financial data is the Gaussianization of the aggregated financial returns. For high-frequency data, the returns are usually modeled by fat-tailed distributions, since the normal distribution has the wrong

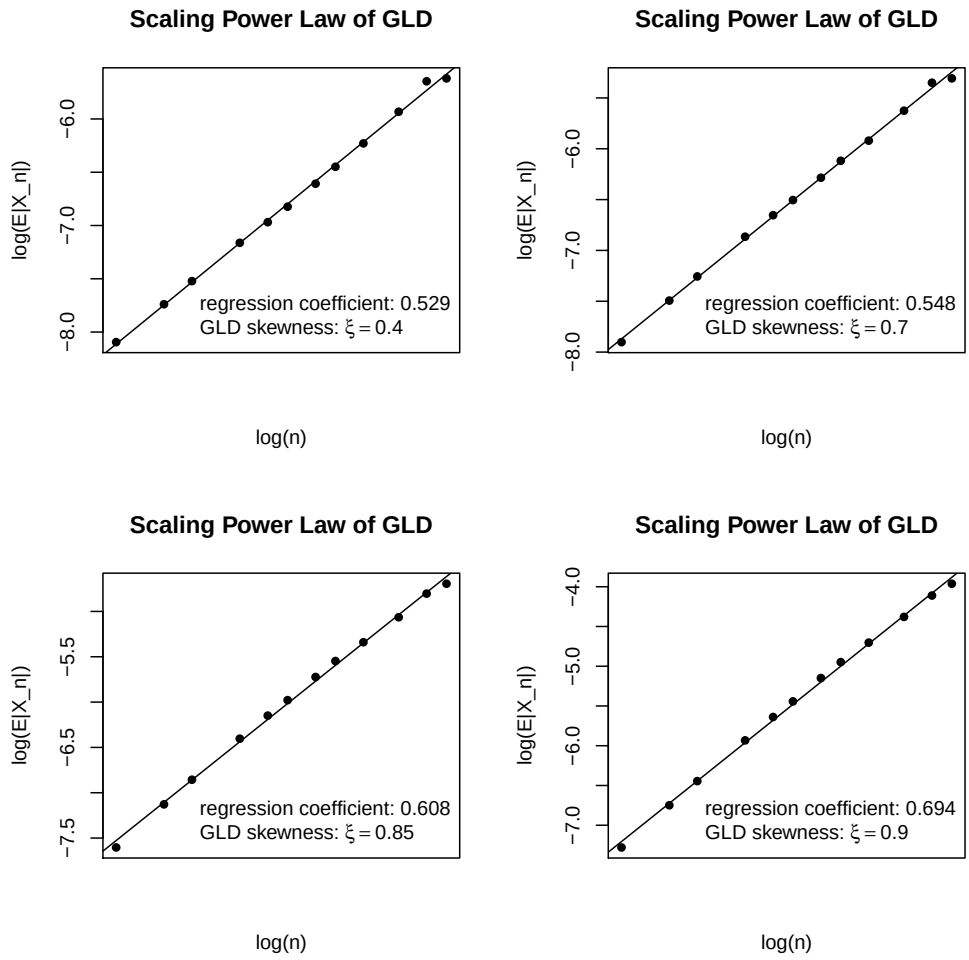


Figure 5.3: Scaling power law of random variates generated by the GLD with location parameter $\tilde{\mu} = 0$, scale parameter $\tilde{\sigma} = 5 \times 10^{-4}$, and shape parameter $\chi = 1$. The remaining shape parameter is reported in the graphics.

shape. However, when the returns are aggregated over long time intervals, the empirical distribution of the aggregated returns is observed to converge onto the normal distribution. Figure 5.4 illustrates this effect with the USD/DEM exchange rate. The data set consists of the hourly returns of USD/DEM from January 1, 1996 at 11 p.m. to December 31, 1996 at 11 p.m. Figure 5.4 shows the fitted GLD shape parameters of the 1, 6, 12, 24, and 60 hour aggregated logarithm returns. The fitted shape parameters appear to be converging onto the normal distribution located at approximately ($\chi = 0$, $\xi = 0.366$) as reported in Table 3.1.

In Fig. 5.5, the same experiment has been performed with US dollar and Swiss Franc (USD/CHF) exchange rate tick data for the years 2004, 2006, 2008 and 2010 [2]. The aggregated returns can be seen to converge in distribution to a normal distribution.

2. The exchange rate tick data was provided by `oanda.com`

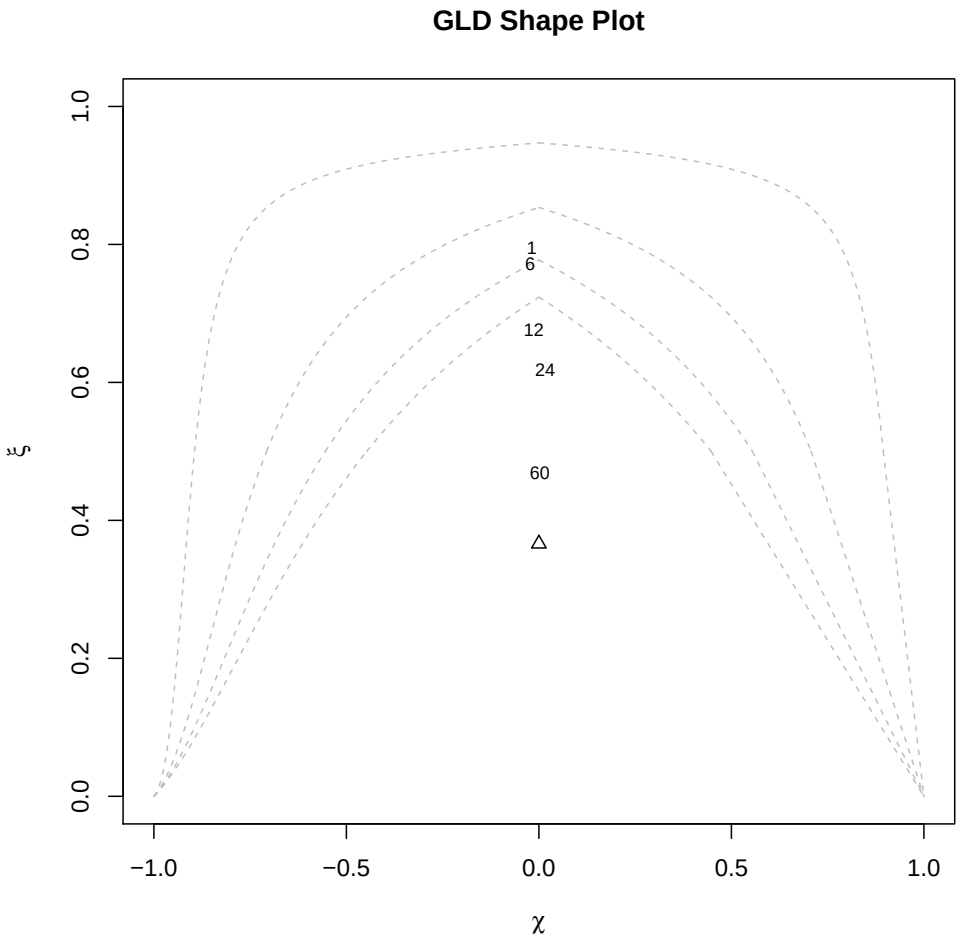


Figure 5.4: Fitted GLD shape parameters of the 1, 6, 12, 24, and 60 hour aggregated logarithm returns of USD/DEM exchange rate. The triangle corresponds to the shape parameters for which the GLD approximate best the normal distribution (see Table 3.1)

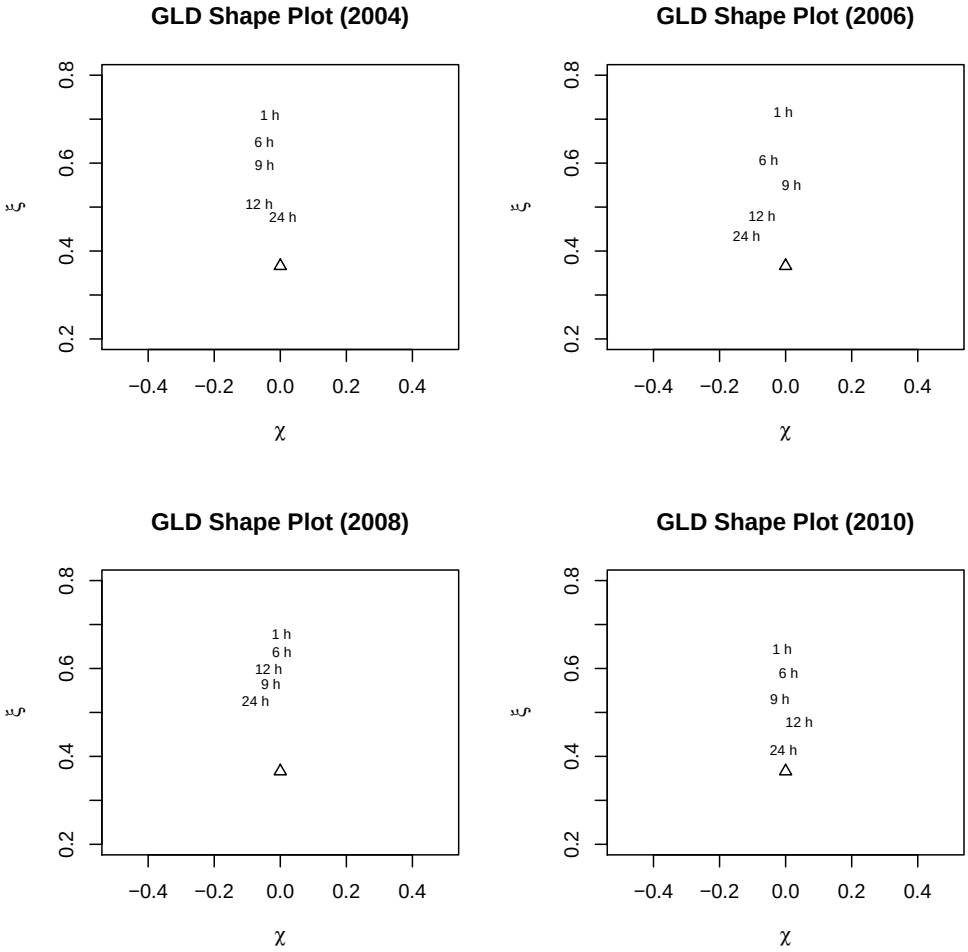


Figure 5.5: Aggregation of the logarithm returns of USD/CHF exchange rate tick data for the years 2004, 2006, 2008 and 2010. The aggregation level used are 1, 6, 12, 24 hour. The quantile matching method was used to estimate the parameter values of the GLD. The triangle corresponds to the shape parameters for which the GLD approximate best the normal distribution (see Table 3.1)

# 6 Conclusion

This first part of the thesis, has introduced a new parameterization of the GLD, which provides an intuitive interpretation of its parameters. The median and interquartile range are respectively synonymous with the location and scale parameters of the distribution. The shape parameters describe the asymmetry and steepness of the distribution. This parameterization differs from previous parameterizations, where the asymmetry and steepness of the GLD is expressed in terms of both tail indices $\lambda_3$ and $\lambda_4$ of Eqs. (2.1) and (2.2). Another advantage of this new parameterization, is that the location and scale parameters can be equated to their sample estimators. This reduces the complexity of the estimator used to fit the remaining two shape parameters. However, this new parameterization comes with the cost of having more intricate expressions for both the conditions of existence for the moments, and the conditions on the shape parameters for the different distribution shapes. Nevertheless, evaluating these expressions remains straightforward. Moreover, the new parameterization enables the use of shape plots that can be used to represent the fitted parameters. Furthermore, value-at-risk, expected shortfall, and the tail indices can be expressed by simple formulae, dependent upon the parameters.

The GLD with unbounded support is an interesting, power-law tailed, alternative to the $\alpha$-stable and the Student $t$-distribution for purposes of modeling financial returns. The key advantage of the GLD is it can accommodate a large range of distribution shapes. A single distribution can therefore be used to model the data for many different asset classes; e.g., bonds, equities, and alternative instruments. This differs from current practice where different distributions must be employed, corresponding to the different distributional shapes of the data sets. Another point, which is often overlooked in practice, is that the same distribution can be applied to different subsets of the data. For example, the distribution of the returns might completely change from year to year. The GLD would then be able to capture this distributional change.

Due to its flexibility, the GLD is a good candidate for studying structural changes of a data set by Bayesian change-point analysis. This may be of use in the identification of market-change regimes.

The R package **gldist** implements the new GLD parameterization, along with the parameter estimation methods that have been presented in this work.

# Part II

# Robust Estimation with the Weighted Trimmed Likelihood Estimator

# 7 Introduction

The first part of this thesis discussed the modeling of financial time series with the flexible generalized lambda distribution. The present part now considers the problem of estimating the parameters of the stochastic models that are widely used in econometrics. One of the most important models is the generalized autoregressive conditional heteroskedasticity model (GARCH), which is used to model the volatility clustering of financial returns. GARCH originates in the ARCH model of Engle [1982], who won the Nobel-memorial prize "for methods of analyzing economic time series with time-varying volatility (ARCH)". The family of GARCH-type models is often used as the foundation of risk measures.

The traditional approach for estimating parameter values of a model such that it is well fitted to a given data set, is to apply the maximum likelihood estimator (MLE). Unfortunately, the MLE can be highly sensitive to any outliers that might be present in the data. This is especially problematic when constructing a risk measure that is reliant upon a distribution model, since the estimated parameters may acquire a bias from the estimator.

The estimation of distributional model parameters in the presence of abnormal points is an active field of research known as robust statistics. The aim of this chapter is to improve the recently introduced weighted trimmed likelihood estimator (WTLE) in order to provide a viable alternative to the maximum likelihood estimator (MLE). To this end, a scheme for automatically computing the parameters of the WTLE estimation is introduced. This auto-WTLE is used to obtain robust estimates of the GARCH parameters. The performance of this new approach is compared to that of other robust GARCH models, in an extensive Monte-Carlo simulation.

The remainder of this part is organized as follows. Chapter 8 recalls the maximum likelihood estimator and its sensitivity to outliers, and then presents the WTLE. Chapter 9 introduces the automatic approach to calculating the weights and the trimming parameter that are required for the WTLE. The Auto-WTLE is applied to the GARCH model in Chapter 10, and its performance is tested against that of the recently introduced robust GARCH models. Concluding remarks, speculation and ideas are presented at the end of this part.

# 8 The weighted trimmed likelihood estimator

The maximum likelihood estimator (MLE), introduced by Fisher [1922], is used to relate parameters of the probability distributions in a statistical model of some data set, to the likelihood of observing the outcomes. Suppose a random sample of $n$ *iid* observations, $\{x_1, x_2, \ldots, x_n\}$, is drawn from an unknown continuous probability density distribution, $f_\theta$, with parameters $\theta$. The estimated parameters, $\hat{\theta}$, are the set of parameters that are most likely, given some assumed model,

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \boldsymbol{\Theta}} \prod_{i=1}^{n} f_\theta(x_i),$$

where $\boldsymbol{\Theta}$ is the set of feasible parameters for $f$.

The above maximization problem is usually transformed to the equivalent log-likelihood, problem which can be expressed in terms of a sum rather than a product. The parameters $\hat{\theta}$ can then be estimated by the maximum log-likelihood function,

$$\mathcal{L}_{\text{MLE}}(\theta) = \sum_{i=1}^{n} \ln f_\theta(x_i). \tag{8.1}$$

It is well known that the MLE is highly sensitive to outliers. This can be understood through a geometrical interpretation. The MLE in Eq. (8.1) is equivalent to

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta \in \boldsymbol{\Theta}} \ \mathcal{L}_{\text{MLE}}(\hat{\theta}) \\
&= \arg\max_{\theta \in \boldsymbol{\Theta}} \ \frac{1}{n} \sum_{i=1}^{n} \ln f_\theta(x_i) \\
&= \arg\max_{\theta \in \boldsymbol{\Theta}} \ \ln \sqrt[n]{f_\theta(x_1)\, f_\theta(x_2)\, \ldots\, f_\theta(x_n)} \\
&= \arg\max_{\theta \in \boldsymbol{\Theta}} \ \sqrt[n]{f_\theta(x_1)\, f_\theta(x_2)\, \ldots\, f_\theta(x_n)}.
\end{aligned}$$

The MLE is therefore equivalent to maximizing the geometric mean of the likely values. This maximum is attained when each of the geometric mean's arguments are of equal size. In other words, the MLE yields estimates for which the likely values are most equal. In the presence of outliers—i.e., in the presence of values for which the likely values in the assumed distributional model are very small—the MLE would yield estimates that reduce the likeliness of the good data points in favor of obtaining more equally likely values across all events. A geometrical interpretation of the objective function can be made with other types of optimization; say, in the maximum product of spacing approach of Section 4.6.

To reduce the impact of outliers on the MLE, Hadi and Luceño [1997] and Vandev and Neykov [1998] introduced the WTLE;

$$\hat{\theta}_{\text{WTLE}} = \arg\min_{\theta\in\Theta} \frac{1}{k} \sum_{i=1}^{k} w_{v(i)}\, g_\theta(x_{v(i)}), \tag{8.2}$$

where $g_\theta(x_{v(1)}) \leqslant g_\theta(x_{v(2)}) \leqslant \cdots \leqslant g_\theta(x_{v(N)})$ are indexed in ascending order for fixed parameters $\theta$ and with permutation index $v(i)$ of $g_\theta(x_i) = -\ln f_\theta(x_i)$, $f_\theta$ is the probability density, and the $w_i$ are weights. The key idea in Eq. (8.2) is to trim the $n-k$ points that are the most unlikely from the estimation of the likelihood function. The WTLE reduces to: (i) the MLE when $k = N$, (ii) the trimmed likelihood estimator when $w_{v(i)} = 1$ for $i \in (1, \ldots, k)$ and $w_{v(i)} = 0$ otherwise, and (iii) the median likelihood estimator, as reported by Vandev and Neykov [1993], when $w_{v(k)} = 1$ and $w_{v(i)} = 0$ for $i \neq k$.

The WTLE is a generalization of the trimmed likelihood estimator (TLE) of Neykov and Neytchev [1990], see also the work of Bednarski and Clarke [1993], and Vandev and Neykov [1993]. The WTLE has been applied to many different fields: Markatou [2000] used the weighted likelihood estimating equations for mixture models, Müller and Neykov [2003] studied related estimators in generalized linear models, and Neykov et al. [2007] employed the WTLE for robust parameter estimation in a finite mixture of distributions.

Bednarski and Clarke [1993] discuss the Fisher consistency, compact differentiability, and asymptotic normality of the TLE. Cizek [2008] explores the consistency and asymptotic properties of the WTLE. Vandev and Neykov [1998]; Müller and Neykov [2003]; Dimova and Neykov [2004] give a derivation of the breakdown point of the WTLE for various models.

The WTLE might can become unfeasible for large data sets, due to its combinatorial nature. Denote by "$k$ sub-sample" the sub-sample of likely values with index $i$ in a sub-set of length $k$ among the full index set $\{1, \ldots, N\}$.

Equation (8.2) then leads to the problem of finding the $k$ sub-sample that minimizes

the estimator. To avoid the optimization of this combinatorial problem, Neykov and Müller [2003] introduced the fast-TLE, which involves repeated iterations of a two-step procedure—a trial step followed by a refinement step. First, a $k$ sub-sample is used to make an initial estimate of the parameters. These estimates are then used to calculate the likelihood values of all points in the data set. Third, the order index of the least likely $N - k$ points is used as a new trimming index. This process is repeated until the convergence criteria are satisfied. Neykov and Müller [2003] showed that the refinement step always yields estimates with an improved or equivalent estimator value.

# 9 The auto-WTLE algorithm

In practice, a fixed value must be chosen for the trimming parameter in Eq. (8.2). If the trimming parameter, $k$, is too small, then the estimator might yield bad estimates due to the sensitivity of the MLE to any outliers that may be present in the data set. If, instead, the chosen trimming parameter is too large, this might result in biased estimates.

This section presents a new method for automatically selecting the trimming parameter, $k$, and the weights, $\omega_i$, of the WTLE (Eq. 8.2). This method is a multi-step iterative procedure. The advantage of the new approach is that it obviates the fine-tuning process necessary to obtain a "robust" parameter in other models.

To start with, an initial set of parameters, $\theta$, is chosen for the distributional model. These initial values could be chosen to be typical values for the given problem, or else could be obtained from another estimator. The probabilities, $F$, of the data points are then calculated in the assumed distributional model with parameters $\theta$. As a consequence of the probability integral transform, if $X$ are random variates with cumulative distribution function $F$, the probabilities $U = F(X)$, are uniformly distributed on $(0, 1)$. The spacings of the probabilities $U$ then give an indication as to which points should be trimmed from, or down-weighted in, the WTLE.

Order statistics and order spacings are the foundations of non-parametric estimation. Goodness-of-fit tests are a good example of such non-parametric estimators. Now, define the spacings, $D_t$, as the differences between the consecutive ordered statistics $U_{(1)}$, $U_{(2)}$, $\ldots$, $U_{(n)}$, where $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ with $U_0 = 0$ and $U_{n+1} = 1$. As noted by Pyke [1965], uniform spacings are interchangeable random variates. The distribution of $D_i$ for any $i$ matches the distribution of the first spacing $D_1$. Moreover, the distribution of the first spacing is, by definition, the same as $F_{U_1}$; the distribution of the first order statistics of the uniform distribution over the interval $(0, 1)$. As described by David and Nagaraja [2003], the cumulative distribution function of the order statistics, $X_{(1)}$, is given by

$$\begin{aligned} F_{X_{(1)}}(x) &= Pr\{X_{(1)} \leq x\} \\ &= 1 - Pr\{X_{(1)} > x\} \end{aligned}$$

$$= 1 - Pr\{\text{all } X_t > x\}$$
$$= 1 - [1 - F_X(x)]^n .$$

where $n$ is the sample size. The distribution of uniform spacings becomes

$$F_{D_t}(x) = F_{D_1}(x) = F_{U_1}(x) = 1 - (1 - x)^n.$$

Hence, the probability distribution function of the uniform spacings is

$$f_{D_t}(x) = n(1 - x)^{n-1}.$$

Given that their theoretical distribution is known, the spacings, $D_i$, can now be filtered by a Monte-Carlo regime-switching model to determine the probabilities of the spacings, $D_t$, to be incorporated within the regime characterized by the density probability $f_{D_t} = n(1 - x)^{n-1}$. In addition to the theoretical regime of the spacings, two alternative regimes are considered. One of these has a density probability that corresponds to a larger sample size, $m_1 = a\,n$ with $a > 1$, while the other corresponds to a smaller sample size, $m_2 = b\,n$ with $0 < b < 1$. This results in a mixture model with three components:

$$f_{D_t}(x) \begin{cases} n(1-x)^{n-1} & \text{if } s_t = 0, \\ m_1(1-x)^{m_1-1} & \text{if } s_t = 1, \\ m_2(1-x)^{m_2-1} & \text{if } s_t = 2, \end{cases}$$

where $s_t$ is the state regime variable of the Markov chain with transition matrix,

$$P = \left[p_{ij}\right],$$

where $p_{ij} = \mathbb{P}(s_t = i \,|\, s_{t-1} = j)$ is the transition probability from state regime $j$ to state regime $i$. Here, a simplified transition matrix

$$\begin{bmatrix} p_{00} & \frac{1}{2}(1-p_{00}) & \frac{1}{2}(1-p_{00}) \\ \frac{1}{2}(1-p_{00}) & p_{00} & \frac{1}{2}(1-p_{00}) \\ \frac{1}{2}(1-p_{00}) & \frac{1}{2}(1-p_{00}) & p_{00} \end{bmatrix}$$

is used, which gives good results in practice as will be seen in the empirical study in Chapter 10.

The probabilities of being in the state $s_t = 0$ at time $t$ are calculated by the scheme

presented in [Kuan, 2002]. Let $\mathcal{D}^t = \{d_1, d_2, \ldots, d_t\}$ denote the collection of sample probability spacings up to index $t$. The conditional density of $d_t$, given information at $t-1$, $\mathcal{D}^{t-1}$, for the simplified model is

$$\begin{aligned} f(d_t \,|\, \mathcal{D}^{t-1}) = n(1-d_t)^{n-1}\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1}) \\ + \frac{1}{2}m_1(1-d_t)^{m_1-1}[1-\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})] \\ + \frac{1}{2}m_2(1-d_t)^{m_2-1}[1-\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})]. \end{aligned}$$

By the Bayes rule, the posterior probability of $s_t$ being in state 0 is given by

$$\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t) = \frac{n(1-d_t)^{n-1}\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})}{f(d_t \,|\, \mathcal{D}^{t-1})},$$

with the prior probability of $s$ at time $t+1$ given information at time $t$, being,

$$\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t) = p_{00}\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t) + \frac{1}{2}(1-p_{00})[1-\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t)].$$

The probabilities of being in state $s_t = 0$ can therefore be obtained by solving the recursive system formed by the previous three equations.

Further, the probabilities $\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)$ can be smoothed to reduce the magnitude of abrupt regime switches. As recommended by Kuan [2002], the method of Kim [1994], which gives the smoothed probabilities, was used. Here, this results in

$$\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^T) = \mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t)\left[\frac{p_{00}\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^T)}{\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)} + \frac{(1-p_{00})\left[1-\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^T)\right]}{1-\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)}\right].$$

The advantage of this simplified Markov chain model, is that there is no need to estimate its parameters by numerical optimization. Instead, only typical values must be provided for the alternative regimes, $m_1$ and $m_2$. The values $m_1 = 10n$ and $m_2 = m/10$, lead to good results in practice. However, the probability of being in state $s_k = 0$ at the starting index $k$, also needs to be specified. In this regards, the spacing at index $\lceil i/n \rceil$ is taken as the starting position because the data point that corresponds to the median of the data set can be expected to be described by the stochastic model under consideration.

The smoothed probabilities of the spacings, $D_t$, in the state assumed by their theoretical distribution are then used as the weights in the WTLE (Eq. 8.2). From this, new set

of parameters, $\theta^+$, are obtained. The procedure is then repeated until the optimized weighted trimmed log-likelihood function reaches a maximum. In practice, it is sufficient to stop the procedure when the objective function has not been improved by more than a factor of 1% from the previous procedure step. This convergence is usually achieved within a few steps.

# 10 Robust GARCH modeling

Generalized autoregressive conditional heteroskedasticity (GARCH) models are widely used to reproduce the stylized facts of financial time series. Today, they play an essential role in risk management and volatility forecasting. It is therefore crucial to develop robust estimators for the GARCH. This section shows how to overcome this limitation by applying the robust weighted trimmed likelihood estimator (WTLE) to the standard GARCH model. The approach is compared with other recently introduced robust GARCH estimators. The results of an extensive simulation study subsequently show that the proposed estimator provides robust and reliable estimates with a small computation cost.

## 10.1 Introduction

Because time-variation of the volatility is a characteristic feature of financial time series, accurate modeling of this variation is critical in many financial applications. It is especially important in risk management. Since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model by Engle [1982], and of its generalization, the GARCH model, by Bollerslev [1986], copious theoretical and applied research work has been performed concerning these models. The success of the GARCH model and its derivations stem mainly from their ability to reproduce the typical properties exhibited by financial time series, particularly, volatility clustering, the fat-tailed return distributions, and the long-term memory effect. Additionally, GARCH processes can be modeled with a wide range of innovation distributions and can be tailored to specific problems. Indeed, Bollerslev [2009] compiled a glossary of more than 150 GARCH models. GARCH modeling is now a common practice, despite the fact that estimation of its parameters involves solving a rather difficult constrained nonlinear optimization problem. Moreover, it is common for different software implementations to produce conflicting estimates [Brooks et al., 2001]. Besides the difficulty in parameter estimation, GARCH models remain, as do any other models, approximations that cannot be expected to encompass all of the complex dynamics of financial markets: Market conditions are strongly affected by factors such as rumor, news, speculation, policy changes, and even data recording errors.

These can result in abnormal points, or outliers, that are beyond the scope of the model. The maximum likelihood estimator (MLE) for GARCH models is very sensitive to these outliers, as was shown by Mendes [2000]; Hotta and Tsay [1998].

Few different methods have been introduced for the robust estimation of GARCH model parameters. Two recent estimators that have been shown to outperform earlier approaches are the recursive robust evaluation of parameters based on outlier criterion statistics [Charles and Darne, 2005], and the robust GARCH model based on a generalized class of M-estimators [Muler and Yohai, 2008]. These methods will be compared with a new estimator introduced here.

The literature usually distinguishes between two families of outliers: additive and innovative. The former are characterized by single abnormal observations, whereas the latter have effects that propagate all along the time series. Here, additive outliers in the conditional volatility of the simple GARCH(1,1) model introduced by Bollerslev [1986], are considered. Note however, that the proposed method can be applied to other GARCH models for which maximum likelihood estimation is possible.

The remainder of this chapter is organized as follows. Section 10.2 first recalls the definition of the GARCH model and its MLE, and then presents the proposed GARCH WTLE. Then, in Section 10.3, the auto-WTLE algorithm presented in Chapter 9 is compared with the recently introduced robust GARCH estimators in an extensive Monte-Carlo simulation.

## 10.2 WTL GARCH(p,q)

For a stationary time series $x_1, x_2, \ldots, x_t, \ldots, x_N$ with mean process $x_t = E(x_t|\Omega_{t-1}) + \varepsilon_t$ and innovation terms $\varepsilon_t$, the GARCH model introduced by Bollerslev [1986] model the innovations as

$$\varepsilon_t = z_t \sigma_t, \tag{10.1a}$$

$$z_t \sim \mathcal{D}_\phi(0,1), \tag{10.1b}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{10.1c}$$

Here, $\Omega_{t-1}$ is the information known at time $t-1$ where $t \in \mathbb{Z}$. $\mathcal{D}_\phi$ is the distribution of the innovations $\mathbf{z}$ with mean zero, variance one, and additional distributional parameters $\phi \in \mathbf{\Phi}^I \subset \mathbb{R}^I$, where $I \in \mathbb{N}$. For example, the additional distributional parameter

of innovations distributed according to Student's $t$-distribution would be the degree of freedom $\nu$. The order of the ARCH and GARCH terms are $q \in \mathbb{N}^*$ and $p \in \mathbb{N}^*$, respectively. Sufficient conditions for the GARCH model to be stationary are $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$, and $\sum_i^q \alpha_i + \sum_j^p \beta_j < 1$. When all $\beta_j = 0$, the GARCH model reduces to the ARCH model of Engle [1982].

Assuming the model in Eq. (10.1), and given an observed univariate financial return series, the MLE can be used to fit the set of parameters $\theta = \{\alpha, \beta, \phi\} \in \mathbf{\Theta}^J \subset \mathbb{R}^J$, where $J = 1+p+q+I$ and $\theta$ includes the parameters of both the GARCH model and innovation distribution. The estimates of the MLE are defined by

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \mathbf{\Theta}^J} \mathcal{L}_{\text{MLE}}(\theta),$$

where the log-likelihood function is

$$\mathcal{L}_{\text{MLE}}(\theta) = \ln \prod_{t=1}^{N} \mathcal{D}_\phi(\varepsilon_t, \sigma_t). \tag{10.2}$$

Equation (10.2) reduces to the so-called quasi-maximum likelihood estimator (QML) when the innovations are assumed to be normally distributed;

$$\mathcal{L}_{\text{QML}}(\theta) = -\frac{1}{2} \sum_{t=1}^{N} \left[ \log(2\pi) + \ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2} \right].$$

The WTLE can be defined for GARCH models by combining Eqs. (8.2), (10.1) and (10.2). The estimates of the WTLE becomes

$$\hat{\theta}_{\text{WTLE}} = \arg\max_{\theta \in \mathbf{\Theta}^J} \frac{1}{k} \sum_{i=1}^{k} w_{v(i)} \ln \mathcal{D}_\phi(\tilde{\varepsilon}_{v(i)}, \tilde{\sigma}_{v(i)}),$$

where

$$\tilde{\sigma}_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \tilde{\varepsilon}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2,$$

and $\mathcal{D}_\phi(\varepsilon_{v(1)}) \geq \mathcal{D}_\phi(\varepsilon_{v(2)}) \geq \cdots \geq \mathcal{D}_\phi(\varepsilon_{v(N)})$ is in descending order with permutation index $v(i)$. However, due to the recursive nature of the GARCH model, care must be taken to ensure that the unlikely innovations are not propagated through the conditional variance. The innovations, $\varepsilon$, are therefore reformulated in terms of their expected values

when they are considered unlikely:

$$\tilde{\varepsilon}_t^2 = \begin{cases} \omega_t \varepsilon_t^2 + (1-\omega_t)\mathbb{E}[\varepsilon_t^2|\Omega_{t-1}] & \text{if } t \leq v(k), \\ \mathbb{E}[\varepsilon_t^2|\Omega_{t-1}] & \text{if } t > v(k). \end{cases}$$

Note that the expected value of the squared innovations at time $t$ given past information $\Omega_{t-1}$ corresponds to the conditional variance at time $t$, $\mathbb{E}[\varepsilon_t^2|\Omega_{t-1}] = \sigma_t^2$, due to the definition of the distribution of the innovations, $\varepsilon_t$, in Eq. (10.1). This gives

$$\tilde{\varepsilon}_t^2 = \begin{cases} \omega_t \varepsilon_t^2 + (1-\omega_t)\tilde{\sigma}_t^2 & \text{if } t \leq v(k), \\ \tilde{\sigma}_t^2 & \text{if } t > v(k). \end{cases}$$

Figures 10.1 and 10.2 illustrates the ability of the auto-WTLE to identify unlikely events for GARCH models. For large outlier scales, the identifier of unlikely points converged to the correct index (Fig. 10.1), whereas for smaller outliers, the approach might consider superfluous points to be unlikely (Fig. 10.2). However, as will be seen in Section 10.3, the impact of mistakenly considering few normal points to be outliers in the GARCH WTLE has negligible impact on the final parameter-value estimates. Figure 10.2 nicely illustrates one of the key advantages of the Auto-WTLE; this approach can identify values that appear too frequently than they ought to for the sample size. Such values are called inner outliers.

## 10.3 Simulation study

All models considered within this study were implemented in the `R` statistical programming language [R Core Team]. The computation times reported offer only an indication of performance and may change with the platform. Regardless, the purpose of this work was not to obtain the most efficient implementation. All models were implemented in `R`, except for the computation of the likelihood function, which was implemented in `C`.

In this section, the GARCH auto-WTLE is compared to the QML, the GARCH M-estimators (M1, M2), with their bounded versions (BM1, BM2) as introduced by Muler and Yohai [2008], and the recursive robust GARCH estimator (REC) of Charles and Darne [2005]. For the M1, M2, BM1, and BM2 estimators, the robust parameters are set to the values recommended by the estimator authors. However, for the REC estimator, stronger threshold statistic ($c = 4$) was used, than those recommended by Charles and Darne [2005]. Indeed, it was noticed that for large outliers, it is crucial to use a low threshold.

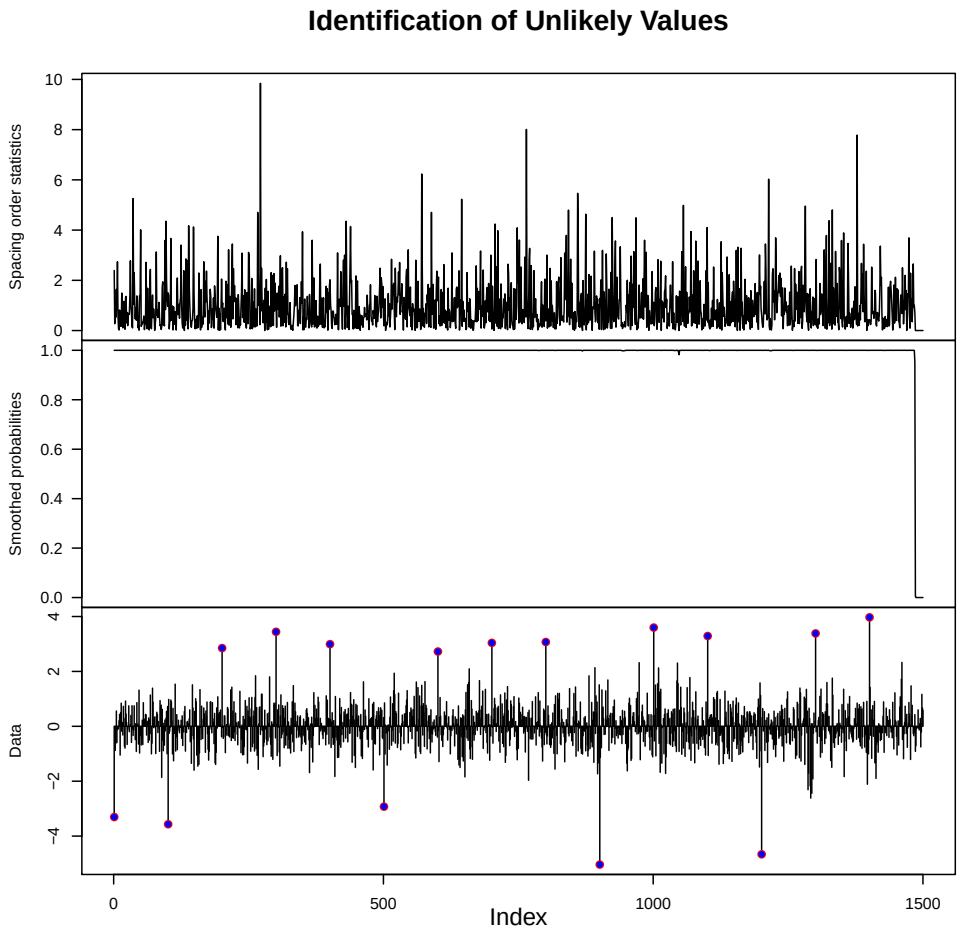


Figure 10.1: Estimation of a contaminated GARCH time series with the auto-WTLE estimator. The GARCH(1,1) series is of length `1500` with parameters $\omega =$ `0.1`, $\alpha =$ `0.2`, and $\beta =$ `0.6` with `15` equidistant outliers of scale d = `5`. The upper figure plots is the spacing order statistics. The middle figure displays the smoothed probabilites after applying the Mone-Carlo switching model filter on the spacings. The lower figure shows the time series. The empty circles are the points for which their likely values have been trimmed in the WTLE and the full cirlces are the exact outliers that were added to the time series.

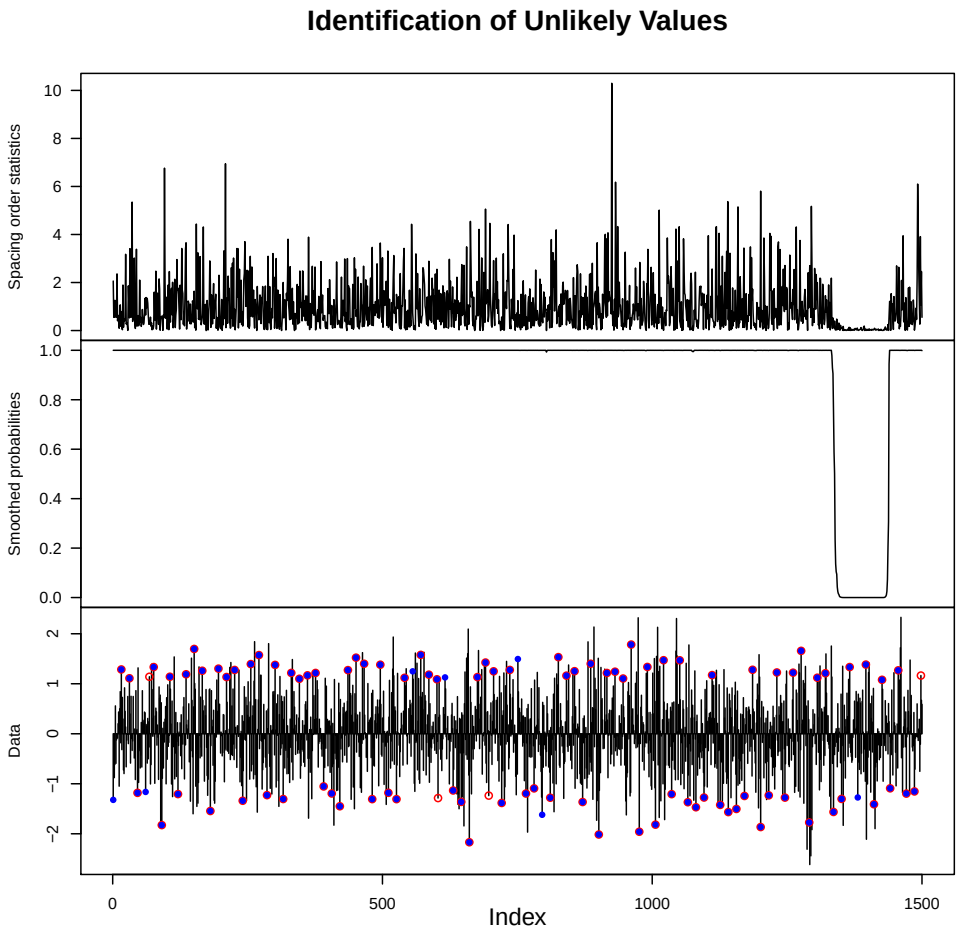


Figure 10.2: Estimation of a contaminated GARCH time series with the auto-WTLE estimator. The GARCH(1,1) series is of length `1500` with parameters $\omega =$ `0.1`, $\alpha =$ `0.2`, and $\beta =$ `0.6` with `15` equidistant outliers of scale d = `5`. The upper figure plots is the spacing order statistics. The middle figure displays the smoothed probabilites after applying the Mone-Carlo switching model filter on the spacings. The lower figure shows the time series. The empty circles are the points for which their likely values have been trimmed in the WTLE and the full cirlces are the exact outliers that were added to the time series.

Otherwise, the unfiltered outliers will lead to poor convergence rates for the optimization routines. The trimming parameter for the GARCH WTLE was automatically defined, as described in Chapter 9.

Since all of the methods compared are based on the MLE, the deviation of the estimates of the robust models with the contaminated series, from the maximum likelihood estimates of the respective uncontaminated series, is reported. The deviations, are defined as $\hat{\theta}^{(y)} - \hat{\theta}^{(x)}$, where $\hat{\theta}^{(y)}$ are the fitted parameters of the contaminated series $y_t$, and $\hat{\theta}^{(x)}$ are the maximum likelihood estimates of the uncontaminated series, $x_t$.

The mean absolute deviation (MAD) of the fitted parameters is defined as

$$\widehat{\text{MAD}} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i^{(y)} - \hat{\theta}_i^{(x)}|,$$

where $N$ is the number of Monte-Carlo runs. Similarly, the mean square deviation (MSD) of the estimates of $y_t$ with the maximum likelihood estimates of the uncontaminated series $x_t$, is expressed as

$$\widehat{\text{MSD}} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\theta}_i^{(y)} - \hat{\theta}_i^{(x)}\right]^2.$$

For all models, the same initial values are used for the conditional variance. Indeed, due to the recursive nature of the GARCH model, initial values must be provided for both $\varepsilon_0^2$ and $\sigma_0^2$. All models use the unconditional variance of the uncontaminated series calculated from the simulation parameters. With normal innovations, the uncontaminated series becomes,

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_i^q \alpha_i - \sum_i^q \beta_i}.$$

Note that the same starting values are used in the routine of each of the different estimators. Starting values that were different from the parameter values used to generate the data set were explicitly chosen in order to assess the ability of the estimator to converge to a solution.

Table 10.1 lists the MADs and MSDs of the models with uncontaminated series. This is in order to show how the estimators are biased from the classical MLE. It is clear that the auto-WTLE deviates very little from the MLE estimates and has the smallest MSD compared to the other models. The auto-WTLE thus has almost no bias to the MLE with uncontaminated series. The table also shows the convergence success rate of the optimization routine. The convergence success rate corresponds to the percentage of Monte-Carlo iterations for which the optimization routine did converge to a solution,

| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. |
|---|---|---|---|---|
| REC | 0.018 (5.10E-04) | 0.244 (6.09E-02) | 0.145 (2.40E-02) | 100% |
| M1 | 0.011 (2.32E-04) | 0.034 (2.13E-03) | 0.026 (1.26E-03) | 100% |
| BM1 | 0.017 (5.62E-04) | 0.054 (4.50E-03) | 0.042 (3.10E-03) | 93% |
| M2 | 0.020 (7.78E-04) | 0.058 (5.40E-03) | 0.045 (3.65E-03) | 99% |
| BM2 | 0.017 (5.45E-04) | 0.116 (1.74E-02) | 0.046 (3.58E-03) | 86% |
| WTL | 0.000 (1.45E-07) | 0.000 (1.48E-06) | 0.000 (9.49E-07) | 100% |

(a) $\omega = 0.1$, $\alpha = 0.5$ and $\beta = 0.4$

| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. |
|---|---|---|---|---|
| REC | 0.047 (6.65E-03) | 0.049 (2.84E-03) | 0.075 (1.29E-02) | 99% |
| M1 | 0.026 (1.64E-03) | 0.014 (3.47E-04) | 0.035 (2.91E-03) | 87% |
| BM1 | 0.044 (5.24E-03) | 0.026 (1.23E-03) | 0.059 (8.71E-03) | 95% |
| M2 | 0.045 (5.62E-03) | 0.024 (9.82E-04) | 0.060 (8.92E-03) | 91% |
| BM2 | 0.044 (5.44E-03) | 0.046 (3.22E-03) | 0.063 (9.46E-03) | 90% |
| WTL | 0.000 (6.10E-07) | 0.000 (1.40E-07) | 0.000 (1.03E-06) | 100% |

(b) $\omega = 0.1$, $\alpha = 0.1$ and $\beta = 0.8$

Table 10.1: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) to assess the estimators' bias to the MLE. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

according to the convergence criterion of the optimization routine. Note these values are specific to the optimization algorithm used (`nlminb()` function in `R`) and might be different for other algorithms.

To make a second comparison, some 1000 sequences were generated, each of 1500 GARCH(1,1) simulated sample variates as described in Eq. (10.1), with 1%, 5%, and 10% outliers. The contaminated time series, $y_t$, were constructed from the uncontaminated series, $x_t$, as $y_t = x_t$ for $t \neq i$ plus outliers $y_i = d\,\sigma_i$ at time index $i$ and scale $d$. A range of outlier scales, $d \in \{2, 4, 6, 10\}$, was used in order to study how the methods perform with slightly and greatly abnormal points. The outliers were taken from the truncated Poisson distribution with a truncation of 10.

Two sets of parameters were used for the GARCH(1,1) model and the starting values in the optimization routines were explicitly set to be different from the optimal values. Tables 10.2 and 10.3 displays results from the simulation using GARCH(1,1) with parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. Although these values are not typically encountered when dealing with financial returns (i.e., a large $\beta$ and small $\alpha$ with a persistence, $\alpha + \beta$, close to 1), they were chosen in order to compare the results with those of Muler and Yohai [2008]. For Tables 10.4 and 10.5, more realistic parameter values of $\omega = 0.1$, $\alpha = 0.1$,

and $\beta = 0.8$, were used. As noted in Charles and Darne [2005], the bounded M-estimators (BM1 and BM2) have a smaller bias than the unbounded M-estimators (M1 and M2), but are subject to a lower convergence success rate. Moreover, M2 and BM2 produce better estimates than do their less-robust counterparts, M1 and BM1. Although the estimates of the REC estimator are similar to those of the other estimators, the REC computation time was much larger than those of the others. Overall, the WTLE yields estimates with the smallest MSE and RMSE, and yet it incurs only a small computational cost.

Besides the MADs and MSDs shown in the tables, also consider the empirical distributions of the deviation of the fitted parameters of the robust models from the fitted parameters obtained with the MLE of the uncontaminated series. Figures 10.3 and 10.4 display results for the REC, BM2 and auto-WTL estimators. Only these three estimators are compared because the parameter distributions of the M2 model are too wide to be included within the same graphics. It is clear that the deviations of the auto-WTLE estimates have the smallest variance, and are closer to zero for the GARCH parameter values.

| | 1% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.005 (3.44E-05) | 0.009 (1.92E-04) | 0.009 (1.45E-04) | 100% | 0.01 |
| REC | 0.017 (4.88E-04) | 0.249 (6.36E-02) | 0.148 (2.47E-02) | 100% | 2.19 |
| M1 | 0.012 (2.71E-04) | 0.036 (2.30E-03) | 0.027 (1.26E-03) | 100% | 0.04 |
| BM1 | 0.020 (7.04E-04) | 0.058 (5.28E-03) | 0.046 (3.44E-03) | 92% | 0.06 |
| M2 | 0.023 (9.18E-04) | 0.063 (6.04E-03) | 0.047 (3.95E-03) | 99% | 0.04 |
| BM2 | 0.020 (6.51E-04) | 0.115 (1.72E-02) | 0.047 (3.50E-03) | 87% | 0.07 |
| WTL | 0.005 (3.37E-05) | 0.009 (1.25E-04) | 0.009 (1.33E-04) | 100% | 0.02 |

| | 1% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.030 (1.12E-03) | 0.056 (5.63E-03) | 0.039 (2.28E-03) | 100% | 0.01 |
| REC | 0.017 (5.45E-04) | 0.259 (6.84E-02) | 0.150 (2.55E-02) | 100% | 2.45 |
| M1 | 0.017 (5.12E-04) | 0.038 (2.38E-03) | 0.040 (2.67E-03) | 100% | 0.04 |
| BM1 | 0.019 (6.34E-04) | 0.050 (3.90E-03) | 0.046 (3.51E-03) | 91% | 0.06 |
| M2 | 0.026 (1.21E-03) | 0.060 (5.63E-03) | 0.061 (6.48E-03) | 99% | 0.05 |
| BM2 | 0.019 (6.22E-04) | 0.099 (1.33E-02) | 0.048 (3.84E-03) | 87% | 0.07 |
| WTL | 0.006 (1.25E-04) | 0.012 (2.81E-04) | 0.011 (2.69E-04) | 100% | 0.03 |

| | 5% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.026 (7.54E-04) | 0.022 (1.50E-03) | 0.022 (9.53E-04) | 100% | 0.01 |
| REC | 0.025 (1.05E-03) | 0.268 (7.35E-02) | 0.136 (2.20E-02) | 100% | 2.20 |
| M1 | 0.034 (1.51E-03) | 0.044 (3.62E-03) | 0.035 (2.63E-03) | 100% | 0.04 |
| BM1 | 0.050 (3.35E-03) | 0.080 (9.73E-03) | 0.051 (5.29E-03) | 93% | 0.05 |
| M2 | 0.049 (3.42E-03) | 0.086 (1.17E-02) | 0.053 (5.58E-03) | 98% | 0.04 |
| BM2 | 0.048 (3.18E-03) | 0.114 (1.75E-02) | 0.056 (5.24E-03) | 87% | 0.06 |
| WTL | 0.022 (5.99E-04) | 0.019 (6.31E-04) | 0.020 (6.62E-04) | 100% | 0.04 |

| | 5% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.240 (5.98E-02) | 0.191 (4.91E-02) | 0.151 (2.75E-02) | 92% | 0.01 |
| REC | 0.049 (5.00E-03) | 0.321 (1.05E-01) | 0.148 (2.78E-02) | 100% | 3.18 |
| M1 | 0.063 (5.85E-03) | 0.158 (2.98E-02) | 0.109 (1.89E-02) | 100% | 0.05 |
| BM1 | 0.023 (9.20E-04) | 0.104 (1.41E-02) | 0.059 (5.94E-03) | 88% | 0.06 |
| M2 | 0.074 (8.15E-03) | 0.174 (4.07E-02) | 0.152 (3.44E-02) | 99% | 0.06 |
| BM2 | 0.020 (6.87E-04) | 0.060 (5.58E-03) | 0.055 (4.84E-03) | 84% | 0.07 |
| WTL | 0.007 (9.33E-05) | 0.015 (4.48E-04) | 0.017 (5.52E-04) | 100% | 0.02 |

| | 10% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.046 (2.36E-03) | 0.029 (2.68E-03) | 0.033 (2.22E-03) | 99% | 0.01 |
| REC | 0.048 (3.31E-03) | 0.285 (8.34E-02) | 0.124 (2.00E-02) | 100% | 2.22 |
| M1 | 0.061 (4.60E-03) | 0.058 (7.78E-03) | 0.046 (4.65E-03) | 100% | 0.04 |
| BM1 | 0.081 (8.00E-03) | 0.103 (1.55E-02) | 0.060 (7.54E-03) | 95% | 0.06 |
| M2 | 0.103 (1.32E-02) | 0.129 (2.39E-02) | 0.061 (7.67E-03) | 99% | 0.04 |
| BM2 | 0.081 (7.93E-03) | 0.121 (2.07E-02) | 0.064 (7.59E-03) | 85% | 0.07 |
| WTL | 0.041 (2.03E-03) | 0.026 (1.32E-03) | 0.030 (1.67E-03) | 100% | 0.03 |

| | 10% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.470 (2.26E-01) | 0.224 (6.50E-02) | 0.224 (5.62E-02) | 82% | 0.01 |
| REC | 0.116 (2.45E-02) | 0.372 (1.40E-01) | 0.170 (4.30E-02) | 99% | 3.46 |
| M1 | 0.079 (1.35E-02) | 0.288 (9.16E-02) | 0.152 (3.38E-02) | 99% | 0.06 |
| BM1 | 0.022 (8.85E-04) | 0.151 (2.65E-02) | 0.065 (6.72E-03) | 90% | 0.06 |
| M2 | 0.073 (9.57E-03) | 0.307 (1.09E-01) | 0.199 (5.26E-02) | 96% | 0.07 |
| BM2 | 0.021 (7.57E-04) | 0.055 (4.64E-03) | 0.054 (5.02E-03) | 83% | 0.07 |
| WTL | 0.007 (1.33E-04) | 0.018 (5.76E-04) | 0.020 (7.39E-04) | 100% | 0.02 |

Table 10.2: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{2, 4\}$ and parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

| | 1% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.076 (7.07E-03) | 0.154 (3.48E-02) | 0.082 (9.72E-03) | 98% | 0.01 |
| REC | 0.018 (5.21E-04) | 0.251 (6.43E-02) | 0.150 (2.52E-02) | 100% | 2.51 |
| M1 | 0.031 (1.39E-03) | 0.038 (2.32E-03) | 0.071 (7.31E-03) | 100% | 0.04 |
| BM1 | 0.019 (6.34E-04) | 0.050 (3.90E-03) | 0.046 (3.51E-03) | 91% | 0.06 |
| M2 | 0.039 (2.49E-03) | 0.063 (6.29E-03) | 0.094 (1.39E-02) | 100% | 0.05 |
| BM2 | 0.019 (6.21E-04) | 0.099 (1.33E-02) | 0.048 (3.84E-03) | 87% | 0.07 |
| WTL | 0.002 (9.09E-06) | 0.006 (6.79E-05) | 0.006 (5.87E-05) | 100% | 0.02 |

| | 1% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.251 (7.95E-02) | 0.272 (9.54E-02) | 0.138 (2.88E-02) | 51% | 0.01 |
| REC | 0.018 (5.47E-04) | 0.251 (6.44E-02) | 0.151 (2.57E-02) | 100% | 2.52 |
| M1 | 0.062 (4.96E-03) | 0.049 (3.72E-03) | 0.146 (2.54E-02) | 100% | 0.04 |
| BM1 | 0.019 (6.34E-04) | 0.050 (3.90E-03) | 0.046 (3.51E-03) | 91% | 0.06 |
| M2 | 0.069 (6.87E-03) | 0.077 (8.88E-03) | 0.170 (3.61E-02) | 99% | 0.05 |
| BM2 | 0.019 (6.21E-04) | 0.099 (1.33E-02) | 0.048 (3.84E-03) | 87% | 0.07 |
| WTL | 0.002 (8.93E-06) | 0.006 (6.77E-05) | 0.006 (5.82E-05) | 100% | 0.02 |

| | 5% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.710 (5.25E-01) | 0.334 (1.29E-01) | 0.254 (7.39E-02) | 29% | 0.01 |
| REC | 0.019 (6.49E-04) | 0.290 (8.59E-02) | 0.175 (3.47E-02) | 100% | 3.51 |
| M1 | 0.150 (2.64E-02) | 0.137 (2.88E-02) | 0.270 (8.48E-02) | 96% | 0.05 |
| BM1 | 0.023 (9.19E-04) | 0.104 (1.41E-02) | 0.059 (5.94E-03) | 88% | 0.06 |
| M2 | 0.155 (2.90E-02) | 0.142 (3.52E-02) | 0.310 (1.06E-01) | 90% | 0.06 |
| BM2 | 0.020 (6.87E-04) | 0.060 (5.58E-03) | 0.055 (4.85E-03) | 84% | 0.07 |
| WTL | 0.004 (3.39E-05) | 0.012 (2.88E-04) | 0.013 (2.85E-04) | 100% | 0.02 |

| | 5% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 1.695 (4.41E+00) | 0.395 (1.63E-01) | 0.346 (1.47E-01) | 1% | 0.02 |
| REC | 0.023 (9.28E-04) | 0.297 (9.07E-02) | 0.183 (3.83E-02) | 100% | 3.52 |
| M1 | 0.210 (4.95E-02) | 0.091 (1.71E-02) | 0.386 (1.52E-01) | 83% | 0.05 |
| BM1 | 0.023 (9.19E-04) | 0.104 (1.41E-02) | 0.059 (5.94E-03) | 88% | 0.06 |
| M2 | 0.191 (4.37E-02) | 0.111 (2.16E-02) | 0.386 (1.52E-01) | 74% | 0.06 |
| BM2 | 0.020 (6.87E-04) | 0.060 (5.58E-03) | 0.055 (4.85E-03) | 84% | 0.07 |
| WTL | 0.004 (3.32E-05) | 0.012 (2.61E-04) | 0.012 (2.52E-04) | 100% | 0.02 |

| | 10% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 1.342 (1.83E+00) | 0.335 (1.30E-01) | 0.314 (1.09E-01) | 14% | 0.01 |
| REC | 0.026 (1.19E-03) | 0.328 (1.10E-01) | 0.209 (4.95E-02) | 100% | 4.19 |
| M1 | 0.125 (2.46E-02) | 0.313 (1.19E-01) | 0.302 (1.01E-01) | 94% | 0.07 |
| BM1 | 0.022 (8.83E-04) | 0.151 (2.65E-02) | 0.065 (6.71E-03) | 90% | 0.07 |
| M2 | 0.125 (2.52E-02) | 0.294 (1.16E-01) | 0.311 (1.08E-01) | 87% | 0.08 |
| BM2 | 0.020 (7.51E-04) | 0.055 (4.62E-03) | 0.054 (4.98E-03) | 83% | 0.07 |
| WTL | 0.005 (5.25E-05) | 0.015 (3.59E-04) | 0.016 (3.99E-04) | 100% | 0.02 |

| | 10% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.579 (8.78E-01) | 0.474 (2.28E-01) | 0.503 (2.59E-01) | 11% | 0.03 |
| REC | 0.059 (1.34E-02) | 0.338 (1.18E-01) | 0.208 (5.29E-02) | 98% | 4.46 |
| M1 | 0.158 (3.66E-02) | 0.263 (1.03E-01) | 0.362 (1.39E-01) | 84% | 0.08 |
| BM1 | 0.022 (8.83E-04) | 0.151 (2.65E-02) | 0.065 (6.71E-03) | 90% | 0.06 |
| M2 | 0.198 (6.53E-02) | 0.159 (4.92E-02) | 0.366 (1.42E-01) | 79% | 0.07 |
| BM2 | 0.020 (7.51E-04) | 0.055 (4.62E-03) | 0.054 (4.98E-03) | 83% | 0.07 |
| WTL | 0.005 (5.08E-05) | 0.015 (3.53E-04) | 0.016 (3.89E-04) | 100% | 0.02 |

Table 10.3: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{6, 10\}$ and parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

| | 1% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.009 (1.44E-04) | 0.005 (4.20E-05) | 0.012 (2.42E-04) | 100% | 0.02 |
| REC | 0.051 (8.53E-03) | 0.051 (3.01E-03) | 0.077 (1.43E-02) | 98% | 1.29 |
| M1 | 0.026 (1.51E-03) | 0.014 (3.42E-04) | 0.034 (2.37E-03) | 90% | 0.05 |
| BM1 | 0.051 (7.71E-03) | 0.029 (1.46E-03) | 0.066 (1.08E-02) | 92% | 0.07 |
| M2 | 0.055 (8.83E-03) | 0.027 (1.24E-03) | 0.068 (1.17E-02) | 88% | 0.06 |
| BM2 | 0.053 (8.82E-03) | 0.046 (3.45E-03) | 0.071 (1.30E-02) | 91% | 0.07 |
| WTL | 0.008 (1.43E-04) | 0.005 (4.19E-05) | 0.012 (2.42E-04) | 100% | 0.02 |

| | 1% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.078 (1.24E-02) | 0.022 (7.67E-04) | 0.063 (9.87E-03) | 100% | 0.01 |
| REC | 0.047 (6.39E-03) | 0.052 (3.16E-03) | 0.076 (1.25E-02) | 99% | 1.69 |
| M1 | 0.049 (6.32E-03) | 0.029 (1.16E-03) | 0.061 (8.39E-03) | 90% | 0.05 |
| BM1 | 0.049 (7.55E-03) | 0.025 (1.13E-03) | 0.066 (1.18E-02) | 93% | 0.06 |
| M2 | 0.071 (1.44E-02) | 0.035 (1.73E-03) | 0.091 (2.00E-02) | 89% | 0.06 |
| BM2 | 0.050 (8.01E-03) | 0.044 (3.14E-03) | 0.071 (1.34E-02) | 91% | 0.07 |
| WTL | 0.008 (4.17E-04) | 0.005 (7.98E-05) | 0.012 (7.47E-04) | 100% | 0.02 |

| | 5% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.032 (1.97E-03) | 0.015 (3.01E-04) | 0.028 (1.44E-03) | 100% | 0.02 |
| REC | 0.111 (3.30E-02) | 0.059 (4.01E-03) | 0.125 (3.35E-02) | 96% | 0.94 |
| M1 | 0.053 (5.94E-03) | 0.017 (4.88E-04) | 0.045 (4.28E-03) | 94% | 0.05 |
| BM1 | 0.104 (3.07E-02) | 0.027 (1.27E-03) | 0.087 (2.08E-02) | 94% | 0.07 |
| M2 | 0.104 (2.92E-02) | 0.025 (1.23E-03) | 0.083 (1.92E-02) | 88% | 0.06 |
| BM2 | 0.105 (3.18E-02) | 0.034 (2.10E-03) | 0.093 (2.30E-02) | 92% | 0.07 |
| WTL | 0.029 (1.79E-03) | 0.013 (2.42E-04) | 0.027 (1.37E-03) | 100% | 0.05 |

| | 5% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.427 (4.21E-01) | 0.085 (8.18E-03) | 0.273 (1.12E-01) | 26% | 0.02 |
| REC | 0.053 (7.20E-03) | 0.054 (3.41E-03) | 0.085 (1.49E-02) | 99% | 2.95 |
| M1 | 0.064 (1.03E-02) | 0.072 (5.68E-03) | 0.105 (1.73E-02) | 90% | 0.07 |
| BM1 | 0.051 (7.45E-03) | 0.026 (1.09E-03) | 0.068 (1.19E-02) | 94% | 0.07 |
| M2 | 0.087 (2.09E-02) | 0.071 (5.60E-03) | 0.123 (2.81E-02) | 89% | 0.08 |
| BM2 | 0.050 (7.46E-03) | 0.036 (2.23E-03) | 0.070 (1.30E-02) | 93% | 0.07 |
| WTL | 0.017 (1.78E-03) | 0.011 (3.10E-04) | 0.026 (3.07E-03) | 100% | 0.02 |

| | 10% with d=2.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.031 (3.62E-03) | 0.030 (1.08E-03) | 0.051 (4.45E-03) | 100% | 0.02 |
| REC | 0.117 (4.62E-02) | 0.071 (5.60E-03) | 0.154 (4.14E-02) | 91% | 0.82 |
| M1 | 0.061 (1.44E-02) | 0.030 (1.21E-03) | 0.071 (1.08E-02) | 84% | 0.06 |
| BM1 | 0.089 (3.28E-02) | 0.034 (1.69E-03) | 0.098 (1.94E-02) | 95% | 0.08 |
| M2 | 0.113 (3.84E-02) | 0.041 (2.79E-03) | 0.093 (1.92E-02) | 82% | 0.06 |
| BM2 | 0.107 (4.73E-02) | 0.035 (1.89E-03) | 0.115 (2.72E-02) | 91% | 0.09 |
| WTL | 0.033 (4.78E-03) | 0.026 (8.39E-04) | 0.048 (4.82E-03) | 100% | 0.04 |

| | 10% with d=4.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.037 (4.67E-03) | 0.105 (1.15E-02) | 0.158 (2.73E-02) | 11% | 0.03 |
| REC | 0.059 (1.02E-02) | 0.050 (3.10E-03) | 0.093 (2.07E-02) | 99% | 3.74 |
| M1 | 0.054 (6.13E-03) | 0.076 (6.25E-03) | 0.096 (1.36E-02) | 94% | 0.07 |
| BM1 | 0.042 (4.06E-03) | 0.026 (1.11E-03) | 0.057 (7.22E-03) | 94% | 0.07 |
| M2 | 0.068 (1.16E-02) | 0.075 (6.20E-03) | 0.104 (1.87E-02) | 89% | 0.08 |
| BM2 | 0.042 (4.65E-03) | 0.034 (1.91E-03) | 0.062 (9.08E-03) | 91% | 0.07 |
| WTL | 0.020 (1.46E-03) | 0.014 (3.87E-04) | 0.032 (2.83E-03) | 100% | 0.02 |

Table 10.4: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{2, 4\}$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

| | 1% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.332 (1.66E-01) | 0.057 (5.15E-03) | 0.224 (8.99E-02) | 87% | 0.02 |
| REC | 0.049 (6.94E-03) | 0.052 (3.11E-03) | 0.077 (1.34E-02) | 99% | 1.67 |
| M1 | 0.114 (3.15E-02) | 0.052 (3.23E-03) | 0.128 (3.54E-02) | 88% | 0.06 |
| BM1 | 0.049 (7.55E-03) | 0.025 (1.13E-03) | 0.066 (1.18E-02) | 93% | 0.06 |
| M2 | 0.133 (4.16E-02) | 0.057 (4.04E-03) | 0.154 (5.01E-02) | 84% | 0.08 |
| BM2 | 0.050 (8.01E-03) | 0.044 (3.14E-03) | 0.071 (1.34E-02) | 91% | 0.07 |
| WTL | 0.004 (6.18E-05) | 0.003 (1.37E-05) | 0.006 (1.09E-04) | 100% | 0.02 |

| | 1% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.834 (8.11E-01) | 0.292 (1.21E-01) | 0.466 (2.60E-01) | 16% | 0.01 |
| REC | 0.053 (8.83E-03) | 0.052 (3.18E-03) | 0.082 (1.63E-02) | 99% | 1.65 |
| M1 | 0.198 (8.19E-02) | 0.082 (7.55E-03) | 0.211 (8.45E-02) | 69% | 0.08 |
| BM1 | 0.049 (7.55E-03) | 0.025 (1.13E-03) | 0.066 (1.18E-02) | 93% | 0.07 |
| M2 | 0.225 (9.67E-02) | 0.087 (8.36E-03) | 0.233 (9.83E-02) | 63% | 0.10 |
| BM2 | 0.050 (8.01E-03) | 0.044 (3.14E-03) | 0.071 (1.34E-02) | 91% | 0.07 |
| WTL | 0.004 (6.24E-05) | 0.003 (1.41E-05) | 0.006 (1.12E-04) | 100% | 0.02 |

| | 5% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | | | | 0% | |
| REC | 0.068 (1.40E-02) | 0.057 (3.72E-03) | 0.106 (2.77E-02) | 98% | 2.89 |
| M1 | 0.099 (2.88E-02) | 0.091 (8.70E-03) | 0.149 (3.65E-02) | 82% | 0.09 |
| BM1 | 0.051 (7.45E-03) | 0.026 (1.09E-03) | 0.068 (1.19E-02) | 94% | 0.07 |
| M2 | 0.133 (4.71E-02) | 0.089 (8.47E-03) | 0.172 (5.19E-02) | 77% | 0.10 |
| BM2 | 0.050 (7.46E-03) | 0.036 (2.23E-03) | 0.070 (1.30E-02) | 93% | 0.07 |
| WTL | 0.009 (2.18E-04) | 0.006 (5.60E-05) | 0.014 (4.24E-04) | 100% | 0.02 |

| | 5% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | | | | 0% | |
| REC | 0.067 (1.29E-02) | 0.058 (3.85E-03) | 0.105 (2.58E-02) | 98% | 2.96 |
| M1 | 0.154 (5.95E-02) | 0.098 (1.01E-02) | 0.184 (5.74E-02) | 64% | 0.10 |
| BM1 | 0.051 (7.45E-03) | 0.026 (1.09E-03) | 0.068 (1.19E-02) | 94% | 0.07 |
| M2 | 0.171 (6.77E-02) | 0.098 (1.02E-02) | 0.198 (6.72E-02) | 64% | 0.11 |
| BM2 | 0.050 (7.46E-03) | 0.036 (2.23E-03) | 0.070 (1.30E-02) | 93% | 0.07 |
| WTL | 0.009 (2.11E-04) | 0.005 (5.22E-05) | 0.013 (4.03E-04) | 100% | 0.02 |

| | 10% with d=6.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | 0.059 (4.35E-03) | 0.109 (1.21E-02) | 0.150 (2.29E-02) | 0% | 0.03 |
| REC | 0.072 (1.44E-02) | 0.058 (4.00E-03) | 0.114 (2.99E-02) | 96% | 3.64 |
| M1 | 0.072 (1.40E-02) | 0.091 (8.83E-03) | 0.126 (2.42E-02) | 89% | 0.08 |
| BM1 | 0.042 (4.06E-03) | 0.026 (1.11E-03) | 0.057 (7.22E-03) | 94% | 0.07 |
| M2 | 0.092 (2.40E-02) | 0.091 (8.84E-03) | 0.137 (3.25E-02) | 79% | 0.10 |
| BM2 | 0.042 (4.65E-03) | 0.034 (1.91E-03) | 0.062 (9.08E-03) | 91% | 0.07 |
| WTL | 0.013 (4.20E-04) | 0.008 (1.26E-04) | 0.021 (8.73E-04) | 100% | 0.02 |

| | 10% with d=10.0 | | | | |
|---|---|---|---|---|---|
| | $\tilde{\omega}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | conv. | time [s] |
| QML | | | | 0% | |
| REC | 0.071 (1.35E-02) | 0.062 (4.49E-03) | 0.114 (2.87E-02) | 93% | 3.63 |
| M1 | 0.114 (3.80E-02) | 0.099 (1.03E-02) | 0.162 (4.38E-02) | 70% | 0.10 |
| BM1 | 0.042 (4.06E-03) | 0.026 (1.11E-03) | 0.057 (7.22E-03) | 94% | 0.07 |
| M2 | 0.139 (5.02E-02) | 0.098 (1.02E-02) | 0.174 (5.38E-02) | 63% | 0.11 |
| BM2 | 0.042 (4.65E-03) | 0.034 (1.91E-03) | 0.062 (9.08E-03) | 91% | 0.07 |
| WTL | 0.012 (3.66E-04) | 0.008 (9.84E-05) | 0.018 (7.14E-04) | 100% | 0.02 |

Table 10.5: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{6, 10\}$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

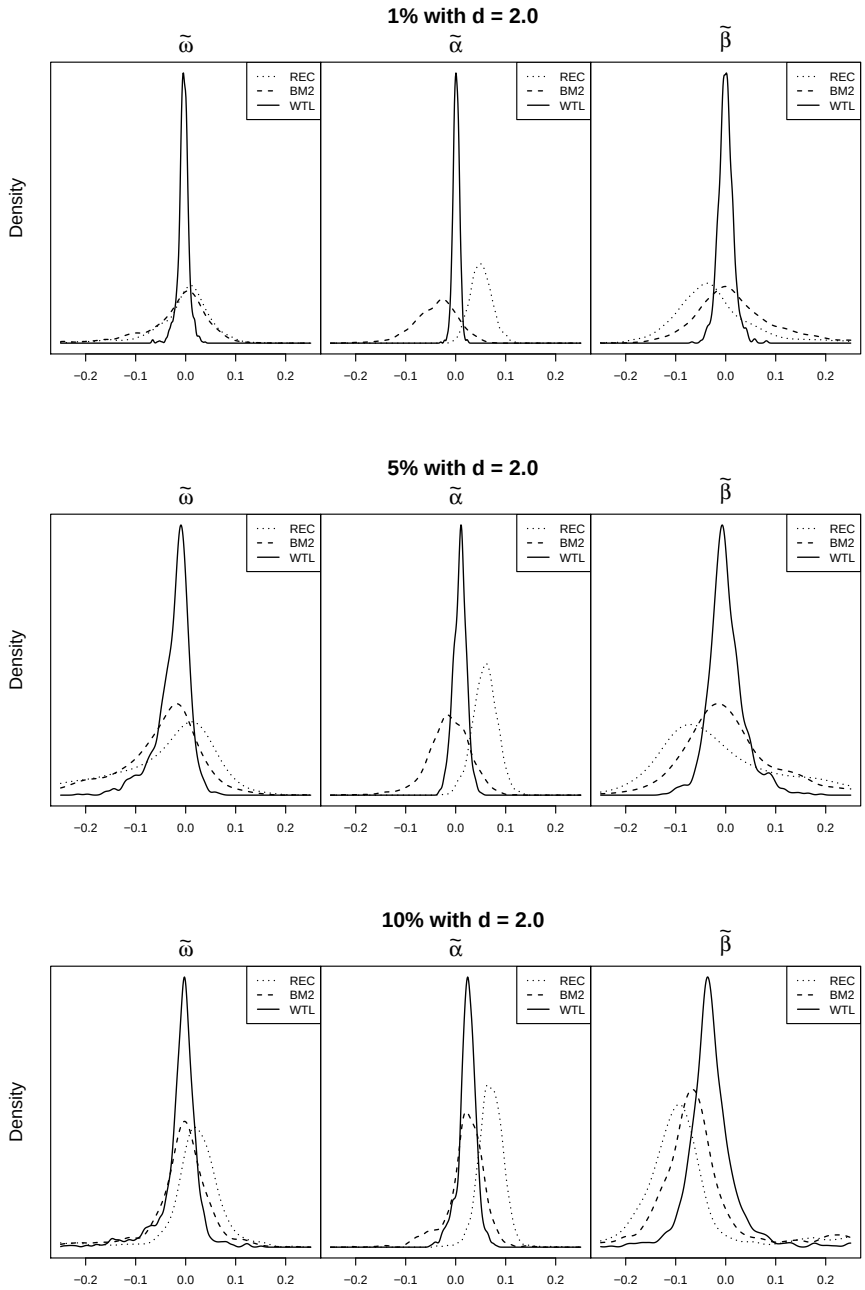


Figure 10.3: Kernel density approximation of the deviations for the simple GARCH(1,1) for REC, BME and WTLE with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d = 2$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 with 500 burn-in sequence. The number of Monte-Carlo replications is 1000.

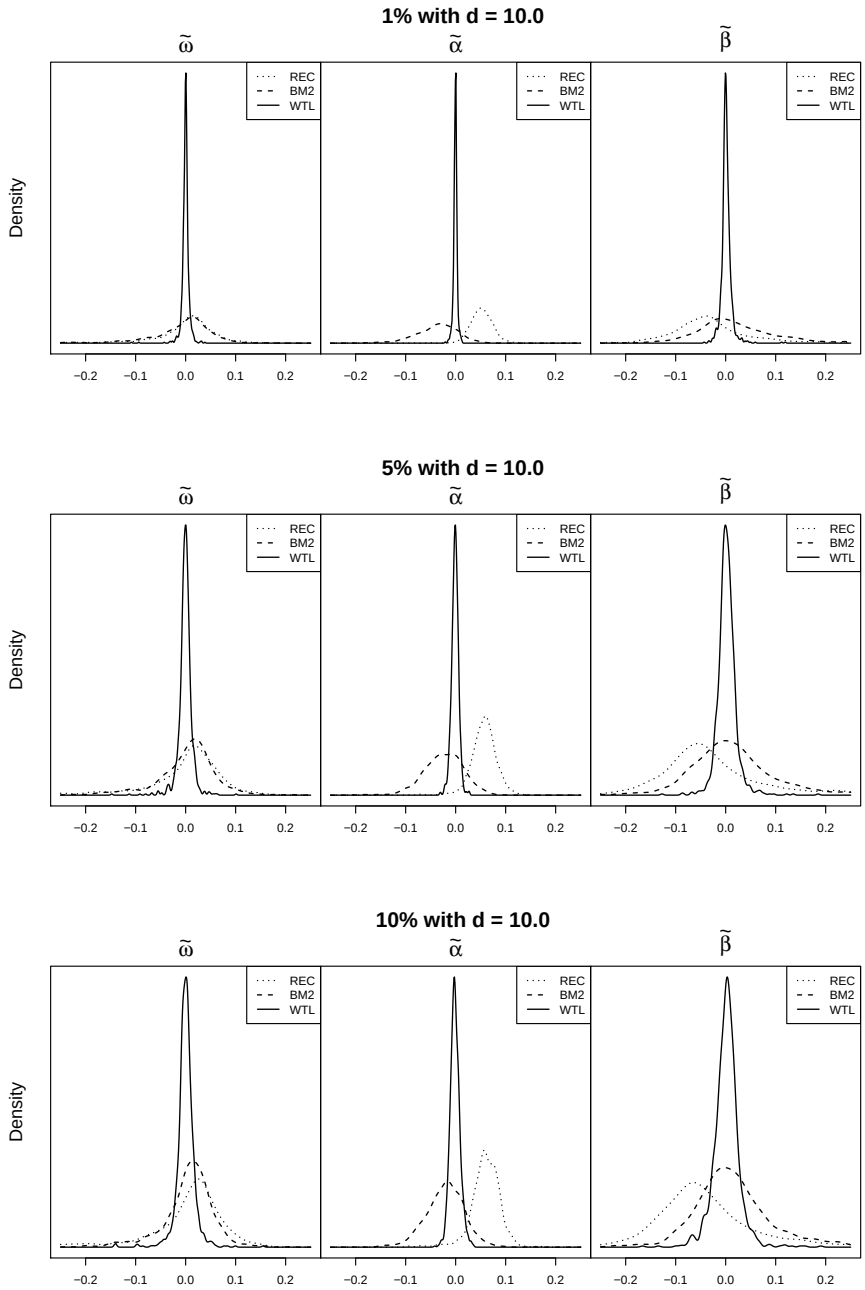


Figure 10.4: Kernel density approximation of the deviations for the simple GARCH(1,1) for REC, BME and WTLE with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d = 10$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 with 500 burn-in sequence. The number of Monte-Carlo replications is 1000.

# 11 Conclusion

The robust estimation of parameters is essential in practice. This is especially true for financial applications where outliers are common in the data. There are many possible origins for these outliers. An outlier may arise from a recording error or from an abrupt regime change due to either a political decision or a rumor. The classical econometric models, such as the GARCH model, can be very sensitive to abnormal points. However, such models are often the foundation of the risk measures that are used by large financial companies to guide their investment strategies. Consequently, there is much to be gained from improving the robustness of econometric models.

In this part, it was shown how to transform the maximum likelihood estimator into a robust estimator. The contribution of the present work, is to construct an automatic selection process for the WTLE parameters. The proposed fully automatic method for selecting the trimming parameter and the weights in the WTLE, obviates the tedious fine tuning process required by other robust models.

The auto-WTLE was successfully applied to GARCH modeling. It was shown, through an extensive simulation study, that the auto-WTLE provides robust and reliable estimates at only a small computational cost. Note that only the simple GARCH(1,1) model was considered. However, the WTLE can be used with any model for which there exists a likelihood estimator.

The automatic calculation procedure for the weights and trimming parameters in the WTLE could be used in other applications. For example, the size of the trimming parameters could be used as stability measure, which would indicate how many of the data points could be accurately represented by the distribution model used. When the trimming parameters significantly increase in a time horizon, this is a clear indication that the underlying model is inadequate and that a new regime might be in effect. Such changes could be represented in terms of a stability measure.

# Part III

# Portfolio Optimization Based on Divergence Measures

# 12 Introduction

The portfolio selection problem consists of finding a repartitioning of a wealth $W$ into $N$ financial instruments with allocation $\mathbf{w} = (w_1, w_2, \ldots, w_N)^{\mathrm{T}}$ such that it produces a feasible portfolio that is most adequate for the investor. For an investor concerned with risk, the criterion of adequacy would be based on a portfolio pseudo-invariant transform, $Y = T_P(\mathbf{w}, V_t, I_t)$, which maps the portfolio selection, $\mathbf{w}$, the instrument values, $V_t$, and the information known at $t$, $I_t$, to a random variate defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, which models the risk perception of the investors. This is a rather general formulation of the portfolio selection problem. A more pragmatic approach of the portfolio selection problem is the risk-reward paradigm, which originated in the Nobel-memorial-awarded mean-variance (MV) portfolio selection of Markowitz [1952]. It consists of finding the weights that produce the smallest variance of the financial returns under specific mean return constraints. The mean-variance paradigm is usually represented as the minimization of any *Risk* measure (the variance, in the MV framework), with respect to a *Reward* target (the mean return, in MV framework). It can be represented by

$$\arg\min_{\mathbf{w}\in\mathcal{W}} \quad \mathit{Risk} \tag{12.1}$$

$$\text{subject to} \quad \mathit{Reward} = c, \tag{12.2}$$

where $\mathcal{W}$ is the set of weights that produces feasible portfolios for the investor, and $c$ is the target value of the *Reward* quantity. Its dual formulation consists of maximizing the *Reward* measure given a target *Risk* value. In the MV framework, the risk measure is represented by the covariance matrix of the assets returns. However, this representation suffers from the fact that it does equally represent risk on both sides of the financial distribution returns. Over the years, new risk measures have emerged, which consider the downside of the return distribution. Examples are the semi-variance introduced by Markowitz [1959], the lower partial risk measure introduced by Fishburn [1977] and further developed by Sortino and Van Der Meer [1991]. Roy [1952] introduced another method of portfolio allocation, known as the safety first principle. It consists of finding the allocation that has the smallest probability of ruin. Over the years, this concept has involved into

the value-at-risk (VaR), advocated by Jorion [1997], and the conditional value-at-risk (CVaR), introduced by Rockafellar and Uryasev [2000]. Another famous portfolio selection framework is that of Black and Litterman [1992], where the investor can include his personal views, on the evolution of the market, in the portfolio criteria. The incorporation of views into the portfolio selection has been extended, by Meucci [2008], into the entropy pooling approach.

Although many portfolio selection models have been introduced since that of Markowitz [1952], the general idea of optimizing a direct quantity of the portfolio has not changed. The investor is assumed to seek the most optimal value of his adequacy function. Nowadays, adequacy functions can be so complicated that general optimization routines might need to be used because the adequacy function has many local minima. Although an investor might obtain weights from these complex optimization problems, in practice, he will not be able to replicate exactly the same allocation due to operational constraints. For example, he might not be able to buy as many contracts of one instrument as requested by his optimization results. The investor is thus left to approximate the solution he has obtained from the optimization, and build a portfolio that is close to his optimal solution. However, if the investor used a complex objective function, the small change might yield a portfolio that is less favorable than that of other local minima.

With regard to this problem, the contribution of this thesis is to go beyond the traditional portfolio optimization idea. A portfolio selection framework is introduced where the investor seeks the allocation that is as close as possible to his "ideal" portfolio, while remaining in agreement with the historical data set. To build such a portfolio selection framework, the $\phi$-divergence measure, from information theory, is used. There are many advantages to using the $\phi$-divergence measure. First, as will be seen in the remainder of this part, the allocation is built such that it is in agreement with the historical data set. Second, the divergence measure is a convex function and enables the use of fast optimization algorithms. Third, the value of the objective value in the minimum portfolio divergence measure provides an indication of how far removed one is from the ideal portfolio. One can therefore construct a statistical test from the value of the objective function. This contrasts with current portfolio optimization schemes. Fourth, with adequate choices of both the target distribution and the divergence measure, the objective function of the $\phi$-portfolios reduces to the expected utility function.

The remainder of this part is organized as follows. Chapter 13 reviews the definitions of the $\phi$-divergence measures, and also their properties. The main result of this part is presented in Chapter 14 where a portfolio selection framework based on $\phi$-divergence measures is constructed. This is followed by a discussion of the dual representation

in Chapter 15. This plays a key role in the numerical resolution of the $\phi$-portfolios. Chapter 16 studies the numerical challenges inherent in the estimation of a portfolio that is as close as possible to a reference portfolio. As will be seen, this challenge reduces to solving a minimax problem. The impact of the choice of divergence measure on the optimized weights is studied in Chapter 17. Since the portfolio framework is to be built upon the divergence measure, the value of the divergence measure can be used as an indication on how far a given portfolio is from the ideal portfolio. In Chapter 18, the value of the objective function is used to calculate statistical tests of divergence from the ideal portfolio. Chapter 19 highlights the connection between the minimum $\phi$-portfolio framework and the expected utility theory.

# 13 Divergence measures

The idea of distance between probability measures goes back to Mahalanobis [1936]. Later, Shannon [1948] introduced an information measure, $I(X,Y)$, defined as the divergence of the joint distribution, $P_{XY}$, of random processes, $X$ and $Y$, from the product, $P_X P_Y$, of the marginal distributions;

$$D(P_{XY}, P_X P_Y) = \int \ln \frac{dP_{XY}}{d(P_X P_Y)} dP_{XY}.$$

Kullback and Leibler [1951] subsequently took the above measure from information theory and used it in probability theory. They used it to measure a difference between any two distributions, $P_1$ and $P_2$. It is nowadays known as the Kullback-Leibler divergence, or as the relative entropy. This divergence measure has been extended, by Csiszár [1963], Morimoto [1963], and Ali and Silvey [1966], to accommodate a whole class of divergences. Let $P$ and $G$ be two probability distributions over a sample space, $\Omega$, such that $P$ is absolutely continuous with respect to $G$. The Csiszár $\phi$-divergence of $P$ with respect to $G$ is

$$D_\phi(P,G) = \int_\chi \phi(\frac{dP}{dG}) dG,$$

where $\phi \in \Phi$ is the class of all convex functions $\phi(x)$ defined for $x > 0$, and satisfying $\phi(1) = 0$. $\phi$ is a convex function twice continuously differentiable in a neighborhood of $x = 1$, with nonzero second derivative at the point $x = 1$. The limiting cases of $\phi$ are defined as: $0\phi(\frac{0}{0}) = 0$, $\phi(0) = \lim_{t\to 0} \phi(t)$, $0\phi(\frac{a}{0}) = lim_{t\to 0} t\phi(\frac{a}{0}) = a \lim_{u\to\infty} \frac{\phi(u)}{u}$, $f(0) = \lim_{t\to 0} f(t)$, and $0f(\frac{a}{0}) = \lim_{t\to 0} tf(\frac{a}{t}) = a \lim_{t\to\infty} \frac{f(u)}{u}$.

When $P$ and $G$ are continuous with respect to a reference measure, $\mu$, on $\Omega$, then the $\phi$-divergence can be expressed in terms of the probability densities $f = dF/d\mu$ and $g = dG/d\mu$;

$$D(P,G) = \int_{-\infty}^{\infty} \phi\left(\frac{f}{g}\right) g d\mu. \tag{13.1}$$

The fundamental properties of the $\phi$-divergence measure are that: it is always positive, $D(P,G) \geq 0$, being zero only if $P = G$; it increases when the two distributions disparity increases; and it has the conjugate divergence, $\phi(G,P) = \tilde{\phi}(P,G)$, where $\tilde{\phi}(x) = x\phi(\frac{1}{x})$.

Table 13.1 lists some divergence measures that have arisen since the introduction of relative entropy as reported in [Pardo, 2006]. These divergence measures were originally used for hypothesis testing, and later used for estimating the parameters of both continuous and discrete models. Statistical applications, and the underlying theory, of these divergences measures can be found in the monographs of Read and Cressie [1988], Vajda [1989], and Pardo [2006], and also in the papers of Morales et al. [1997, 2003].

Table 13.1: Examples of $\phi$-divergence functions as reported in [Pardo, 2006].

| $\phi$-function | Divergence type |
|---|---|
| $x \log x - x + 1$ | Kullback-Leibler (1959) |
| $-\log x + x - 1$ | Minimum discrimination information |
| $(x-1)\log x$ | J-divergence |
| $\frac{1}{2}(x-1)^2$ | Pearson (1900), Kagan (1963) |
| $\frac{(x-1)^2}{(x+1)^2}$ | Balakrishnan and Sanghvi (1968) |
| $\frac{-x^s+s(x-1)+1}{1-s}, \quad s \neq 1$ | Rathie and Kannappan (1972) |
| $\frac{1-x}{2} - \left(\frac{}{2}(1+x^{-r})\right)^{-1/r} + 1$ | Harmonic mean (Mathai and Rathie (1975)) |
| $\frac{(1-x)^2}{2(a+(1-a)x)}, \quad 0 \leq a \leq 1$ | Rukhin (1994) |
| $\frac{ax\log x-(ax+1-a)\log(ax+1-a)}{a(1-a)}$ | Lin (1991) |
| $\frac{x^{\lambda+1}-x-\lambda(x-1)}{\lambda(\lambda+1)}$ | Cressie and Read (1984) |
| $\lvert 1-x^a\rvert^{1/a}, \quad 0 < a < 1$ | Matusita (1964) |
| $\lvert 1-x\rvert^a, \quad a > 1$ | $\chi$-divergence of order $a$ (Vajda 1973) |
| $\lvert 1-x\rvert$ | Total variation (Saks 1937) |

Liese and Vajda [2006] elegantly demonstrated, by a new generalized Taylor expansion, that all $\phi$-divergences can be represented as an average statistical information with a measure of *a priori* probabilities dependent on the $\phi$ function. This suggests a Bayesian interpretation of parameter estimators based on $\phi$-divergence measures. The estimators seek the parameter values that build an empirical distribution that is the closest to the reference distribution, while remaining in agreement with the data set.

An important family of $\phi$-divergence measures, is the one studied by Cressie and Read [1984]. It is frequently used in statistics because it includes a large set of divergence measures. The power $\phi$-divergence can be expressed as

$$\phi_\gamma(x) = \frac{x^\gamma - 1 - \gamma(x-1)}{\gamma(\gamma-1)}$$

for $-\infty < \gamma < \infty$. The power-divergence measures are undefined for $\gamma = 1$ and $\gamma = 0$.

However, in the continuous limits, $\phi_\gamma$, becomes $\lim_{\gamma\to 1}\phi_\gamma(x) = x\log(x) - x + 1$, and $\lim_{\gamma\to 0}\phi_\gamma(x) = -\log(x) + x - 1$. The power divergence measure encompasses several well-known divergence measures. When $\gamma = \frac{1}{2}$, it corresponds to the Hellinger divergence measure, $2H^2(P,G)$; when $\gamma = 2$, it is the Pearson divergence measure, $\chi^2(P,G)/2$; for $\gamma = 1$ it equates to the Kullback-Leibler divergence, I(P,G); when $\gamma = -1$, it becomes the reverse Pearson (or Neyman) divergence, $\chi^2(G,P)/2$; and with $\gamma = 0$, it is the Leibler (or reverse Kullback) divergence.

The idea of using information theory in finance is not new. For example, the entropy measures have been used by Stutzer [1996] and Kitamura and Stutzer [1997] to obtain a predictive distribution that has minimum relative entropy with respect to a prior distribution. Their method inspired Foster and Whiteman [1999] and Robertson et al. [2005] to assign probabilities to scenarios in order to forecast financial distributions. Meucci [2008] also used relative entropy in the context of portfolio optimization, where the relative entropy is used to combine investor's views to the empirical views. The use of entropy measures in finance differs from the approach taken here. The only exception found was the paper of Yamada and Rajasekera [1993], which used the relative entropy to find an allocation of a portfolio such that it is as close as possible to an existing portfolio.

# 14 Phi-portfolios

The optimal portfolio allocation, $\mathbf{w}^* \in \mathcal{W}$, defined on some set of feasible weights for the investors, $\mathcal{W}$, is the allocation that minimizes the $\phi$-divergence between $P_{\mathbf{w}}$, the empirical portfolio distribution with allocation $\mathbf{w}$, and the target portfolio distribution, $P_r$. That is,

$$\mathbf{w}^* = \arg\min_{\mathbf{w}\in\mathcal{W}} \int_a^b \phi\left(\frac{dP_{\mathbf{w}}}{dP_r}\right) dP_r, \tag{14.1}$$

where $a = P_r^{-1}(\alpha)$ and $b = P_r^{-1}(\beta)$ are, respectively, the lower and upper quantiles of the reference distribution, $P_r$, over which the divergence measure should be estimated. Note that the range over which the divergence measure should be estimated has been restricted. This enables the focus to be brought upon specific parts of the target distribution; for example, when the interest is in getting as close as possible to only the lower part of the target distribution. This circumstance would be expressed as the lower and upper quantile bounds of Eq. (14.1), with $\{\alpha = 0,\ \beta = \frac{1}{2}\}$. Hereinafter, the name $\phi$-portfolio will be used to denote the allocation that is most appropriate to the investor in the sense of the $\phi$-divergence measure in Eq. (14.1).

In the general case, a multivariate distribution would be used for the reference portfolio, $P_r$. However, here, to simplify the approach, the problem is reduced to the univariate distribution of the target distribution, i.e., the expected distribution of the returns of the final portfolio. Note that the inter-connectivity among the components of the portfolio could be modeled as additional constraints. For example, the risk budget constraints could be used to introduce, in the portfolio selection, the notion of risk relation among the components, in terms of the covariance. Moreover, a notion of the temporality of financial returns could be included by the use of maximum drawdown. Risk budgeting would be an adequate approach to model risk, while remaining simple in the optimization.

# 15 Dual phi-portfolios

## 15.1 Concept

The estimation of the $\phi$-divergence, as expressed in Eq. (13.1), requires the estimation of the empirical portfolio distribution, $P_{\mathbf{w}}$. In practice, the empirical distribution can be estimated by non-parametric estimators such as the kernel density estimators. However, doing so introduces the problem of the bias originating from the density estimator on the divergence measure. For example, in the case of kernel density estimators, the bias introduced by the choice of the kernel bandwidth must be taken into account. With regard to this, several attempts have been made to lower the influence of the empirical density estimation in the $\phi$-divergence measure. Lindsay [1994] suggests applying the same kernel estimator to the target distribution in order to reduce the biasing effect of the kernel estimator. However, as pointed out by Keziou [2003], Liese and Vajda [2006] and Broniatowski and Keziou [2006], the $\phi$-divergence estimator can be expressed in a dual representation that does not require an empirical density estimator. This dual representation is based on the properties of convex functions. For all $a$ and $b$ in the domain of a continuously differentiable convex function $f$, it holds that

$$f(b) \geq f(a) + f'(a)(b-a), \tag{15.1}$$

with $f(b) = f(a)$ if and only if $a = b$. Geometrically, this means that a convex function always lies above its tangents. This is known as the first-order convex condition.

Given a set of weights, $\mathbf{w}$, the $\phi$-portfolio measure, $D_\phi\left(P_{\mathbf{w}}, P_r\right)$, can be expressed in terms of an auxiliary distribution, $P_\theta$, that shares the same support as both the reference distribution and the empirical distribution. By substituting $\frac{dP_\theta}{dP_r}$ for $a$ and $\frac{dP_{\mathbf{w}}}{dP_r}$ for $b$, in Eq. (15.1), and by integrating with respect to $P_r$ over the interval $(a, b)$, yields

$$D_\phi(P_{\mathbf{w}}, P_r) \geq \int_a^b \phi\left(\frac{dP_\theta}{dP_r}\right) dP_r + \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_{\mathbf{w}} - \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\theta. \tag{15.2}$$

It directly follows that the inequality in Eq. (15.2) reduces to an equality when $P_\theta = P_{\mathbf{w}}$.

The dual representation of the $\phi$-portfolio measure for an auxiliary probability $P_\theta$ is

$$D_\phi(P_{\mathbf{w}}, P_r) = \sup_{P_\theta \in \mathbb{P}} d_\phi(P_{\mathbf{w}}, P_r, P_\theta),$$

where

$$d_\phi(P_{\mathbf{w}}, P_r, P_\theta) = \int_a^b \phi\left(\frac{dP_\theta}{dP_r}\right) dP_r + \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_{\mathbf{w}} - \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\theta. \quad (15.3)$$

The expected value of the portfolio distribution with allocation $\mathbf{w}$ in Eq. (15.3), is $\mathbb{E}_{P_{\mathbf{w}}}[x] = \int_a^b x\, dP_{\mathbf{w}}$. This can be replaced by its empirical value, since $\mathbb{E}_{P_{\mathbf{w}}} \to \frac{1}{n}\sum P_{\mathbf{w}}$ for large $n$. The dual form for continuous distributions thus becomes

$$d_\phi(P_{\mathbf{w}}, P_r, P_\theta) = \frac{1}{n}\sum_i \phi'\left(\frac{f_\theta(x_i)}{f_r(x_i)}\right) + \int_a^b \left[\phi\left(\frac{f_\theta(x)}{f_r(x)}\right) f_r(x) - \phi'\left(\frac{f_\theta(x)}{f_r(x)}\right) f_\theta(x)\right] dx, \quad (15.4)$$

for all indices $i$ such that $a \leq x_i \leq b$ and $n$ is the number of points that fulfill the index condition.

## 15.2 Dual form with the GLD

The generalized lambda distribution (GLD) of Part I, now becomes of special interest. Since the GLD can approximate a large range of distributions, it becomes a natural candidate for the auxiliary distribution, $P_\theta$, in Eq. (15.4). A particular aspect of the GLD is that it is described by its quantile function. The dual form of Eq. (15.4) is now expressed in terms of the quantile density function of the auxiliary distribution, $P_\theta$. Recall that the quantile density function is the derivative of the quantile function with respect to the probabilities,

$$q(u) = \frac{d}{du}Q(u),$$

for probabilities $u$. Given that the probability density function, $f$, is related to the cumulative probability function, $F$, by $f(x) = dF(x)/dx$ by definition, the quantile density is related to the probability density by

$$q(u) = \frac{d}{du}F^{-1}(u) = \frac{1}{f(F^{-1}(u))}.$$

In terms of the quantile function of the auxiliary distribution, the dual form of the $\phi$-divergence measure between the portfolio reference distribution, $P_r$, and the empirical

portfolio distribution, $P_{\mathbf{w}}$, given a set of weights, $\mathbf{w}$, is

$$D_\phi(P_{\mathbf{w}}, P_r) = \arg\sup_{\theta\in\mathbf{\Theta}} d_\phi(P_{\mathbf{w}}, P_r, Q_\theta),$$

where

$$d_\phi(P_{\mathbf{w}}, P_r, Q_\theta) = \frac{1}{n}\sum_{i=1}^{n} \phi'\left(A_i^{-1}\right) + \int_\alpha^\beta \left[\phi\left(B^{-1}\right)B - \phi'\left(B^{-1}\right)\right] du,$$

and $A_i = q_\theta\left(Q_\theta^{-1}(x_i)\right) f_r(x_i)$ for all $x_i$ such that $Q_r(\alpha) \leq x_i \leq Q_r(\beta)$, and $B = q_\theta(u) f_r\left(Q_\theta(u)\right)$. $Q_\theta$ and $q_\theta$ are the quantile function and quantile density function of $P_\theta$, respectively. As in the previous section, take the power divergence measure of Cressie and Read [1984] for the $\phi$-divergence measure. Recall that the power divergence function is defined as

$$\phi_\gamma(x) = \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$$

for $\gamma \in \mathbb{R}$ and has the limiting cases

$$\phi_0(x) = -\log x + x - 1, \tag{15.5}$$

$$\phi_1(x) = x\log x - x + 1. \tag{15.6}$$

The dual representation of the power divergence is now expressed in terms of the quantile density functions. After a straightforward mathematical derivation,

$$d_\gamma(P_{\mathbf{w}}, P_r, Q_\theta) = \frac{1}{\gamma - 1}\sum_{i=1}^{n}\left(A_i^{1-\gamma} - 1\right) - \frac{1}{\gamma}\int_\alpha^\beta B\left(B^{-\gamma} - 1\right) du.$$

When $\gamma \to 0$;

$$\lim_{\gamma\to 0} d_\gamma(P_{\mathbf{w}}, P_r, Q_\theta) = \sum_{i=1}^{n}(1 - A_i) + \int_\alpha^\beta B\ln B\, du,$$

and when $\gamma \to 1$;

$$\lim_{\gamma\to 1} d_\gamma(P_{\mathbf{w}}, P_r, Q_\theta) = -\sum_{i=1}^{n}\ln A_i + \int_\alpha^\beta (B - 1)\, du.$$

When the GLD is considered as a candidate for the auxiliary distribution for $Q_\theta$, the divergence, $d_\gamma(P_{\mathbf{w}}, P_r)$, becomes the supremum of the dual form over the range of

parameters of the GLD:

$$\begin{aligned} \arg\sup_{\tilde{\mu},\tilde{\sigma},\chi,\xi} \quad & d_\gamma(P_{\mathbf{w}}, P_r, Q_{\text{GLD}}) \\ \text{subject to} \quad & 0 < \tilde{\sigma} \\ & -1 < \chi < 1 \\ & 0 < \xi < 1 \\ & Q_{\text{GLD}}(0) \leq \min(\mathbf{x}) \\ & \max(\mathbf{x}) \leq Q_{\text{GLD}}(1), \end{aligned}$$

where $\tilde{\mu}$, $\tilde{\sigma}$, $\chi$ and $\xi$ are, respectively, the median, interquartile range and two shape parameters of the GLD. The last two restrictions in the equation block above ensure that all values of the portfolio are included in the support of the GLD.

## 15.3 Primal-dual interior point implementation

The previous section showed how the divergence measure can be transformed to a maximization problem. The dual form of the divergence measure in Eq. (15.4) becomes a maximization problem with a concave objective function and simple bound constraints for the GLD parameters. This section presents the method used to solve that problem. The method presented here follows that of Boyd and Vandenberghe [2004]. The solution of the optimization problem in Eq. (15.4) must thus fulfill the modified Karush–Kuhn–Tucker (KKT) conditions. Hence,

$$\nabla d_\gamma(P_{\mathbf{w}}, P_r, Q_\theta) - D^T\lambda = 0, \tag{15.7a}$$

$$-\Gamma_\lambda C - (1/t)\mathbf{I} = 0, \tag{15.7b}$$

where $\lambda$ are the KKT multipliers with $t > 0$, $\Gamma_v$ is the diagonal operator that maps a vector $v$ to the diagonal matrix diag$[v]$, $\mathbf{I}$ is the identity matrix, and

$$C = \begin{bmatrix} -\tilde{\sigma} \\ -1-\chi \\ \chi - 1 \\ -\xi \\ \xi - 1 \\ Q_0 - a \\ b - Q_1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ \frac{\partial Q_0}{\partial \tilde{\mu}} & \frac{\partial Q_0}{\partial \tilde{\sigma}} & \frac{\partial Q_0}{\partial \chi} & \frac{\partial Q_0}{\partial \xi} \\ -\frac{\partial Q_1}{\partial \tilde{\mu}} & -\frac{\partial Q_1}{\partial \tilde{\sigma}} & -\frac{\partial Q_1}{\partial \chi} & -\frac{\partial Q_1}{\partial \xi} \end{bmatrix}. \tag{15.8}$$

$a$ and $b$ are respectively defined as the smallest point and largest point of the portfolio returns that should be included in the support of the GLD (i.e., $a = \min(\mathbf{x})$ and $b = \max(\mathbf{x})$). $Q_0$ and $Q_1$ are the GLD quantiles at probabilities 0 and 1 respectively, and $\{\tilde{\mu}, \tilde{\sigma}, \chi, \xi\}$ are the parameters of the GLD.

The gradient of the dual power divergence becomes

$$\nabla d_\gamma(P_{\mathbf{w}}, P_r, Q_\theta) = \frac{1}{\gamma} \int_\alpha^\beta (\gamma - 1 + B^\gamma)\, B^{-\gamma} B'\, du - \sum_i^n A_i^{-\gamma} A_i',$$

where

$$A_i' = f_r(x_i) \left[ \nabla q_\theta \left( Q_\theta^{-1}(x_i) \right) + \frac{1}{q_\theta \left( Q_\theta^{-1}(x_i) \right)} \frac{dq_\theta}{du} \left( Q_\theta^{-1}(x_i) \right) \right],$$

and

$$B' = f_r(Q_\theta(u)) \nabla q_\theta(u) + q_\theta(u) \nabla Q_\theta(u) \frac{df_r}{dx}(Q_\theta(u)).$$

The first KKT condition (15.7a) is denoted hereafter as the *dual residual*, $r_{\text{dual}}$. The second condition (15.7b) shall be called the *centrality residual*, $r_{\text{cent}}$.

The KKT conditions (15.7) can be solved by Newton steps. The Newton algorithm constructs a sequences of $x_n$ with steps that converge to the extremum $x$. Let $\Delta x = x - x_n$ be the iteration step, the second order Taylor expansion of the objective function, $f$, about the point $x_n$ is

$$f(x + \Delta x) = f(x_n) + f'(x_n)\Delta x + \frac{1}{2} f''(x_n) \Delta x^2.$$

The function $f$ attains its extremum when its derivatives with respect to $\Delta x$ are equal to zero. We thus have to solve the linear equation

$$f''(x_n)\Delta x = -f'(x_n).$$

Given the modified KKT conditions (in 15.7), the Newton steps, $\Delta x = (\Delta\theta, \Delta\lambda)$, are the solution of the following linear system of equations,

$$\begin{bmatrix} \nabla^2 d_\gamma + \lambda^T C & D^T \\ -\Gamma_\lambda D & -\Gamma_C \end{bmatrix} \begin{bmatrix} \Delta\theta \\ \Delta\lambda \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \end{bmatrix},$$

where $C$ and $D$ are defined in Eq. (15.8).

Once the new primal-dual direction, $\Delta x$, has been calculated, it is used in a simple line search. This consists of determining the largest step $s$, toward the supremum, that is deemed feasible. The resulting successive iterations of $x^+$ and $\lambda^+$ can be defined as

$$\mathbf{x}^+ = x + s\Delta x \quad \text{and} \quad \lambda^+ = \lambda + s\Delta\lambda.$$

To solve the problem, the backtracking approach presented in [Boyd and Vandenberghe, 2004] is used. The approach starts with the largest feasible $s$ that keeps all constraints satisfied, $\lambda^+ \geq 0$. This can be expressed as

$$s^{\max} = \sup\{s \in [0,1] \,|\, \lambda + s\Delta\lambda_{pd} \geq 0\} \tag{15.9}$$

$$= \min\{1, \min\{-\lambda_1/\Delta\lambda_i | \Delta\lambda_i < 0\}\}. \tag{15.10}$$

$s$ is then backtracked: starting at $s = .99\, s^{\max}$, $s$ is multiplied by $\beta \in (0,1)$ (to obtain $C^+ < 0$). and then the scaling of $s$ by $\beta$ is repeated until

$$\|KKT^+\|_2 \leq (1 - \alpha s)\|KKT\|_2.$$

Typical choices of the scaling parameters are $\alpha \in [0.01, 0.1]$ and $\beta \in [0.3, 0.8]$.

The interior point method requires a feasible starting point. A typical choice is to use the median and interquartile range of the target distribution for the initial (GLD) parameter-values of $\tilde{\mu}$ and $\tilde{\sigma}$ respectively. To ensure that all points of the portfolio are included in the support of the distribution, the two shape parameters are set to values that produce a distribution with infinite support (see Section 3.2); say, for example $\chi = 0$ and $\xi = 0.6$. Given the starting values, the following procedure is repeated until the primal-dual residual, centrality residual, and the surrogate duality gap, $\nu = -C^T\lambda$, have reached satisfactory values: $\|r_{\text{dual}}\|_2 \leq \varepsilon_{\text{dual}}$ and $\hat{\nu} \leq \varepsilon$.

The interior point algorithm can be summarized in the following steps. Using the set of parameters $\theta = (\hat{\mu}, \hat{\sigma}, \chi, \xi)$ as starting values with $\lambda > 0$, $\mu > 1$, $\varepsilon_{feas} > 0$ and $\varepsilon > 0$; the following steps are performed until a convergence criterion has been satisfied:

1. Set $t = \mu m/\hat{\nu}$,
2. Calculate the primal-dual direction, $\Delta x$,
3. Perform a line search on $y = y + s\Delta y$, where $s > 0$ has to be determined.

# 16 Minimum phi-portfolios

For the dual form of the divergence measure in Chapter 15, the optimal weights, $\mathbf{w}^*$, in terms of the minimum $\phi$-portfolio are

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathcal{W}} \sup_{\theta} d_\phi(P_{\mathbf{w}}, P_r, P_\theta).$$

A simplistic approach to solving this optimization problem would be to perform inner and outer optimizations that solves the minimization and maximization problems respectively. Such an approach would be time consuming and might be numerically unstable. However, by definition, the divergence measure is a continuous convex function. Moreover, the dual form is a concave function with bounded constraints. The minimum and supremum of the objective function are attained when the derivatives with respect to the weights, $\mathbf{w}$, and the auxiliary distribution parameters, $\theta$, are equal to zero. This problem is the equivalent of finding the saddle point of the objective function. Many different methods have been introduced to find the saddle point of constrained functions. Here, a primal-dual interior point solution to the above minimax problem is introduced.

The general form of the minimax problem can be formulated as:

$$\begin{aligned}
\min_x \max_y \quad & f(x, y) \\
\text{subject to} \quad & c_i(x) = 0, \qquad i \in \mathcal{E} \\
& c_i(x) \leq 0, \qquad i \in \mathcal{I} \\
& Ax = a \\
& h_j(y) = 0, \qquad j \in \mathcal{E} \\
& h_j(y) \leq 0, \qquad j \in \mathcal{I} \\
& By = b,
\end{aligned}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $f(x, y)$ is convex in $x$ and concave in $y$, the constraints $h$ and $c$ are both convex, $\mathcal{E}$ is the index set of equality constraints, $\mathcal{I}$ is the index set of inequality

constraints, and $A$, $a$, $B$, and $b$ describe the linear constraints of $x$ and $y$ respectively. Likewise in Section 15.3, the modified KKT conditions to solve this problem are

$$R = \begin{bmatrix} \delta_x f(x,y) + Dc^T\lambda + A^T\nu \\ -\Gamma_\lambda c - (1/t)\mathbf{I}_c \\ Ax - a \\ \delta_y f(x,y) - Dh^T\xi - B^T\mu \\ -\Gamma_\xi h(y) - (1/t)\mathbf{I}_h \\ By - b \end{bmatrix} = 0,$$

where $\lambda$, $\nu$, $\xi$, $\mu$ are the KKT multipliers, $\Gamma_v$ is the diagonal operator that maps a vector $v$ to the diagonal matrix diag$[v]$, $\mathbf{I}_c$ and $\mathbf{I}_h$ are diagonal matrices with zeros when $i \in \mathcal{E}$ and ones when $i \in \mathcal{I}$, and

$$c = \begin{bmatrix} c_1(x) \\ \vdots \\ c_n(x) \end{bmatrix}, \; Dc = \begin{bmatrix} \nabla c_1(x) \\ \vdots \\ \nabla c_n(x) \end{bmatrix}, \; h = \begin{bmatrix} h_1(y) \\ \vdots \\ h_m(y) \end{bmatrix}, \text{ and } Dh = \begin{bmatrix} \nabla h_1(y) \\ \vdots \\ \nabla h_m(y) \end{bmatrix}.$$

The solution of the modified KKT conditions can be found by Newton steps. Each iteration are solutions of the linear system of the form

$$\begin{bmatrix} H^2_{xx} & Dc & A^T & H^2_{xy} & O & O \\ -\Gamma_\lambda Dc & -\Gamma_c & O & O & O & O \\ A & O & O & O & O & O \\ H^2_{yx} & O & O & H^2_{yy} & -Dh & -B^T \\ O & O & O & -\Gamma_\nu Dh & -\Gamma_h & O \\ O & O & O & B & O & O \end{bmatrix} \begin{bmatrix} \delta x \\ \delta\lambda \\ \delta\nu \\ \delta y \\ \delta\xi \\ \delta\mu \end{bmatrix} = -R,$$

where $H^2_{xx} = \delta^2_{xx} f(x,y) + \sum_{i=1}^n \lambda_i \nabla^2 c_i(x)$, $H^2_{yy} = \delta^2_{yy} f(x,y) - \sum_{i=1}^m \xi_i \nabla^2 h_i(y)$, $H^2_{xy} = \delta^2_{xy} f(x,y)$ and $H^2_{yx} = \delta^2_{yx} f(x,y)$. The sequence of Newton steps are iterated until a convergence criterion has been satisfied.

In this thesis, the minimum $\phi$-divergence estimator with the dual power divergence function and the GLD as the auxiliary distribution has been implemented. Given the experiments that will be covered in the following chapters, the implementation of the optimization problem was reduced to:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \max_{\theta} \; d_\phi(P_{\mathbf{w}}, P_r, P_\theta)$$

$$
\begin{aligned}
\text{subject to} \quad & C_i(\mathbf{w}) = 0, \qquad i \in \mathcal{E} \\
& C_i(\mathbf{w}) \geq 0, \qquad i \in \mathcal{I} \\
& 0 < \tilde{\sigma} \\
& -1 < \chi < 1 \\
& 0 < \xi < 1 \\
& Q_\theta(0) \leq \min(\mathbf{x}) \\
& \max(\mathbf{x}) \leq Q_\theta(1).
\end{aligned}
$$

$C$ defines equality and inequality constraints on the portfolio weights $\mathbf{w}$. $\tilde{\mu}$, $\tilde{\sigma}$, $\chi$ and $\xi$ are the median, interquartile range and two shape parameters of the GLD, and $\theta = \{\tilde{\mu}, \tilde{\sigma}, \chi, \xi\}$.

# 17 Choice of divergence measures

This chapter seeks to understand the effect of the divergence measure on the weights allocation. Indeed, the choice of the divergence measure can produce different optimal weights. Figure 17.1 displays the divergence surface between the normal distribution with mean 0 and standard deviation 1, and normal distributions of varying mean and standard deviation. The power divergence family was used, with different sets of exponent values as shown in the figure. It is clear that the divergence has different shapes, depending upon the power parameter. This is a strong indication that portfolio estimators based on different divergence measures will yield different weight parameter estimates.

The supremum of the dual form of the divergence measure is used to estimate the divergence measure itself. Recall that the supremum of the dual form of the power divergence is obtained when its derivatives are equal to zero, where

$$d_\gamma(P_{\mathbf{w}}, P_r, P_\theta) = \frac{1}{n(\gamma-1)} \sum_{i=1}^{n} \left[ \left( \frac{f_\theta}{f_r} \right)^{\gamma-1} - 1 \right] - \frac{1}{\gamma} \int_a^b f_r \left[ \left( \frac{f_\theta}{f_r} \right)^{\gamma} - 1 \right] du,$$

$$\delta_\theta d_\gamma(P_{\mathbf{w}}, P_r, P_\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\nabla f_\theta(x_i)}{f_r(x_i)} \left( \frac{f_\theta(x_i)}{f_r(x_i)} \right)^{\gamma-1} - \int_a^b \nabla f_\theta \left( \frac{f_\theta}{f_r} \right)^{\gamma-1} du.$$

As noted by Toma and Broniatowski [2011], the summation corresponds to the weighted likelihood score function. The weight function is

$$\left( \frac{f_\theta(x_i)}{f_r(x_i)} \right)^{\gamma-1}.$$

The effect of the choice of the power divergence parameter, $\gamma$, upon the optimized weights, can be understood by studying this weight function. Note that for $\gamma = 1$, there are no weights, and so the estimator reduces to the maximum likelihood estimator.

Two cases must be distinguished. The first case is when the reference distribution can be described by the data set. The second case is when the data set does not converge in distribution to the reference model.

When the portfolio data set converges in distribution to the reference portfolio distribu-

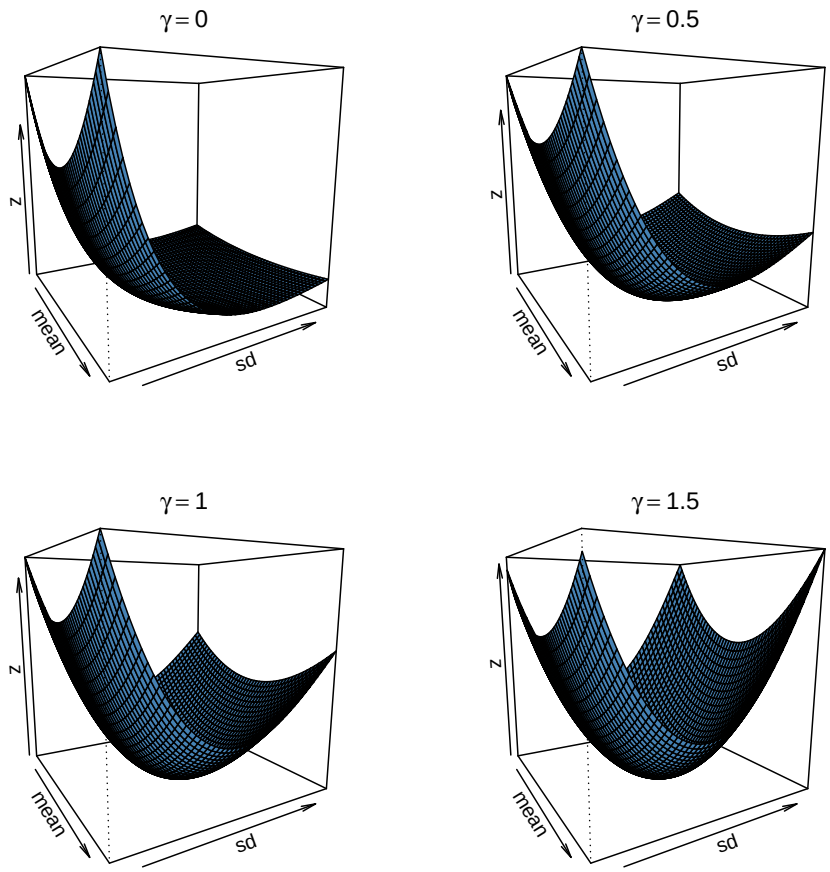


Figure 17.1: Examples of shapes of the power divergence measure with different $\gamma$ values.

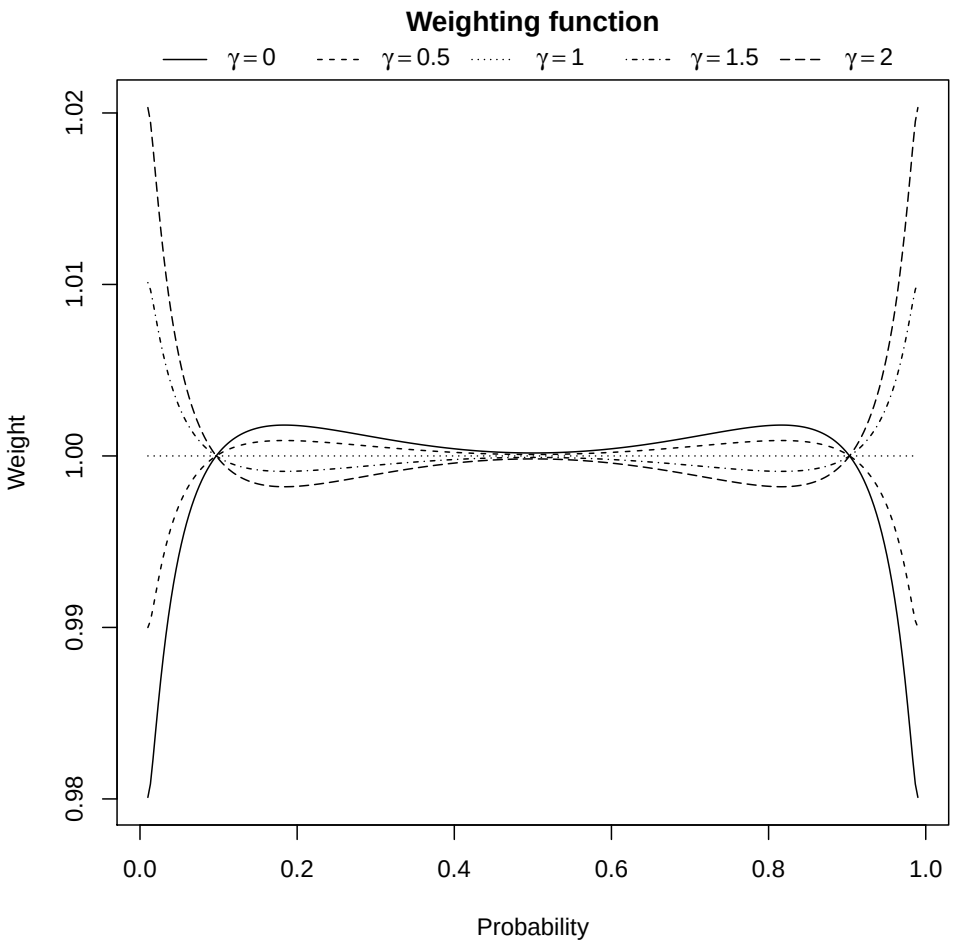


Figure 17.2: Weighing function of the likelihood score function when the empirical portfolio converges in distribution to the target distribution for different $\gamma$ values.

tion, the auxiliary distribution in the dual form converges to the reference distribution as well; i.e., $f_\theta \to f_r$. The weighting function, $[f_\theta(x)/f_r(x)]^{\gamma-1}$, converges to unity, and thus all points are equally weighted. Figure 17.2 illustrates this effect; the weight function is close to one for most of the distribution range, irrespective of the gamma factor.

Regarding portfolio allocation, when the portfolio data set can converge in distribution to the reference distribution, the choice of the divergence measure does therefore not change the weights allocation. To illustrate this, two time series were simulated: both from the normal distribution with a mean of 0; one having a standard deviation of 1 and the other having a standard deviation of 10. The reference distribution was taken to be the normal distribution with mean 0 and standard deviation 5. The optimized portfolio

should therefore be the equally weighted portfolio. Figure 17.3 displays the weight plot for a range of power exponents between 0 and 1. The weights remain constant over the range of power exponents.

However, when the data is significantly different from the reference distribution, the likelihood score function is weighted by the likelihood ratio. At the supremum of the dual divergence measure, the auxiliary distribution, $f_\theta$, converges in distribution to the portfolio distribution, $f_r$, as seen in Chapter 15. The value of the weight function depends on the power exponent $\gamma$, if it is less or greater than 1, and on the likelihood ratio, $\frac{f_\theta(x_i)}{f_r(x_i)}$, if the likely value of the auxiliary distribution is less or greater than the one of the reference distribution at $x_i$. Consider the points with index $i$ where $f_\theta(x_i) > f_r(x_i)$; that is when the reference distribution has thinner tails than the empirical portfolio. When $\gamma > 1$ ($\gamma < 1$), the score function is then upweighted (downweighted) at these points. The effect is inverted for values at index $i$ where $f_\theta(x_i) < f_r(x_i)$.

To illustrate this, three cases are considered. The first case is when the reference distribution scale is larger than the data set scale. The second case is when the scale parameter of the reference distribution is smaller than that of the data set. The third case is when the reference distribution has a larger mean than does the data set.

Figure 17.4 displays the results for $\gamma < 1$. When the reference distribution has larger scale than the data set, the divergence portfolio privileges the allocation that produces the best fit in the tails. This effect is increased when reducing $\gamma$ toward 0. In the case when the target distribution has a smaller scale than the reference distribution, reducing $\gamma$ towards zero has the effect of increasing the weighting around the center of the distribution.

When $\gamma > 1$, Fig. 17.5, the effect is inverted as is expected from the form of the weight function, which is to the power $\gamma - 1$. When the reference distribution has a larger scale than the data set does, the center of the distribution gains higher weighting as $\gamma$ increases. When the reference distribution has a smaller scale than the data set does, the tail of the distribution has higher weight as $\gamma$ increases.

Now to examine the effect of the power parameter $\gamma$ in the context of portfolio optimization with a real data set; the Swiss Pension Fund Benchmarks [1] of Pictet is used. The family of Pictet LPP indices was created in 1985 with the introduction of new regulations in Switzerland governing the investment of pension fund assets. Since then it has established itself as the authoritative pension fund index for Switzerland. In 2000, a family of three indices, called LPP25, LPP40, and LPP60, where the increasing numbers denote increasingly risky profiles, were introduced to provide a suitable benchmark for Swiss pension funds. It is composed of 6 instruments. In this example, the objective

1. The data set (LPP2005) is available in the `R` package **fEcofin**

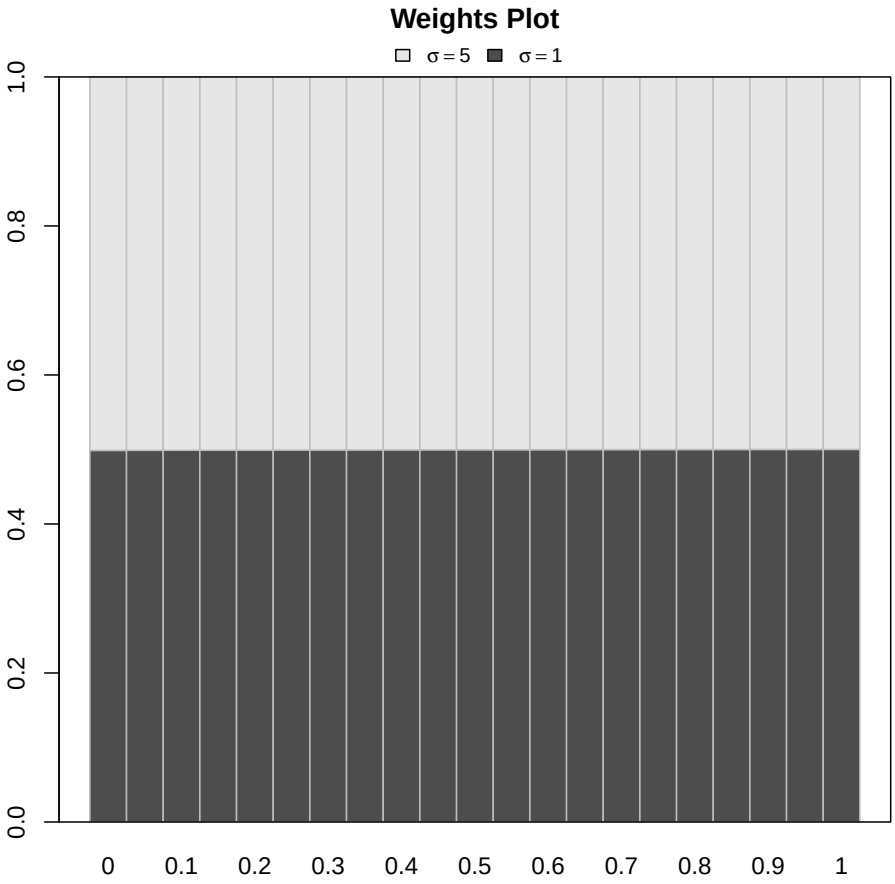


Figure 17.3: Weights plot for a range of power exponents between 0 and 1 for a two time series portfolio where the equally weighted portfolio distribution is set as the target distribution.

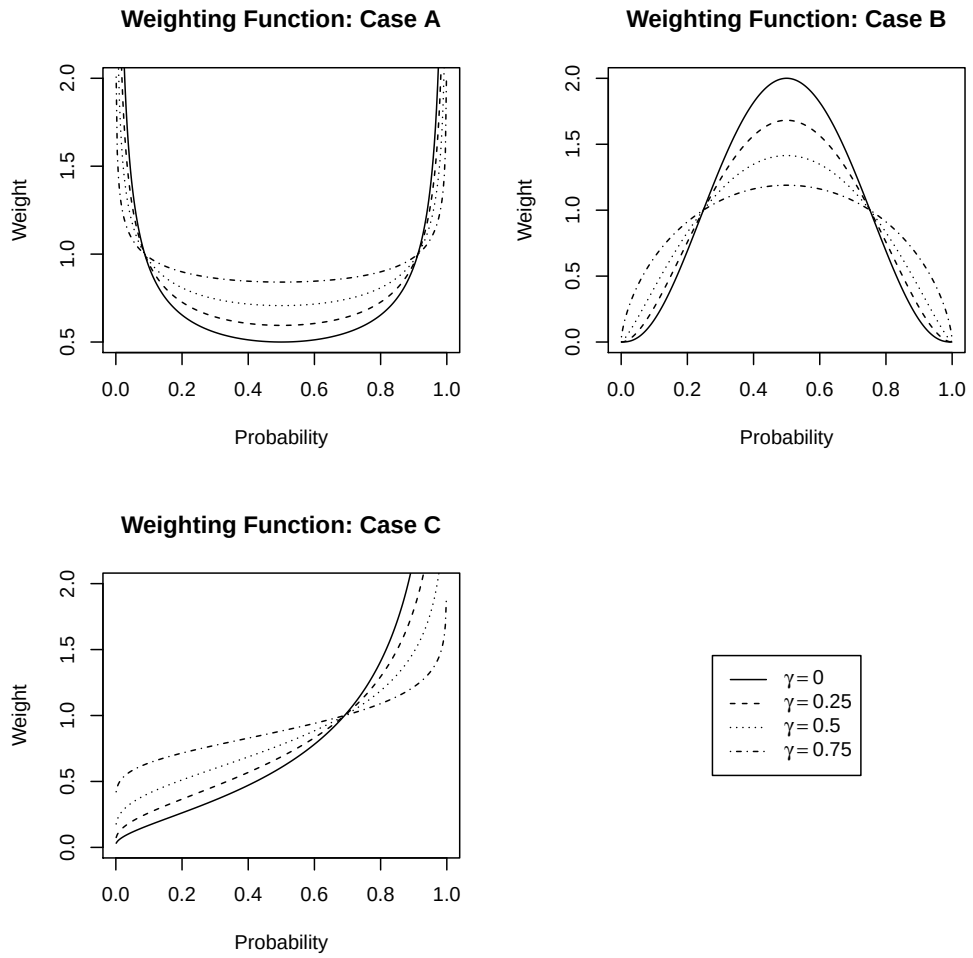


Figure 17.4: Weighting function when: (A) the reference distribution scale is larger than the data set scale, (B) the scale parameter of the reference distribution is smaller than that of the data set, (C) the reference distribution has a larger mean than does the data set.

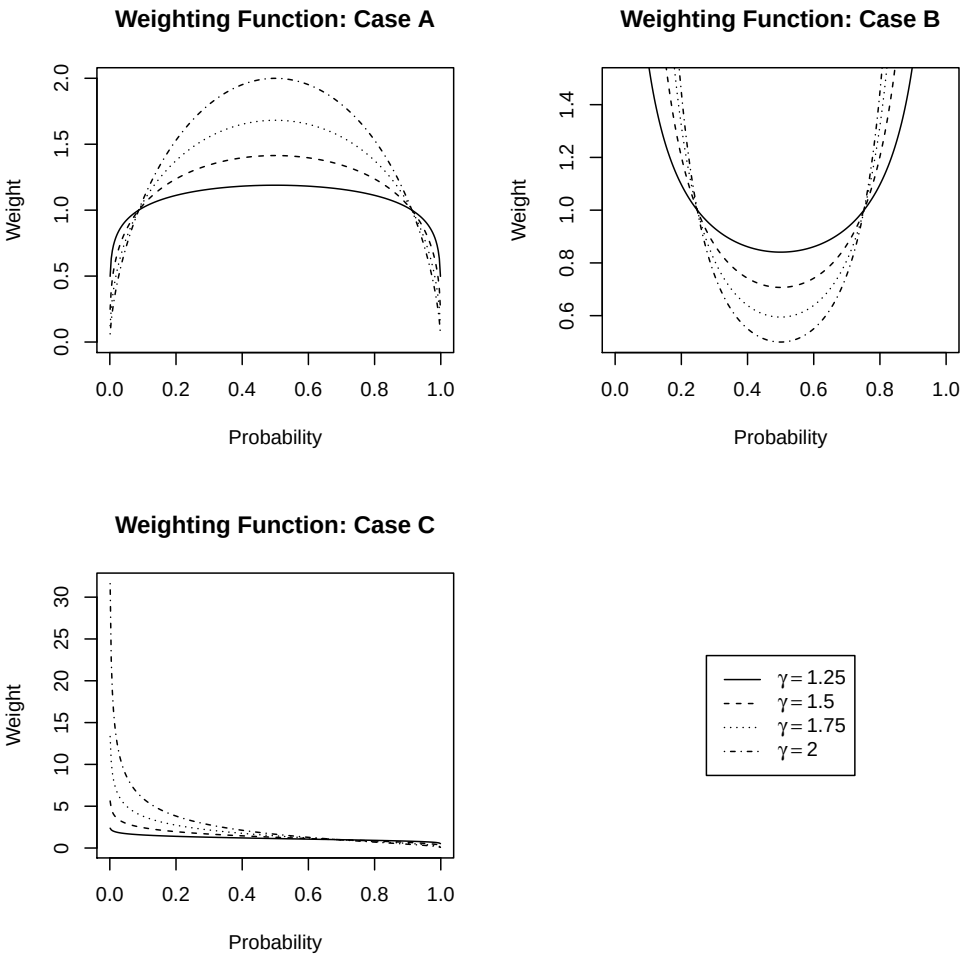


Figure 17.5: Weighting function when: (A) the reference distribution scale is larger than the data set scale, (B) the scale parameter of the reference distribution is smaller than that of the data set, (C) the reference distribution has a larger mean than does the data set.

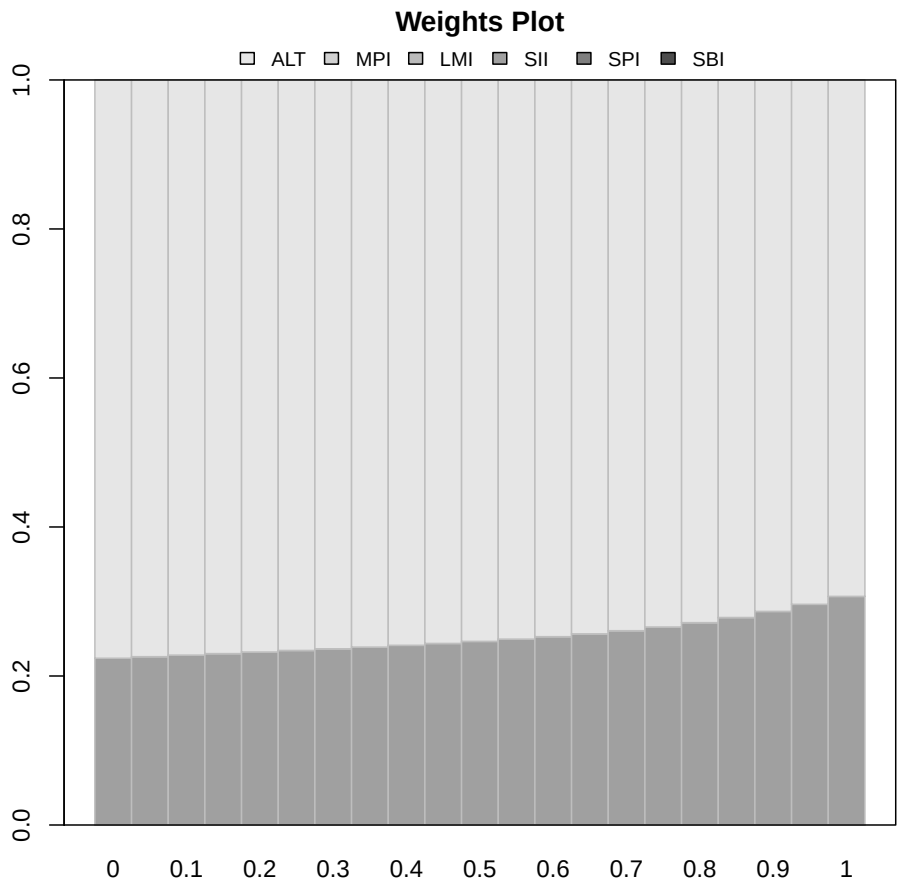


Figure 17.6: Weights plot for the LPP2005 data set. the reference distribution is the normal distribution fitted to the LPP60 benchmark.

is to find the asset allocation of these 6 assets such that the resulting portfolio is as close as possible to the normal distribution with the mean and standard deviation of the benchmark LPP60. Figure 17.6 displays the optimal weights obtained for different values of the power divergence.

Given that the allocation might change, depending on the choice of the divergence measure, it is of interest to see if the shape of the divergence measure between the empirical portfolio and the reference portfolios might change as well. In this regard, Fig. 17.7 displays the ternary plot of the divergence measure for the portfolio, with components SBI, SPI and SII, with respect to the reference portfolio. SBI is the Swiss bond index, SPI is the Swiss performance index and SII is the Swiss real estate index. The logarithm returns of

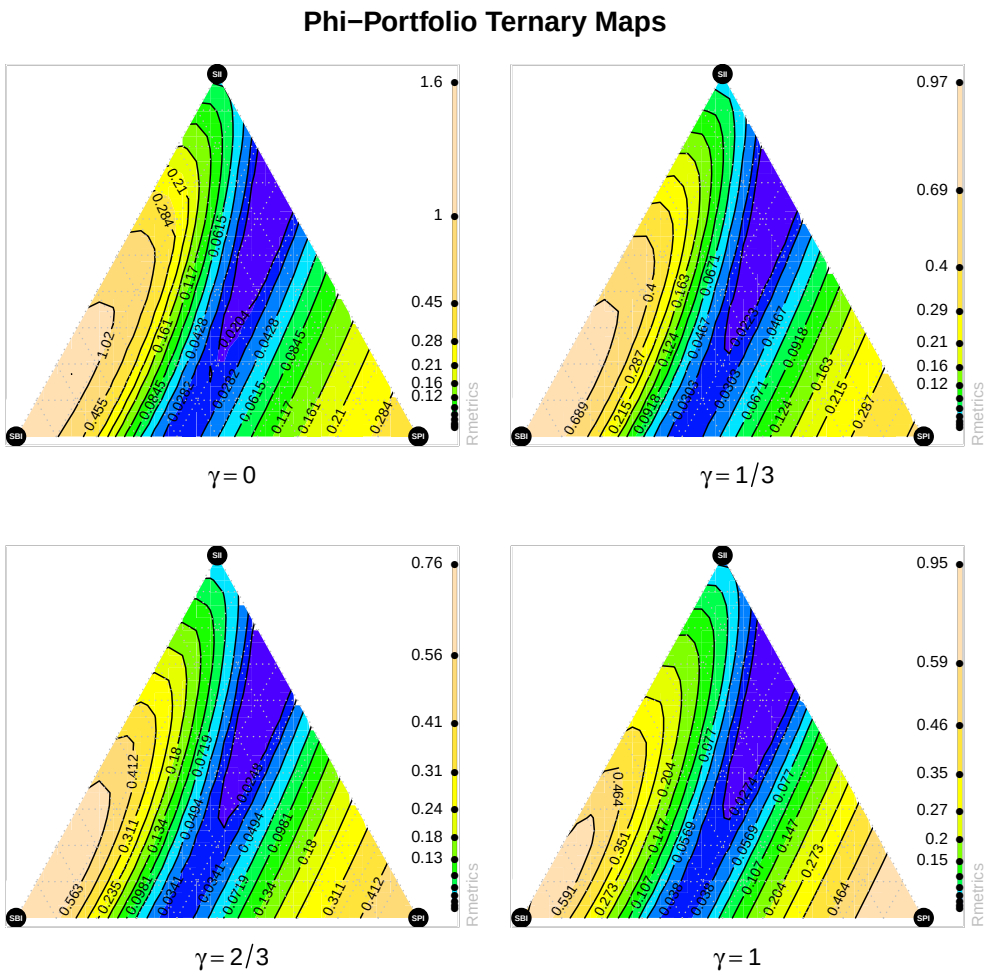


Figure 17.7: Ternary plot to illustrate the shape of the divergence measure for the phi-portfolio with three assets portfolio with different $\gamma$ values.

these indices with daily records from 2000–01–03 to 2007–05–04 [2] were considered. The reference distribution used is the Student $t$-distribution with its parameter-values fitted to the equally weighted portfolio with the SBI, SPI and SII. The four plots have been generated with different power divergence parameters as indicated in the figure. The shape of the divergence measure does not significantly change, although the divergence values themselves are quite different. This is important because it shows that the overall shape of the divergence measure is insensitive to the choice of divergence estimator. This provides an indication of the stability of the approach.

2. The data set (SWXLP) is available in the R package **fEcofin**

# 18 Statistical test

Recall that the $\phi$-portfolio seeks the allocation that produces a portfolio that is as close as possible to the reference distribution. The closeness is expressed by the divergence measure. The value of the objective function in the $\phi$-portfolio therefore carries valuable information. It says how close one can get to the reference model. This is not the case for traditional portfolio allocation.

Cressie and Read [1984] established the asymptotic null distribution of the goodness-of-fit statistics within the power divergence family. They show that the $t$-test is linked to the $\chi^2$ distribution. Broniatowski and Keziou [2009] proved that the asymptotic null distribution of the power divergence family remains a $\chi^2$ distribution when the divergence measure is expressed in its dual form. An important result in [Broniatowski and Keziou, 2009, Theorem 3.2.b] is that at the supremum of the dual divergence, which occurs when the auxiliary distribution tends to the distribution of the sample data, the statistic

$$\frac{2n}{\phi''(1)}\hat{d}_\phi$$

convergences in distribution to the $\chi^2$ distribution where $\hat{d}_\phi$ is the supremum of the dual divergence. In the case of the power divergence family, we have $\phi''(1) = 1$ and the statistics reduces to $2n\,\hat{d}_\gamma$.

The asymptotic distribution of $\phi$-divergence statistical tests remains an open research topic. In this context, Jager and Wellner [2007] introduced confidence bands for the power divergence family. The interested reader is referred to their paper and their `R` package **phitest**. Lately, Toma and Leoni-Aubin [2010] and Toma and Broniatowski [2011] have introduced a more accurate approximation of the $p$-values of the dual divergence, based on the saddle point approximation of the distribution of the supremum dual divergence as introduced by Robinson et al. [2003] for $M$-estimators.

Figure 18.1 displays the histogram of the statistical tests of the divergence for 1000 time series with the same distribution as the reference distribution that being the normal distribution with mean 0 and standard deviation 1. It is clear from the plot that the statistical test converges in distribution to the $\chi^2$ distribution.

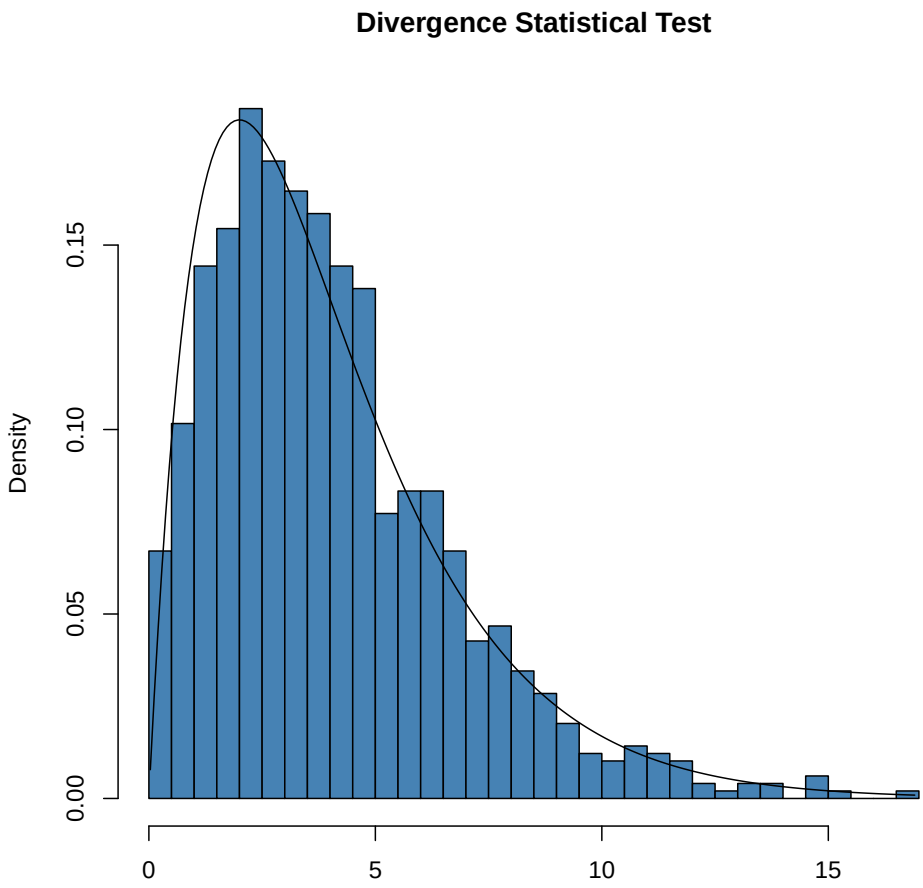


Figure 18.1: Histogram of the statistical test of divergence measures together with the $\chi^2$ density distribution.

As a second illustration, Fig. 18.2 shows how the $p$-values increase while the distribution of the data converges in distribution to the reference model. The data is distributed according to the Student $t$-distribution with mean 0, standard deviation 1 and different degrees of freedom. The normal distribution with mean 0 and standard deviation 1 is used as the reference distribution. It is clear that when Student's $t$-distribution's degrees of freedom increases, the $p$-values increase as well since the distribution of data set will converge to a normal distribution.

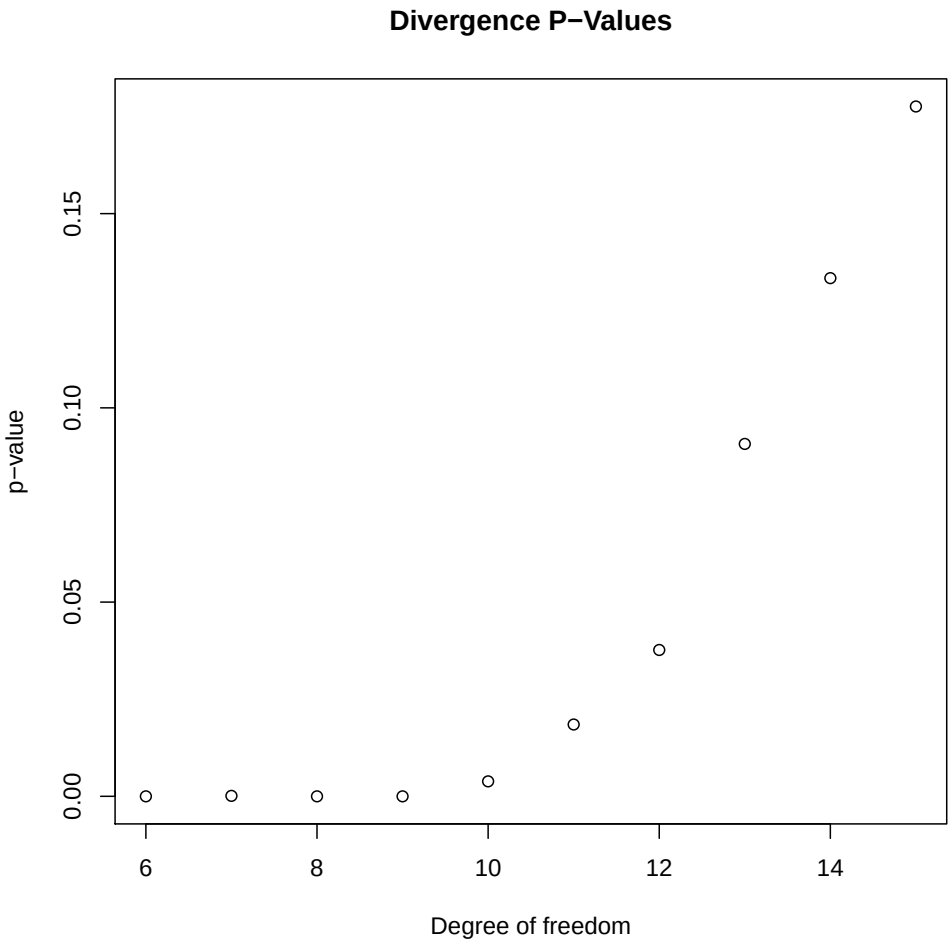


Figure 18.2: $p$-values of the statistical test divergence measure when the generated data set converges in distribution to the reference distribution. Here, the normal distribution is used as the reference model, whereas the samples have Student's $t$-distribution with increasing degree of freedom.

# 19 Expected utility theory

In the expected utility framework, the decision maker constructs a utility function that represents how much the investment would be of interest for him relative to his wealth. The utility function is usually modeled by a concave function that reflects that a smaller gain would be of greater interest to an investor with a smaller wealth. A gain of \$1 is of greater interest to an investor who has \$10, than to an investor who has \$1000. The investor is assumed to be risk-averse and would therefore seek a portfolio allocation that is optimal in a global sense. the investor preference $\succ$ is modeled by the stochastic dominance. For two random outcomes, $X$ and $Y$, that belong to a subset of the measurable probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, the preference $\succ_u$ is defined as

$$X \succ_u Y \Longleftrightarrow \mathbb{E}[u(X)] > \mathbb{E}[u(Y)].$$

The functional form of the expected utility $\nu(.) = \mathbb{E}[u(.)]$ is known as the expected utility function. When the utility function expresses the risk that the investors are willing to take, then the expected utility functions give a way to measure investor's preferences.

This chapter draws the connection between the $\phi$-portfolios and the expected utility theory. Recall that the $\phi$-portfolio is defined as the allocation, $\mathbf{w}^*$, for which divergence between the portfolio distribution and a reference distribution is smallest:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \int \phi\left(\frac{f_{\mathbf{w}}(x)}{f_r(x)}\right) f_r(x)\, dx, \tag{19.1}$$

where $\phi$ is the divergence measure, $f_{\mathbf{w}}$ is the empirical distribution of the portfolio given by the set of weights, $\mathbf{w}$, and $f_r$ is the reference distribution. When the uniform distribution is taken as the reference distribution, the optimal allocation minimizes the expected value of the composition of the divergence $\phi$ function and the distribution of the empirical portfolio distribution, $\phi \circ f_{\mathbf{w}}$. In this case, the optimal weights, $\mathbf{w}^*$, are

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \int u(x)dx = \mathbb{E}[u(x)] = \nu(x), \tag{19.2}$$

where $u = -\phi \circ f_{\mathbf{w}}$. By construction, the function $\nu$ is both increasing and concave, and

corresponds to the utility function defined above. The $\phi$-portfolio is thus equivalent to the maximum expected utility portfolio when the reference distribution is the uniform distribution.

It is interesting to note that the Legendre-Fréchet transformation, used here to express to the divergence measure in its dual form, has also been used within the scope of the utility theory [Föllmer and Schied, 2002], where it has been shown that for special choice of the loss function, it reduces to the relative entropy.

The traditional risk-reward portfolio optimization approach can be linked to the utility theory when the risk measures are described as convex functions as shown by De Giorgi [2005]. The $\phi$-portfolios thus also inherit this connection to the risk-reward paradigm.

# 20 Conclusion

This part has introduced a portfolio allocation scheme based on the $\phi$-divergence measure. The approach uses the dual representation of the divergence measure to avoid using a non-parametric estimation of the empirical distribution. The advantages of the approach are multifold.

As was seen, the scheme seeks the allocation that produces a portfolio that is the closest to a reference model while remaining as compatible as possible to the empirical data, which can be interpreted as a type of Bayesian estimator. This is in contrast to the traditional risk-reward paradigm where an attempt is made to optimize one quantity given constraints. It is believed to be a more viable approach because it reduces the risk estimation error induced by the risk-reward model. In the approach devised in this part, the weights selected are those that are in greatest accord with the empirical data. This could lead to a situation where the selected weights might have a slightly increased deviation, but a much better mean vector. This is also important in the context of robust estimators. As in the case of the power divergence measure, the power exponent can make the estimator more robust to outliers. Given the choice of the divergence measure, the $\phi$-portfolios can be viewed as robust estimators.

Moreover, the value of the optimized $\phi$-portfolio function provides information on how close the investor is to his reference model. It can be used as a statistical test; providing a warning that the reference model might have been too ambitious for the empirical data.

Another important aspect is that the divergence measure is convex and fast optimization routines can be used in practice. This chapter has presented the primal-dual interior optimization for estimating the supremum of the dual representation of the divergence measure and to solve the minimum $\phi$-portfolio problem.

The relationship between this new framework and the utility theory has been highlighted. It has been shown that the utility function becomes a particular case of the $\phi$-portfolio when the uniform distribution is chosen as the reference distribution.

Divergence measure is an active research topic. Many problems remain open. There is no clear comparison of the properties for the different functions that have been introduced over the last few decades. Also, a complete asymptotic theory of $\phi$-divergences has yet to

be found. All these open problems constitute a favorable outlook for the practical use of $\phi$-portfolios. An exhaustive comparison of $\phi$-portfolios to common portfolio optimization schemes would be a valuable contribution to the field.

Another possible direction of research in portfolio allocation based on divergence measure would be the selection of assets such that the allocated portfolio is as far as possible from a worst-case reference distribution. Preliminary studies have been performed, in the context of the safety first principle, by Stutzer [2000]; Haley and Whiteman [2008]. A study based on general $\phi$-divergence measures would be interesting.

# Bibliography


J. Aldrich. R. A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science*, 12(3):162–176, Aug. 1997.

S. Ali and S. Silvey. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc., Ser. B*, 28:131–142, 1966.

W. H. Asquith. L-moments and TL-moments of the generalized lambda distribution. *Computational Statistics & Data Analysis*, 51(9):4484–4496, May 2007.

O. E. Barndorff-Nielsen and K. Prause. Apparent scaling. *Finance and Stochastics*, 5: 103–113, 2001.

T. Bednarski and B. Clarke. Trimmed likelihood estimation of location and scale of the normal distribution. *Australian Journal of Statistics*, 35(2):141–153, 1993.

M. Bigerelle, D. Najjar, B. Fournier, N. Rupin, and A. Iost. Application of lambda distributions and bootstrap analysis to the prediction of fatigue lifetime and confidence intervals. *International Journal of Fatigue*, 28(3):223–236, Mar. 2006.

F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, pages 28–43, 1992.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, Apr. 1986.

T. Bollerslev. Glossary to ARCH (GARCH). Feb. 2009.

A. Bowley. *Elements of Statistics*. Scribner, New York, 1920.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.

M. Broniatowski and A. Keziou. Minimization of $\phi$-divergences on sets of signed measures. *Stud. Sci. Math. Hung.*, 43(4):403–442, 2006.

M. Broniatowski and A. Keziou. Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis*, 100(1):16–36, 2009.

C. Brooks, S. Burke, and G. Persand. Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting*, 17(1):45–56, 2001.

A. Charles and O. Darne. Outliers and GARCH models in financial data. *Economics Letters*, 86(3):347–352, Mar. 2005.

R. C. H. Cheng and N. A. K. Amin. Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):394–403, 1983.

P. Cizek. Greneral trimmed estimation: robust approach to nonlinear and limited dependent variable models. *Econometric Theory*, 24(06):1500–1529, 2008.

C. J. Corrado. Option pricing based on the generalized lambda distribution. *Journal of Futures Markets*, 21(3):213–236, 2001.

N. Cressie and T. R. Read. Multinomial goodness-of-fit tests. *J. R. Stat. Soc., Ser. B*, 46: 440–464, 1984.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hung. Acad. Sci., Ser. A*, 8:85–108, 1963.

H. David and H. Nagaraja. Order statistics. *NJ: John Wiley & Sons*, 7:159–61, 2003.

E. De Giorgi. Reward-risk portfolio selection and stochastic dominance. *Journal of Banking & Finance*, 29(4):895–926, 2005.

B. Dengiz. The generalized lambda distribution in simulation of M/M/1 queue systems. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 3:161–171, 1988.

R. Dimova and N. Neykov. Generalized d-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications. *Theory and applications of recent robust methods*, pages 83–91, 2004.

F. Downton. Linear estimates with polynomial coefficients. *Biometrika*, 53(1/2):129–141, 1966.

E. A. Elamir and A. H. Seheult. Trimmed L-moments. *Computational Statistics & Data Analysis*, 43(3):299–314, 2003.

E. Elton, M. Gruber, and C. Blake. Survivor bias and mutual fund performance. *Review of Financial Studies*, 9(4):1097–1120, 1996.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extreme events for insurance and finance*. Springer, Berlin, 1997.

R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

J. J. Filliben. The probability plot correlation coefficient test for normality. *Technometrics*, 17(1):111–117, Feb. 1975.

P. Fishburn. Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, 67(2):116–126, 1977.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Lond. Phil. Trans. (A)*, 223:1–33, 1922.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

F. Foster and C. Whiteman. An application of Bayesian option pricing to the Soybean market. *American journal of agricultural economics*, 81(3):722–727, 1999.

B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, and A. Iost. Application of the generalized lambda distributions in a statistical process control methodology. *Journal of Process Control*, 16(10):1087–1098, 2006.

B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, and R. Wilcox. Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis*, 51(6):2813–2835, 2007.

D. Freedman and P. Diaconis. On the histogram as a density estimator: $L_2$ theory. *Probability Theory and Related Fields*, 57(4):453–476, Dec. 1981.

M. Freimer, G. Kollia, G. Mudholkar, and C. Lin. A study of the generalized Tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10):3547–3567, 1988.

W. Gilchrist. *Statistical modelling with quantile functions.* CRC Press, 2000.

J. Glattfelder, A. Dupuis, and R. Olsen. Patterns in high-frequency FX data: discovery of 12 empirical scaling laws. *Quantitative Finance*, 11(4):599–614, 2011.

A. Hadi and A. Luceño. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis*, 25(3):251–272, 1997.

M. R. Haley and C. H. Whiteman. Generalized safety first and a new twist on portfolio performance. *Econometric Reviews*, 27(4–6):457–483, 2008.

J. Hastings, Cecil, F. Mosteller, J. W. Tukey, and C. P. Winsor. Low moments for small samples: A comparative study of order statistics. *The Annals of Mathematical Statistics*, 18(3):413–426, 1947.

D. Hogben. Some properties of Tukey's test for non-additivity. Unpublished Ph.D. thesis, Rutgers-The State University, 1963.

L. Hotta and R. Tsay. Outliers in GARCH processes, 1998.

L. Jager and J. A. Wellner. Goodness-of-fit tests via phi-divergences. *Ann. Stat.*, 35(5): 2018–2053, 2007.

A. H. Jessen and T. Mikosch. Regularly varying functions. *Publ. Inst. Math., Nouv. Sér.*, 80(94):171–192, 2006.

B. L. Joiner and J. R. Rosenblatt. Some properties of the range in samples from Tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66 (334):394–399, 1971.

P. Jorion. *Value at risk: the new benchmark for controlling market risk*, volume 2. McGraw-Hill New York, 1997.

Z. Karian and E. Dudewicz. *Fitting statistical distributions: the generalized lambda distribution and generalized bootstrap methods.* Chapman & Hall/CRC, 2000.

Z. Karian and E. Dudewicz. Comparison of GLD fitting methods: superiority of percentile fits to moments in L2 norm. *Journal of Iranian Statistical Society*, 2(2):171–187, 2003.

Z. A. Karian and E. J. Dudewicz. Fitting the generalized lambda distribution to data: a method based on percentiles. *Communications in Statistics - Simulation and Computation*, 28(3):793–819, 1999.

J. Karvanen and A. Nuutinen. Characterizing the generalized lambda distribution by L-moments. *Computational Statistics & Data Analysis*, 52(4):1971–1983, Jan. 2008.

J. Karvanen, J. Eriksson, and V. Koivunen. Maximum likelihood estimation of ICA model for wide class of source distributions. In *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, volume 1, pages 445–454, 2000.

A. Keziou. Dual representation of $\phi$-divergences and applications. *Comptes Rendus Mathematique*, 336(10):857–862, 2003.

C.-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60 (1–2):1–22, 1994.

T. Kim and H. White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73, Mar. 2004.

R. King and H. MacGillivray. Fitting the generalized lambda distribution with location and scale-free shape functionals. *American Journal of Mathematical and Management Sciences*, 27(3-4):441–460, 2007.

R. A. R. King and H. L. MacGillivray. A starship estimation method for the generalized $\lambda$ distributions. *Australian & New Zealand Journal of Statistics*, 41(3):353–374, 1999.

Y. Kitamura and M. Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, July 1997.

C. Kuan. Lecture on the Markov switching model. Institute of Economics Academia Sinica, 2002.

S. Kullback. *Information theory and statistics*. John Wiley and Sons, New York, 1959.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

A. Lakhany and H. Mausser. Estimating the parameters of the generalized lambda distribution. *Algo Research Quarterly*, 3(3):47–58, 2000.

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *Information Theory, IEEE Transactions on*, 52(10):4394–4412, Oct. 2006.

B. Lindsay. Efficiency versus robustness - the case for minimum Hellinger distance and related methods. *Annals of Statistics*, 22(2):1081–1114, June 1994.

A. Luceño. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2):904–917, Nov. 2006.

P. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2, pages 49–55. New Delhi, 1936.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36 (4):394–419, 1963.

M. Markatou. Mixture models, robustness, and the weighted likelihood methodology. *Biometrics*, 56(2):483–486, 2000.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

H. Markowitz. *Portfolio Selection*. John Wiley & Sons, New York, 1959.

B. Mendes. Assessing the bias of maximum likelihood estimates of contaminated GARCH models. *Journal of Statistical Computation and Simulation*, 67(4):359–376, 2000.

A. Meucci. Fully flexible views: theory and practice. *Risk*, 21(10):97–102, 2008.

J. J. A. Moors. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(1):25–32, 1988.

D. Morales, L. Pardo, and I. Vajda. Some new statistics for testing hypotheses in parametric models. *Journal of Multivariate Analysis*, 62(1):137–168, July 1997.

D. Morales, L. Pardo, M. Pardo, and I. Vajda. Limit laws for disparities of spacings. *Journal of Nonparametric Statistics*, 15(3):325–342, June 2003.

T. Morimoto. Markov processes and the H-theorem. *J. Phys. Soc. Jap.*, 18:328–331, 1963.

U. A. Mueller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance*, 14(6):1189–1208, 1990.

N. Muler and V. Yohai. Robust estimates for GARCH models. *Journal of Statistical Planning and Inference*, 138(10):2918–2940, 2008.

C. Müller and N. Neykov. Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference*, 116(2):503–519, 2003.

D. Najjar, M. Bigerelle, C. Lefebvre, and A. Iost. A new approach to predict the pit depth extreme value of a localized corrosion process. *ISIJ international*, 43(5):720–725, 2003.

A. Negiz and A. Çinar. Statistical monitoring of multivariable dynamic processes with state-space models. *AIChE Journal*, 43(8):2002–2020, 1997.

N. Neykov and C. Müller. Breakdown point and computation of trimmed likelihood estimators in generalized linear models. *Developments in robust statistics. Physica-Verlag, Heidelberg*, pages 277–286, 2003.

N. Neykov and P. Neytchev. A robust alternative of the maximum likelihood estimators. *COMPSTAT 1990 - Short Communications*, pages 99–100, 1990.

N. Neykov, P. Filzmoser, R. Dimova, and P. Neytchev. Robust fitting of mixtures using the trimmed likelihood estimator. *Computational Statistics & Data Analysis*, 52(1): 299–308, 2007.

A. Ozturk and R. Dale. A study of fitting the generalized lambda distribution to solar-radiation data. *Journal of Applied Meteorology*, 21(7):995–1004, 1982.

A. Ozturk and R. Dale. Least-squares estimation of the parameters of the generalized lambda-distribution. *Technometrics*, 27(1):81–84, 1985.

S. Pal. Evaluation of nonnormal process capability indices using generalized lambda distribution. *Quality Engineering*, 17(1):77–85, 2004.

L. Pardo. *Statistical Inference Based on Divergence Measures.* Taylor & Francis Group, LLC, 2006.

R. Pyke. Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 395–449, 1965.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

J. S. Ramberg and B. W. Schmeiser. An approximate method for generating asymmetric random variables. *Commun. ACM*, 17(2):78–82, 1974.

J. S. Ramberg, P. R. Tadikamalla, E. J. Dudewicz, and E. F. Mykytka. A probability distribution and its uses in fitting data. *Technometrics*, 21(2):201–214, May 1979.

B. Ranneby. The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 11(2):93–112, 1984.

T. R. Read and N. A. Cressie. *Goodness-of-fit statistics for discrete multivariate data.* New York etc.: Springer-Verlag, 1988.

J. Robertson, E. Tallman, and C. Whiteman. Forecasting using relative entropy. *Journal of Money Credit and Banking*, 37(3):383–401, June 2005.

J. Robinson, E. Ronchetti, and G. Young. Saddlepoint approximations and tests based on multivariate $M$-estimates. *Ann. Stat.*, 31(4):1154–1169, 2003.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):431–449, 1952.

R. Schnidrig and D. Würtz. Investigation of the volatility and autocorrelation function of the USD/DEM exchange rate on operational time scales. In *High Frequency Data in Finance Conference (HFDF-I), Zürich*, 1995.

D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

S. Shapiro, M. Wilk, and H. Chen. A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324):1343–1372, 1968.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(Part 3–4):591–611, Dec. 1965.

H. Shore. Comparison of generalized lambda distribution (GLD) and response modeling methodology (RMM) as general platforms for distribution fitting. *Communications In Statistics-Theory and Methods*, 36(13-16):2805–2819, 2007.

F. Sortino and R. Van Der Meer. Downside risk. *The Journal of Portfolio Management*, 17(4):27–31, 1991.

H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, Mar. 1926.

M. Stutzer. A simple nonparametric approach to derivative security valuation. *Journal of Finance*, 51(5):1633–1652, Dec. 1996.

M. Stutzer. A portfolio performance index. *Financial Analysts Journal*, 56(3):52, 2000.

S. Su. A discretized approach to flexibly fit generalized lambda distributions to data. *Journal of Modern Applied Statistical Methods*, 4(2):408–424, 2005.

S. Su. Numerical maximum log likelihood estimation for generalized lambda distributions. *Computational Statistics & Data Analysis*, 51(8):3983–3998, May 2007.

S. Su. Fitting GLD to data via quantile matching method. In Z. Karian and E. Dudewicz, editors, *Handbook of Fitting Statistical Distributions with R*. CRC Press/Taylor & Francis, 2010.

A. Tarsitano. Fitting the generalized lambda distribution to income data. In *COMPSTAT'2004 Symposium*, pages 1861–1867. Physica-Verlag/Springer, 2004.

A. Tarsitano. Comparing estimation methods for the FPLD. *Journal of Probability and Statistics*, 2010.

A. Toma and M. Broniatowski. Dual divergence estimators and tests: Robustness results. *Journal of Multivariate Analysis*, 102(1):20–36, 2011.

A. Toma and S. Leoni-Aubin. Robust tests based on dual divergence estimators and saddlepoint approximations. *Journal of Multivariate Analysis*, 101(5):1143–1155, 2010.

J. W. Tukey. The practical relationship between the common transformations of percentages or fractions and of amounts. Technical Report Technical Report 36, Statistical Research Group, Princeton, 1960.

J. W. Tukey. The future of data analysis. *The Annals of Mathematical Statistics*, 33(1): 1–67, Mar. 1962.

I. Vajda. *Theory of statistical inference and information*. Dordrecht etc.: Kluwer Academic Publishers, 1989.

J. Van Dyke. Numerical investigation of the random variable y $= c(u^\lambda - (1-u)^\lambda)$. Unpublished working paper, National Bureau of Standards, 1961.

D. Vandev and N. Neykov. *New directions in statistical data analysis and robustness*, chapter Robust maximum likelihood in the gaussian case, pages 257–264. Birkhauser Verlag, 1993.

D. Vandev and N. Neykov. About regression estimators with high breakdown point. *Statistics*, 32(2):111–129, 1998.

M. Yamada and J. Rajasekera. Portfolio re-balancing with the entropy criteria. Technical report, Report No. 310, QUICK Research Institute Corp, 1993.