Royal Institute of Technology

Master Thesis

Spurious Heavy Tails

Author:
Ted Segerfors

Supervisor:
Filip Lindskog

May 26, 2015

Abstract

Since the financial crisis which started in 2007, the risk awareness in the financial sector is greater than ever. Financial institutions such as banks and insurance companies are heavily regulated in order to create a harmonic and resilient global economic environment. Sufficiently large capital buffers may protect institutions from bankruptcy due to some adverse financial events leading to an undesirable outcome for the company. In many regulatory frameworks, the institutions are obliged to estimate high quantiles of their loss distributions. This is relatively unproblematic when large samples of relevant historical data are available. Serious statistical problems appear when only small samples of relevant data are available. One possible solution would be to pool two or more samples that appear to have the same distribution, in order to create a larger sample.

This thesis identifies the advantages and risks of pooling of small samples. For some mixtures of normally distributed samples, with what is considered to be the same variances, the pooled data may indicate heavy tails. Since a finite mixture of normally distributed samples has light tails, this is an example of spurious heavy tails.

Even though two samples may appear to have the same distribution function it is not necessarily better to pool the samples in order to obtain a larger sample size with the aim of more accurate quantile estimation. For two normally distributed samples of sizes m and n and standard deviations s and v, we find that when v/s is approximately 2, $n+m$ is less than 100 and $m/(m+n)$ is approximately 0.75, then there is a considerable risk of believing that the two samples have equal variance and that the pooled sample has heavy tails.

Keywords: Small samples, Tail index estimation, Normal mixture models, Heavy tails

Sammanfattning

Efter den finansiella krisen som hade sin start 2007 har riskmedvetenheten inom den finansiella sektorn ökat. Finansiella institutioner så som banker och försäkringsbolag är noga reglerade och kontrollerade för att skapa en stark och stabil världsekonomi. Genom att banker och försäkringsbolag enligt regelverken måste ha kapitalbuffertar som ska skydda mot konkurser vid oväntade och oönskade händelser skapas en mer harmonisk finansiell marknad. Dessa regelverk som institutionerna måste följa innebär ofta att de ansvariga måste skatta höga kvantiler av institutionens förväntade förlustfunktion. Att skapa en pålitligt modell och sedan skatta höga kvantiler är lätt när det finns mycket relevant data tillgänglig. När det inte finns tillräckligt med historisk data uppkommer statistiska problem. En lösning på problemet är att poola två eller flera grupper av data som ser ut att komma från samma fördelningsfunktion för att på så sätt skapa en större grupp med historisk data tillgänglig.

Detta arbetet går igenom fördelar och risker med att poola data när det inte finns tillräckligt med relevant historisk data för att skapa en pålitlig modell. En viss mix av normalfördelade datagrupper som ser ut att ha samma varians kan uppfattas att komma från tungsvansade fördelningar. Eftersom normalfördelningen inte är en tungsvansad fördelning kan denna missuppfattning skapa problem, detta är ett exempel på falska tunga svansar.

Även fast två datagrupper ser ut att komma från samma fördelningsfunktion så är det inte nödvändigtvis bättre att poola dessa grupper för att skapa ett större urval. För två normalfördelade datagrupper med storlekarna m och n och standardavvikelserna s och v, är det farligaste scenariot när v/s är ungefär 2, $n+m$ är mindre än 100 och $m/(m+n)$ är ungefär 0.75. När detta inträffar finns det en signifikant risk att de två datagrupperna ser ut att komma från samma fördelningsfunktion och att den poolade datan innehar tungsvansade egenskaper.

Acknowledgements

I would like to thank my supervisor Filip Lindskog at the Department of Mathematics at KTH for continuous guidance and valuable support throughout the entire process of the thesis.

Stockholm, May 2015

Ted Segerfors

Contents

Chapter 1

Introduction

One of the main factors of the financial crisis that started in 2007 was the fact that banks engaged in too much risk compared to their capital stock. When a bank became insolvent it resulted in either bankruptcy or a bailout by the government. The problem with the bailouts was that it was such large banks that needed the bailouts which put the governments under a great amount of stress. Ever since the crisis the attitude towards financial risk has completely changed. Today the financial institutions such as banks and insurance companies are heavy regulated by financial authorities, as well as investigated by clients and potential investors. Two of the regulatory frameworks that are widely used today are the Basel III and the Solvency II frameworks. One of the common factors with these frameworks is that the institutions are supposed to estimate high quantiles of their loss distributions. This is easily done with common risk measures such as value at risk (VaR) and expected shortfall (ES) when there is a lot of data present. The problem that occurs is when the risk managers do not have sufficient historic data at hand in order to construct accurate models. One important part is to understand if the underlying distribution of the data set is regularly varying or not, i.e. if it is heavy-tailed. If the data set comes from a heavy-tailed distribution the estimated high quantile is much larger than the same quantile of a light-tailed distribution.

When dealing with small samples the accuracy of an estimated high quantile may be questioned. A solution to the lack of data could be to pool a data set with another that looks like it come from the same distribution. By pooling two or more data sets the sample size increases and a more accurate model can be built. But in order to pool data sets there has to be a good understanding of the sets, if this is done wrong there could be devastating consequences. Some examples of when the risk managers have insufficient data sets and pooling of data could be an option in order to estimate high quantiles could be when dealing with operational risk or fire insurance for industries.

The purpose with this thesis is to understand what can happen when pooling two or more data sets and what can be misinterpreted. What combinations of samples are easy to misinterpret and how should a data pool be composed in order to get an accurate model. It also aims at presenting necessary and helpful tools in order to perform extreme value analysis of pooled data. The importance with this report is to truly understand the underlying distributions and what the result will be when pooling the data. Hopefully this degree project helps banks and insurance companies to get a more accurate estimation of the high quantiles of their loss distributions.

In chapter 2 the theoretical background will be described. It will help the reader to get a deeper understanding of the underlying problem and how the analysis is composed. It contains standard topics used in risk management in the financial industry. In chapter 3 the tail index estimators are presented and analysed. These estimators have different properties and it is important to get a good understanding of these estimators in order to make accurate estimations of the underlying model. In chapter 4 the composition of the normal mixture model is presented. How can a mixture be constructed and what can be difficult when dealing with mixtures of such kinds? In chapter 5 the results of the simulations and analysis is presented. What kind of mixtures of pooled data are easy to misinterpret and how should we pool data in order to make good models? In chapter 6 the conclusion of the thesis is presented.

Chapter 2

Theoretical Background

2.1 Extreme Value Theory

The problem with extreme value theory is that the events that are studied are not the usual events that occur, it is rather the unusual events. However this is not a problem if there are a lot of data available, then the extreme events could be found within the data. The bigger problem with the extremal events modelling is when there is a lack of data, when the data set is small. According to Coles (2000) extreme values are scarce by definition, which means that the levels of the estimates are much greater than those who have already been observed, then one have to extrapolate from observed data to unobserved data.

The work by de Haan and Ferreira (2006) describes the asymptotic theory of sample extremes as a parallel theory to the central limit theory, where they both do have similarities. The central limit theorem describes what happens with the sum $X_1 + X_2 + \cdots + X_n$ as $n \to \infty$ contrary to the asymptotic theory of sample extremes which describes the asymptotic behaviour of $\max(X_1, X_2, \ldots, X_n)$ or $\min(X_1, X_2, \ldots, X_n)$ as $n \to \infty$, where $X_1, X_2, \ldots, X_n$ are independent and identically distributed variables. If you compare two different insurance companies where one of the companies are strictly insuring personal vehicles such as cars, and the other company insures large industries from fires. For the insurance company that insures cars, the sum of all expected future paid out claims is the interesting part since all payments they make are more or less alike in size. For the other company that work with fire insurances for industries, one large fire could mean an insurance claim of a massive amount of money contrary to the "normal" claim sizes.

2.2 Heavy Tails

Heavy tails are present in all sorts of fields, it is known that by 2016 the top 1% of the richest people in the world will own more than the rest of the population. Another example is where people decides to live on the planet. There are great amounts of land which is not populated and there are some relatively small locations with great amounts of people. Within finance and risk management the heavy tails may be represented by volatile log-returns and large insurance claims. According to Hult et al. (2012) there is no definition of "heavy tail", but it is common to consider the right tail of a distribution as heavy if

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{e^{-\lambda x}} = \infty \quad \text{for every } \lambda > 0,$$

where $\bar{F}(x) = 1 - F(x)$, and $F(x)$ is the underlying distribution function. This means that a distribution with a heavier right tail than any exponential distribution is considered as a distribution with a heavy tail. Examples of heavy-tailed distributions would be the Pareto distribution as well as the Student's t distribution. An example of a non heavy-tailed distribution is the Normal distribution. One problem with misinterpreting the data as coming from a heavy-tailed distribution would be that you expect an extremal value as a worst case scenario when in fact there will most likely not occur such a case. Especially when estimating large quantiles of an underlying distribution, it is of great importance that you have an accurate estimation when it comes to if the distribution is heavy-tailed or not.

2.3 Peaks over Threshold Method

Since from a risk managers point of view it is of great importance to understand if the data comes from a heavy-tailed distribution, a suitable model for fitting the data is the peaks over threshold model (POT). This model is a generalized model that analyses the data over a specific high threshold which is of interest when gaining an understanding of the tail of the distribution. Since the focus is on the distribution of the tail it is not logical to include the whole data set, that is why the specific high threshold is set.
Suppose that we have i.i.d. random variables $X_1, \ldots, X_n$ with common unknown distribution function F. Say that we are interested in the excesses over a specific high threshold u, we define the distribution function for the excesses over threshold as

$$F_u(x) = P\{X - u \leq x | X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)},$$

for $0 \leq x < x_0 - u$ where x_0 is the finite or infinite endpoint of the distribution F. From Hult et al. (2012) we know that the distribution of appropriately scaled excesses $X_k - u$ over a high threshold u is typically well approximated by the Generalized Pareto Distribution (GPD). The two parameter GPD has the distribution function

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\sigma})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\sigma) & \text{if } \xi = 0 \end{cases} \tag{2.1}$$

where $\sigma \geq 0$ and $x \geq 0$. This GPD can be grouped in to three classes. When $\xi > 0$ the distribution is a reparametrized regular Pareto distribution. When $\xi < 0$ the distribution is the type II Pareto and when $\xi = 0$ it is called the Gumbel distribution. Take $\xi > 0$ and $\sigma > 0$ then from (2.1) we get

$$G_{\xi,\sigma}(x) = 1 - (1 + \xi x/\sigma)^{-1/\xi}.$$

We now consider X as a random variable with distribution function F that has a regularly varying right tail

$$\lim_{u\to\infty} \frac{\bar{F}(\lambda u)}{\bar{F}(u)} = \lambda^{-\alpha}, \tag{2.2}$$

for all $\lambda > 0$ and some $\alpha > 0$. Where $\bar{F} = 1 - F$. Then we get

$$\lim_{u\to\infty} P\left(\frac{X-u}{u/\alpha} > x \mid X > u\right) = \lim_{u\to\infty} \frac{P(X > u(1 + x/\alpha))}{P(X > u)}$$

$$= (1 + x/\alpha)^{-\alpha} = \bar{G}_{1/\alpha,1}(x).$$

2.4 Inference Theory

One of the most essential parts within inference theory is the development of efficient estimators of model parameters. Boos and Stefanski (2013) mentions that the likelihood of the model along with the maximum likelihood estimation of the parameters of that same model is an approach to statistical inference that applies to a wide variety of problems. One simple example they describe in their work is the maximum likelihood parameter estimation of an independent and identically normally distributed sample $X_1, \ldots, X_n$. The sample comes from the $\mathrm{N}(\mu, \sigma^2)$ family with the density function $f(x; \mu, \sigma)$. Then the likelihood of the i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)^T$ is

$$L(\mu, \sigma | \, \mathbf{X}) = \prod_{i=1}^{n} f(X_i; \mu, \sigma),$$

which gives the maximum likelihood estimation of the parameters μ and σ as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\big)^2.$$

When your model is complete and fully specified the maximum likelihood estimations are efficient and an easy tool to use. One crucial part of the likelihood approach is that the model is accurate. In order to make good inferences from the statistical model one has to fully understand the different parts of the model as well as how these parts handles the analysed data.

2.5 Univariate Normal Distribution

In this degree project mixtures of univariate normal distributions are composed. In order to understand the behaviour of the mixture it is important to know the theory behind the underlying distributions. What normally defines a normally distributed random variable X is its density $f(x)$. The density of the standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}},$$

and the density function of $N(\mu, \sigma)$, where μ is the mean and σ is the standard deviation is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The univariate normal distribution is not considered as a heavy-tailed distribution since it is of exponential order.

2.6 Univariate Student's t Distribution

The student's t distribution is a heavy-tailed distribution which is used in this thesis. We do want to construct a mixture of normally distributed variables that behaves as a heavy-tailed distribution and in this case the student's t distribution. In order to do so we need a theoretical understanding of the distribution. A student's t distributed random variable X has the density function

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

where ν is the degrees of freedom and $\Gamma(\cdot)$ is the gamma function. A student's t distributed random variable X with ν degrees of freedom can also be expressed as

$$X = Z\sqrt{\frac{\nu}{V}},$$

where Z is a standard normal distributed random variable and V has a chi-squared distribution with ν degrees of freedom and is independent of Z.

2.7 Risk Measures

A risk measure is a measure of the riskiness of a position quantified in monetary units. It can be viewed as a buffer of capital that should be kept in reserve in order to act as protection if the investment gives an undesirable return. The financial authorities wants to regulate the risk taken by financial institutions such as banks and insurance companies. By using common risk measures the risk taken by the institutions can be regulated by the authorities. The risk measure do often imply high quantiles of the loss distribution. This is a problem when dealing with small samples which will be dealt with later on. There are two common risk measures that are widely used, value at risk (VaR) and expected shortfall (ES).

2.7.1 Value at Risk

VaR indicates in its most common form, the monetary amount at risk of an investment with a given probability over a certain time period. This is a quantified measure used by investors and risk controllers to measure the risk of a specific asset or a portfolio of assets. According to Hult et al. (2012) the value at risk at level $p \in (0, 1)$ of a portfolio with value X at time 1 is

$$\mathrm{VaR}_p(X) = \min\{m \ : \ \mathrm{P}(mR_0 + X < 0) \leq p\},$$

where R_0 is the percentage return of a risk-free asset. This means that the value at risk of the portfolio value X at time 1 is the least amount of money that should be invested in a risk-free asset at time 0, in order to get a negative return at time 1 with probability less than or equal to p. Value at risk is often criticized because it doesn't take into consideration what happens in the tail beyond the selected confidence level. Ignoring the shape of the tail could be costly because of the fact that huge losses can be hidden in the tail. When using VaR in practice the underlying distribution is not necessarily known, then one uses the empirical VaR which is defined as

$$\widehat{\mathrm{VaR}}_p(X) = L_{[np]+1,n},$$

where $L_{1,n} \geq \cdots \geq L_{n,n}$ are the ordered historical losses.

2.7.2 Expected Shortfall

As an alternative to the VaR measure there is expected shortfall (ES). ES do take the tail beyond the level p into account when calculating the risk. It is defined as follows

$$\mathrm{ES}_p(X) = \frac{1}{p} \int_0^p \mathrm{VaR}_u(X) du.$$

This risk measure is often considered as a superior alternative to VaR because it takes the whole tail of the distribution into account. The empirical expected shortfall is defined as

$$\widehat{\text{ES}}_p(X) = \frac{1}{p}\int_0^p L_{[nu]+1,n}du = \frac{1}{p}\left(\sum_{k=1}^{[np]}\frac{L_{k,n}}{n} + \left(p - \frac{[np]}{n}\right)L_{[np]+1,n}\right).$$

2.8 Risk Regulations

Within the financial industry, institutions manages other institutions' and private persons' money. Banks have been criticized when they have been making to risky investments with other people's money in order to make a huge profit for themselves. To prevent banks and other institutions to risk others money, some regulatory framework has to exist. Standardized guidelines on what capital buffer financial institutions have to set a side when taking risks is not totally new. Already in the 80's the Basel Committee on Banking Supervision (BCBS) published the Basel I framework, which was a set of minimum requirements on capital for banks. These sets of requirements sooner developed in to the Basel II framework. Within the financial industry there are mostly two regulatory frameworks that are in use, Basel III for the banking industry and Solvency II for the insurance industry.

BCBS (2010) states that Basel III's reforms are developed in order to strengthen the global capital market. With stronger capital and liquidity rules the banking sector will more likely absorb financial stress and shocks. Basel III is mostly composed by the issues that was raised after the most recent financial crisis, where the committee believes that a strong and resilient banking sector is the foundation of a global sustainable economic growth. With the Basel III framework banks are free to choose between many different risk measures when quantifying their risks in monetary terms. The most common measure is value at risk since it is easy to use. When calculating risk with the VaR measure with high confidence, a high quantile of the loss distribution has to be estimated.

Solvency II is a framework for the insurance and reinsurance industry in the European Union. It serves as a tool to harmonize the industry. European Commission (2014) states that Solvency II is based on three pillars:

Pillar 1: Harmonised valuation and capital requirements

Pillar 2: Harmonised governance, internal control and risk management requirements

Pillar 3: Harmonised supervisory reporting and public disclosure

Contrary to the Solvency I framework the new solvency requirements are more sophisticated and risk-sensitive. This will contribute to a better coverage of the risk taking, since the model is not a ”one-model-fits-all” way of estimating the size of capital buffers. As for the Basel III framework and other different regulation requirements, the standard risk measure used in Solvency II is VaR.

Chapter 3

Tail Index Estimators

There is a difference when dealing with extremal events contrary to "usual" events. As described in the previous chapter the extremal events are considered unusual or scarce. One important part when dealing with extremal events are the understanding of the data. Embrechts et al. (2001) stresses the importance of looking at graphs and plots of the data and your estimations in order to get a greater understanding of the underlying problem. In this chapter the methods that are used later on in this thesis are described. In extreme value theory there are many different methods that can be used in many different ways, here we have narrowed it down to the most relevant for our case.

In this section the different methods used in this paper when fitting a GPD to simulated data are described.

3.1 Maximum Likelihood Estimation

Consider the excesses over threshold (u), $X_1, \dots, X_n$, as independent variables having the Generalized Pareto Distribution with $\xi \neq 0$. Then the log-likelihood for the GPD parameters is

$$l(\xi, \sigma) = -n \log \sigma - (1 + 1/\xi) \sum_{i=1}^{n} \log \big[1 + \xi \frac{x_i}{\sigma}\big],$$

provided that

$$1 + \xi \frac{x_i}{\sigma} > 0 \quad \text{for } i = 1, \dots, n.$$

With $\xi = 0$ the log-likelihood equation is

$$l(\sigma) = -n \log \sigma - \sum_{i=1}^{n} \frac{x_i}{\sigma}.$$

Maximization of the log-likelihood function leads to the maximum likelihood estimation of the parameters for the GPD function. This is not done analytically, but for a given data set $\{x_1, \ldots, x_n\}$ this is easily done numerically.

3.1.1 Large Sample Asymptotics

One of the benefits with the Maximum Likelihood parameter estimation is the wide applicability for many different sampling distributions. For the Maximum Likelihood method approximations of standard errors and confidence intervals can be calculated. As the sample size n increases to infinity these results can be strictly determined. From Coles (2000) we get the following Theorem.

Theorem 3.1 *Let $x_1, \ldots, x_n$ be independent realizations from a distribution with a parametric family $\mathcal{F}$, and let $l(\cdot)$ and $\tilde{\theta}_0$ denote respectively the log-likelihood function and the maximum likelihood estimator of the d-dimensional model parameter θ_0. Then, under suitable regulatory conditions, for large n*

$$\tilde{\theta}_0 \sim MVN_d(\theta_0, I_E(\theta_0)^{-1}),$$

where

$$I_E(\theta) = \begin{bmatrix} e_{1,1}(\theta) & \cdots & & e_{1,d}(\theta) \\ \vdots & \ddots & e_{i,j}(\theta) & \vdots \\ & e_{j,i}(\theta) & \ddots & \\ e_{d,1}(\theta) & \cdots & & e_{d,d}(\theta) \end{bmatrix},$$

with

$$e_{i,j}(\theta) = E\left\{ -\frac{\partial^2}{\partial\theta_i \partial\theta_j} l(\theta) \right\}.$$

□

The matrix I_E is called the expected information matrix. Now the approximated confidence intervals for $\theta_0 = (\theta_1, \ldots, \theta_d)$ when n is large follows a normal distribution

$$\theta_i \sim N(\theta_i, \psi_{i,i}),$$

where $\psi_{i,i}$ is the i:th component in the diagonal of the inverted expected information matrix. By the known behaviour of the normal distribution we know that the approximated $(1-\alpha)$ confidence interval of θ_i is

$$\hat{\theta}_i \pm z_{\frac{\alpha}{2}} \sqrt{\psi_{i,i}},$$

where $z_{\frac{\alpha}{2}}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. Since the values of θ_0 often are unknown Coles (2000) presents the observed information matrix I_O as a complement to the expected information matrix. This observation matrix is formed as

$$I_O(\theta) = \begin{bmatrix} -\frac{\partial^2}{\partial\theta_1^2}l(\theta) & \cdots & & -\frac{\partial^2}{\partial\theta_1\partial\theta_d}l(\theta) \\ \vdots & \ddots & -\frac{\partial^2}{\partial\theta_i\partial\theta_j}l(\theta) & \vdots \\ & -\frac{\partial^2}{\partial\theta_j\partial\theta_i}l(\theta) & \ddots & \\ -\frac{\partial^2}{\partial\theta_d\partial\theta_1}l(\theta) & \cdots & & -\frac{\partial^2}{\partial\theta_d^2}l(\theta) \end{bmatrix}.$$

With the observed information matrix evaluated at $\theta = \hat{\theta}$ the elements of the inverted information matrix is denoted by $\bar{\psi}_{i,j}$. Then the approximated $(1-\alpha)$ confidence interval for θ_i is

$$\hat{\theta}_i \pm z_{\frac{\alpha}{2}}\sqrt{\bar{\psi}_{i,i}}.$$

By using this for the two parameter generalized pareto distribution Hosking and Wallis (1987) shows that

$$\text{var}\begin{bmatrix}\hat{\sigma} \\ \hat{\xi}\end{bmatrix} \sim \frac{1}{n}\begin{bmatrix} 2\sigma^2(1+\xi) & \sigma(1+\xi) \\ \sigma(1+\xi) & (1+\xi)^2 \end{bmatrix}, \quad \xi > -1/2.$$

3.1.2 Threshold Selection

When performing extreme value analysis through the peaks over threshold method, the threshold selection is important in order to achieve reliable results. In the previous section it was shown that the maximum likelihood estimator converges at the rate n^{-1}, but this is only true for a suitable choice of the threshold u. Here we study how the maximum likelihood estimator behaves for different choices of thresholds. The first approach is to pick the number of exceedances k as $k(n) = n^{\alpha}$, where $\alpha \in (0,1)$. To the left in Figure 3.1 the mean of the estimated tail index of 1,000 samples where the sample size n increases from 40 to 1,000 is shown. The data comes from Student's t distributed samples with $\nu = 3$, which corresponds to a true tail index of $\xi = 1/3$. To the right in Figure 3.1 the the tail index is estimated through the same method but here $k(n) = \beta n$.

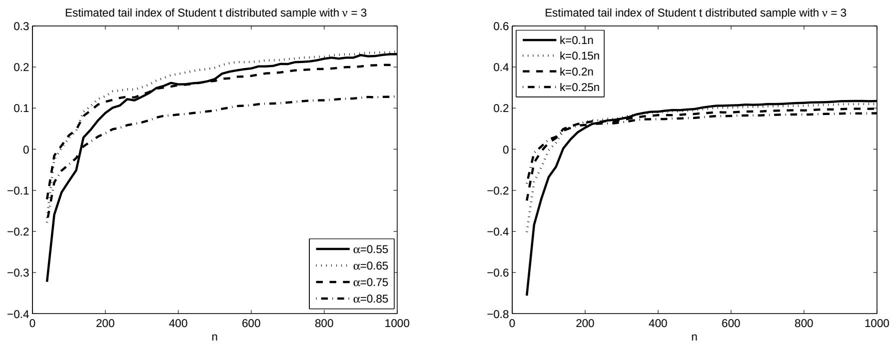


Figure 3.1: *Mean of 1,000 estimations of the tail index with the maximum likelihood estimator for Student's t distributed samples with threshold selection as* Left: $k(n) = n^{\alpha}$ *and* Right: $k(n) = \beta n$.

As we can see in Figure 3.1 the different threshold selections do not differ much except for the selection where $\alpha = 0.85$.

3.1.3 Small Sample Behaviour

In order to understand the estimators better when small sample sizes are considered we simulate samples from a heavy-tailed distribution and increase the sample size and investigate how the bias ($E[\hat{\xi} - \xi]$) behaves and how the mean squared error ($E[(\hat{\xi} - \xi)^2]$) behaves. As the heavy-tailed distribution we choose the Student's t-distribution where we vary the degrees of freedom as well as the sample size. In order to determine the bias and MSE of the MLE we first have to choose the optimal threshold u. From previous section we know that the estimated tail index is not sensitive to the threshold selection for a small sample. We choose the threshold u as the top 20% of the data, which corresponds to $k(n) = 0.2n$. Then we use maximum likelihood estimation in order to estimate ξ from the excesses over the specific threshold u. Here we use MATLAB's built in function *gpfit* which is a maximum likelihood estimator. The bias and MSE of the maximum likelihood estimation is shown in Figure 3.2. Here the Bias is considered as

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\xi} - \xi),$$

and the MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\xi} - \xi)^2.$$

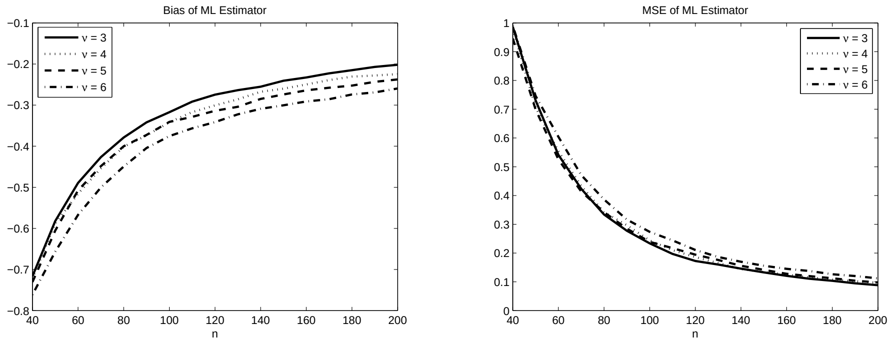


Figure 3.2: *Plots showing the bias and MSE of the estimated ξ's by the Maximum Likelihood. Data comes from Student's t-distribution with ν degrees of freedom. Threshold set as* 20%. *Left: Bias of MLE. Right: MSE of MLE.*

From the plots we can tell that the estimated bias of the maximum likelihood estimation is large for small samples of the Student's t distribution. To further investigate the bias of the estimator the density plots of $\hat{\xi}$ for the Student t distribution with $\nu = 3$ degrees of freedom are shown in Figure 3.3. We can tell that for small $n < 90$ there actually is a large estimated bias (the small bump at around -1.2 do contribute much). As n increases $\hat{\xi}$ moves towards a more stable value but still centred left of the true value ($\xi = 1/3$). As presented earlier one of the criteria of the maximum likelihood estimator is that $\xi > -0.5$. This must be taken into consideration when dealing with small samples and small tail indices as we can see from Figure 3.3.

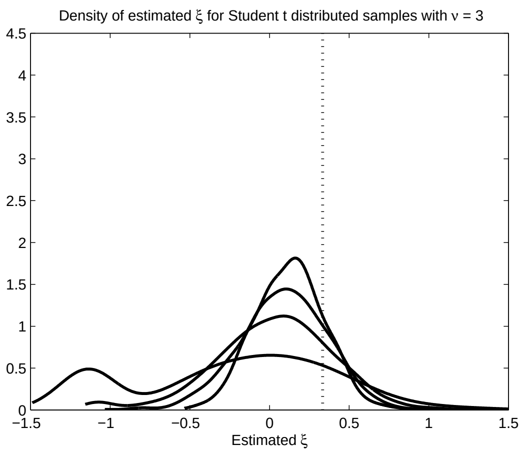


Figure 3.3: *Density plots of $\hat{\xi}$ estimated by the ML-method for* $50 \leq n \leq 200$ *from 1000 samples of Student's t distribution with* $\nu = 3$ *degrees of freedom, where the dashed line represent the true* $\xi = 1/3$.

3.2 Method of Moments

Suppose we have a sample from the two-parameter GPD

$$G_{\xi,\sigma}(x) = 1 - (1 + \xi \frac{x}{\sigma})^{-1/\xi}.$$

From Migdalas et al. (2013) we get that the moments of the random variable X is

$$E\left[\left(1 + \xi \frac{X}{\sigma}\right)^r\right] = \frac{1}{1 - r\xi}, \quad 1 - r\xi > 0,$$

which leads to

$$E(X^r) = r! \frac{\sigma^r}{\xi^{r+1}} \frac{\Gamma\left(\frac{1}{\xi} - r\right)}{\Gamma\left(1 + \frac{1}{\xi}\right)}, \quad \xi < \frac{1}{r},$$

where $\Gamma(.)$ is the Gamma function. Then the mean and variance are

$$E(X) = \frac{\sigma}{1 - \xi}, \quad \xi < 1,$$

and

$$\text{Var}(X) = \frac{\sigma^2}{(1 - \xi)^2 (1 - 2\xi)}, \quad \xi < \frac{1}{2}.$$

If we simulate a sample from the two-parameter GPD with parameters (ξ, σ) and let $\bar{x}$ be the sample mean and s^2 be the sample variance. Then by replacing the mean and variance by $\bar{x}$ and s^2 the Method of Moments (MOM) estimators are

$$\hat{\xi}_{MOM} = -\frac{1}{2}\left(\frac{\bar{x}^2}{s^2} - 1\right) \quad \text{and} \quad \hat{\sigma}_{MOM} = \frac{1}{2}\bar{x}\left(\frac{\bar{x}^2}{s^2} + 1\right).$$

With the sample mean and the sample variance explicitly expressed for the estimators we get

$$\hat{\xi}_{MOM} = -\frac{1}{2}\left(\frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_i\right)^2} - 1\right),$$

and

$$\hat{\sigma}_{MOM} = \frac{1}{2n}\sum_{i=1}^n x_i \times \left(\frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_i\right)^2} + 1\right).$$

3.2.1 Large Sample Asymptotics

Hosking and Wallis (1987) shows that the MOM estimators are asymptotically stable for $\xi < 1/4$ (when there exists a fourth moment) and normally distributed with variances

$$\text{var}\begin{bmatrix}\hat{\sigma}\\ \hat{\xi}\end{bmatrix} \sim \frac{1}{n}\frac{(1-\xi)^2}{(1-2\xi)(1-3\xi)(1-4\xi)} \times \begin{bmatrix} 2\sigma^2(1-6\xi-12\xi^2) & \sigma(1-2\xi)(1-4\xi-12\xi^2) \\ \sigma(1-2\xi)(1-4\xi-12\xi^2) & (1-2\xi)^2(1-\xi-6\xi^2)\end{bmatrix}.$$

3.2.2 Threshold Selection

As for the maximum likelihood estimator the threshold selection is a part of the extreme value analysis for the method of moments estimator. The same analysis is performed for this estimator as for the MLE. The results are shown in Figure 3.4.

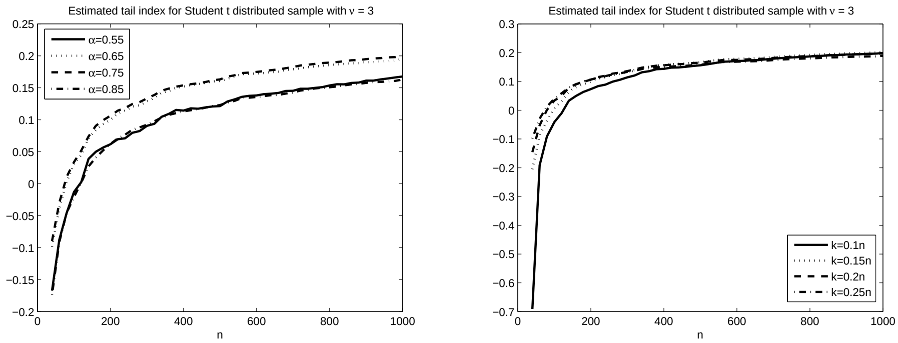


Figure 3.4: *Mean of 1,000 estimations of the tail index with the method of moments estimator for Student's t distributed samples with threshold selection as* Left: $k(n) = n^{\alpha}$ *and* Right: $k(n) = \beta n$.

From Figure 3.4 we can see that the threshold selections that perform best is when $\alpha = 0.65,\ \alpha = 0.75,\ \beta = 0.15,\ \beta = 0.2,\ \beta = 0.25$.

3.2.3 Small Sample Behaviour

We do the same analysis as for the MLE. We choose the same threshold as before and look at the bias and MSE of the MOM-estimator when the sample size n is increased. We use the same data as for the maximum likelihood estimation and only the excesses are considered. The result is shown in Figure 3.5.

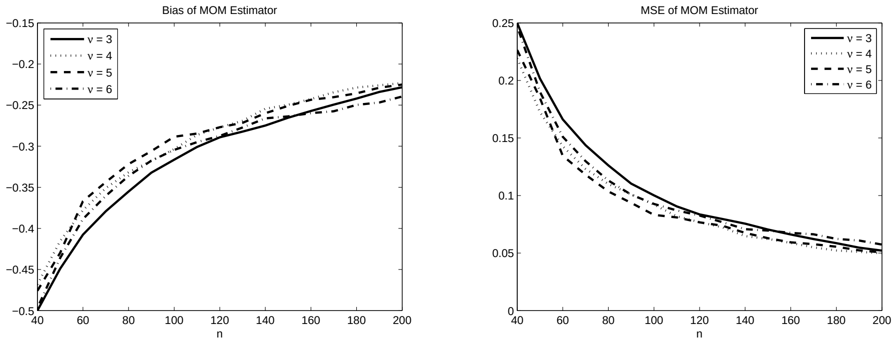


Figure 3.5: *Plots showing the bias and MSE of the estimated ξ's by the Method of Moments method. Data comes from Student's t-distribution with ν degrees of freedom. Threshold set as 20%. Left: Bias of MOME. Right: MSE of MOME.*

As we did for the MLE we plot the densities of $\hat{\xi}$ for 1000 samples of the Student's t distribution as the sample size n increases from 50 to 200. The results are plotted in Figure 3.6. By the density plots we can see that the estimated bias is negative as the previous plots show. As the sample size increases the mean of the density moves towards a certain value, but for $n = 200$ it is still not around the true value ($\xi = 1/3$).

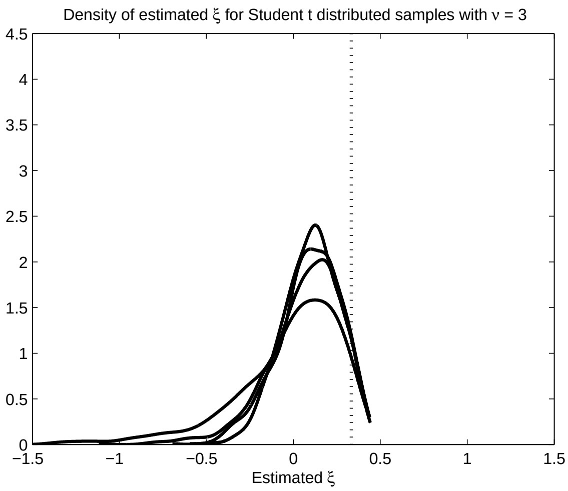


Figure 3.6: *Density plots of $\hat{\xi}$ estimated by the MOM-method for* $50 \leq n \leq 200$ *from 1000 samples of Student's t distribution with* $\nu = 3$ *degrees of freedom, where the dashed line represent the true* $\xi = 1/3$.

3.3 Hill's Estimator

Suppose we have $X_1, \ldots, X_n$ iid with distribution function $F_X(x)$, $x \geq u > 0$. We sort our simulated sample $x_{1,n} \geq \cdots \geq x_{n,n}$. Then the Hill estimator is

$$\hat{\alpha}^{(H)} = \hat{\alpha}^{(H)}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \ln x_{j,n} - \ln x_{k,n} \right)^{-1},$$

where $\hat{\xi} = (\hat{\alpha}^{(H)})^{-1}$. The difficulty with Hill's estimator is the choice of k (number of exceedances of the threshold). Dacorogna et al. (1995) shows that the asymptotic properties of the Hill estimator can be developed for the following class of distributions

$$F(x) = 1 - ax^{-\alpha}(1 + bx^{-\beta}), \tag{3.1}$$

where $\alpha, \beta > 0$ and $a, b \in \mathbb{R}$. Note that this class of distributions satisfies condition (2.2). According to Dacorogna et al. (1995) the asymptotic expected value of the Hill estimator for this class of distributions for a given number of exceedances, k, can be approximated by

$$E(\hat{\xi}(k)) \approx \frac{1}{\alpha} - \frac{b\beta}{\alpha(\alpha+\beta)} a^{-\frac{\beta}{\alpha}} \left(\frac{k}{n}\right)^{\frac{\beta}{\alpha}}, \tag{3.2}$$

and the variance

$$\mathrm{Var}(\hat{\xi}(k)) \approx \frac{1}{k\alpha^2}. \tag{3.3}$$

Lets say that we have a sample $\{x_1, \ldots, x_{100}\}$ from the class of distributions in (3.1), where $\alpha = \beta = 4$ and $a = b = 1$. We calculate $E[\hat{\xi}(k)]\hat{\alpha}^{(H)}_k$ and plot if for $k = 1, \ldots, 20$. This is shown in Figure 3.7.

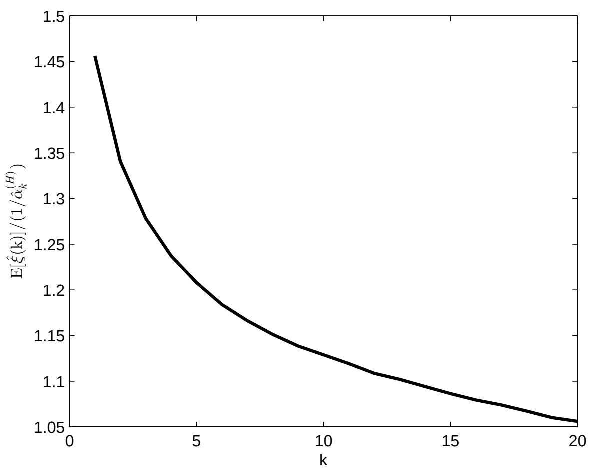


Figure 3.7: *The quotient $E[\hat{\xi}(k)]\hat{\alpha}^{(H)}_k$ for a sample of size 100 as a function of number of exceedances for the distribution $F(x) = 1 - x^{-4}(1 + x^{-4})$.*

From (3.2) and (3.3) we can tell that a small k is preferable in an unbiasedness perspective, and a large k is preferable in a minimize variance perspective. There has to be a trade-off between these two perspectives. As Huisman et al. (2012) we impose the restriction $\alpha = \beta$, which implicitly makes the bias linear in k. This restriction is only true for the limiting Extreme Value Distribution and not for small samples. According to Dacorogna et al. (1995) for the Student's t distribution α equals the number of degrees of freedom and $\beta = 2$. But after experimental simulations Dacorogna et al. (1995) shows that the tail-index estimations are not very sensitive for different choices of β. This makes the assumption $\alpha = \beta$ valid. By

$$\hat{\xi}(k) = E[\hat{\xi}(k)] - \sqrt{\mathrm{Var}\big(\hat{\xi}(k)\big)}\frac{\hat{\xi}(k) - E[\hat{\xi}(k)]}{\sqrt{\mathrm{Var}\big(\hat{\xi}(k)\big)}}, \tag{3.4}$$

equations (3.2)-(3.4) gives

$$\hat{\xi}(k) = c_0 + c_1 k + \frac{c_2}{\sqrt{k}}\epsilon(k), \tag{3.5}$$

where $k = 1, \dots, K$. Equation (3.5) can be composed on matrix form as

$$\hat{\xi} = ZC + V\epsilon, \tag{3.6}$$

Where Z is a $(K \times 2)$ matrix with ones in the first column and the k-values in the second. V is a $(K \times 1)$ vector with $\sqrt{\mathrm{Var}\big(\hat{\xi}(k)\big)}$ as its elements. This equation can be solved by the least squares method. But since the variance in equation (3.3) is inversely related to k the error term $\sqrt{\mathrm{Var}\big(\hat{\xi}(k)\big)}\epsilon(k)$ is heteroscedastic. By multiplying with a $(K \times K)$ weighting matrix W, which has $\sqrt{1}, \dots, \sqrt{K}$ in the diagonal and zeros elsewhere, equation (3.6) can be written as

$$\hat{\xi}^m = (Z'W'WZ)^{-1}Z'W'W\hat{\xi},$$

where the first element in $\hat{\xi}^m$ is the estimated tail-index. This modified Hill estimator can also be viewed as the weighted average of the traditional Hill estimators for $k = 1, \dots, K$

$$\hat{\xi}^m_{1,1}(K) = \sum_{k=1}^{K} w(k)\hat{\xi}(k).$$

After some matrix algebra the weights $w(k)$ can be determined as

$$w(k) = \frac{k\sum_{j=1}^{K} j^3 - k^2\sum_{j=1}^{K} j^2}{\sum_{j=1}^{K} j\sum_{j=1}^{K} j^3 - \Big(\sum_{j=1}^{K} j^2\Big)^2}.$$

3.3.1 Large Sample Asymptotics

According to Deheuvels et al. (1988) one reason for choosing the Hill estimator is that the estimate $\hat{\xi}$ is close to optimal when it comes to minimizing the asymptotic mean squared error. They also tells us that for any $0 < \xi < \infty$ the following statements are equivalent:

(i) The distribution function F has an upper tail of form

$$1 - F(x) = P(X_1 > x) = x^{-1/\xi} l(x) \quad \text{for} \quad x > 0,$$

where $l(x)$ is a slowly varying function at infinity and $\{X_1, \ldots, X_n\}$ is a sequence of i.i.d. with distribution function F.

(ii) For all sequences satisfying

$$1 \leq k_n \leq n - 1, \quad k_n \to \infty \quad \text{and} \quad n^{-1} k_n \to 0 \quad \text{as} \quad n \to \infty,$$

$\hat{\xi}_n \xrightarrow{p} \xi$ as $n \to \infty$.

(iii) For any sequence of the form $k_n = [n^a]$, (with $0 < a < 1$, and where $[x]$ denotes the integer part of x)
$\hat{\xi}_n \xrightarrow{a.s.} \xi$ as $n \to \infty$.

After further investigations it can be shown, under certain regulatory conditions on F and k_n that

$$k_n^{1/2} \hat{\xi} \xrightarrow{d} N(\xi, \xi^2) \quad \text{as} \quad n \to \infty.$$

3.3.2 Threshold Selection

In this section we investigate for which threshold u the Hill estimator and the modified Hill estimator gives the best results. We conduct the same simulations as for the previous estimators and the results are shown in Figure 3.8 and Figure 3.9.

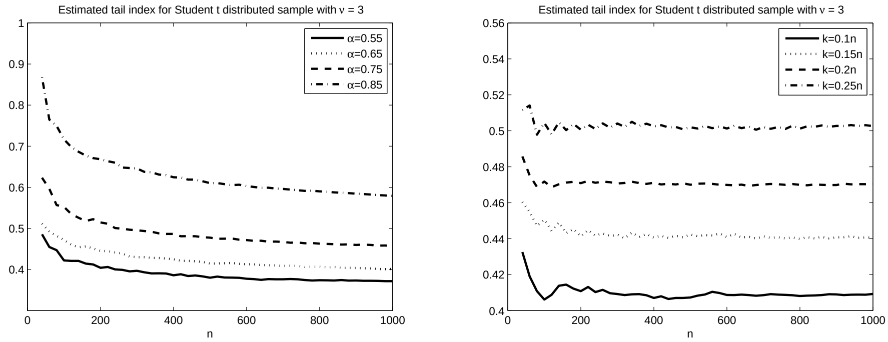


Figure 3.8: *Mean of 1,000 estimations of the tail index with the Hill estimator for Student's t distributed samples with threshold selection as* Left: $k(n) = n^{\alpha}$ *and* Right: $k(n) = \beta n$.

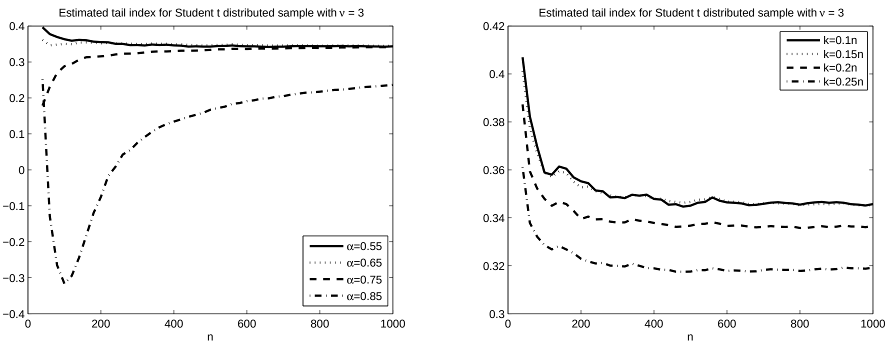


Figure 3.9: *Mean of 1,000 estimations of the tail index with the modified Hill estimator for Student's t distributed samples with threshold selection as* Left: $k(n) = n^{\alpha}$ *and* Right: $k(n) = \beta n$.

As the graphs shows the Hill estimator do behave better for a higher threshold which is the same as a lower amount of exceedances. The modified Hill estimator do perform well for small samples for all the different threshold selections except for $k(n) = n^{0.85}$.

3.3.3 Small Sample Behaviour

The same analysis for the small samples is done for the Hill estimator. In order to compare Hill's estimator with the modified Hill's estimator which has a weighted average of the number of exceedances we set the threshold for the standard Hill's estimator as the top 10% of the ordered data, which we could see in the previous section is the best choice. The bias and MSE of the standard Hill estimator is shown in Figure 3.10 and for the modified Hill estimator in Figure 3.11.

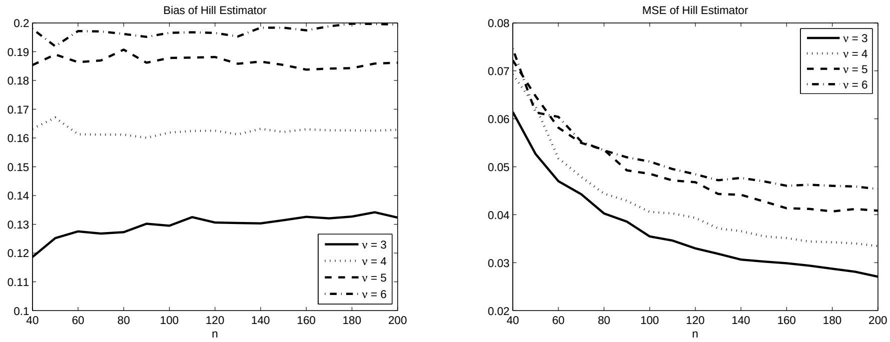


Figure 3.10: *Plots showing the bias and MSE of the estimated ξ's by the Hill estimator. Data comes from Student's t-distribution with ν degrees of freedom. Threshold set as* 10%. *Left: Bias of Hill estimator. Right: MSE of Hill estimator.*

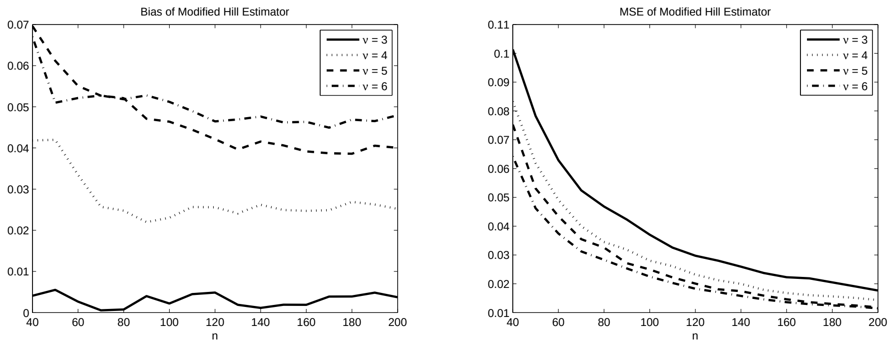


Figure 3.11: *Plots showing the bias and MSE of the estimated ξ's by the modified Hill estimator. Data comes from Student's t-distribution with ν degrees of freedom. Threshold set as* 20%. *Left: Bias of modified Hill estimator. Right: MSE of modified Hill estimator.*

The densities of the estimators are shown in Figure 3.12.

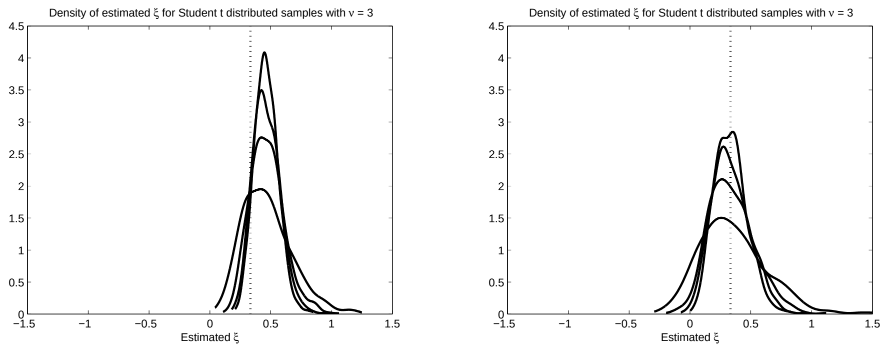


Figure 3.12: *Density plots of $\hat{\xi}$ estimated by the Hill estimator (left) and the modified Hill estimator (right), for $50 \leq n \leq 200$ from 1000 samples of Student's t distribution with $\nu = 3$ degrees of freedom, where the dashed line represent the true $\xi = 1/3$.*

As we can see from the results the modified Hill estimator do perform better for the samples. The density of the modified Hill estimator is centred around the true value while the regular Hill estimator is right skewed and centred slightly to the right of the true value, although it performs well for larger samples.

3.4 The QQ-Estimator

Suppose we have $X_{1,n} \leq \cdots \leq X_{n,n}$ iid with distribution function G from the Pareto family with shape index $\alpha > 0$, and location parameter μ. We know that the pareto distribution function is

$$F_\alpha(x) = 1 - x^{-\alpha}, \quad x \geq 1.$$

From Kratz and Resnick (1996) we know that since we have data from a location-scale family

$$G_{\mu,\sigma}(x) = G_{0,1}\big(\frac{x-\mu}{\sigma}\big)$$

where μ and σ are unknown. Then for $y > 0$

$$G_{0,\alpha}(y) := P\big(\log X_1 > y\big) = e^{-\alpha y}.$$

We have that $G_{0,1}^{\leftarrow}\big(\frac{i}{n+1}\big)$ is the theoretical quantile and that $X_{i,n}$ is the empirical quantile. Then the plot

$$\{(G_{0,1}^{\leftarrow}\big(\frac{i}{n+1}\big), \log X_{i,n}), 1 \leq i \leq n\} = \{(-\log\big(1 - \frac{i}{n+1}\big), \log X_{i,n}), 1 \leq i \leq n\}$$

should approximately be linear with the slope α^{-1}. If we set

$$y_i = -\log\left(1 - \frac{i}{n+1}\right),$$

and

$$z_i = \log X_{i,n},$$

then the slope of the least squares line is

$$SL(\{(y_i, z_i), 1 \leq i \leq n\}) = \frac{\bar{S}_{yz} - \bar{y}\bar{z}}{\bar{S}_{yy} - \bar{y}^2},$$

where

$$\bar{S}_{yz} = \frac{1}{n}\sum_{i=1}^{n} y_i z_i, \quad \bar{S}_{yy} = \frac{1}{n}\sum_{i=1}^{n} y_i^2,$$

and

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

This leads to the estimated slope of

$$\widehat{\alpha^{-1}} = \frac{\sum_{i=1}^{n} -\log\left(\frac{i}{n+1}\right)\{n\log X_{n-i+1,n} - \sum_{j=1}^{n}\log X_{n-j+1,n}\}}{n\sum_{i=1}^{n}\left(-\log\left(\frac{i}{n+1}\right)\right)^2 - \left(\sum_{i=1}^{n} -\log\left(\frac{i}{n+1}\right)\right)^2}. \tag{3.7}$$

But since we are interested in the right tail and not the center of the distribution this is not necessarily very accurate. If we instead suppose that we have a sample $Z_1, \ldots, Z_n$ from a distribution function F with a regularly varying right tail i.e. for large t

$$\frac{1 - F(tx)}{1 - F(x)} \approx x^{-\alpha}.$$

We want to look at the k largest Z's which are excesses over a threshold, u, this gives the qq-plot as

$$\left(-\log(1 - \frac{i}{k+1}), \log(\frac{Z_{n-k+i}}{Z_{n-k}})\right) \quad \text{for } i = 1, \ldots, k.$$

Then the new estimated tail index is

$$\widehat{\alpha^{-1}} = SL(\{(-\log(1 - \frac{i}{k+1}), \log(\frac{Z_{n-k+i}}{Z_{n-k}})), 1 \leq i \leq k\}).$$

Since Z_{n-k} is constant we can write

$$\widehat{\alpha^{-1}} = SL(\{(-\log(1-\frac{i}{k+1}), \log Z_{n-k+i}), 1 \le i \le k\}).$$

We can now rewrite equation (3.7) as the dynamic qq-estimator described by Kratz and Resnick (1996)

$$\widehat{\alpha^{-1}} = \frac{\frac{1}{k}\sum_{i=1}^{k}\big(-\log\big(1-\frac{i}{k+1}\big)\big)\log\big(\frac{Z_{n-k+i}}{Z_{n-k}}\big) - \frac{1}{k}\sum_{i=1}^{k}\big(-\log\big(1-\frac{i}{k+1}\big)\big)H_k}{\frac{1}{k}\sum_{i=1}^{k}\big(-\log\big(1-\frac{i}{k+1}\big)\big)^2 - \big(\frac{1}{k}\sum_{i=1}^{k}-\log\big(1-\frac{i}{k+1}\big)\big)^2},$$

where H_k is the Hill estimator

$$H_k = \frac{1}{k}\sum_{i=1}^{k}\log\big(\frac{Z_{n-k+i}}{Z_{n-k}}\big).$$

3.4.1 Large Sample Asymptotics

From Kratz and Resnick (1996) we get that the dynamic qq-estimator can be expressed as

$$\widehat{\alpha^{-1}} = \frac{\sum_{i=1}^{k}-\log\big(\frac{i}{k+1}\big)\Big\{k\log(Z_{n-i+1,n}) - \sum_{j=1}^{k}\log(Z_{n-j+1,n})\Big\}}{k\sum_{i=1}^{k}\big(-\log(\frac{i}{k+1})\big)^2 - \Bigg(\sum_{i=1}^{k}-\log(\frac{i}{k+1})\Bigg)^2}.$$

We use the non-decreasing function

$$U(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t), \quad t > 0,$$

which we suppose that the second order regular variation condition holds for. We set $\gamma = \alpha^{-1}$, and suppose that there is a $\rho \le 0$ and a function $0 < A(t) \to 0$ s.t. for all $x > 1$

$$\frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} \to cx^{\gamma}\left(\frac{x^{\rho}-1}{\rho}\right) \quad (t \to \infty), \tag{3.8}$$

for some real number c. The function $A(\cdot)$ is regularly varying of index ρ, and $U(\cdot)$ is regularly varying of index γ. We need a restriction on the growth of the number of exceedances as n increases. This restriction is

$$k \to \infty, \quad k/n \to 0, \quad \sqrt{k}A(n/k) \to 0 \quad \text{as} \quad n \to \infty. \tag{3.9}$$

Kratz and Resnick (1996) proofs that if 3.8 and 3.9 holds, then

$$\sqrt{k}\widehat{\alpha^{-1}} \xrightarrow{d} N(\alpha^{-1}, 2\alpha^{-2}) \quad \text{as} \quad n \to \infty.$$

3.4.2 Threshold Selection

As for the previous estimators we have to choose a threshold for the ordered data in order for the qq-estimator to estimate the tail index of the data. We conduct the same simulations and analysis as before, the results are shown in Figure 3.13.

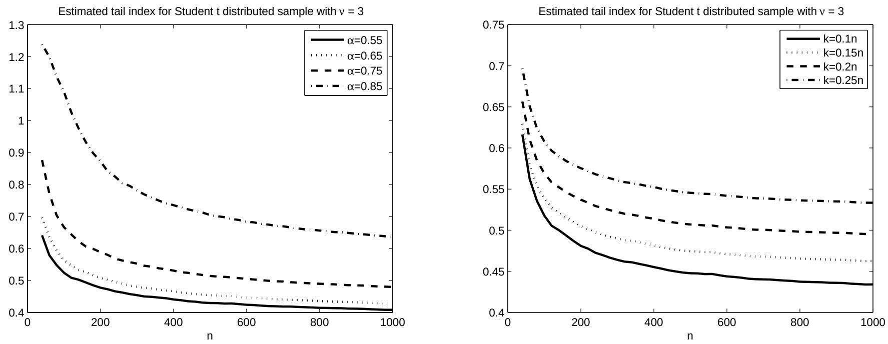


Figure 3.13: *Mean of 1,000 estimations of the tail index with the qq estimator for Student's t distributed samples with threshold selection as* Left: $k(n) = n^{\alpha}$ *and* Right: $k(n) = \beta n$.

From Figure 3.13 it looks like the best performing threshold selection corresponds to $k(n) = n^{0.55}$ and $k(n) = 0.1n$.

3.4.3 Small Sample Behaviour

The same analysis as for the previous estimators is done, with the threshold set as 10% and the sample size, n, is increased from 40 to 200. The result is shown in Figure 3.14.

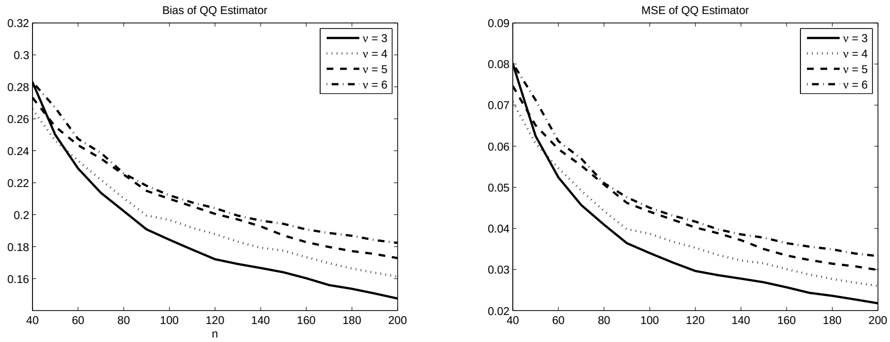


Figure 3.14: *Plots showing the bias and MSE of the estimated ξ's by the dynamic qq-estimator. Data comes from Student's t-distribution. Threshold set as* 20%. *Left: Bias of dynamic qq-estimator. Right: MSE of dynamic qq-estimator.*

The density of the qq-estimated $\hat{\xi}$ is shown in Figure 3.15. As we can tell the qq-estimator performs poor for the small samples as it has a large estimated bias.

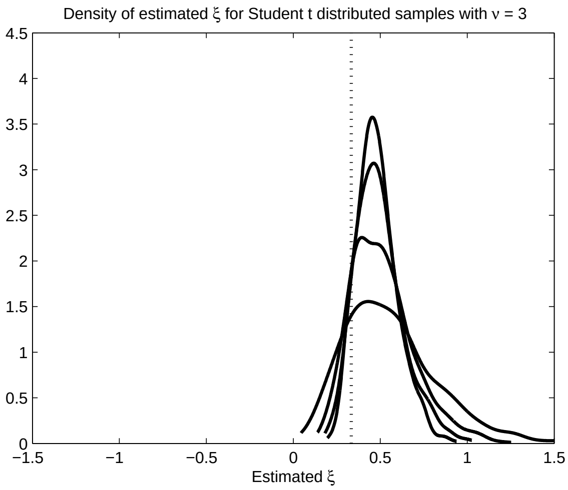


Figure 3.15: *Density plots of $\hat{\xi}$ estimated by the qq-method for* $50 \leq n \leq 200$ *from 1000 samples of Student's t distribution with* $\nu = 3$ *degrees of freedom, where the dashed line represent the true* $\xi = 1/3$.

3.5 Comparison of Estimators Asymptotics

During this chapter the different estimators used in this thesis have been presented. Some are more alike than others and some are constructed in a different way. The similarity of the estimators is that the variance of the tail-index estimates are all converging at the same rate (n^{-1}). By converging at the same rate one might think that they all behave similarly for a large sample, but this is not the truth. The importance for a large sample is the selection of the threshold u, the optimal selection might be different for different estimators. By simulating 1,000 samples of size 50,000 from a Student's t distribution with $\nu = 5$ degrees of freedom, we can see how the estimators behave for a large sample size. The results are shown in Figure 3.16.

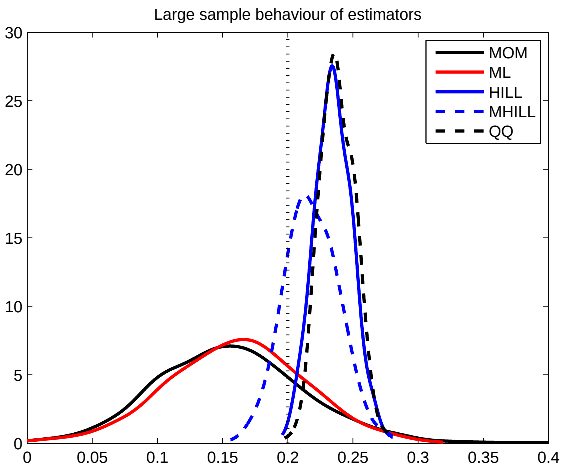


Figure 3.16: *Plot showing the densities of the different estimators for 1,000 samples of size 50,000 of a Student's t distribution with* $\nu = 5$ *degrees of freedom.*

It looks like the Hill estimator, the modified Hill estimator and the dynamic qq-estimator has a smaller variance than the MOM-estimator and the MLE. The most accurate estimator of the five under study is the modified Hill estimator, which do not deviate much from the true value. These estimators might react differently for different threshold selections, in this case the threshold is set as the top 1% of the sample.

Chapter 4

Normal Mixture Model

We want to understand how to construct a mixture of two normally distributed samples, with variances that are considered equal, so that the mixture can be interpreted as a heavy-tailed distribution. One commonly used heavy-tailed distribution that is constructed by normally distributed values is the Student's t distribution. In this chapter we investigate how to compose a mixture such that it behaves as a Student's t distribution.

4.1 Levene's Test

By Levene's test one can test if two or more different samples have a common variance. It tests the null hypothesis that the samples under study have equal variances. Suppose you have k random samples, $x_{i1}, ..., x_{in_i}$ each of sizes n_i, where $i = 1, \ldots, k$, and that you want to test if these samples have equal variances. By Levene's test, the test statistic F according to Gastwirth et al. (2009), is defined as

$$F = \frac{N-k}{k-1} \frac{\sum_{i=1}^{k} n_i(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_{i\cdot})^2},$$

where,

N is the total size of the pooled sample, $N = \sum_{i=1}^{k} n_i$.

d_{ij} is the absolute value of the deviation between each data point and the sample mean of the belonging group, $d_{ij} = |x_{ij} - \bar{x}_i|$, with $\bar{x}_i$ as the sample mean of the i:th sample.

$\bar{d}_{i\cdot}$ is the mean of the absolute value of the deviations d_{ij} in each group, $\bar{d}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} d_{ij}$.

$\bar{d}_{..}$ is the mean of all the absolute values of the deviations $\bar{d}_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} d_{ij}$.

The test statistic, F, is tested against the α-quantile of the F-distribution with $k-1$ and $N-k$ degrees of freedom. If F is smaller than the α-quantile of the F-distribution the null hypothesis can be rejected at a $1-\alpha$ significance level.

One interesting aspect of the test is whether the F-statistics depends on just the quotient σ_2/σ_1 or the independent σ-values. Say if we have two samples of normally distributed variables with different standard deviations and both with mean zero $\{\sigma_1 Z_{11}, \ldots, \sigma_1 Z_{1n_1}\}$ and $\{\sigma_1 \frac{\sigma_2}{\sigma_1} Z_{21}, \ldots, \sigma_1 \frac{\sigma_2}{\sigma_1} Z_{2n_2}\}$. Then

$$\bar{d}_{1j} = \sigma_1 \big| Z_{1j} - \frac{1}{n_1}\sum_{i=1}^{n_1} Z_{1i} \big| \quad \text{and} \quad \bar{d}_{2j} = \sigma_1 \frac{\sigma_2}{\sigma_1} \big| Z_{2j} - \frac{1}{n_2}\sum_{i=1}^{n_2} Z_{2i} \big|.$$

We also get that

$$\bar{d}_{1\cdot} = \frac{\sigma_1}{n_1}\sum_{j=1}^{n_1} \big| Z_{1j} - \frac{1}{n_1}\sum_{i=1}^{n_1} Z_{1i} \big| \quad \text{and} \quad \bar{d}_{2\cdot} = \frac{\sigma_1}{n_2}\frac{\sigma_2}{\sigma_1}\sum_{j=1}^{n_2} \big| Z_{2j} - \frac{1}{n_2}\sum_{i=1}^{n_2} Z_{2i} \big|,$$

while

$$\bar{d}_{..} = \frac{\sigma_1}{n_1+n_2}\Bigg(\sum_{j=1}^{n_1} \big| Z_{1j} - \frac{1}{n_1}\sum_{i=1}^{n_1} Z_{1i} \big| + \frac{\sigma_2}{\sigma_1}\sum_{j=1}^{n_2} \big| Z_{2j} - \frac{1}{n_2}\sum_{i=1}^{n_2} Z_{2i} \big| \Bigg).$$

As we can see σ_1^2 will appear both in the nominator and the denominator for the F-statistics. This will result in that the F-statistics will only be dependent of the quotient σ_2/σ_1 as well as of the sample sizes n_1 and n_2.

4.1.1 Example

Consider two samples, $X \sim N(0,1)$, $Y \sim N(0,\sigma_2^2)$. Let's look at what happens if we change the sample sizes n_1 and n_2 and if we vary σ_2.

The simulation is set up as follows,

- Simulate 1000 samples of $\{x_1, \ldots, x_{40}\}$ from $N(0,1)$ and 1000 samples of $\{y_1, \ldots, y_j\}$ from $N(0,\sigma_2^2)$ for $j = 10$ and $\sigma_2 = 1$.

- Use the samples and perform Levene's test. Look at the ratio of the samples that passes the test.

- Vary σ_2 up to 2, with a step size of 0.01.

- Repeat the steps above for $j = 11, \ldots, 20$.

In Figure 4.1 the results from the simulation is shown.

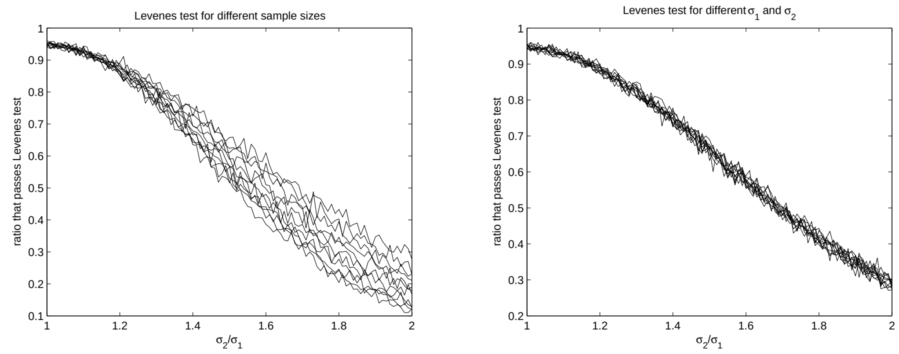


Figure 4.1: Left: *Levene's test for two samples of random normally distributed variables, where one of the sample sizes is 40 and the other increases from 10 to 20. The line at the top corresponds to the sample size of 10 whilst the line on the bottom corresponds to the sample size of 20. The y-axis shows the ratio of the 1000 samples where the null hypothesis couldn't be rejected at a significance level of* 5%. Right: *Levene's test for two samples of random normally distributed variables, where* $n_1 = 40$ *and* $n_2 = 10$. *Each line corresponds to a value of* $\sigma_1 = 1, 10, 20, \ldots, 100$. *The y-axis shows the ratio of the 1000 samples where the null hypothesis couldn't be rejected at a significance level of* 5%, *and the x-axis corresponds to the quotient* σ_2/σ_1.

As we can see from the example the sample sizes and the ratio σ_2/σ_1 do matter in Levene's test as could be expected. What happens if we vary σ_1 and σ_2 but keep σ_2/σ_1 constant. To the right in Figure 4.1 the result of Levene's test for σ_1 varying from 1 to 100 is shown. This result confirms the study we did previously in this section where we showed that the F-statistics is only dependent of the quotient σ_2/σ_1 and the sample sizes, and not by the individual values of σ_1 and σ_2.
Since we compare the value of F to the $(1-\alpha)$ quantile of the F-distribution it could be useful to understand the empirical density of the F-value for some certain cases. As explained before the F-value should follow the F-distribution when $\sigma_1 = \sigma_2$. If we let $\sigma_1 = \sigma_2$, $n_1 = 40$, $n_2 = 10$ and simulate 1000 samples of $\{X_{1,1}, \ldots, X_{1,n_1}, X_{2,1}, \ldots, X_{2,n_2}\}$ and calculate the corresponding F-value we get the empirical density shown in Figure 4.2. The dashed vertical lines corresponds to the 90, 95 and 99 percent quantiles for the F-distribution and the solid vertical lines is the corresponding quantiles of the empirical F-value. As we can see from the figure these quantiles are very close to each other which tells us that it is a good fit.

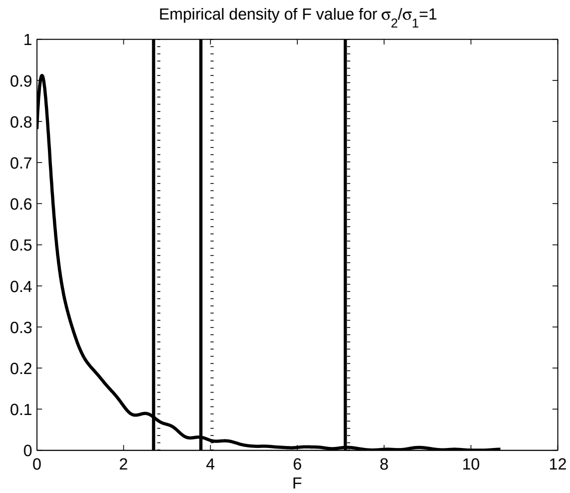


Figure 4.2: *Empirical density of the F-value for $\sigma_1 = \sigma_2$, $n_1 = 40$, $n_2 = 10$. The dotted lines corresponds to the $(1-\alpha)$ quantile of the F-distribution with 1 and n_1+n_2-1 degrees of freedom, where from left to right $\alpha = 0.1,\ 0.05,\ 0.01$. The solid lines shows the corresponding empirical quantiles.*

Now we let $\sigma_2 = 2\sigma_1$ and perform the same analysis. This result is shown in Figure 4.3. Here we can see that the empirical quantiles from the F-value is far from the corresponding quantiles from the F-distribution. But since we only reject or not reject the null hypothesis this deviation can be hidden in the analysis. From the simulation we get that the probability that the null hypothesis couldn't be rejected at a significance level of 1%, 5%, 10% is 0.39, 0.22, 0.14. The same simulations are done for different sample sizes and the results are shown in Table 4.1.

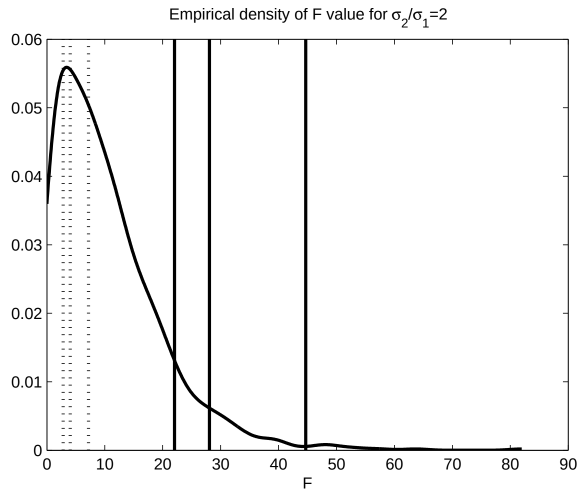


Figure 4.3: *Empirical density of the F-value for* $\sigma_2 = 2\sigma_1$*,* $n_1 = 40$*,* $n_2 = 10$*. The dotted lines corresponds to the* $(1-\alpha)$ *quantile of the F-distribution with 1 and* n_1+n_2-1 *degrees of freedom, where from left to right* $\alpha = 0.1,\ 0.05,\ 0.01$*. The solid lines shows the corresponding empirical quantiles.*

		$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$n_1 = 30$	$n_2 = 10$	0.16	0.25	0.47
	$n_2 = 15$	0.12	0.19	0.40
	$n_2 = 20$	0.09	0.15	0.35
$n_1 = 40$	$n_2 = 10$	0.14	0.22	0.39
	$n_2 = 15$	0.09	0.14	0.31
	$n_2 = 20$	0.06	0.10	0.25
$n_1 = 50$	$n_2 = 10$	0.14	0.20	0.37
	$n_2 = 15$	0.08	0.13	0.27
	$n_2 = 20$	0.05	0.08	0.18
$n_1 = 60$	$n_2 = 10$	0.13	0.20	0.35
	$n_2 = 15$	0.07	0.11	0.24
	$n_2 = 20$	0.05	0.08	0.18

Table 4.1: *Table showing the probabilities that the null hypothesis couldn't be rejected for different sample sizes where* $\sigma_2 = 2\sigma_1$*.*

By Levene's test we get an understanding of what type of normal mixtures that can be composed while not being viewed as a mixture. We can tell from our simulations that for some certain sample sizes, a combination of two normally distributed samples with $\sigma_2 = 2\sigma_1$ the chance to interpret the samples as having equal variances is as high as 50%.

4.2 Maximum Likelihood Parameter Estimation

A mixture of two normal distributions, $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ has the distribution function

$$F(x) = p\Phi(\frac{x}{\sigma_1}) + (1-p)\Phi(\frac{x}{\sigma_2}),$$

where p corresponds to the probability to draw a value from the $N(0, \sigma_1^2)$ distribution, and the probability $1-p$ from the $N(0, \sigma_2^2)$ distribution. The maximum likelihood estimates of p, σ_1, σ_2 are the values that maximize the log-likelihood function

$$l(x|p, \sigma_1, \sigma_2) = \sum_{i=1}^{n} \log f(x_i|p, \sigma_1, \sigma_2),$$

where $\{x_1, \ldots, x_n\}$ is an i.i.d. sample that comes from a distribution with unknown distribution function. From

$$f(x|p, \sigma_1, \sigma_2) = \frac{d}{dx}F(x) = \frac{p}{\sigma_1}\phi(\frac{x}{\sigma_1}) + \frac{1-p}{\sigma_2}\phi(\frac{x}{\sigma_2}),$$

where $\phi(x)$ is the probability density function of the standard normal distribution, we get that

$$l(x|p, \sigma_1, \sigma_2) = \sum_{i=1}^{n} \log\left(\frac{p}{\sigma_1}\phi(\frac{x}{\sigma_1}) + \frac{1-p}{\sigma_2}\phi(\frac{x}{\sigma_2})\right).$$

Since the Maximum Likelihood estimation wants to fit the estimation to the entire data there might be a worse fit in the tail than for the center of the data. Since our interest is in the tail we can perform the ML-method on the conditional density for a suiting threshold. Then we have the distribution conditional that $x > y$ as

$$F(x|x>y) = \frac{F(x) - F(y)}{1 - F(y)} = \frac{p\Phi(\frac{x}{\sigma_1}) + (1-p)\Phi(\frac{x}{\sigma_2}) - \left(p\Phi(\frac{y}{\sigma_1}) + (1-p)\Phi(\frac{y}{\sigma_2})\right)}{1 - \left(p\Phi(\frac{y}{\sigma_1}) + (1-p)\Phi(\frac{y}{\sigma_2})\right)},$$

this gives that the conditional density is

$$f(x|x>y) = \frac{dF(x|x>y)}{dx} = \frac{\frac{p}{\sigma_1}\phi(\frac{x}{\sigma_1}) + \frac{1-p}{\sigma_2}\phi(\frac{x}{\sigma_2})}{1 - \left(p\Phi(\frac{y}{\sigma_1}) + (1-p)\Phi(\frac{y}{\sigma_2})\right)}.$$

Thus the log-likelihood function that is to be minimized is

$$l(x|x>y|p, \sigma_1, \sigma_2) = \sum_{i=1}^{n} \log\left(\frac{\frac{p}{\sigma_1}\phi(\frac{x}{\sigma_1}) + \frac{1-p}{\sigma_2}\phi(\frac{x}{\sigma_2})}{1 - \left(p\Phi(\frac{y}{\sigma_1}) + (1-p)\Phi(\frac{y}{\sigma_2})\right)}\right).$$

We can investigate the behaviour of the maximum likelihood estimates of the parameters σ_1, σ_2 and p by simulating 1,000 samples of sizes $n = 50, 100, 150, 200$ and perform the algorithm. The densities of the estimates are shown in Figure 4.4.

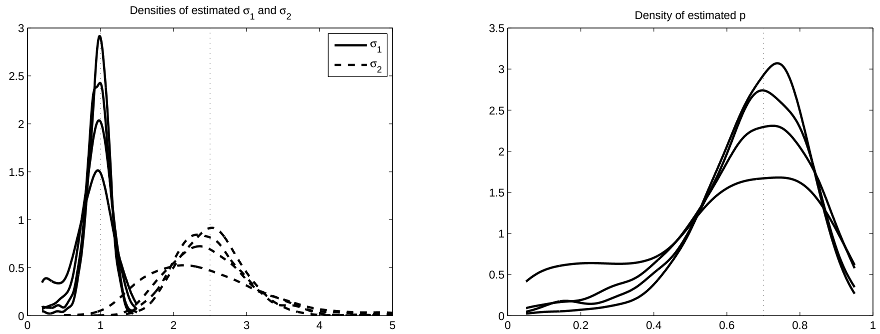


Figure 4.4: *Densities of* $\hat{\sigma}_1, \hat{\sigma}_2$ *(left) and* $\hat{p}$ *(right) of 1000 samples with* $n = 50, 100, 150, 200$ *by using maximum likelihood, where the vertical lines corresponds to the true values.*

4.3 EM-Algorithm

We have a two component mixture model constructed in the following way

$$\begin{aligned} X_1 &\sim N(\mu_1, \sigma_1^2) \\ X_2 &\sim N(\mu_2, \sigma_2^2) \\ X &= IX_1 + (1 - I)X_2, \end{aligned}$$

where $I \in \{0, 1\}$ and $P(I = 1) = p$. The density of the mixture model is

$$f_X(x) = p\phi_{\theta_1}(x) + (1 - p)\phi_{\theta_2}(x),$$

where $\phi_{\theta_i}(x)$ is the pdf of the normal distribution with mean μ_i and standard deviation σ_i. With a sample size of N the log-likelihood function is

$$l(\theta; \mathbf{x}) = \sum_{i=1}^{N} \log \left(p\phi_{\theta_1}(x_i) + (1 - p)\phi_{\theta_2}(x_i)\right)$$

This log-likelihood is though to maximize numerically, but according to Kamaruzzaman et al. (2012) we instead use the latent variables I_i, with $I_i = 1$ we say that X_i comes from X_1, and with $I_i = 0$ it comes from X_2. Then we can construct the log-likelihood equation as

$$l(\theta; \mathbf{x}, \mathbf{I}) = \sum_{i=1}^{N} [I_i \log \phi_{\theta_1}(x_i) + (1 - I_i) \log \phi_{\theta_2}(x_i)] + \sum_{i=1}^{N} [I_i \log p + (1 - I_i) \log(1 - p)].$$

The I_i's are unknown, so we replace them with the expected value, also called the responsibility of model 1

$$\gamma_i(\theta) = E[I_i|\theta, \mathbf{x}].$$

The EM algorithm then takes the responsibility as a weight and by iterations the maximum likelihood estimates can be found.

- First take starting guesses of $\hat{\mu_1},\ \hat{\mu_2},\ \hat{\sigma_1},\ \hat{\sigma_2},\ \hat{p}$.

- The expectation step (E-step): Calculate the responsibilities

$$\hat{\gamma}_i = \frac{\hat{p}\phi_{\hat{\theta}_1}(x_i)}{\hat{p}\phi_{\hat{\theta}_1}(x_i) + (1-\hat{p})\phi_{\hat{\theta}_2}(x_i)}, \quad i = 1, \ldots, N.$$

- The maximization step (M-step): Calculate the weighted means and variances using the responsibilities $\hat{\gamma}_i$'s.

$$\hat{\mu}_1 = \frac{\sum_{i=1}^N \hat{\gamma}_i x_i}{\sum_{i=1}^N \hat{\gamma}_i}, \quad \hat{\mu}_2 = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i) x_i}{\sum_{i=1}^N (1-\hat{\gamma}_i)},$$

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^N \hat{\gamma}_i (x_i - \hat{\mu}_1)^2}{\sum_{i=1}^N \hat{\gamma}_i}, \quad \hat{\sigma}_2^2 = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i)(x_i - \hat{\mu}_2)^2}{\sum_{i=1}^N (1-\hat{\gamma}_i)},$$

and

$$\hat{p} = \sum_{i=1}^N \frac{\hat{\gamma}_i}{N}.$$

By iteration over the E-step and the M-step until convergence of the parameters, the maximum likelihood estimated values are obtained.

As an illustration of the performance of the EM-algorithm for small samples, the densities of the estimated σ_1, σ_2 and p is shown in Figure 4.5. In this simulation the parameter values are $\sigma_1 = 1, \sigma_2 = 2.5$ and $p = 0.7$, which are also shown in the figure.

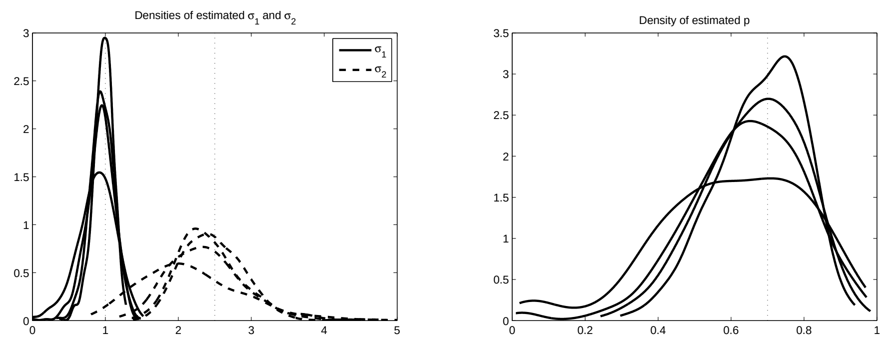


Figure 4.5: *Densities of* $\hat{\sigma}_1, \hat{\sigma}_2$ *(left) and* $\hat{p}$ *(right) of 1000 samples with* $n = 50, 100, 150, 200$ *by using the EM algorithm, where the vertical lines corresponds to the true values.*

As we can see from the density curves the algorithm performs well even for small sample sizes, and the peaks of the densities move towards the true values.

Chapter 5

Results

In this chapter we use the methods described earlier in order to obtain our results from simulations and analysis. We investigate some interesting cases that could occur and present how these cases should be taken care of.

5.1 Normal Mixture Model

We want to construct a mixture of two normal distribution both with mean 0. The two distributions should be considered to have equal variances and behave like a Student's t distribution. From Section 4.1 we can see that a mixture of two normally distributed samples with $\sigma_2/\sigma_1 = 2$ have a 30% chance to pass the Levene's test with sample sizes $n_1 = 40$ and $n_2 = 10$. We conduct a Maximum Likelihood estimation of σ_1, σ_2 and p. From equation 2.1 we determine the best fit through numerical methods. Our sample $\{x_1, \ldots, x_n\}$ are constructed as $t_\nu^{-1}(\frac{1:n}{n+1})$. The result are shown in Figure 5.1.

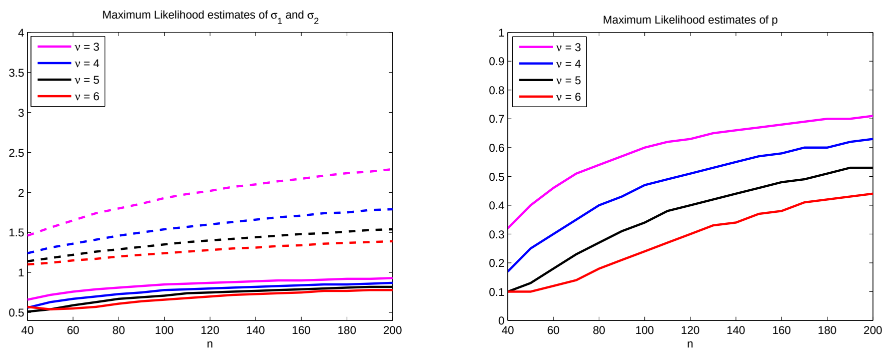


Figure 5.1: Left: *Maximum Likelihood estimation of* σ_1 *and* σ_2 *with a sample from Student's t distribution. The solid lines represents* σ_1 *and the dashed lines represent* σ_2. Right: *Maximum Likelihood estimation of* p *with a sample from the Student's t distribution.*

We can see that if we choose a sample of Student's t distributed variables with degrees of freedom $\nu = 4$, the maximum likelihood estimated parameters of the normal mixture is $(\sigma_1,\ \sigma_2,\ p) \approx (0.9,\ 1.8,\ 0.65)$. Now we take a look at the ML-estimation of the parameters of the largest 30% of the data. This time our sample $\{x_{0.7n}, \ldots, x_n\}$ are constructed as $t_\nu^{-1}\left(\frac{0.7n:n}{n+1}\right)$ and we use the conditional log-likelihood function in order to estimate the parameters. The results are shown in Figure 5.2.

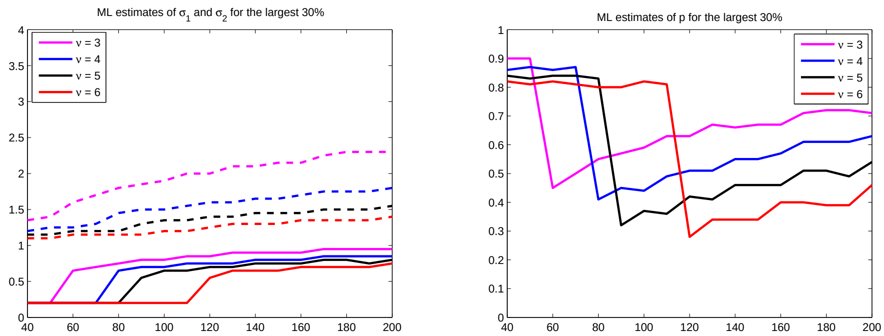


Figure 5.2: Left: *Maximum Likelihood estimation of* σ_1 *and* σ_2 *with the largest* 30% *of the data from Student's t distribution with* ν *degrees of freedom. The solid lines represents* σ_1 *and the dashed lines represent* σ_2. Right: *Maximum Likelihood estimation of* p *with the largest* 30% *of the data from Student's t distribution with* ν *degrees of freedom.*

There is only a small difference in the results for the larger samples as we can see. The problem is for the smaller samples where $n < 120$. Since we only choose the top 30% of the sample the actual sample that is used is rather small which makes it difficult for the maximum likelihood algorithm.

If we instead use the EM-Algorithm to estimate the parameters of the mixture model the result is different. We simulate 1000 samples of $\{x_1, \ldots, x_n\}$ from Student's t distribution with ν degrees of freedom. We then use the EM-Algorithm in order to estimate the parameters. The results shown in Figure 5.3 are the mean of the 1000 simulations for each sample size.

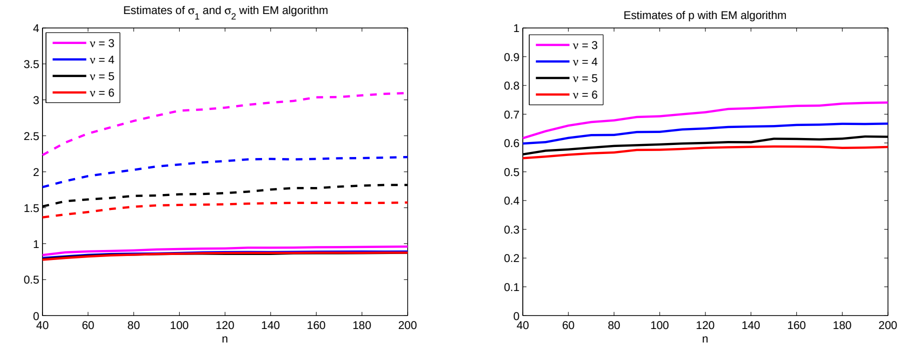


Figure 5.3: Left: *Estimation of σ_1 and σ_2 with a sample from Student's t distribution by the EM-Algorithm. The solid lines represents σ_1 and the dashed lines represent σ_2.* Right: *Estimation of p with a sample from the Student's t distribution by the EM-Algorithm.*

As we can see the EM-algorithm gives larger values on σ_2 than the maximum likelihood estimation. It also looks like the EM-algorithm behaves better for smaller sample sizes, where the estimates are more consistent contrary to the ML method. We can compare the two different methods by testing both models against the same samples. We simulate 500 samples of size 100 from a Student's t distribution with 4 degrees of freedom. Then we estimate the parameters σ_2/σ_1 and p for the normal mixture model by the maximum likelihood estimation and the EM-algorithm. The results are shown in Figure 5.4.

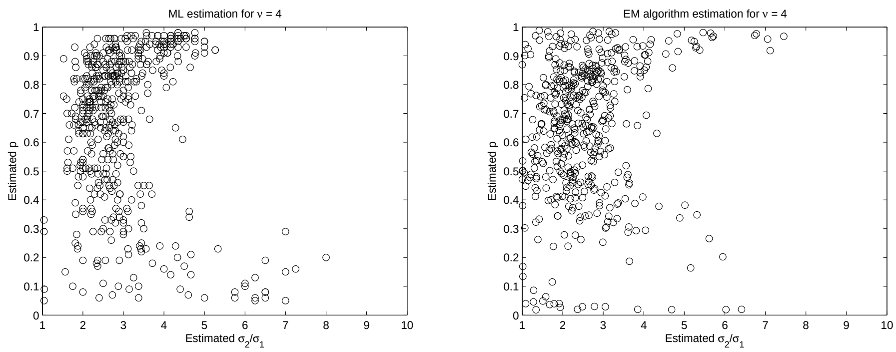


Figure 5.4: *Estimation of σ_2/σ_1 and p with 1000 samples of size 100 from Student's t distribution with $\nu = 4$ degrees of freedom.* Left: *Estimated parameters from the ML method.* Right: *Estimated parameters from the EM-Algorithm.*

As we can see from the results both look similar in their results with a mass around $\sigma_2/\sigma_1 \approx 2.5$ and $p \approx 0.75$. Neither of the methods give any stable estimates of the parameters, which could be due to the randomness of the small samples.

5.2 Tail Index Estimation of Normal Mixture Model

Now we conduct the same tests as we did in section 3.1.3, but here we use the normal mixture model as the data. We use the parameters from the maximum likelihood method for the mixture model. Since $p = \frac{n_1}{n_1+n_2} = 0.65$, the threshold cannot be at 10% since the excesses will most likely exclusively come from $N(0, \sigma_2^2)$ which will result in a light-tailed behaviour of the data. We set the threshold at 20% and plot the densities of the estimated ξ values for the 1000 samples. The results are shown in Figure 5.5.

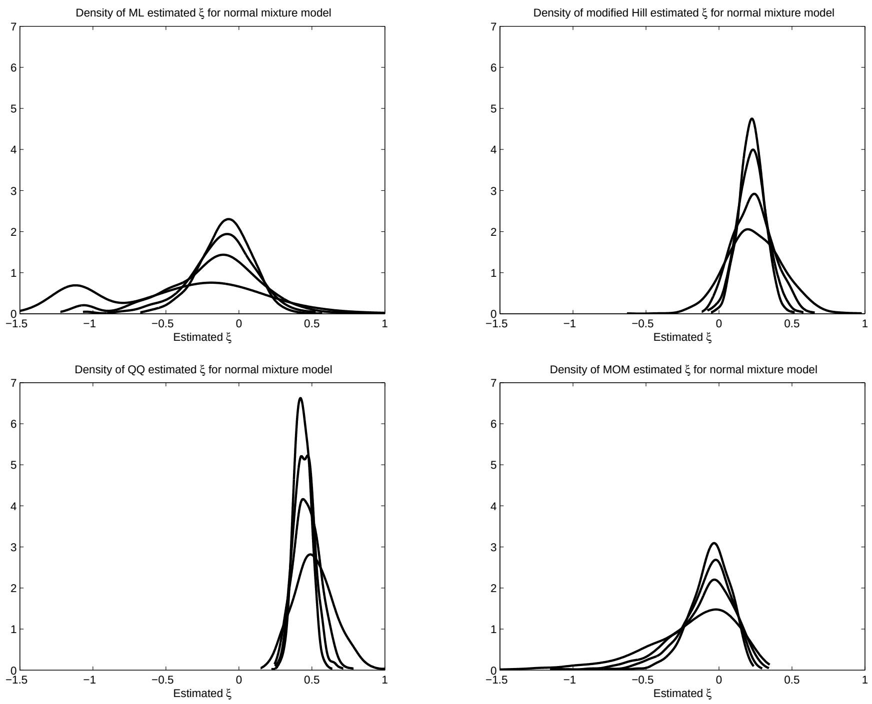


Figure 5.5: *Densities of the estimated ξ's for the normal mixture model with $\sigma_1 = 0.9$, $\sigma_2 = 1.8$ and $p = 0.65$, where $n = 50, 100, 150, 200$.* Top left: *Maximum Likelihood estimator.* Top right: *Modified Hill estimator.* Bottom left: *Quantile-Quantile estimator.* Bottom right: *Method of Moments estimator.*

As we can see from the results all the different estimators are stabilizing around a certain value for larger n. The MLE behave a lot different for smaller n than for larger, while the other estimators more or less behaves the same but are more concentrated around a certain value as the sample size increases. We can compare these results with the densities of the estimators for the corresponding Student's t distribution with $\nu = 4$ degrees of freedoms.

The results are presented in Figure 5.6. As we can see the results do not differ much from the normal mixture case. This indicates that the mixture of two normally distributed samples can behave as the Student's t distribution.

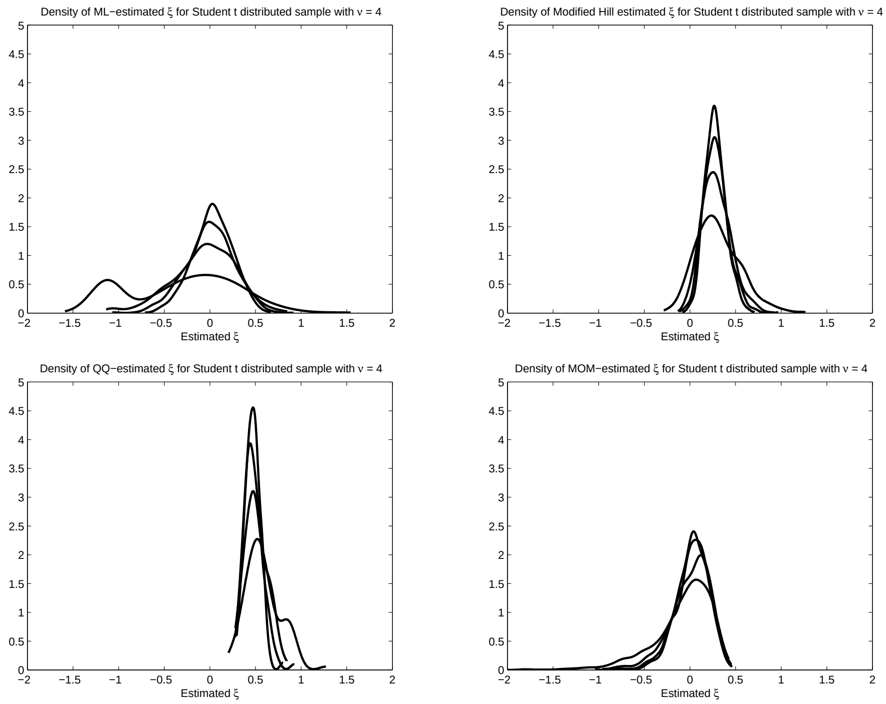


Figure 5.6: *Densities of the estimated* ξ*'s for the Student's t distributed samples with* $\nu = 4$ *degrees of freedoms, and* $n = 50, 100, 150, 200$. Top left: *Maximum Likelihood estimator.* Top right: *Modified Hill estimator.* Bottom left: *Quantile-Quantile estimator.* Bottom right: *Method of Moments estimator.*

In this specific case and throughout the report it looks like the modified Hill estimator is the best performing estimator. It is the estimator with the smallest bias and the smallest mean squared error. From now on we use this estimator for further more general investigations of the normal mixture model. We know that the mixture model can appear to behave as a heavy-tailed distribution from the previous case. Now we look at all possible combinations of different sample sizes. In order to limit the simulations two boundaries are set.

- $p = \frac{n_1}{n_1+n_2} \geq 0.6$

- $40 \leq n_1 + n_2 \leq 200$

From earlier studies we know that Levene's test as well as the tail index estimation is not dependent on the individual standard deviations of the mixture model, it is dependent on the quotient σ_2/σ_1. First we look at the case when $\sigma_2/\sigma_1 = 1.5$. The simulation is constructed as

· Simulate 1,000 samples of size 190 from the standard normal distribution and 1,000 samples of size 80 from the normal distribution with standard deviation 1.5.

· Let $30 \leq n_1 \leq 190$ and $5 \leq n_2 \leq 80$ and check that the sample sizes fulfils the boundaries set earlier.

· Calculate the mean tail index by the modified Hill's estimator and the length of the interval which contains 90% of the data.

In Figure 5.7 the tail index estimates are shown for different combinations of sample sizes.

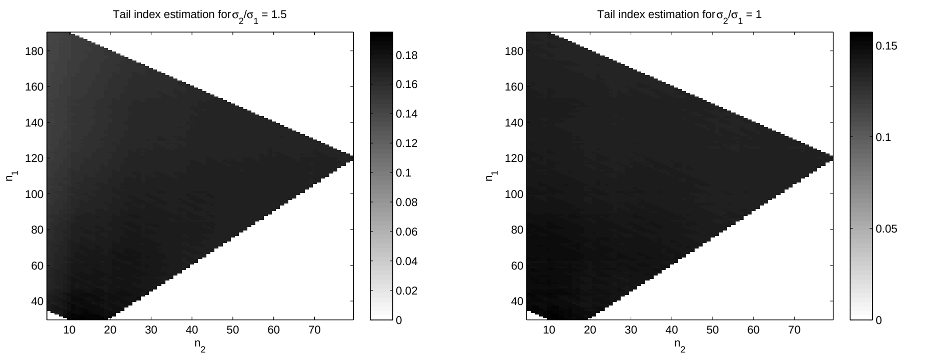


Figure 5.7: Left: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.5$ *for different combinations of sample sizes.* Right: *Tail index estimation for standard normally distributed samples.*

As we can see in Figure 5.7, there is not much difference for the estimated tail indices of the mixture model as for the standard normal model. This is due to the fact that the modified Hill estimator is biased for small tail indices. We can compare the two index estimations by subtracting the estimated index from the standard normally distributed samples from the mixture samples. By doing this we can evaluate if the mixture is considered to have a heavier tail than the standard normal model. This is shown in Figure 5.8.

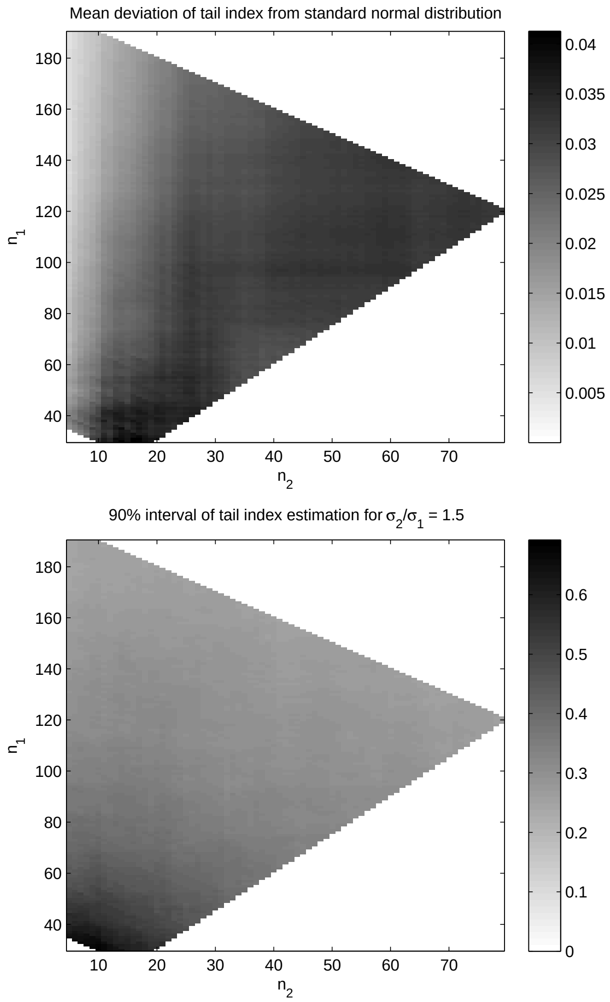


Figure 5.8: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.5$ *and standard normally distributed samples.* Bottom: *The length of the interval that contains 90% of the data.*

As we can see on the top of Figure 5.8 there is not much difference between the estimated tail indices for the mixture model and the standard normal model. Thus the data from the mixture model cannot be interpreted as heavy-tailed data. Since this plot only shows the mean of the estimated tail indices this does not reflect 100% of the outcomes. In the bottom of Figure 5.8 the length of the interval that contains 90% of the data is shown. As we can see there is a wider interval for smaller sample sizes since the variance and the mean squared error is larger in those cases. In Figure 5.9 the ratio that passes Levene's test for 1,000 samples is shown. In this context there is no idea to investigate tail behaviour of a mixture model where the variances of the two components are considered unequal. The plot suggests that for $n_2 > 30$ in most cases the chance that the two samples that construct the mixture has equal variances is below 30%. Lets conduct the same analysis with $\sigma_2/\sigma_1 = 1.75, 2, 2.25$.

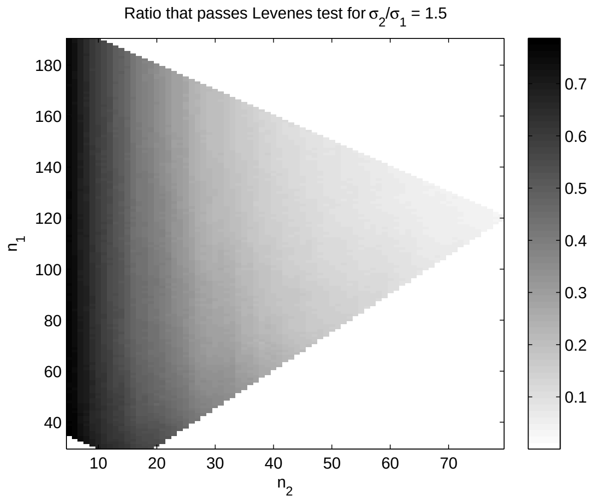


Figure 5.9: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 1.5$.

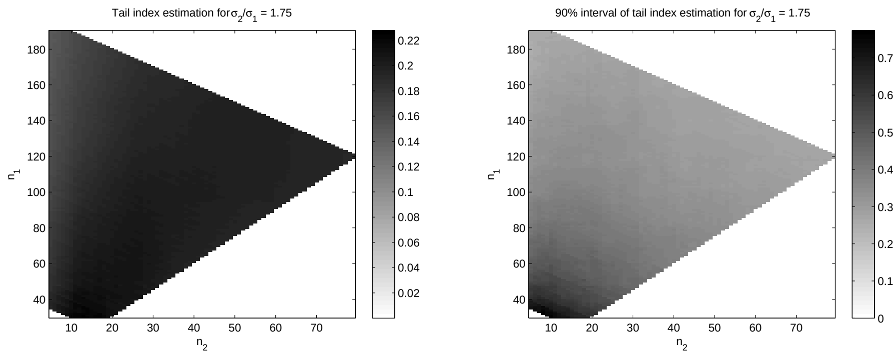


Figure 5.10: Left: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *for different combinations of sample sizes.* Right: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *and standard normally distributed samples.*

As expected Figure 5.10 tells us that the mean of the estimated tail indices is around 0.2, which tells us that the mixture is interpreted as coming from a heavy-tailed distribution function. To the right in the figure the 90% interval lengths are shown. From Figure 5.11 we can tell that there is a small difference between the mean of the estimated tail indices for the mixture model and the standard normal model. The area where the difference is greatest is when both n_1 and n_2 is small. But is the mixture model constructed so that the two components variances are considered equal in this area? In the bottom of Figure 5.11 the ratio that passes Levene's test is shown.

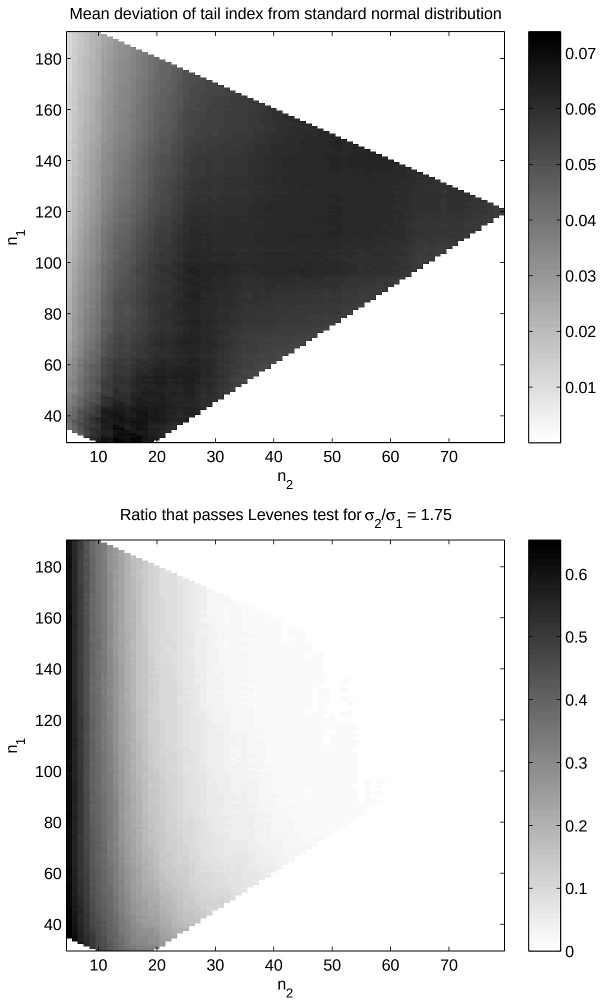


Figure 5.11: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 1.75$.

As the figure shows, it looks like for a mixture with $n_1 < 50$ and $n_2 < 20$ there is about 30% chance that the mixture passes the Levene's test as a non-mixture. This combinations of sample sizes are also the ones that gives the largest tail-index estimations.

Now we take a look at the case when $\sigma_2/\sigma_1 = 2$. The mean of the tail index estimations are shown in Figure 5.12 along with the 90% interval.

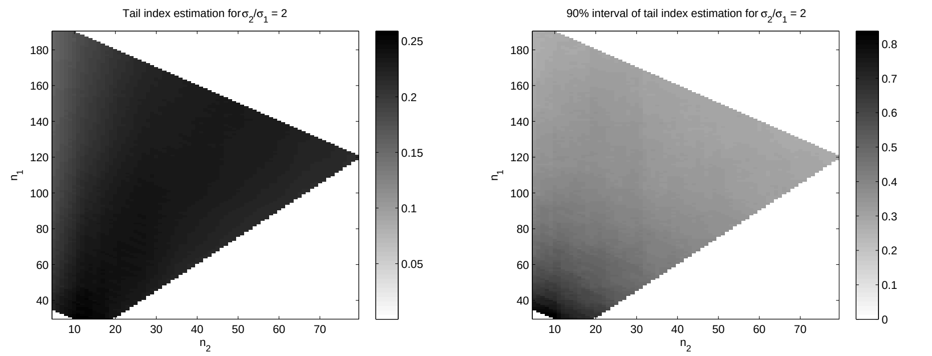


Figure 5.12: Left: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *for different combinations of sample sizes.* Right: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *and standard normally distributed samples.*

As we can see from Figure 5.13, the worst case scenario is along a straight line where $n_1/(n_1 + n_2) \approx 0.75$. These results indicates that a normal mixture with $\sigma_2/\sigma_1 = 2$ and $p = 0.75$ could be interpreted as a heavy-tailed model. Can we compose such a mixture without knowing it? We take a look at the Levene's test for a normal mixture with $\sigma_2/\sigma_1 = 2$, this is shown in the bottom of Figure 5.13.

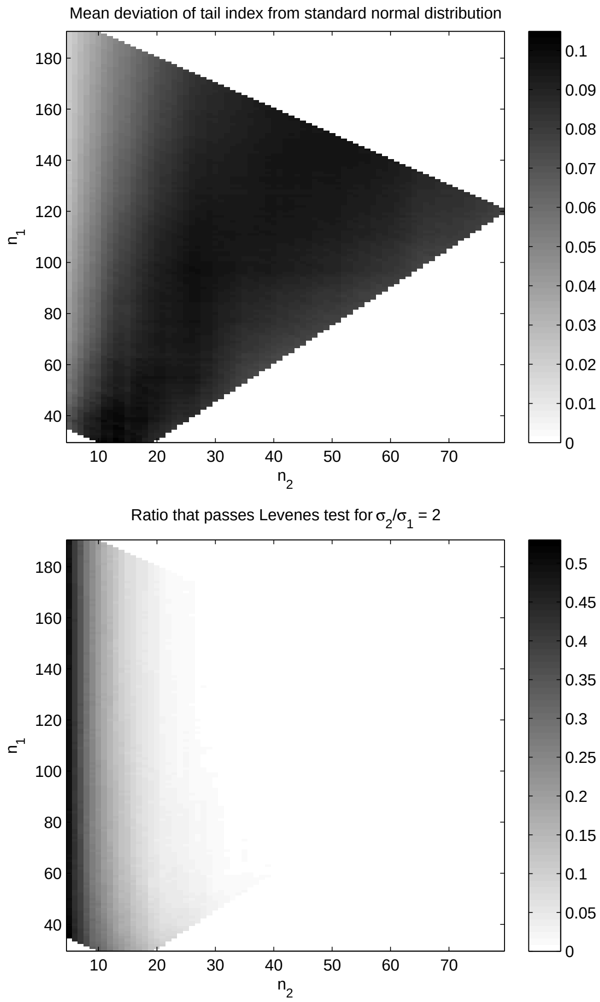


Figure 5.13: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 2$.

As expected the ratio that passes Levene's test is smaller than for $\sigma_2/\sigma_1 = 1.75$. There is still a 20% chance that the components of the normal mixture is considered having the same variances for $n_2 < 15$.

Finally we will take a look at the case when $\sigma_2/\sigma_1 = 2.25$. The mean of the estimated tail indices and the 90% interval is shown in Figure 5.14.

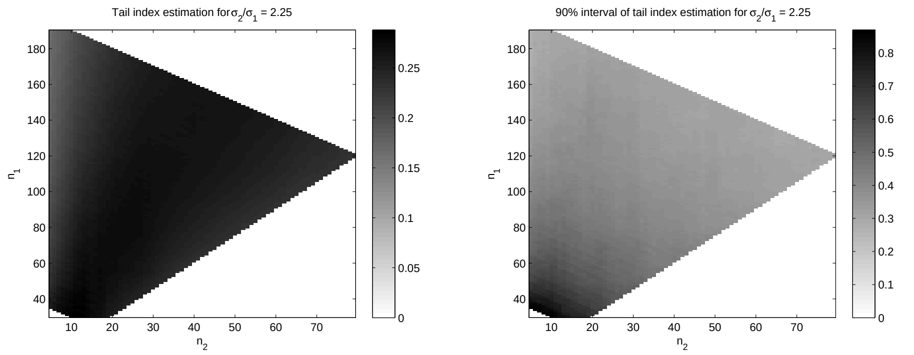


Figure 5.14: Left: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *for different combinations of sample sizes.* Right: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *and standard normally distributed samples.*

In Figure 5.15 we can see that this type of mixture behaves the same as the previous cases, but with a maximum mean deviation from the standard normal model by 0.12. It looks like the worst mean deviation occurs for the same $p \approx 0.75$. In the bottom of the same figure the ratio that passes Levene's test for $\sigma_2/\sigma_1 = 2.25$ is shown.

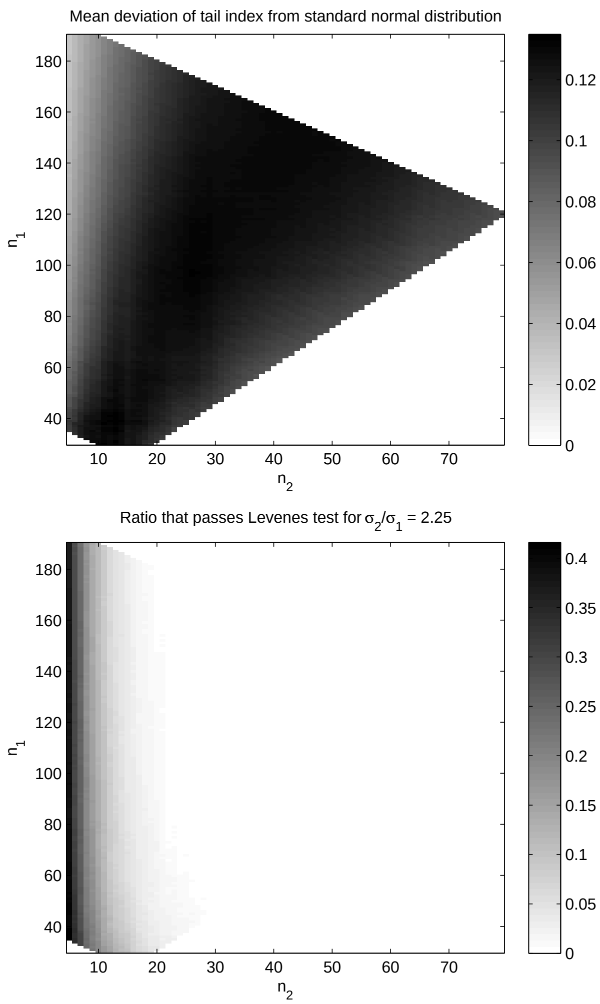


Figure 5.15: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 2.25$.

Here, even fewer samples passes Levene's test. This time it is a 20% that the mixture passes the test when $n_2 \leq 10$.

Chapter 6

Summary and Conclusion

When there is a lot of data available it is much easier to perform tail index analysis. Since the estimators in this thesis all converges, and do it at the same rate, there is no difference on which approach to use when investigating the tail indices of large samples. The problem is when there are small samples under investigation. If everything is smooth and neat when there are large samples, it is the opposite when dealing with small samples. The different estimators behave differently for small samples, all with different bias and different mean squared error. Another problem on top of the small sample problem is the bias for small tail indices. Even though you think you have a good understanding of the sample, there exists a bias for small tail indices for all the estimators. One way to understand if the sample under investigation comes from a heavy-tailed distribution is to compare it with the standard normal distribution. Since the standard normal distribution has a tail index of 0, this is a good model to use as a benchmark.

The most dangerous construction of a normal mixture is when the quotient $\sigma_2/\sigma_1 \geq 2$. For certain combinations of sample sizes n_1 and n_2 such that $n_1 + n_2 < 100$ and $n_1/(n_1 + n_2) \approx 0.75$ there is a significant chance that the components of the mixture is considered to have equal variances. When believing that the two samples that composes the mixture have the same variance in combination with a positive estimated tail index one can be misled by the result. By believing that the underlying distribution is heavy-tailed an estimation of a large quantile can be far from the true quantile which can be devastating for the firm or risk controller. Thus when dealing with this combination of data samples one has to be cautious. It is important to use plots in order to understand the empirical distribution of the data. A good way of understanding the data set is to plot the empirical probability density function along with the theoretical density function of the estimated mixture model. By comparing the two density curves one can get a good understanding of how good of a fit the model is. Since the estimators are sensitive to a change in the threshold it is a good idea to change the threshold as well, and see what happens with the estimated tail index. By using different estimators one can get a second opinion on the estimation.

Bibliography

BCBS (2010). Basel III: A global regulatory framework for more resilient banks and banking systems.

Boos, D. and Stefanski, L. (2013). *Essential Statistical Inference*. Springer.

Coles, S. (2000). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Dacorogna, M., Müller, U., Pictet, O., and de Vries, C. (1995). The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer.

Deheuvels, P., Haeusler, E., and Mason, D. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(02):371–381.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2001). *Modelling Extremal Events for Insurance and Finance*. Springer.

European Commission (2014). Draft of impact assessment report of solvency II.

Gastwirth, J., Gel, Y., and Miao, W. (2009). The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science*, 24(3):343–360.

Hosking, J. and Wallis, J. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):339–349.

Huisman, R., Koedijk, K., Kool, C., and Palm, F. (2012). Tail-Index Estimates in Small Samples. *Journal of Business & Economic Statistics*, 19(2):208–216.

Hult, H., Lindskog, F., Hammarlid, O., and Rehn, C. (2012). *Risk and Portfolio Analysis*. Springer.

Kamaruzzaman, Z., Isa, Z., and Ismail, M. (2012). Mixtures of Normal Distributions: Application to Bursa Malaysia Stock Market Indices. *World Applied Sciences Journal*, 16(6):781–790.

Kratz, M. and Resnick, S. (1996). The Q-Q Estimator and Heavy Tails. *Stochastic Models*, 12(4):699–724.

Migdalas, A., Georgiadis, C., Stiakakis, E., Sifaleras, A., and Papathanasiou, J. (2013). *Optimization Theory, Decision Making, and Operations Research Applications*. Springer.

Appendices

Appendix A

Coloured 3D Graphs of Estimated Tail Indices

In this appendix the 3D plots from chapter 5 is presented in colour.

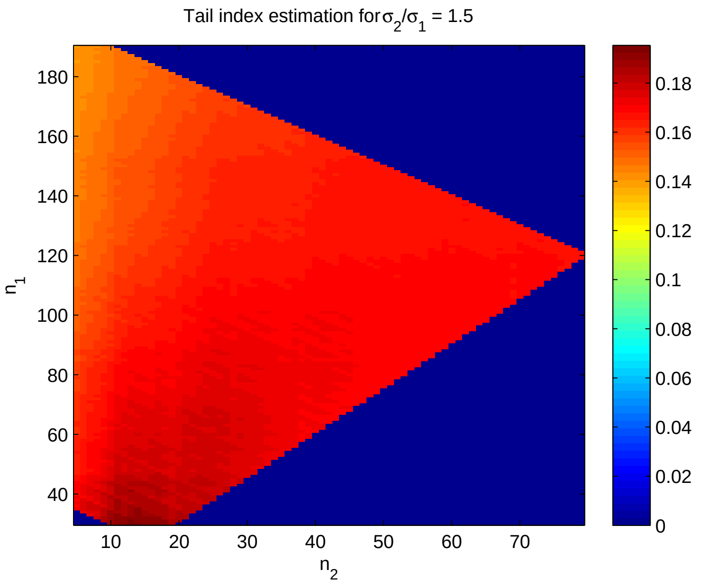


Figure A.1: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.5$ *for different combinations of sample sizes.*

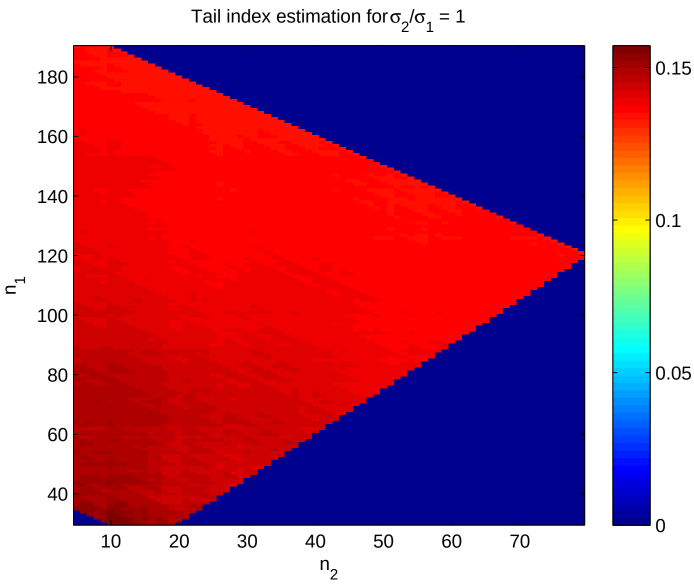


Figure A.2: *Tail index estimation for a mixture of normally distributed samples where $\sigma_2/\sigma_1 = 1$ for different combinations of sample sizes.*

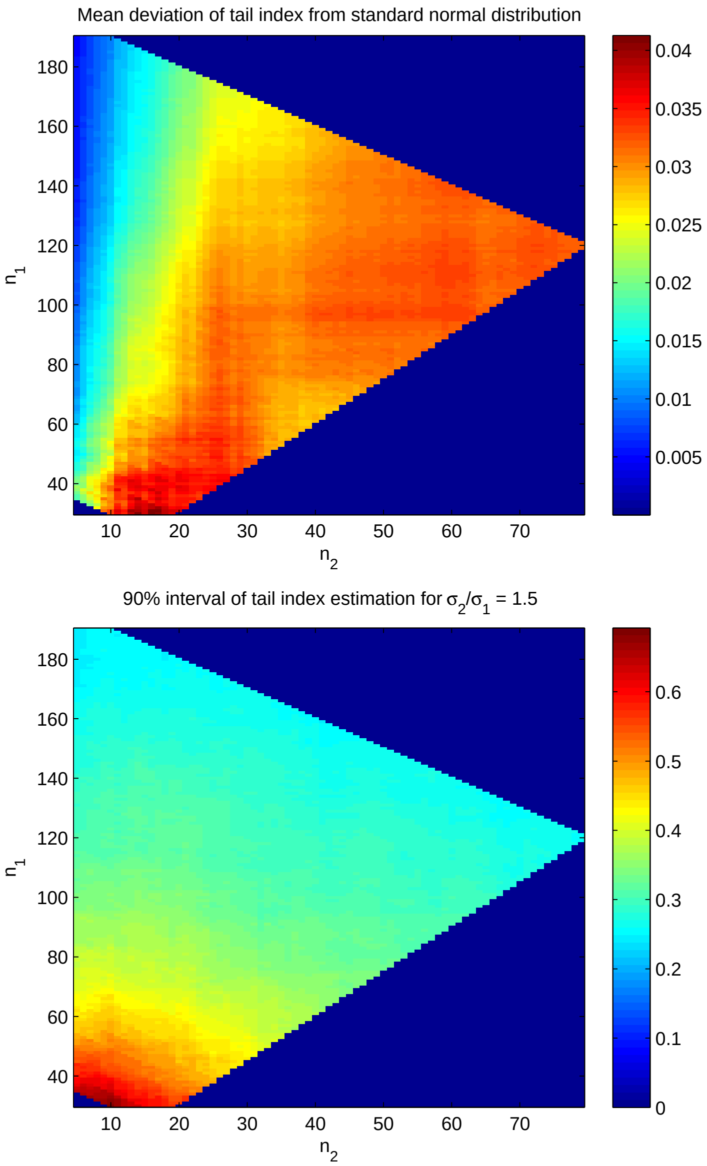


Figure A.3: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.5$ *and standard normally distributed samples.* Bottom: *The length of the interval that contains 90% of the data.*

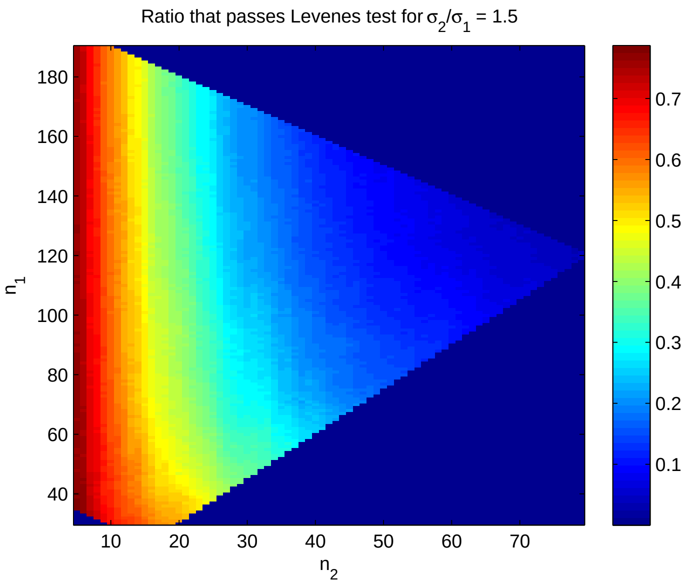


Figure A.4: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 1.5$.

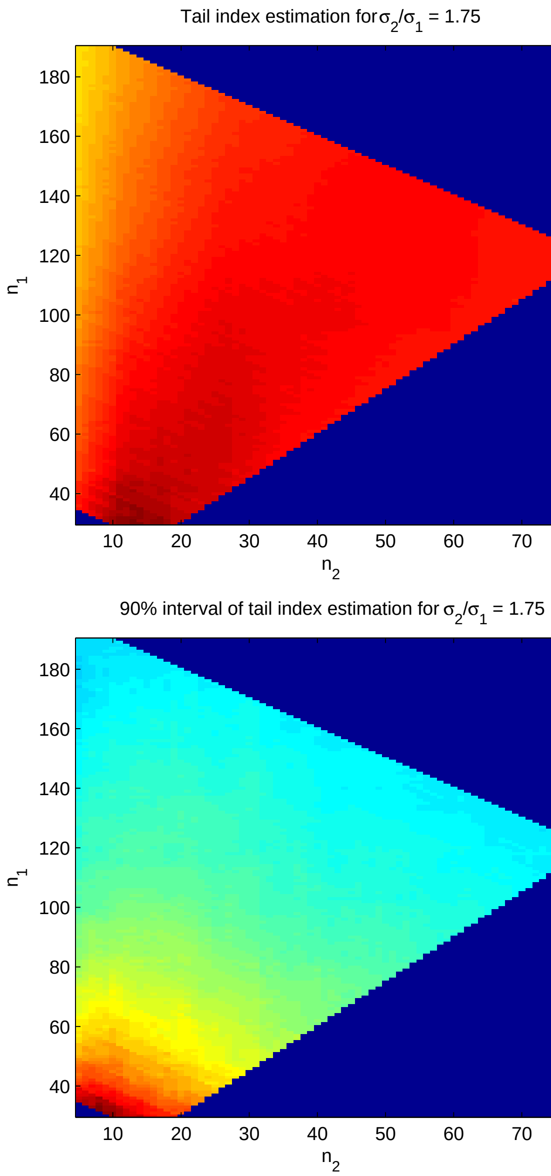


Figure A.5: Top: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *for different combinations of sample sizes.* Bottom: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *and standard normally distributed samples.*

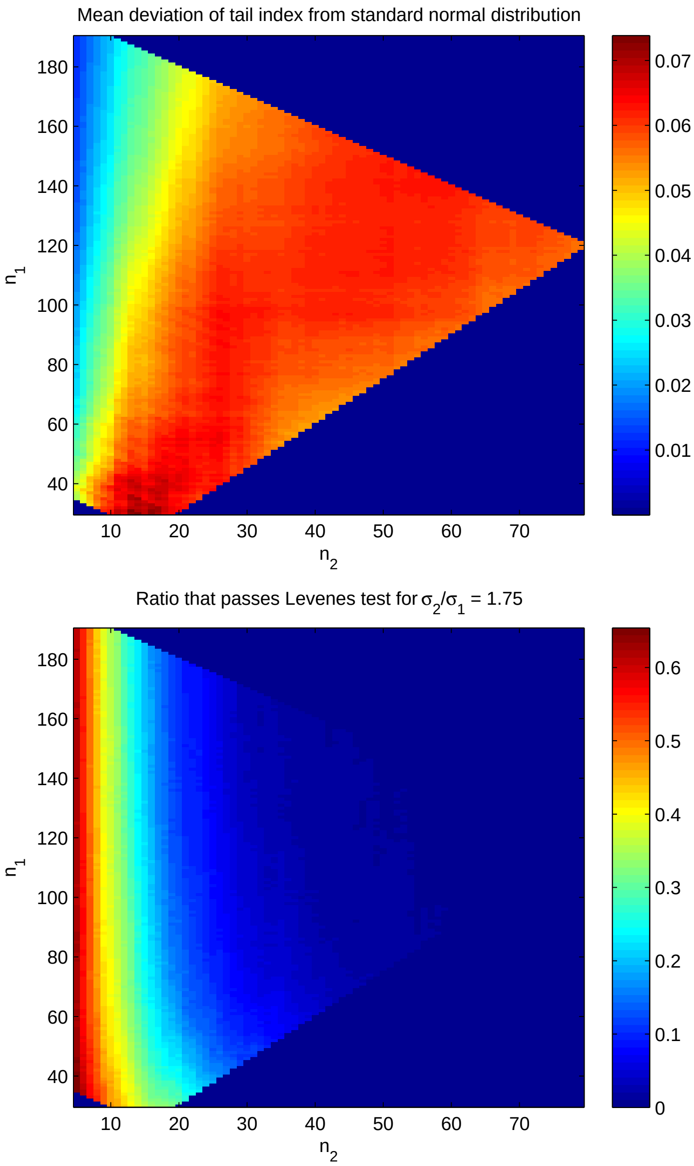


Figure A.6: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 1.75$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 1.75$.

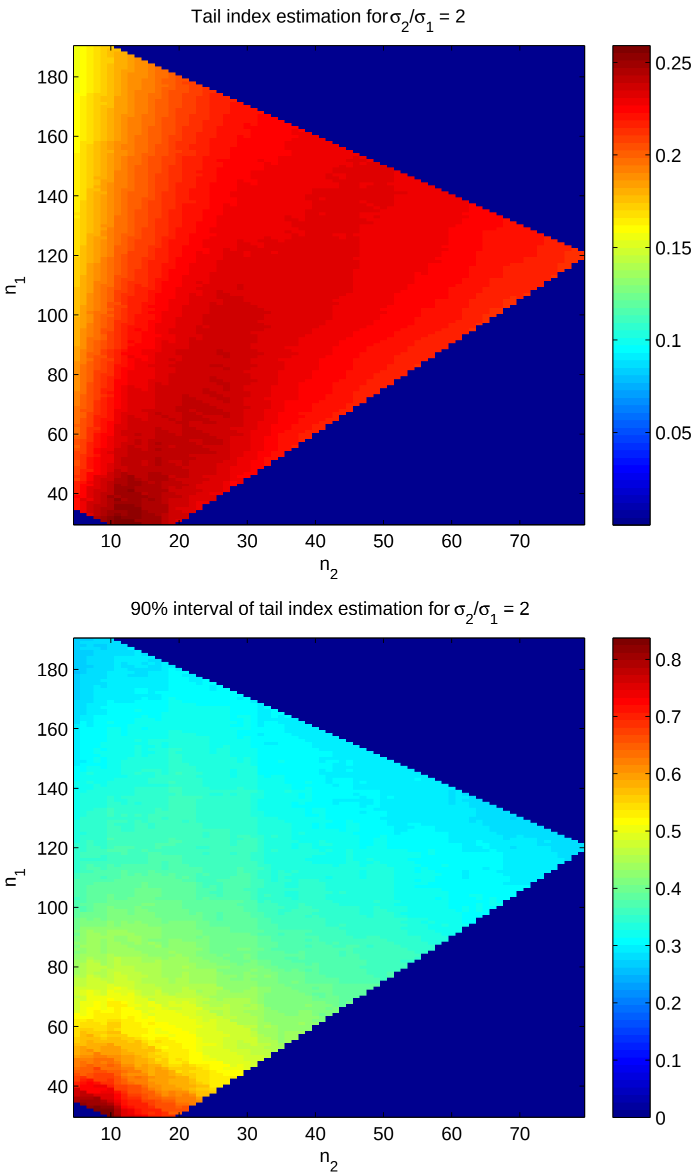
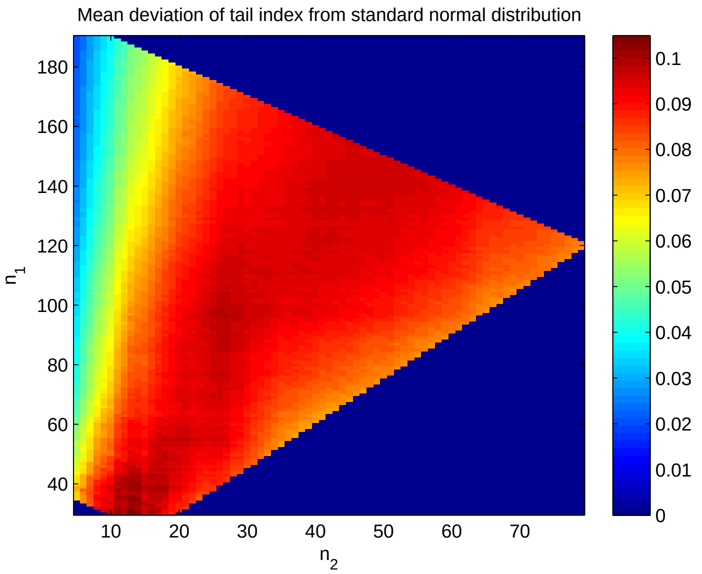


Figure A.7: Top: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *for different combinations of sample sizes.* Bottom: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *and standard normally distributed samples.*



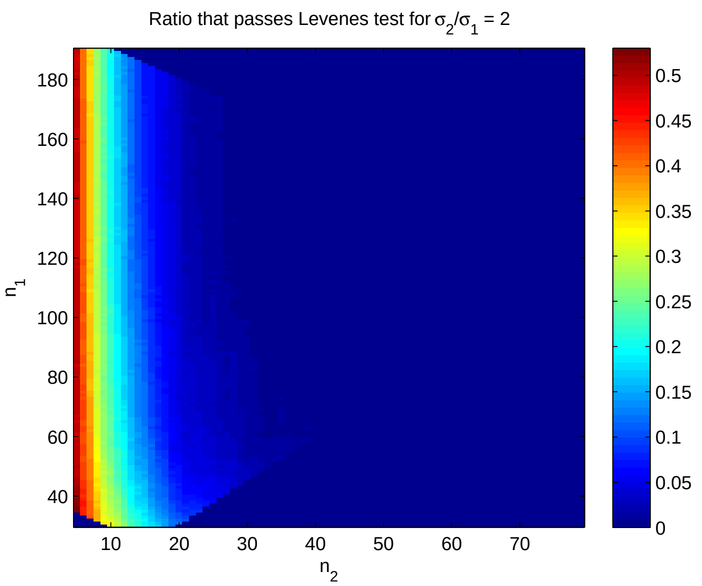


Figure A.8: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 2$.

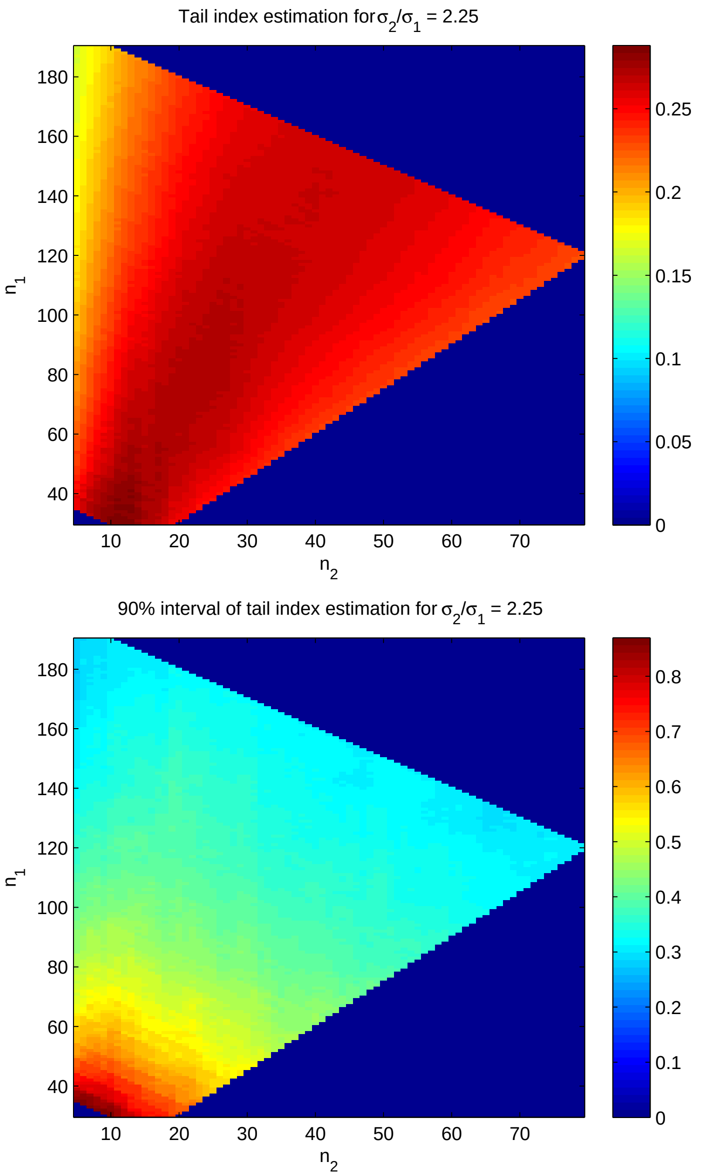


Figure A.9: Top: *Tail index estimation for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *for different combinations of sample sizes.* Bottom: *The length of the interval that contains 90% of the data of the estimated tail indices for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *and standard normally distributed samples.*

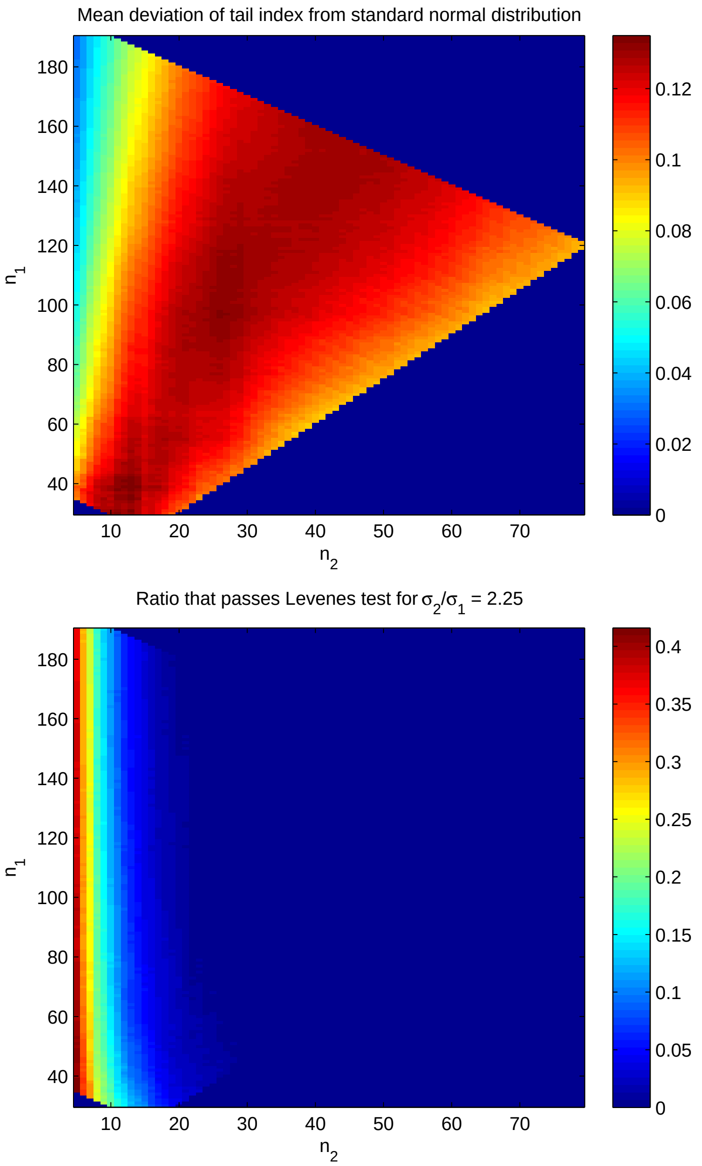


Figure A.10: Top: *Mean difference of the estimated tail index for a mixture of normally distributed samples where* $\sigma_2/\sigma_1 = 2.25$ *and standard normally distributed samples.* Bottom: *Ratio that passes Levene's test for* $\sigma_2/\sigma_1 = 2.25$.