

AN EXTREME VALUE APPROACH TO PRICING CREDIT RISK

SOFIA LANDIN

Master's thesis
2018:E69



LUND INSTITUTE OF TECHNOLOGY
Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Abstract

An Extreme Value Approach To Pricing Credit Risk will outline the possibility to investigate a company's price of risk over different time periods given a pre-defined risk level. With help of credit default swap (CDS) prices and extreme value theory the credit risk can be estimated for different return levels. This is incorporated by investigating monthly CDS data from Deutsche Bank AG EUR CDS 5Y between the time periods of August 2001 to April 2018 and applying the data to extreme value theory.

Keywords: *Credit Risk, Credit Default Swap, Credit Valuation Adjustment, Extreme Value Theory, Generalized Extreme Value Distribution, Gumbel Distribution, Generalized Pareto Distribution, Block Maxima, Peak-over-Threshold, Probable Maximum Loss*

Acknowledgements

The idea behind *An Extreme Value Approach To Pricing Credit Risk* originates from the experience of pricing credit risk at the corporate FX sales desk in London. This knowledge, encouraged the idea to combine this practice with the course; *Statistical Modeling of Extreme Value*, held at Lund University at the Centre of Mathematical Sciences, taught by Nader Tajvidi.

N. Tajvidi is the supervisor of this thesis and has been patient, supportive and very helpful during the process. Therefore, I would like to give my deepest gratitude to him.

In order to handle the data and make statistical interpretations, MATLAB R2017b and RStudio have been used in line with the package *in2Extreme*. The CDS data is public available and collected from Bloomberg.

Outline

In the first part, the background behind the subject will be outlined followed by a short introduction how counterparty risk is priced and applied on the over-the-counter (OTC) market. Thereafter a summary of the CDS-derivative is defined.

In the second part, general statistical and extreme value theory with methods applicable with the theory are explained followed by statistical interpretations which are necessary in order to make those assumptions. By apply the data to the family of generalized extreme value distributions, three different methods will be evaluated to find the most accurate fit and how to find an appropriate approximation of pricing credit risk for a range of years. The methods outlined are; Block Maxima, Peaks-over-Threshold and Probable Maximum Loss.

The theory will investigate monthly historical CDS data from Deutsche Bank AG EUR CDS 5Y between the time periods of August 2001 to April 2018. The ambition is to investigate what methods, based on CDS prices considered extreme, are most appropriate describing how credit risk should be priced for different time periods given pre-defined risk level in the future. Finally, conclusions of the results are presented.

Content

1. Introduction	4
Credit Valuation Adjustment	4
Credit Default Swap	5
2. General Statistical Theory	6
Stationary Stochastic Process	6
Autocorrelation Function: ACF	6
Delta Method	7
Deviance Function	8
Profile Likelihood	8
Likelihood Ratio Test	9
Model Valuation: Goodness of Fit, Diagnostic Plots	9
Empirical Distribution Function	9
Probability Plot	10
Quantile Plot	10
3. Extreme Value Theory	11
Extremal Types Theorem	11
Generalized Extreme Value Distribution: GEV	12
Block Maxima	13
Likelihood Function	14
Log-Likelihood Function	14
Peak over Threshold: POT	15
Generalized Pareto Distribution: GPD	15
Theorem	16
Modeling Threshold Exceedances	17
Mean Residual Life Plot	17
Parameter Stability Plot: Model Based Approach	18
Parameter Estimation with Maximum Likelihood Method	18
Return Level	19
Return Level Plot	20
Restriction for Return Levels	20
Model Comparison	21
Dependency and Declustering	21
Declustering Approach	22
Probable Maximum Loss: PML	24
4. Analysis	25
Block Maxima	28
Generalized Extreme Value Distribution: GEV	28
Gumbel Distribution	33
Peak-over-Threshold Method	36
Generalized Pareto Distribution: GPD	37
Exponential Distribution	42
Declustering	45
Return Level Plots	48
Probable Maximum Loss: PML	52
Parameter Estimates	52
Return Level	52
5. Conclusion	53
Bibliography	55

1. Introduction

Credit Valuation Adjustment

Credit valuation adjustment, CVA, is the valuation of counterparty risk and is a valuation adjustment used in the financial market for pricing in counterparty risk to financial derivatives. CVA can be regarded as the discounted value of a derivative that the holder is charged given the probability of counterparty default. Conclusively, CVA is considered to be a market risk of the trading book of the bank. During the financial crisis, only one-third of losses due to counterparty risk were related to actual default. Remaining losses were market-to-market related and have raised the awareness of holding sufficient counterparty risk capital. Consequently, the Basel III framework incorporates further parameters to CVA to sufficiently capitalise the counterparty risk.

In order to price counterparty risk and hence estimate the CVA capital charge, the market price for the derivative needs to be approximated. When examine over-the-counter, OTC, derivatives, which is not traded on a publicly open market such as a stock exchange, the most sufficient way to do so is by incorporating the price of credit default swaps, CDSs. It is a derivative used for hedging purposes against counterparty risk against the reference entity. CDS-spreads are therefore an important parameter into models when pricing CVA.

The Basel Committee on Banking Supervision states; *“Whenever the CDS spread of the counterparty is available, this must be used. Whenever such a CDS spread is not available, the bank must use a proxy spread that is appropriate based on the rating, industry and region of the counterparty.”* [4]

Credit Default Swap

Credit default swaps, CDS, were introduced to the market in mid-1990s and have become an efficient hedging instrument. The instrument involves an exchange between a protection buyer and a protection seller with a reference entity as underlying instrument. The reference entity could be a corporate or a sovereign. The protection buyer pays a continuous premium to the protection seller during the tenor of the contract or until the reference entity faces a credit event or becomes insolvent. If the reference entity faces a such an event, the protection buyer can settle the CDS contract and the protection seller is obliged to cover the loss associated with the event to the protection buyer. The loss is approximated as the difference between the face value of the underlying instrument and the sum that can be recovered. The type of settlement is determined when entering the contract and can be of physical nature or cash-settled.

When pricing the CDS-premium and consequently estimation of the expected loss, the probability of default, PD, and the recovery rate, RR, are incorporated. The exact relationship can be seen in the following equation:

$$CDS\ premium = PD * (1 - RR).$$

The CDS-premium is quoted in basis points, 0.01%, on an annual basis. However, the premium is often paid quarterly. [8]

2. General Statistical Theory

Stationary Stochastic Process

An independent and identically distributed (i.i.d.) random process constitutes of a series of independent and identically distributed random variables X_1, X_2, \dots . However, real-life processes rarely express this kind of behaviour since a variable often tend to depend on the outcome of the previous variable. Not least do many real-life processes express cyclical and behavioural patterns.

For a weakly stationary, or covariance stationary, stochastic process two main characteristics holds; Firstly, the mean value function of the process is constant and do not change with time.

$$m(t) = E[X(t)] = m$$

Secondly, the covariance function only depends on the time-lag between two random variables.

$$r(s, t) = r(t - s) = r(\tau)$$

where $\tau = t - s$.

Consequently, the variance function equals

$$v(t, t) = r(t - t) = r(0)$$

hence independent of t . [5]

Financial data tend to be of non-stationary nature due to dependency between the variables $t, t+1, t+2, \dots$ but can sometimes be transformed into a weakly stationary process by e.g. detrending.

Autocorrelation Function: ACF

The auto-correlation function measures a process dependency with respect to different periods in time. By assuming X_t and X_s are stochastic variables for a process X where t and s are different time-points, the auto-correlation function, R , is defined accordingly;

$$R(s, t) = \frac{E[X_t - \mu_t]E[X_s - \mu_s]}{\sigma_t \sigma_s}$$

where μ_t, μ_s and σ_t, σ_s is the mean and standard deviation for the stochastic variables, respectively.

If R is well defined, the auto-correlation function ranges between $[-1,1]$ where -1 is completely negative correlation and 1 perfect correlation. If we assume that the process is normal then the random variables are independent if and only if the auto-correlation function takes the value of 0 which implies that the process is non-correlated between different time-lags. Accordingly, a functions dependency can be depicted graphically in auto-correlation graph with time on the x-axis and the correlation value on the y-axis. Hence the time-lag where the random variables of the process behave independently can easily be illustrated.

Maximum Likelihood

Maximum likelihood is an estimation method to find the unknown parameter θ_0 for a distribution family F . The idea is find the model with the highest likelihood, i.e. the model with highest probability given the population data. The likelihood function is hence the probability function, f , for the sample data with respect to θ , $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$. For simplification, applying the logarithm to the likelihood function is widely used and derives the log-likelihood function $l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta)$. Since the logarithm is a monotonic function, the maximum value of the log-likelihood function is identical to the maximum of the likelihood function. [1]

The benefit of the maximum likelihood estimation is that the MLE of a parameter can be substituted into a function hence gives the MLE of that function, e.g. the maximum likelihood estimation of $\eta = g(\theta)$ equals $\hat{\eta} = g(\hat{\theta})$.

Delta Method

The delta method is applied when estimating confidence intervals for functions which has multiple parameters and where the maximum likelihood estimates are assumed to be normally distributed. By taking the asymptotic normality of the MLE, the delta method proves the relation can be derived to a function of MLE with continuous derivative with respect to the parameters. Hence, if the parameter θ is a d-dimensional parameter and $\eta = g(\theta)$ is a scalar function, then the maximum likelihood function of η can be assumed to be normally distributed with a variance derived from the delta method. The

approximate variance is also known as the Gaussian approximation. Further, it can be proved that below relation holds.

$$\sqrt{n}(\hat{\eta} - \eta) \rightarrow N(0, Var(\hat{\eta}))$$

According to the delta method, the variance of a scalar function η is derived by matrix multiplication with respect to the gradient of the η and the covariance matrix of the MLE of θ . [1]

$$Var(\hat{\eta}) = \bar{v}_\eta^T \sigma_\theta \bar{v}_\eta$$

Deviance Function

The deviance function, $D(\theta) = 2\{l(\hat{\theta}_0) - l(\theta)\}$, estimates the accuracy of the maximum likelihood estimator by assuming models with high likelihood have small deviances with respect to θ . Given a threshold, c , a confidence region C can be defined as the probability that the deviance is within the specified region. Accordingly, this is defined by

$$C = \{ \theta : D(\theta) \leq c \}$$

where c is determined such the probability that C contains the true parameter value θ_0 by a pre-defined probability $(1 - \alpha)$. The deviance function follows an asymptotically chi-square distribution. [1]

Profile Likelihood

When normality assumption no longer applies, which often is the case when examine return levels of higher magnitudes, the profile likelihood can be an alternative to the maximum likelihood estimation. Assuming the parameter θ is a d-dimensional parameter vector, the profile log-likelihood for θ_i is written as:

$$l_p(\theta_i) = \max l(\theta_i, \theta_{-i})$$

where $l(\theta_i, \theta_{-i})$ is the log-likelihood for the parameter θ and θ_{-i} includes all parts of θ excluding θ_i . In turn the profile log-likelihood, $l_p(\theta_i)$, is for every value of θ_i the maximisation of the log-likelihood

with respect to all parameters θ . To derive confidence levels the deviance function is applied to the profile log-likelihood which in turn follows a chi-square distribution. [1]

Likelihood Ratio Test

The log-profile likelihood is useful for comparisons and selections of models. Let M_0 be a model with fewer parameters compared to M_1 . Assume l_0 is the log-likelihood function of M_0 and l_1 is the log-likelihood function of M_1 and then apply to the deviance function D ,

$$D = 2\{l_1(M_1) - l_0(M_0)\}$$

Then the model M_0 can be rejected in favour of M_1 at a significance level α if and only if $D > c_\alpha$, where c_α is the $(1 - \alpha)$ quantile of the chi-square distribution. [1]

Model Valuation: Goodness of Fit, Diagnostic Plots

Once a model has been assigned to describe a population of data it is of importance to validate how well the model fits and describes the data. In absence of alternative methods, the simplest way to test the accuracy is to compare the model with the populated data in which the model was derived.

Empirical Distribution Function

Definition

Given an ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ from a population with distribution function F , the empirical distribution function is defined by $\hat{F}(x) = \frac{i}{n+1}$ for $x_{(i)} \leq x \leq x_{(i+1)}$. [1]

By plotting the estimated distribution function, \tilde{F} , against the empirical distribution function, \hat{F} , two different goodness-of-fit tests can be visualised graphically to test the significance of the model; the *probability plot* and the *quantile plot*. If the data outlines in the unit diagonal, then the \tilde{F} can be accepted to be a valid model for the populated data. Likewise, if linearity is weak \tilde{F} should be rejected as an appropriate model. [1]

Probability Plot

Definition

Given an ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ from a population with estimated distribution function \tilde{F} , a probability plot consists of the points

$$\left\{ \left(\hat{F}(x_{(i)}), \frac{i}{n+1} \right) : i = 1, \dots, n \right\}. [1]$$

Quantile Plot

Definition

Given an ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ from a population with estimated distribution function \tilde{F} , a quantile plot consists of the points

$$\left\{ \left(\hat{F}^{-1} \left(\frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, n \right\}. [1]$$

As seen in the definitions above, the both models contain the same information just applied in different ways. The quantile plot takes the assumption that the quantiles of $x_{(i)}$ and $\hat{F}^{-1} \left(\frac{i}{n+1} \right)$ make up the approximation of the $\frac{i}{n+1}$ quantile of the underlying distribution F . The probability plot has the drawback that it reaches 1 for large values of $x_{(i)}$. This is the headwind of the model since large values of $x_{(i)}$ is what is of interest as far as extreme value theory is concerned.

3. Extreme Value Theory

Extreme value modeling examines the stochastic behaviour of extreme events. Let's define $M_n = \max\{X_1, \dots, X_n\}$ where X_1, \dots, X_n are independent random variables with a distribution function F in common. Therefore, M_n is the maximum stochastic variable defined for any sequence of random variables. The distribution function of M_n can be derived according to

$$\Pr (M_n \leq x) = \Pr(X_1 \leq x, \dots, X_n \leq x) = \Pr(X_1 \leq x) * \dots * \Pr(X_n \leq x) = F^n(x).$$

However, the result here is not very helpful since the distribution function F is usually unknown. To find an alternative solution, the distribution function is approximated by incorporating the normalising constants in line with the central limit theorem and letting $n \rightarrow \infty$, which gives

$$M_n^* = \frac{M_n - b_n}{a_n}, \{a_n > 0\} \text{ and } \{b_n\}.$$

As $n \rightarrow \infty$, the a_n and b_n parameters stabilise the location and scale of M_n resulting in a non-degenerate hence limiting the distribution of M_n^* . It can be shown that the limit distribution, non-degenerate of M_n^* , if it exists, must belong to one of the three possible extreme value distributions; Gumbel, Fréchet or Weibull, no matter what the underlying distribution of F is for the population. The distributions are defined as the family I, II and III, according to the theorem below. [1]

Extremal Types Theorem

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr \left\{ \frac{(M_n - b_n)}{a_n} \leq z \right\} \rightarrow G(z) \text{ as } n \rightarrow \infty, \text{ where } G \text{ is a non-degenerate distribution function, then } G$$

belongs to one of the following families:

$$I: G(z) = \exp \left\{ - \exp \left[- \left(\frac{z-b}{a} \right) \right] \right\}, \quad -\infty < z < \infty;$$

$$II: G(z) = \begin{cases} 0, & z \leq b \\ \exp \left\{ - \left(\frac{z-b}{a} \right)^{-\alpha} \right\}, & z > b \end{cases}$$

$$III: G(z) = \begin{cases} \exp \left\{ - \left[- \left(\frac{z-b}{a} \right)^\alpha \right] \right\}, & z < b \\ 1, & z \geq b \end{cases}$$

For parameters $a > 0, b$ and, in the case of families II and III, $\alpha > 0$. [1]

What differentiates the 3 families from each other is the behaviour of the tail of the distribution function F and hence it gives rise to different behaviour in terms of extreme values. The upper endpoint of the Weibull distribution is finite, meanwhile Gumbel and Fréchet upper endpoints go towards infinity. [1]

Generalized Extreme Value Distribution: GEV

The three families can be substituted into a single distribution, the Generalized Extreme Value distribution, GEV.

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z-\mu}{\sigma} \right) \right]^{-1/\xi} \right\},$$

$$-\infty < \mu < \infty, \sigma > 0, -\infty < \xi < \infty;$$

$$\text{defined for } \left\{ z: 1 + \xi \left(\frac{z-\mu}{\sigma} \right) > 0 \right\}.$$

The parameters μ, σ, ξ define the location, scale and shape of the distribution, respectively. The shape parameter, ξ , affects the shape of the tail. Hence its value in the GEV distribution corresponds to one of 3 families of distribution: $\xi > 0$ corresponds to the Fréchet distribution, $\xi < 0$ to the Weibull and $\xi = 0$, defined as the limit in which $\xi \rightarrow 0$, is the Gumbel distribution. [1]

$$G(z) = \exp \left\{ - \exp \left[- \left(\frac{z-\mu}{\sigma} \right) \right] \right\}, \quad -\infty < z < \infty.$$

By combining the three families into one distribution enables the underlying data to determine the tail behaviour and hence be assigned to the most appropriate family. [1]

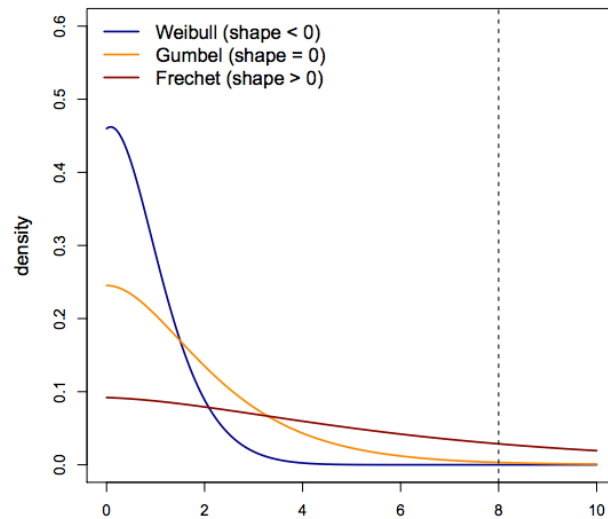


Figure 1. The tail behaviour with respect to the different families; Weibull, Gumbel and Fréchet [3]

The *extremal types theorem* in combination with the findings above let the maximum values hold the *max-stable* property which can only be satisfied if and only if the underlying random variable is a GEV distribution. The max-stable property implies that the maximum and underlying random variables are of the same type meaning that one is the location and scale transformation of the other. This property is vital when examine extreme values and applying the different approaches, Block Maxima, Peaks-over-Threshold and Probable Maximum Loss, which are the models this thesis will examine.

Block Maxima

The result above leads on to the estimation and modeling of maximum values by introducing the Block Maxima approach. The sample are assumed to be independent observations and grouped into blocks. The size of each block is determined to correspond to a monthly or annual size and hence the maxima from each block, the block maxima, M_1, \dots, M_m , is modelled. Since the extremes are of interest, investigating the quantiles of the data, $G(z_p) = 1 - p$, will be of great importance, where $1/p$ is the *return period*.

The *return level*, z_p , namely the quantile of the GEV distribution, is interpreted as the level which the annual maximas are exceeded with probability p for a given year.

Specifically, we have

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{\log(1 - p)\}^{-\xi}], & \xi \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\}, & \xi = 0. \end{cases}$$

Choosing the right block size might be a headwind. Extreme by its definition is a rare event happening, hence the size should be large and generate few maxima from the sampled data. However, a too large block size hence few maxima will lead to a larger variance. Conversely, a too small block size violates the asymptotic assumption of the extremal types theorem, but gives more data to minimize the variance error. To summarise, the block size determines a trade-off between bias and variance. [1]

Likelihood Function

By assuming the block maxima are independent variables and are assigned to the GEV distribution, the likelihood function for the GEV parameters are defined as below. [1]

$$L(\mu, \sigma, \xi) = \prod_{i=1}^m \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right]^{-1 - \frac{1}{\xi}} \exp\left\{-\left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \quad \xi \neq 0$$

Log-Likelihood Function

By taking the logarithm of the likelihood function of the GEV parameters the log-likelihood function is derived.

$$l(\mu, \sigma, \xi) = m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^m \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \quad \xi \neq 0$$

given that $1 + \xi \left(\frac{z_i - \mu}{\sigma}\right) > 0, i = 1, \dots, m$

When the shape parameter equals zero, $\xi = 0$, the Gumbel limit is used for defining the GEV distribution. Hence below equation are derived under such scenario.

$$l(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left(\frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ \left(\frac{z_i - \mu}{\sigma} \right) \right\}, \quad \xi = 0$$

The estimated parameters' $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ distribution is asymptotically multivariate normal with mean (μ, σ, ξ) and variance-covariance matrix which is the inverse of the information matrix derived from maximum likelihood estimate (MLE), assuming $\xi > -0.5$. [1]

The advantage with the Block Maxima approach is that it satisfies the assumption of independence. Meanwhile, the disadvantage is that it may include data points which should not be considered extreme due to lack of any extreme value in a given block. Additionally, the method might miss an extreme value because another more extreme already occurred within the same block. By incorporating the Peak-over-Threshold model the issues described can be made redundant. The Peak-over-Threshold model is therefore considered to be an alternative and more suitable model if the whole dataset is available (as opposed to having only extreme value data).

Peak over Threshold: POT

The Peak-over-Threshold, POT, approach only considers the extreme events over a predefined threshold u . The number of exceedances above the threshold, N , is considered to follow a Poisson process and the sizes of the arbitrary exceedances, $X_i - u$, are independent from N and Generalized Pareto distributed. The advantage of the POT-model is that the parameters are stable as of an increasing threshold and more extreme exceedances have the same shape parameters as less extreme ones. This implies that parameters from higher levels can be derived from parameters from lower levels, and most importantly the shape parameter, ξ , is the same for the GP-family as of for the GEV family. See below precise mathematical formulations of these properties.

Generalized Pareto Distribution: GPD

Let u define a threshold in which exceedances, X_i , above that threshold is considered extreme. By applying conditional probability, given that the parent distribution F is known, the distribution of the exceedances can be derived. However, in most cases the parent distribution F is unknown hence the GEV distribution is assigned to the exceedances. Accordingly, following theorem is defined.

Theorem

Let X_1, X_2, \dots be a sequence of independent random variables with common distribution function F , and let $M_n = \max\{X_1, \dots, X_n\}$. Denote an arbitrary term in the X_i sequence by X , and suppose that F satisfies Theorem 3.1.1, so that for large n ,

$$\Pr\{M_n \leq z\} \approx G(z),$$

where

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for some $\mu, \sigma > 0$ and ξ . Then, for large enough u , the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \xi\frac{y}{\tilde{\sigma}}\right)^{-1/\xi}$$

defined on $\{y: y > 0 \text{ and } (1 + \xi\frac{y}{\tilde{\sigma}}) > 0\}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$. [1]

If $\xi < 0$ the exceedances have an upper limit of $\mu - \frac{\tilde{\sigma}}{\xi}$. Otherwise, if $\xi \geq 0$, there do not exist an upper limit. [1]

The importance of this theorem is that if the Block Maxima can be approximated by the $G(z)$ function above, then the exceedances over the predefined threshold u can be approximated to the function $H(y)$, namely *the Generalized Pareto Family, GP-family*. The advantage of this conclusion is that the shape parameter, ξ , which plays an important part of the behaviour of the distribution and extreme data, is the same for the *GP-family* as of for the *GEV family*.

The case $\xi = 0$, interpreted as $\xi \rightarrow 0$, in which $H(y)$ is equivalent to an exponential distribution with parameter $\frac{1}{\tilde{\sigma}}$, is defined accordingly: [1]

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y > 0$$

Modeling Threshold Exceedances

In line with the Block Maxima, choosing the right size of the threshold is a trade-off between the bias of the model versus the variance of the extreme data. The benchmark is generally to choose a threshold low as possible given that the limit theorem approximates from enough data points. Below follow two methods regarding how to choose an appropriate threshold. The first method, the mean residual life plot, derives from the conditional mean of GPD and is applied before model approximation. The second method applies a spectrum of thresholds where the shape parameter, ξ , is assumed to be constant above u_0 and approximations of the scale parameter, σ_u , are linear in u . [1]

Mean Residual Life Plot

The mean residual life plot derives from the mean of the GPD. Let's assume that GPD is sufficient describing the exceedances u_0 originated from the stochastic series X_1, \dots, X_n . Further assume that the $\xi < 1$. Then the mean of the GPD is:

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi}$$

If GPD is sufficient explaining the exceedances over the threshold u_0 , then implicitly GPD is sufficient explaining exceedances with a larger threshold u , such that $u > u_0$. Let $\sigma_u = \sigma_{u_0} + \xi(u - \mu)$.

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}$$

Hence, as seen in the equation above, the conditional mean is a linear function with regards to u . The result above implicates by taking the empirical sample mean of the threshold exceedances, a certain locus of points can be used when graphically determining the threshold. If the GPD is sufficient explaining the exceedances, then the mean residual plot should be linear in u above the threshold.

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{max} \right\}$$

The $x_{(1)}, \dots, x_{(n_u)}$ is the n_u observations that violets u and x_{max} is the maximum of the X_i .

Confidence intervals are included in the plot to measure the uncertainty in the estimates of the theoretical sample means which follow approximately a normal distribution. By examining the mean residual life plot, the threshold u is to be chosen so that there is a linear line up to the chosen threshold before it decays sharply. [1]

Parameter Stability Plot: Model Based Approach

An alternative method to the mean residual life plot is to model multiple thresholds where the estimated parameters are approximately constant and then pick the smallest threshold u in which linearity is constant to higher values and within the confidence levels.

As in the previous method, if GPD is sufficient describing the exceedances over the threshold u_0 , then GPD is sufficient describing the exceedances over $u > u_0$ or $u + u_0$. The shape parameter is constant regardless of the chosen threshold meanwhile the scale parameter changes as the threshold changes; $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$. By reparametrize the scale parameter and letting it be constant as the threshold changes this issue is encountered:

$$\sigma^* = \sigma_u - \xi u.$$

Confidence intervals for the estimated $\tilde{\xi}$ are derived from the variance-covariance matrix and the delta method. [1]

Parameter Estimation with Maximum Likelihood Method

After the threshold has been determined, parameter estimation is obtained by applying the maximum likelihood method. Let k be the number of exceedances over the threshold u in which is defined as x_1, \dots, x_k then the below log-likelihood functions are defined: [1]

$$l(\sigma, \xi) = -k \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \log \left(1 + \xi \frac{x_i}{\sigma}\right), \quad \xi \neq 0$$

given that $(1 + \xi \frac{x_i}{\sigma}) > 0, i = 1, \dots, k$

$$l(\sigma) = -k \log \sigma - \frac{1}{\sigma} \sum_{i=1}^k x_i, \quad \xi = 0$$

Return Level

The behaviour of the quantiles is of most interest as far as extreme value modeling is concerned, not just the derived parameters. By assuming the GPD, the shape and the scale parameters are sufficient explaining the threshold exceedances, the *return level*, x_m , can be derived. As with the Block Maxima approach, the *return level*, x_m , is the level in which on average is exceeded once every m observation. The *return level*, x_m can be derived accordingly:

$$Pr(X > x | X > u) = \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

Given $x > u$,

$$Pr(X > x | X > u) = \frac{Pr(X > x \cap X > u)}{Pr(X > u)} = \frac{Pr(X > x)}{Pr(X > u)}$$

Let $\zeta_u = Pr(X > u)$

$$Pr(X > x) = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

The *m-observation return level* x_m is then,

$$x_m = u + \frac{\sigma}{\xi} \left[(m\zeta_u)^\xi - 1 \right], \quad \xi \neq 0$$

$$x_m = u + \sigma \log(m\zeta_u), \quad \xi = 0$$

Plotting the *m-observation return level* x_m on a logarithmic scale against m , the shape parameter is evaluated in the same way as with the GEV model. Linearity satisfies $\xi = 0$, convexity $\xi < 0$ and concavity $\xi > 0$. [1]

Return Level Plot

Expressing data in terms of quantiles gives a simply way to graphically estimate the tail behaviour and assign the data to the right family. Let's define $y_p = -\log(1 - p)$ in the equation expressing the *return level*, z_p . By plotting y_p against the *return level*, z_p , the *return level plot* is derived and graphically depicts different shape parameters. The plot is linear when $\xi = 0$, convex when $\xi < 0$ and concave when $\xi > 0$. The dashed lines in the plot are the confidence intervals which are increasing at larger return levels implying more uncertainty and risk are incorporated at higher return levels. [1]

Restriction for Return Levels

Examining small values of p hence investigating long return periods is of interest as far as extreme modeling is concerned. Further, if $\hat{\xi} < 0$, investigation of the upper end-point of the distribution is also of interest since the upper-end point then can then be estimated as the "*infinite observation period*" which is the return period z_p when $p = 0$. Below follows the maximum likelihood of the upper end-point when $\hat{\xi} < 0$.

$$\hat{z}_0 = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}$$

If $\hat{\xi} \geq 0$ the upper end-point goes to infinity.

The delta method gives:

$$Var(\hat{z}_0) \approx \nabla z_0^T V \nabla z_0,$$

where

$$\nabla z_0^T = \left[\frac{\partial z_0}{\partial \mu}, \frac{\partial z_0}{\partial \sigma}, \frac{\partial z_0}{\partial \xi} \right] = [1, -\xi^{-1}, \sigma \xi^{-2}]$$

$$V = Var(\hat{\mu}, \hat{\sigma}, \hat{\xi})$$

under the assumption of the estimated parameters $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. [1]

Model Comparison

The advantage with the Peak-over-Threshold model compared to the Block Maxima approach is that the POT analyses more of the available data if the threshold is set sufficiently. It is vital not to reject any data considered extreme since sufficient data points are necessary to apply the model. Too few data points will not give as good approximation of the underlying distribution and moreover lead to a too narrow confidence interval for the return level. POT can therefore be regarded to be a better approach. The disadvantage of the POT model is that exceedances tend to cluster above high thresholds.

Dependency and Declustering

The models described so far assume the extreme values are independent which might not be the case in reality. If the underlying observations are dependent, the extreme events tend to cluster and not fulfil the properties of a stochastic process where statistical properties are unchanged with time. As far as independence of extreme values is concerned, the Block Maxima approach fulfils the independent property better than the POT-model. Hence when applying the POT-model, the extreme values above the threshold need to be tested for independency and clustering. If exceedances are dependent and clustered, *declustering* methods need to be applied. Only thereafter, the cluster maxima can be considered independent over a predefined threshold level. This is applied by empirically define clusters considered extreme, extract the maximum observation from each cluster and fit the maxima to a GPD. This reasoning is mathematically valid since a stationary process of independent variables with the same marginal distribution as another stationary process will have a limit distribution associated with the other stationary process' limit distribution.

$$G(z) \approx G^\theta(z)$$

The difference between the two distributions are an extremal index, θ , defined as $0 < \theta \leq 1$. The extremal index takes the value of 1 for independent series. The above relation is an important assumption, consequently, if $G(z)$ belongs to a GEV distribution, $G^\theta(z)$ will similarly do. Likewise, if $G(z)$ belongs to a Gumbel distribution, $G^\theta(z)$ will also do. Interestingly, the shape parameter, ξ , is the same for the two distributions meanwhile the location and scale parameters will have following relations. [1]

$$\mu^* = \mu - \frac{\sigma}{\xi} (1 - \theta^{-\xi}) \text{ and } \sigma^* = \sigma \theta^\xi, \quad \xi \neq 0$$

$$\mu^* = \mu + \sigma \log \theta \text{ and } \sigma^* = \sigma, \quad \xi = 0$$

Declustering Approach

The *declustering approach* is made in the following steps. Firstly, clusters considered extreme are empirically defined. Secondly, the maximum observation in every cluster is established. Thirdly, the maximum observation within each cluster is assumed to be independent and belong to the Generalized Pareto distribution. Finally, the maxima in each cluster is fitted to the GPD.

Declustering is applied by defining a threshold, u , and let the exceedances above the threshold be populated to the same cluster. When a predefined number of observations are below the threshold the cluster is considered to have ended and a new cluster is defined. Consequently, a cluster is considered to be open until a predetermined amount of observations, r , no longer exceeds the threshold u . Let's call r "the minimum gap between clusters". Defining a new cluster is a trade-off between how many observations to be below the threshold and at what level is appropriate to set the threshold. Hence, as with the POT-model, choosing the right value of r is a trade-off between independency versus loss of valuable data, a trade of between bias and variance. If r is chosen too small the independency between the clusters are violated. However, a too large r will make the variance large and clusters, which could be considered independent, are not fully used and loss of valuable data is a fact. The best practise is to check the sensitivity of the data for different values of r in order to find the most suitable value.

The declustering approach is not straightforward when taking both u and r into account since the scale parameter, σ , will change in line with threshold meanwhile the shape parameter, ξ , is stable regards to u and r respectively. Therefore, an alternative approach is to check the model significance by applying the return level plot. Moreover, a higher u and r make the standard errors less significant and need to be considered when comparing different return levels.

Since the intensity of when clusters occur is of importance, the *m-observation return level* is defined as:

$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u\theta)^\xi - 1]$$

The scale parameter, σ , and shape parameter, ξ , above are the parameters of the threshold exceedances and distributed according to the Generalized Pareto distribution. The extremal index, $\hat{\theta}$, is defined as the ratio of *number of clusters above the threshold u* , n_c , with respect to *number of exceedances of the threshold u* , n_u . Further, ζ_u , is the *probability of an exceedance of u* .

$$\hat{\theta} = \frac{n_c}{n_u}$$

$$\hat{\zeta}_u = \frac{n_u}{n}$$

Hence $\zeta_u \theta \sim \frac{n_c}{n}$. [1]

Probable Maximum Loss: PML

In order to predict future probable worse events, *the Probable Maximum Loss, PML*, is a statistical approach which has been developed and allows for considering risk levels with respect to a range of different time horizons. Implicitly return levels will differ depending on risk preferences, r , and time periods, T . Hence plotting the upper quantile versus time will depict the PML function.

Let M_T be the maximum observation over a given threshold u and under time period T . Then there are two possible outcomes under the time period T ; either $M_T \leq u$ or there exist at least one observation which violates the threshold, hence $M_T = (u + \text{the largest exceedance})$.

If the Poisson process does not have any points in the time period $[0, T]$, following relation $M_T \leq u + v$ applies and below equation from the POT-model is derived:

$$P(M_T \leq u + v) = \exp \left\{ -\lambda T \left(1 + \xi \frac{v}{\sigma} \right)_+^{-1/\xi} \right\} = \exp \left\{ - \left(1 + \xi \frac{v - ((\lambda T)^\xi - 1) \frac{\sigma}{\xi}}{\sigma (\lambda T)^\xi} \right)_+^{-1/\xi} \right\}$$

The lambda, λ , is defined as the *ratio of number of clusters, n_c , with respect to number of years of observation, n_{years}* , $\hat{\lambda} = \frac{n_c}{n_{years}}$.

Since the p^{th} - quantile is of interest the probability density function of $M_T \leq u + v$ equals $1 - p$.

$$P(M_T \leq u + v) = 1 - p$$

The return level with respect to the risk level p and the time-period T is derived by solving for v . Hence $x_{T,p}$ is the quantile for the PML with respect to risk level p and time period T and is solved by inserting the parameters derived from the POT-model. Consequently, the PML return level is defined as:

$$x_{T,p} = u + \frac{\sigma}{\xi} \left(\frac{(\lambda T)^\xi}{(-\log(1-p))^\xi} - 1 \right)$$

The PML return level, $x_{T,p}$, derived above is an simple and intuitive method to estimate a given worst loss with regards to chosen risk level and time horizon. [6]

4. Analysis

Following analysis is based on monthly CDS data from Deutsche Bank AG EUR CDS 5Y between the time periods of August 2001 to April 2018. In order to handle the data and make statistical assumptions, MATLAB R2017b and RStudio is used in line with the package in2Extreme [2][3][7].

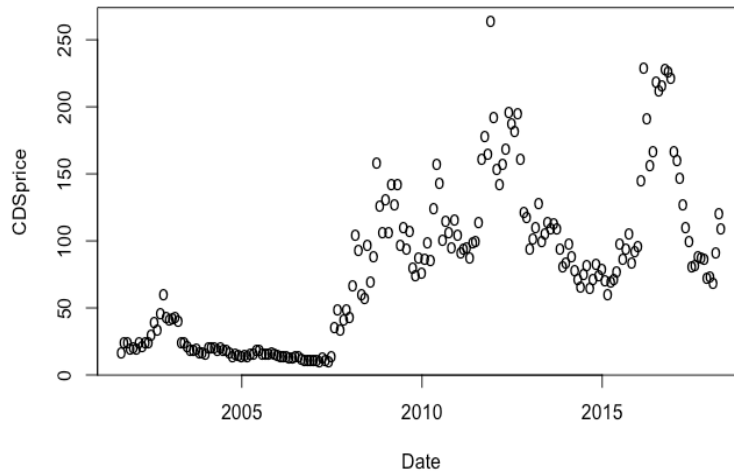


Figure 2. Scatter plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread. CDS price in EUR versus monthly data from August 2001 to April 2018

By investigating the CDS price distributions in Figure 2, one can assume dependency and non-stationarity between the CDS prices on a monthly basis, where one lag is equivalent to the CDS price of one month. To prove the assumption of dependency mathematically, the autocorrelation function is applied to the time series, from August 2001 to April 2018 and can be seen in Figure 3 below.



Figure 3. Auto-correlation function of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread. Monthly data from August 2001 to April 2018

By further investigation of the autocorrelation function, even a time lag higher than 20 months is significant and correlation of monthly CDS prices more than 20 months is confirmed. Assumption of independence and non-stationarity are vital as far as mathematical models and extreme value models are concerned and need to be interpreted in order to make sufficient conclusions of the provided outcomes. However, financial data rarely demonstrates those characteristics due to the dependency between the variables $t, t+1, t+2, \dots$ and dividing up the time series might be necessary to achieve weakly stationarity in the data series. As proven in Figure 3, independency between the monthly CDS prices is rejected and the time series need to be divided up accordingly to solve the issue of non-stationarity between the data points.

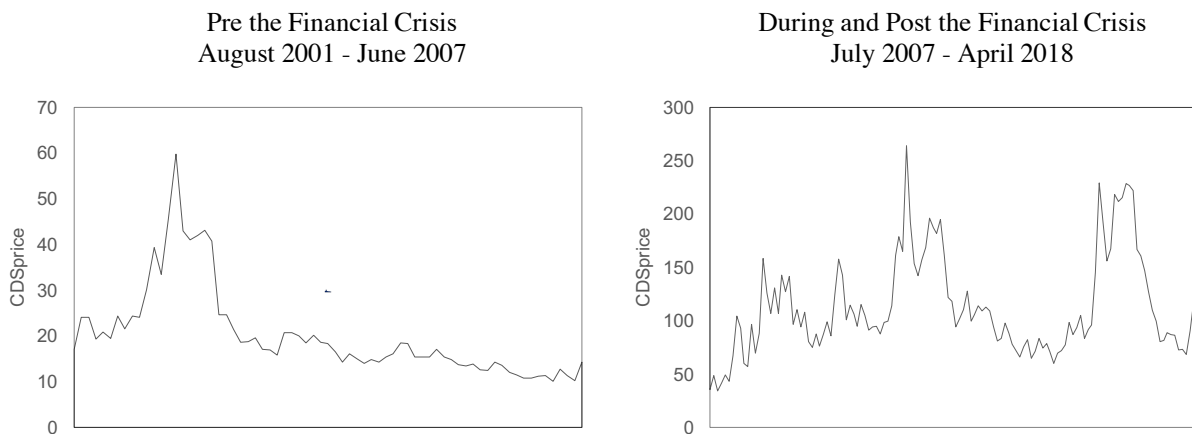


Figure 4 and 5. Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread. Monthly data pre- and during/post the financial crisis

In Figure 4 and 5, the CDS prices pre- and post the financial crisis is virtually depicted and below in Figure 6 and 7 its autocorrelation functions are presented.

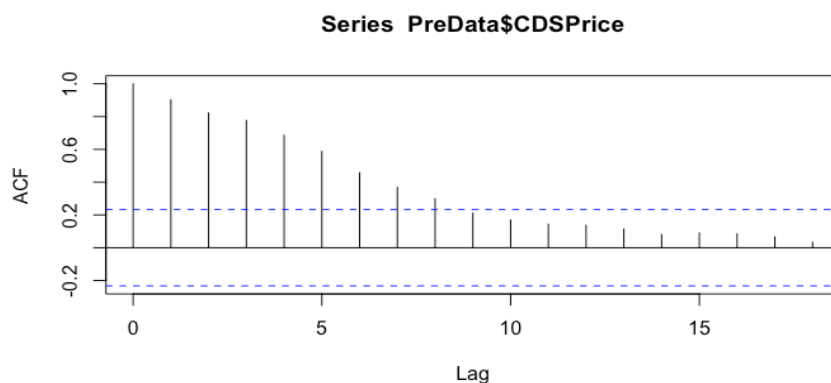


Figure 6. Auto-correlation function of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread before the Financial crisis (August 2001 to June 2007)

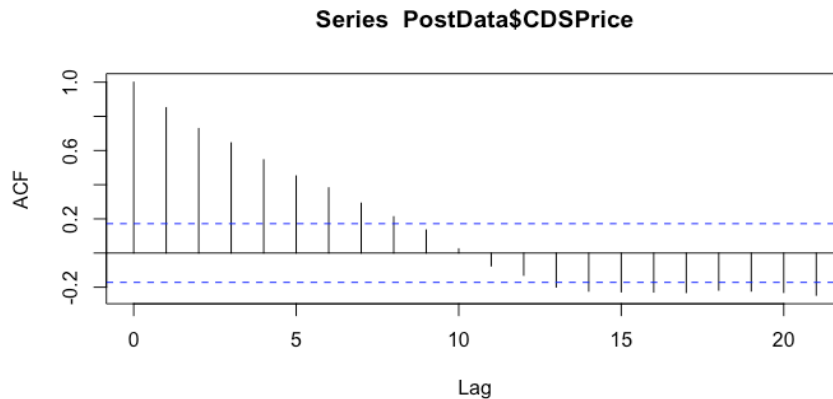


Figure 7. Auto-correlation function of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Interestingly, the autocorrelation functions indicate that the dependency in the CDS prices are more significant before the financial crisis relative to during and after the crisis. Analysis of the autocorrelation functions for both before, during and after the crisis, indicate that CDS prices can be assumed to be independent after 9 months. What is noteworthy, the autocorrelation for the CDS prices decays more sharply during and after the crisis, even to become negative correlated after a time lag of 11 months, compared to before the crisis.

Independency and non-stationarity of the CDS prices are the first and far most important assumption to be established and can now be assumed to be valid for time lags of 9 months. The CDS data during and post the financial crisis is what is of most interest as far as this thesis is concerned. Hence only the post data will be considered going forward.

Block Maxima

Block Maxima in tandem with GEV distribution is the first approach to describe the behaviour of the underlying data. Hence the GEV distribution can be seen as the modeled density function in Figure 8. The linearity in the probability and quantile plots below indicates that the GEV distribution could be considered an appropriate fit to the underlying CDS price data for shorter return periods.

Generalized Extreme Value Distribution: GEV

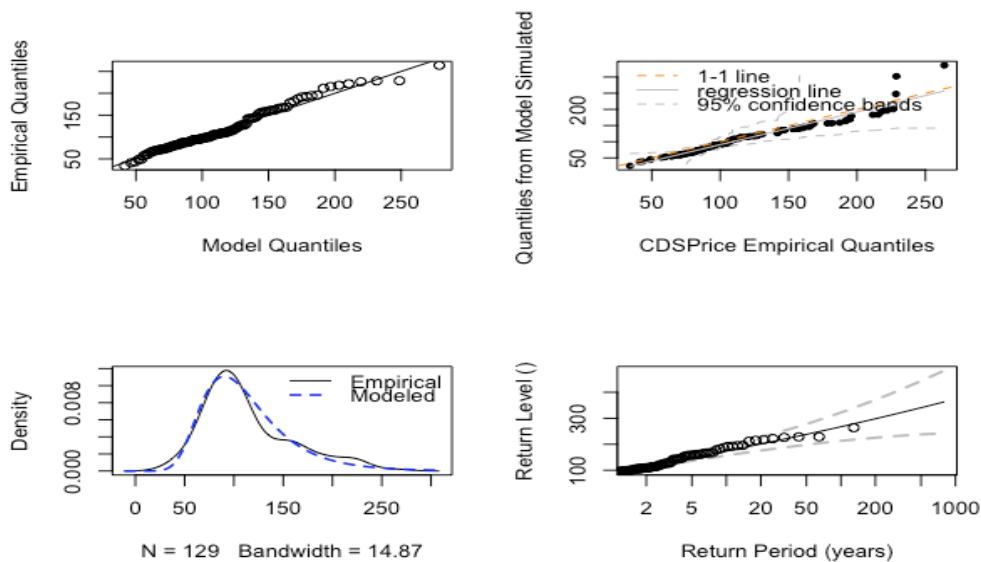


Figure 8. GEV Distribution
 Probability Plot, Quantile Plot, Density and Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Parameter Estimates

In order to find suitable parameter estimates of the location, scale and shape parameters, the maximum likelihood, the normal approximation, the delta method and the profile likelihood are all applied. The normal approximation is calculated with 95% confidence interval where μ is the estimate of the location, scale and shape parameter, $z_\alpha = 1.96$ and Var is derived from the estimated parameter covariance matrix. The confidence interval apply as the equation below.

$$CI = \mu \pm z_\alpha \sqrt{Var}$$

Maximum Likelihood

Table 1. Parameter estimates with Maximum Likelihood

	Location	Scale	Shape
Parameter Estimates	92.1144	33.2539	0.0575

Standard Error Estimates	3.2817	2.4188	0.0650
---------------------------------	--------	--------	--------

Table 2. Estimated parameter covariance matrix with Maximum Likelihood

	Location	Scale	Shape
Location	10.7697	3.4043	-0.0697
Scale	3.4043	5.8508	-0.0293
Shape	-0.0697	-0.0293	0.0042

Table 3. Estimated Hessian Matrix with Maximum Likelihood

Hessian Matrix	Location	Scale	Shape
Location	0.1233	-0.0638	1.5885
Scale	-0.0638	0.2101	0.4049
Shape	1.5885	0.4049	265.3682

Normal Approximation

Table 4. Parameter estimates with Normal Approximation

	95% Lower Bound	Mean	95% Upper Bound
Location	85.6822	92.1144	98.5466
Scale	28.5130	33.2539	37.9948
Shape	-0.0695	0.0575	0.1845

Delta Method

Table 5. Parameter estimates with Delta Method

	95% Lower Bound	Mean	95% Upper Bound
Location	85.6823	92.1144	98.5464
Scale	28.5131	33.2539	37.9947
Shape	-0.0699	0.0575	0.1850

Profile Likelihood of Shape Parameter

Table 6. 95 % Confidence intervals for shape parameter based on Profile Likelihood

	95% Lower Bound	Mean	95% Upper Bound
Shape	-0.0644	0.0580	0.1842

The token of the shape parameter is what is of greatest importance as far as the behaviour of the tail of the distribution is concerned. In all four cases, the lower bound of the confidence interval gives a negative value of the shape parameter, which implies finite upper bound of the underlying distribution.

Remarkably, the confidence interval between the normal approximation, delta and profile likelihood do not differ from each other in a wider perspective.

Return Levels

To find an accurate approximation of the price of risk for a range of years, namely indicative return levels for a range of years, below return levels are estimated with respect to the delta method and the profile likelihood respectively in line with the above estimated parameters. Return levels are interpreted as the CDS price in EUR for a range of years.

Delta Method

Table 7: Estimated return levels based on Delta Method

Return Levels	95% Lower Bound	Mean	95% Upper Bound
2 Years	97.16	104.43	111.70
5 Years	132.90	144.21	155.52
8 Years	148.55	163.10	177.66
10 Years	155.51	172.01	188.51
15 Years	167.48	188.23	208.99
20 Years	175.46	199.83	224.20
30 Years	186.04	216.36	246.69

Profile Likelihood

Table 8: Estimated return levels based on Profile Likelihood

Return Levels	95% Lower Bound	Mean	95% Upper Bound
2 Years	97.47	104.43	111.96
5 Years	133.96	144.21	156.94
8 Years	150.32	163.10	180.50
10 Years	157.92	172.01	192.24
15 Years	171.14	188.23	214.79
20 Years	180.21	199.83	231.62
30 Years	192.73	216.36	256.97

The mean value of the return levels with respect to the delta and the profile likelihood do not differ as much, meanwhile a wider difference is noticed in the confidence intervals between the two approaches. This should be expected since normal distribution assumption may not apply and therefore the profile likelihood gives a better estimation of the mean and confidence intervals of the different return levels. Profile likelihood plots for respective return levels are graphically depicted in Figure 9 and the return level plot for different return levels in Figure 10.

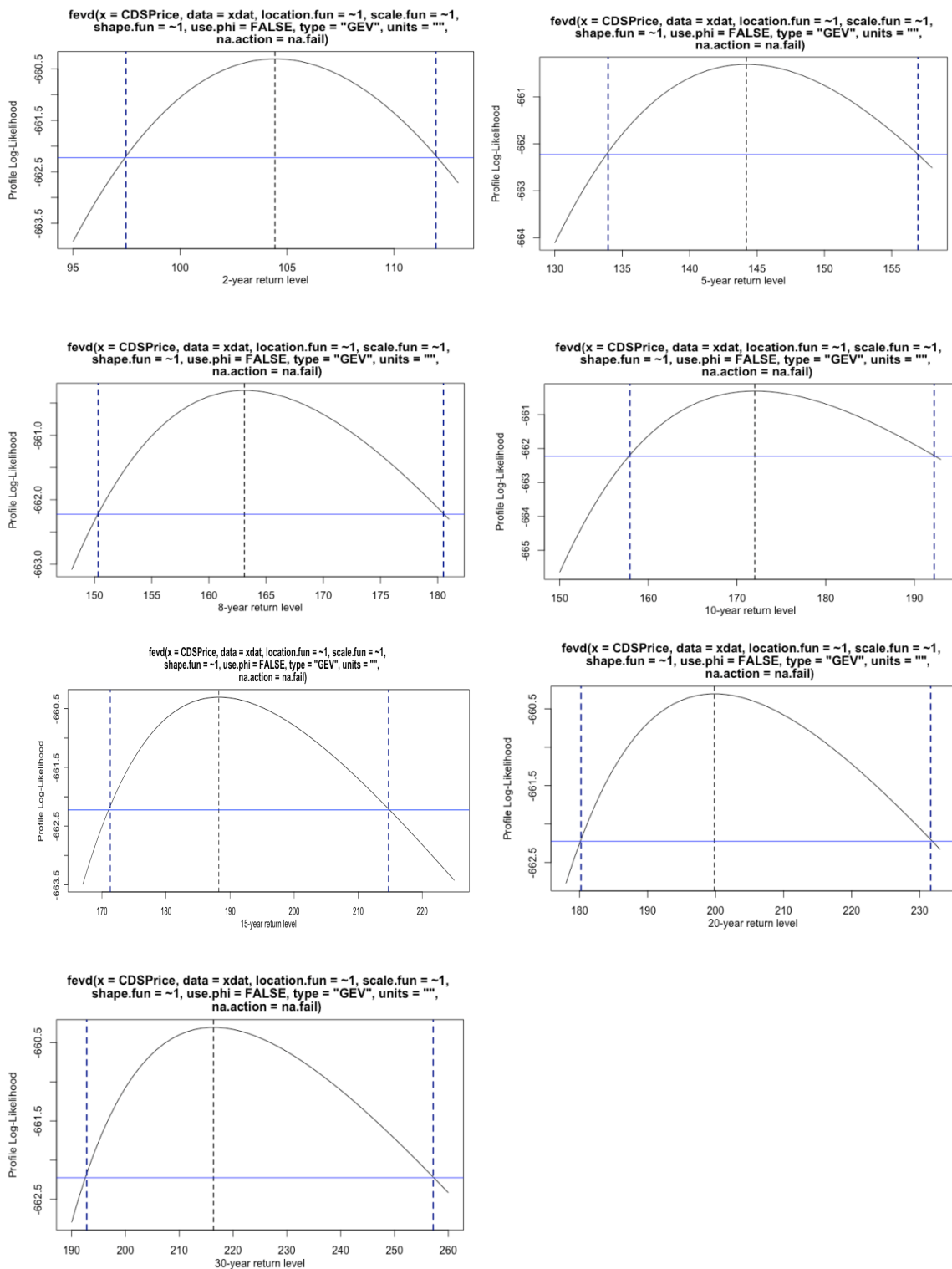


Figure 9. GEV Distribution. Profile Likelihood Plots for return levels 2-years, 5-years, 8-years, 10-years, 15-years, 20-years, 25-years and 30-years of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

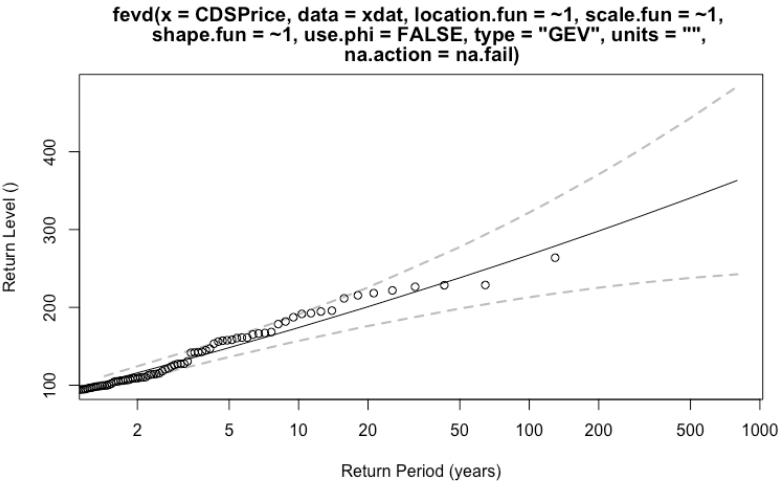


Figure 10. GEV Distribution. Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Gumbel Distribution

When examining the shape parameter's mean value and confidence interval it might take the value of zero. Not least since the lower bound of the confidence interval gives rise to a negative value of the shape parameter. Applying the Gumbel distribution with a shape parameter equal to zero is therefore also considered when applying the Block Maxima and is depicted as the modeled density function in Figure 11. The moderate linearity in the probability and quantile plot supports this hypothesis.

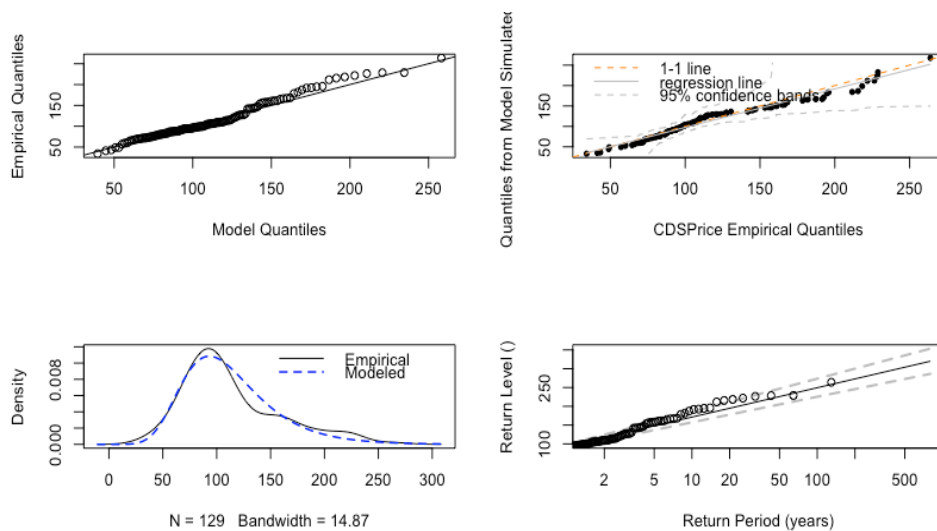


Figure 11. Gumbel Distribution Probability Plot, Quantile Plot, Density and Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Parameter Estimates

As with the GEV distribution, parameter estimates are performed with respect to maximum likelihood, normal approximation and the delta method when applying the Gumbel distribution. As before, the normal approximation is done with a 95% confidence interval where μ is the estimate of the location and scale parameter, $z_\alpha = 1.96$ and Var is derived from the estimated parameter covariance matrix.

$$CI = \mu \pm z_\alpha \sqrt{Var}$$

Maximum Likelihood

Table 9. Parameter estimates with Maximum Likelihood

	Location	Scale
Parameter Estimates	93.1489	33.9252
Standard Error Estimates	3.1361	2.3758

Table 10. Estimated parameter covariance matrix with Maximum Likelihood

	Location	Scale
Location	9.8353	2.2715
Scale	2.2714	5.6443

Table 11. Estimated Hessian Matrix with Maximum Likelihood

	Location	Scale
Location	0.1121	-0.0451
Scale	-0.0451	0.1953

Normal Approximation

Table 12. Parameter estimates with Normal Approximation

	95% Lower Bound	Mean	95% Upper Bound
Location	85.6822	92.1144	98.5466
Scale	28.5130	33.2539	37.9948

Delta Method

Table 13. Parameter estimates with Delta Method

	95% Lower Bound	Mean	95% Upper Bound
Location	85.6823	92.1144	98.5464
Scale	28.5131	33.2539	37.9947

The maximum likelihood method for the location and scale parameter give different values for the Gumbel distribution compared to the GEV distribution. Meanwhile the normal approximation and the delta method give the same estimates of the parameters and confidence intervals. The return level plot for a range of return periods is depicted in Figure 12.

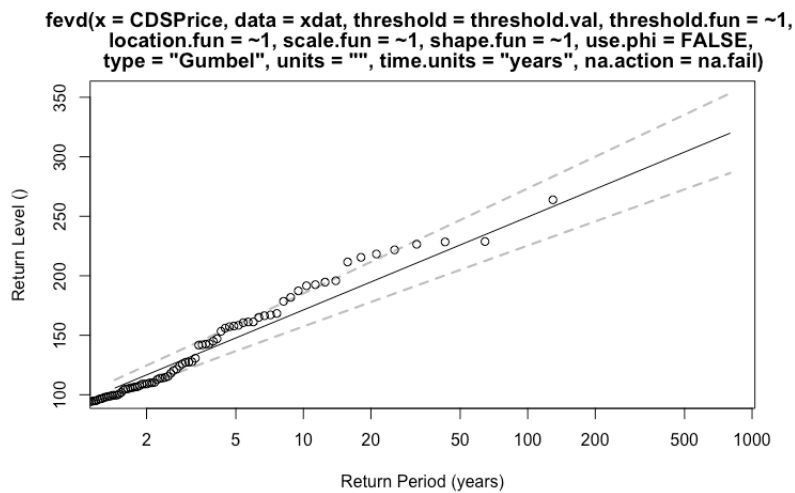


Figure 12. Gumbel Distribution. Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Likelihood Ratio Test

The likelihood ratio test is performed to find out which distribution, the GEV or the Gumbel, is of most relevance and describes the CDS price data most accurately. The test is set up with the Gumbel distribution as the null hypothesis and the GEV distribution as the alternative hypothesis. The likelihood ratio test rejects the alternative hypothesis since the p-value is above the significance level of 0.05. Conclusively, the distribution of the CDS price under and after the financial crisis behave as a Gumbel distribution with a shape parameter equal to zero.

$$H_0: \gamma = 0$$

$$H_1: \gamma \neq 0$$

Table 14. Likelihood Ratio Test with Gumbel Distribution as null hypothesis and GEV Distribution as alternative hypothesis

Likelihood Ratio	0.8081
Chi-Square Critical Value	3.8415
Alpha	0.05
Degrees of Freedom	1
P-value	0.3687
Alternative Hypothesis is Greater	

Peak-over-Threshold Method

The Peak-over-Threshold method gives another interpretation of describing the underlying data compared to the Block Maxima approach. Firstly, an appropriate threshold is set up with help of the mean residual life plot and the model based approach. Then the observations over the threshold is fitted to a Generalized Pareto distribution.

Mean Residual Life Plot

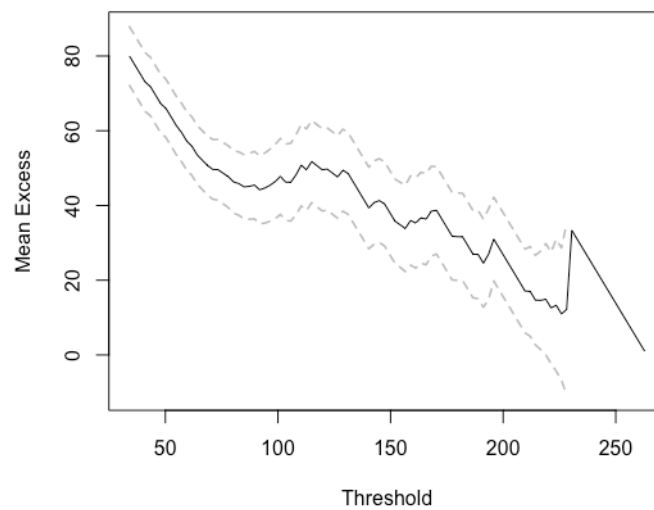


Figure 13. Mean Residual Life Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Appropriate Threshold:

130

Model Based Approach

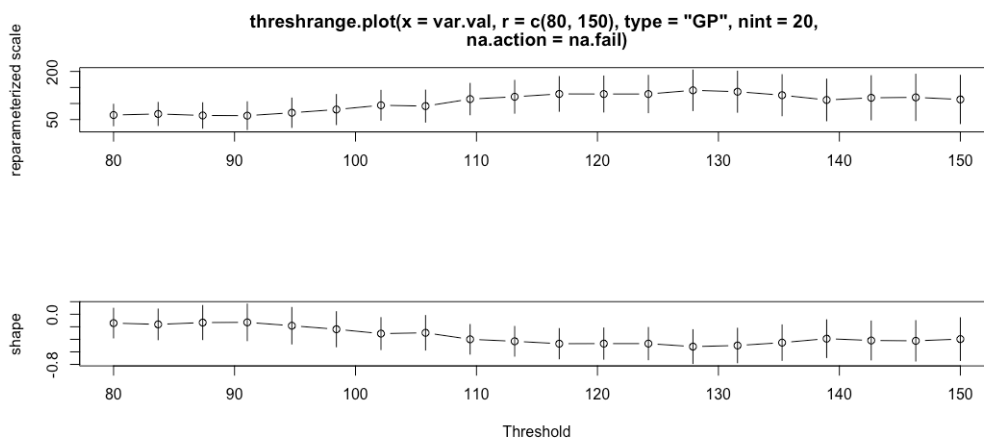


Figure 14. Model Based Approach of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Appropriate Threshold:

130

In both cases, when applying the mean residual plot and the model based approach a threshold value of 130 is considered to be appropriate due to the linearity in the models. Hence the parameter estimations are constructed from that perspective accordingly.

Generalized Pareto Distribution: GPD

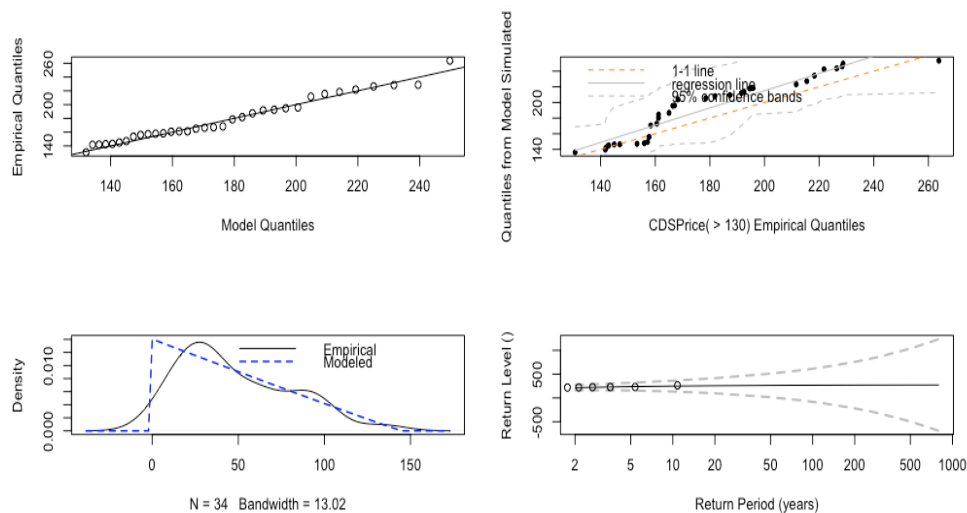


Figure 15. GP Distribution Probability Plot, Quantile Plot, Density and Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

The exceedances, which is the observations above a threshold value of 130, is fitted to a GP distribution and depicted as the modeled density function in Figure 15. The relative linearity in the probability plot is an indication that the exceedances can be fitted to a GP distribution. However, this can be questionable when investigating the quantile plot since the linearity is not as substantial there. Hence the POT model could seem ambiguous for higher return levels.

Parameter Estimates

Parameter estimates of the exceedances with respect to GP distribution is done with maximum likelihood, normal approximation and profile likelihood. As before, normal approximation is done with 95% confidence interval based on asymptotic normal distribution of the maximum likelihood estimates. μ is the estimate of the scale and shape parameters, $z_\alpha = 1.96$ and Var is derived from the estimated parameter covariance matrix.

Maximum Likelihood

Table 15. Parameter estimates with Maximum Likelihood and threshold value of 130

	Scale	Shape
Parameter Estimates	71.2500	-0.4911
Standard Error Estimates	14.4758	0.1382

Table 16. Estimated parameter covariance matrix with Maximum Likelihood and threshold value of 130

	Scale	Shape
Scale	209.5479	-1.8804
Shape	-1.8804	0.0191

Table 17. Estimated Hessian Matrix with Maximum Likelihood and threshold value of 130

	Scale	Shape
Scale	0.0410	4.0400
Shape	4.0400	450.0500

Normal Approximation

Table 18. Parameter estimates with Normal Approximation and threshold value of 130

Normal Approximation	95% Lower Bound	Mean	95% Upper Bound
Scale	42.8775	71.2500	99.6225
Shape	-0.7620	-0.4911	-0.2202

Delta Method

Table 19. Parameter estimates with Delta Method and threshold value of 130

Delta Method	95% Lower Bound	Mean	95% Upper Bound
Scale	42.8780	71.2500	99.6220
Shape	-0.7619	-0.4911	-0.2202

Profile Likelihood and Estimation of Upper Endpoint

Table 20. 95 % Confidence intervals for shape parameter based on Profile Likelihood with threshold value of 130

	95% Lower Bound	Mean	95% Upper Bound
Scale	71.4461	71.2500	107.8543
Shape	-0.4848	-0.491	-0.1643

The shape parameter is negative when applying the maximum likelihood, normal approximation, the delta method and the profile likelihood. Hence the upper end-point of the return levels, \hat{z}_0 , should exist

and is therefore estimated with maximum likelihood. Additionally, confidence intervals are set up accordingly by applying normal approximation with 95% confidence interval. $z_\alpha = 1.96$ and Var is derived from the delta method where V is the variance-covariance matrix of the estimated parameters. The equations are shown below.

$$\hat{z}_0 = -\frac{\hat{\sigma}}{\hat{\xi}}$$

$$Var(\hat{z}_0) \approx \nabla z_0^T V \nabla z_0$$

$$V = Cov(\hat{\sigma}, \hat{\xi})$$

$$\nabla z_0^T = \left[\frac{\partial z_0}{\partial \sigma}, \frac{\partial z_0}{\partial \xi} \right] = [-\xi^{-1}, \sigma \xi^{-2}]$$

$$CI = \pm z_\alpha \sqrt{Var(\hat{z}_0)}$$

Table 21. Maximum Likelihood estimate of upper endpoint of distribution with threshold value of 130

	Estimate
Scale	71.2500
Shape	-0.4910
Upper End-Point	145.1120

Table 22. Covariance estimates of the scale and shape parameter with a threshold value of 130

	Scale	Shape
Scale	209.5479	-1.8804
Shape	-1.8804	0.0191

$$Var(\hat{z}_0) = 4,801.2$$

$$CI = \pm z_\alpha \sqrt{Var(\hat{z}_0)} = \pm 1.96 \sqrt{4,801.2} = 135.81$$

Table 23. Parameter estimates with Normal Approximation of the upper endpoint and threshold value of 130

Normal Approximation	95% Lower Bound	Mean	95% Upper Bound
Upper Endpoint	9.3019	145.1120	280.9221

Return Levels

Return levels, namely the approximation of the price of credit risk, are estimated for a range of years by applying the POT model based for the delta and the profile likelihood method. The mean of the return levels and its confidence intervals are seen in Table 24 and 25. As mention before, return levels are the CDS price in EUR for a range of years.

Delta Method

Table 24: Estimated return levels based on Delta Method with threshold value of 130

Return Levels	95% Lower Bound	Mean	95% Upper Bound
2 Years	162.42	216.45	270.47
5 Years	156.15	237.70	319.25
8 Years	143.60	245.41	347.21
10 Years	135.20	248.49	361.77
15 Years	115.51	253.29	391.07
20 Years	97.72	256.16	414.60
30 Years	66.53	259.58	452.63

Profile Likelihood

Table 25: Estimated return levels based on Profile Likelihood with threshold value of 130

Return Levels	95% Lower Bound	Mean	95% Upper Bound
2 Years	210.05	216.45	235.37
5 Years	229.49	237.70	262.96
8 Years	236.41	245.41	276.60
10 Years	239.46	248.49	282.94
15 Years	244.07	253.29	293.65
20 Years	246.73	256.16	301.17
30 Years	249.72	259.58	310.82

As with Block Maxima, the confidence intervals for the return levels differ between the delta method and the profile likelihood method due to normal distribution assumptions, meanwhile the mean values are constant between the two methods. The profile likelihood plots for the different return levels with respect to the POT method are depicted in Figure 16. As seen in Figure 16, due to numerical difficulties when implementing the two-sided intervals, the profile likelihood plots need to be interpreted as one-sided. Further, the return level plot for a range of return periods is seen in Figure 17.

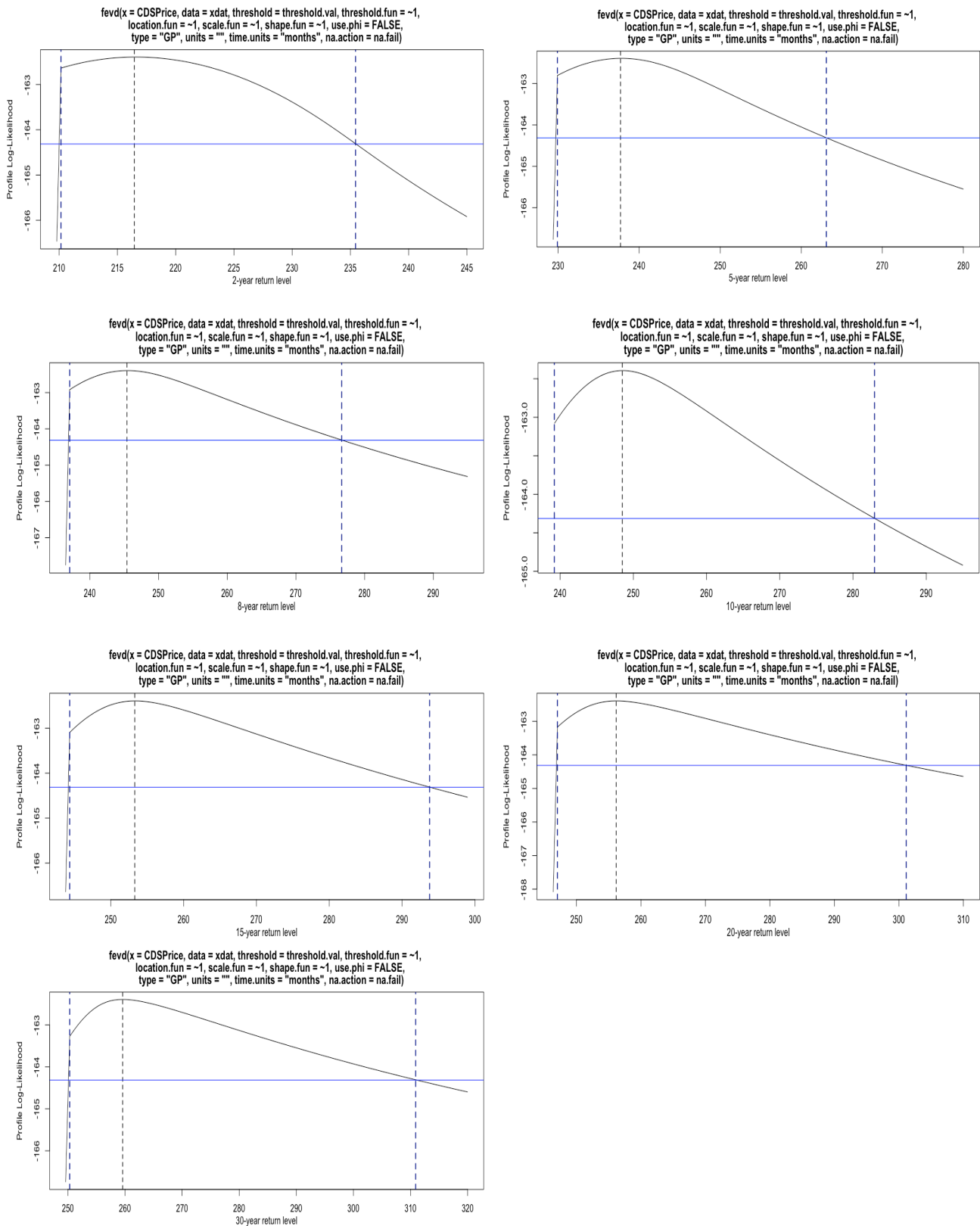


Figure 16. GP Distribution. Profile Likelihood Plots for return levels 2-years, 5-years, 8-years, 10-years, 15-years, 20-years, 25-years and 30-years of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

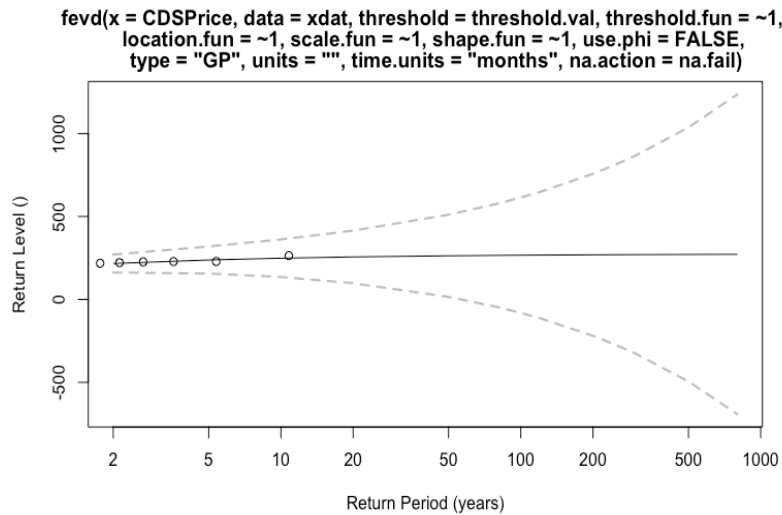


Figure 17. GP Distribution. Return Level Plot of Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018)

Exponential Distribution

It is of interest to investigate the behaviour of the shape parameter when equal to zero, namely investigate if the exceedances can be fitted to an exponential distribution. However, by valuating the above parameter estimations, it is unlikely that the shape parameter takes a value of zero. To prove this is the case, maximum likelihood estimation of the scale parameter and 95% confidence intervals is done under this assumption where $z_\alpha = 1.96$ μ and Var are the mean and variance estimates of the scale parameter over the threshold $u = 130$.

Maximum Likelihood of Scale Parameter

Maximum likelihood estimate of the scale parameter, σ , under $H_0: \gamma = 0$.

Table 26. Maximum likelihood estimates of the mean and variance of the scale parameter based on the exceedances over the threshold $u = 130$. Estimations are made under the hypothesis that the exceedances follow an exponential distribution, $H_0: \gamma = 0$

	Mean	Variance
Scale	178.2689	1063.7290

Normal Approximation of Confidence Interval for Scale Parameter

$$CI(\sigma) = \sigma \pm z_\alpha \sqrt{Var(\sigma)}$$

Table 27. Scale parameter estimates with Normal Approximation with threshold value of 130

Confidence Interval	95% Lower Bound	95% Upper Bound
Scale	114.3438	242.1940

Likelihood Ratio Test

Likelihood ratio test is performed with the exponential distribution as null hypothesis and the GP distribution as the alternative hypothesis.

$$H_0: \gamma = 0$$

$$H_1: \gamma \neq 0$$

l_0 is the log-likelihood function maximum for the exponential distribution and l_1 is the log-likelihood function for the GP distribution. The likelihood ratio test is then carried out as below.

$$D = 2\{l_1(M_1) - l_0(M_0)\}$$

Table 28. Likelihood Ratio Test with Exponential Distribution as null hypothesis and GP Distribution as alternative hypothesis

Log-likelihood function under Exponential distribution: l_0	165.8108
Log-likelihood function under GP distribution: l_1	237.8639
Chi-Square Critical Value	3.8415
Alpha	0.05
Degrees of Freedom	1
$D = 2\{l_1(M_1) - l_0(M_0)\}$	144.1062

As seen in Table 28, the likelihood ratio test rejects the null hypothesis since D is above the critical value of the chi-square distribution. Hence the CDS prices during and after the financial crisis are not distributed as an exponential distribution with shape parameter equal to zero.

To further prove that the null hypothesis should be rejected a quantile plot is done under the assumption that the underlying CDS price data follow an exponential distribution. As seen in Figure 18, the linearity is weak hence a shape parameter equal to zero should be rejected as an appropriate model.

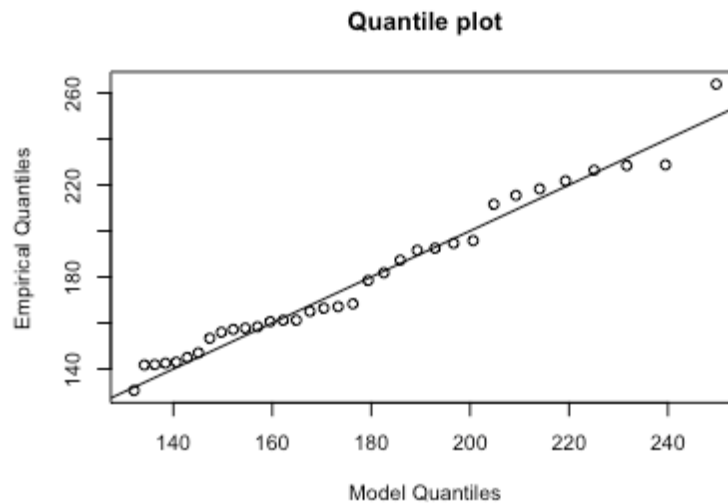


Figure 18. Quantile Plot when $H_0: \gamma = 0$

In the above two methods, namely the Block Maxima approach and the POT model, it has been assumed that the data follows a sequence of independent random variables. As mentioned, the autocorrelation function gave the implication that the CDS prices cannot be assumed to be considered independent until a lag of 9 months. The parameter estimates and the return level results of the Block Maxima and the POT model therefore lacks this important property, in which the outcomes of those models should be questioned. Both models used for this thesis should therefore be seen as suggestions for further analysis of this type of data based on much larger samples.

Declustering

In order to incorporate a higher sophistication into the models and observe clusters where the data is assumed to be independent, the declustering method is applied. Unlike in the Block Maxima approach and the POT model, a return level of 10 years is the only level evaluated going forward for the declustering method.

Table 29. Estimated characteristics of the threshold approach which is fitted to Deutsche Bank AG EUR CDS Curve 5Y - Mid Spread during and after the Financial crisis (July 2007 – April 2018). Below are estimated parameters of the threshold u and minimum gap r . $\hat{\sigma}$ and $\hat{\xi}$ are the scale and shape estimates of the Generalized Pareto distribution. $\hat{\theta}$ is the extremal index, \hat{x}_{10} is the estimated 10-year return levels and n_c are the number of clusters above the threshold u . The standard errors are expressed in parentheses

	$u = 80$			$u = 110$		
	$r = 2$	$r = 3$	$r = 4$	$r = 2$	$r = 3$	$r = 4$
n_c	6	4	3	8	5	5
$\hat{\sigma}$	52.8432 (8.2110)	52.8432 (8.2110)	52.8432 (8.2110)	70.9863 (12.5392)	70.9863 (12.5392)	70.9863 (12.5392)
$\hat{\xi}$	-0.1412 (0.1207)	-0.1412 (0.1207)	-0.1412 (0.1207)	-0.4095 (0.120)	-0.4095 (0.120)	-0.4095 (0.1197)
\hat{x}_{10}	257.47 (206.99, 307.94)	257.47 (206.99, 307.94)	257.47 (206.99, 307.94)	247.07 (169.66, 324.48)	247.07 (169.66, 324.48)	247.07 (169.66, 324.48)
$\hat{\theta}$	0.06	0.04	0.03	0.16	0.10	0.10

	$u = 130$			$u = 140$		
	$r = 2$	$r = 3$	$r = 4$	$r = 2$	$r = 3$	$r = 4$
n_c	5	4	4	5	5	5
$\hat{\sigma}$	71.2500 (14.4758)	71.2500 (14.4758)	71.2500 (14.4758)	54.9513 (12.4259)	54.9513 (12.4259)	54.9513 (12.4259)
$\hat{\xi}$	-0.4911 (0.1382)	-0.4911 (0.1382)	-0.4911 (0.1382)	-0.3714 (0.1583)	-0.3714 (0.1583)	-0.3714 (0.1583)
\hat{x}_{10}	248.49 (135.20, 361.77)	248.49 (135.20, 361.77)	248.49 (135.20, 361.77)	246.48 (185.09, 307.87)	246.48 (185.09, 307.87)	246.48 (185.09, 307.87)
$\hat{\theta}$	0.15	0.12	0.12	0.15	0.15	0.15

When investigating the empirical results above, neither the scale nor the shape parameter change in line with the thresholds and with respect to the different minimum gaps. Hence the return level does not change between the same thresholds. This is graphically depicted in Figures 23-26. To further investigate the different declustering groupings with respect to different thresholds and minimum gaps graphs are illustrated in Figures 19-22. Finding the balance between u and r , as mention before, is a trade-off between variance and bias. A too low threshold will give a weak GPD approximation of cluster maxima meanwhile give a smaller variance in the exceedances, and vice versa.

The problem of the scale and the shape parameter not changing is a concern due to the number of underlying data points. To make empirically correct assumptions, the threshold and the number of the minimum gap shall have an impact of the scale, shape and extremal index, which above results do not meet. Noteworthy, the return level can be stable even though different combinations of the threshold and the minimum gap index, but that do not seem to be the reason here.

The above findings even more stress the importance of taking non-stationarity into consideration. Due to the strong short term dependence in the data, one needs a much longer series of observations to make inference on the extremes of the future CDS prices.

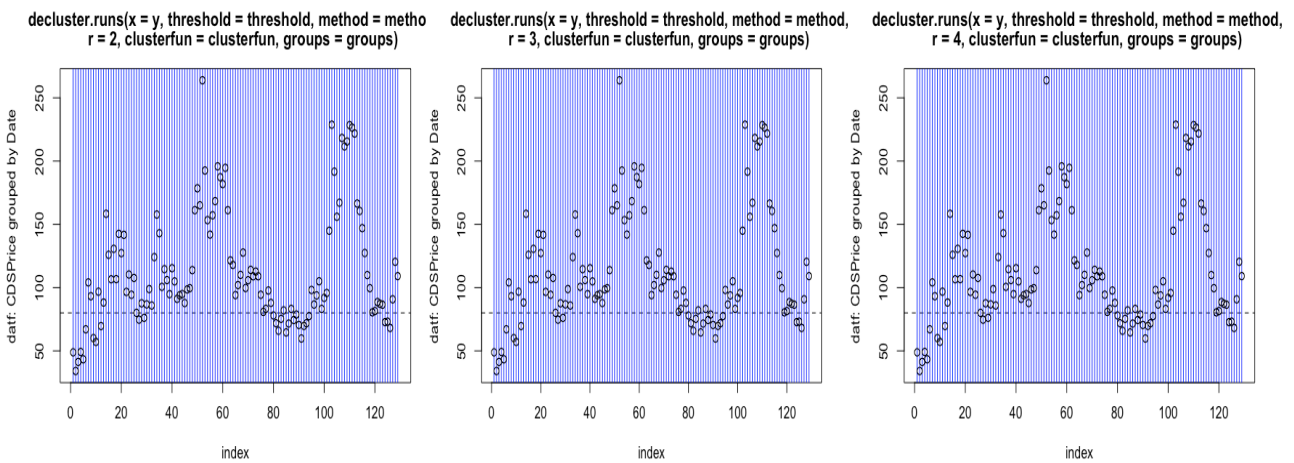


Figure 19. Declustering Groupings with $u = 80$, $r=2$, $r=3$ and $r=4$

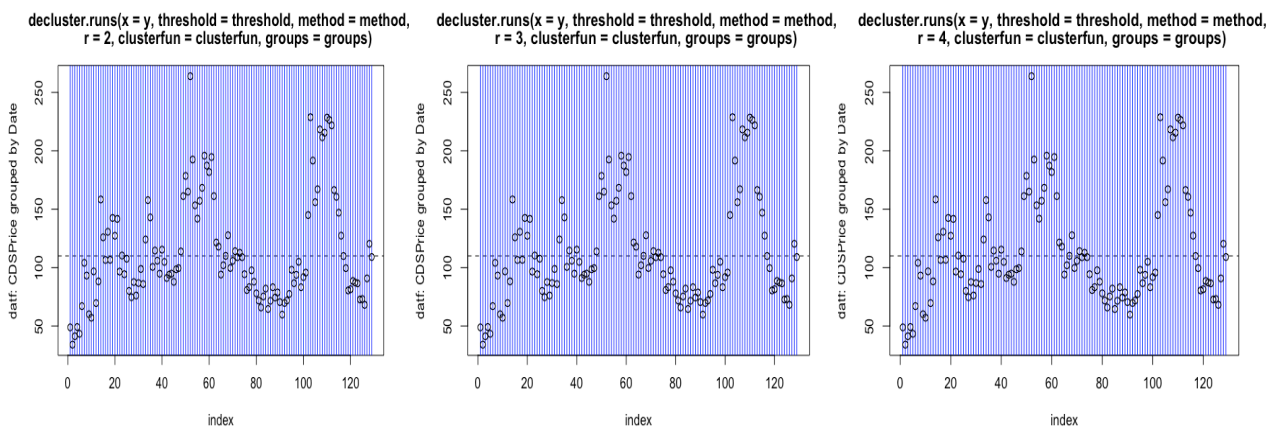


Figure 20. Declustering Groupings with $u = 110$, $r=2$, $r=3$ and $r=4$

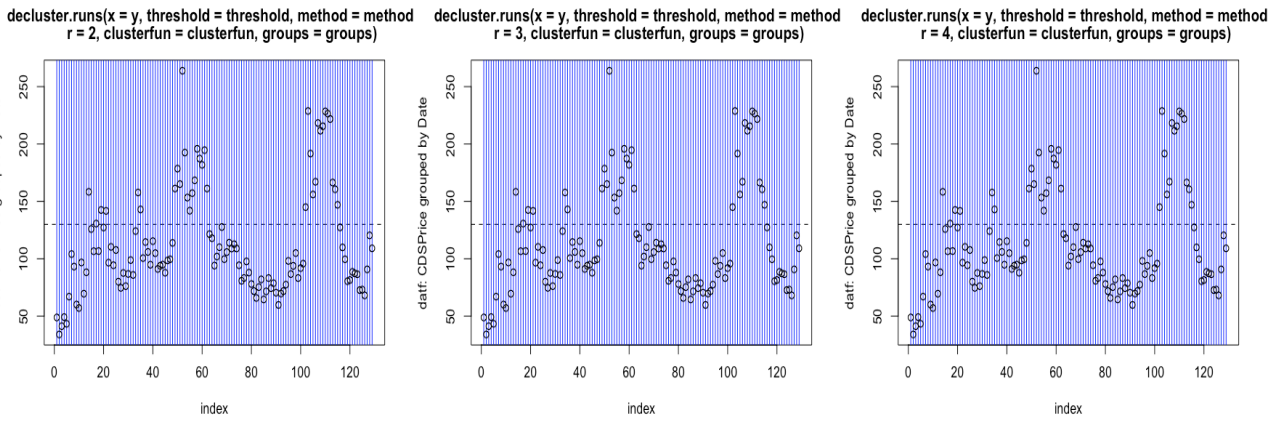


Figure 21. Declustering Groupings with $u = 130$, $r=2$, $r =3$ and $r=4$

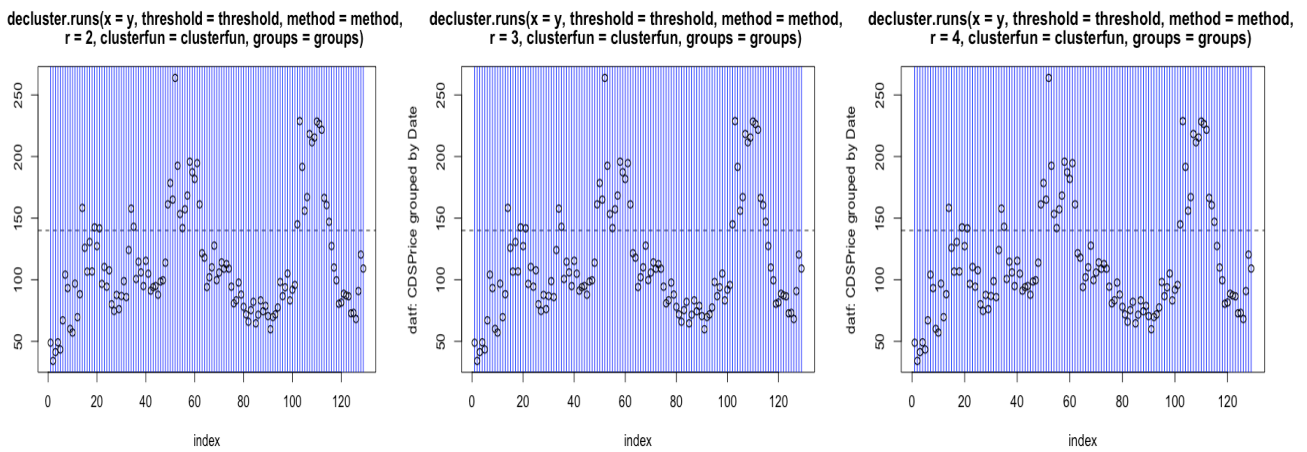


Figure 22. Declustering Groupings with $u=140$, $r=2$, $r =3$ and $r=4$

Return Level Plots

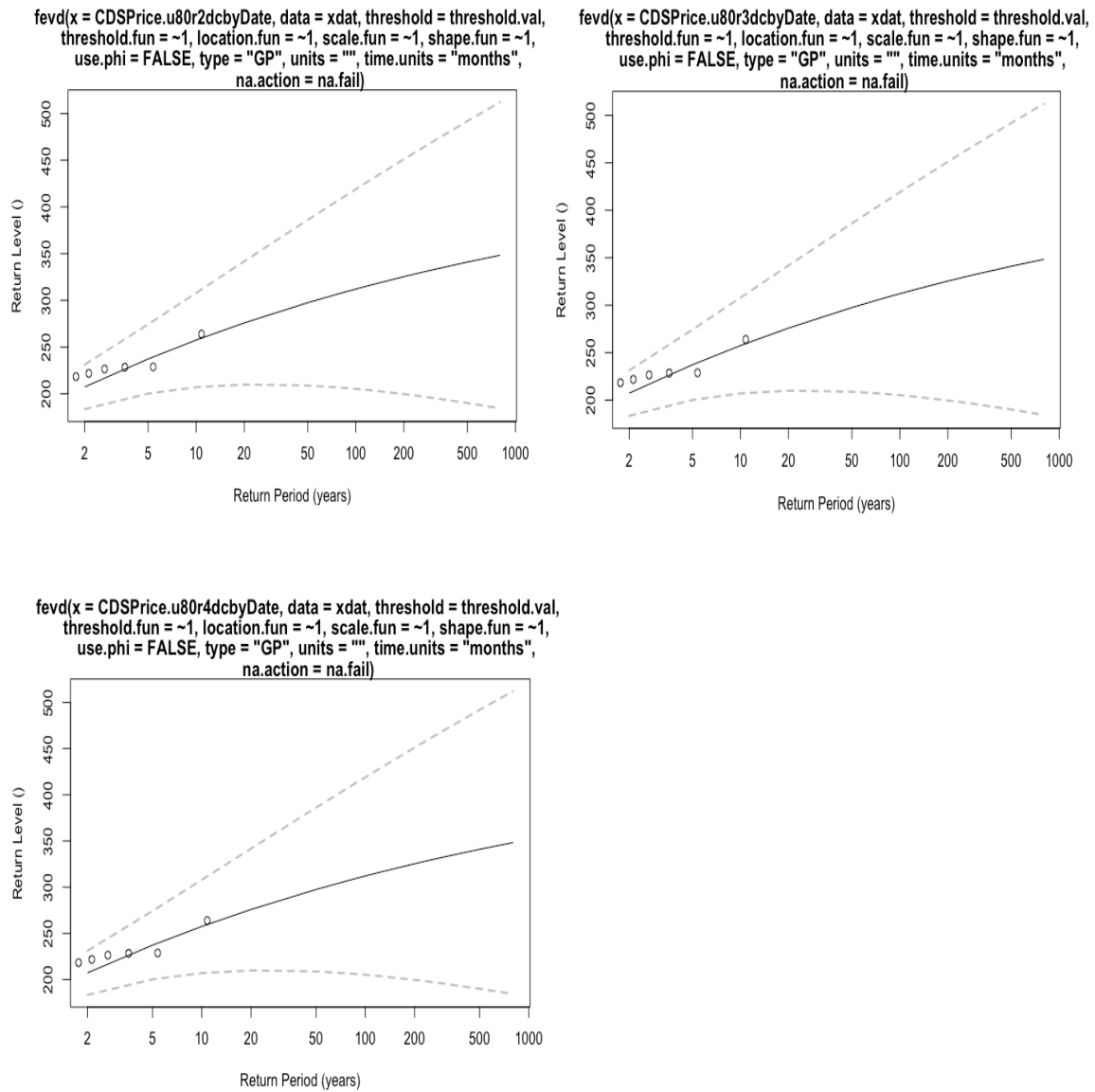


Figure 23. Declustering Return Level Plots with $u=80$, $r=2$, $r=3$ and $r=4$

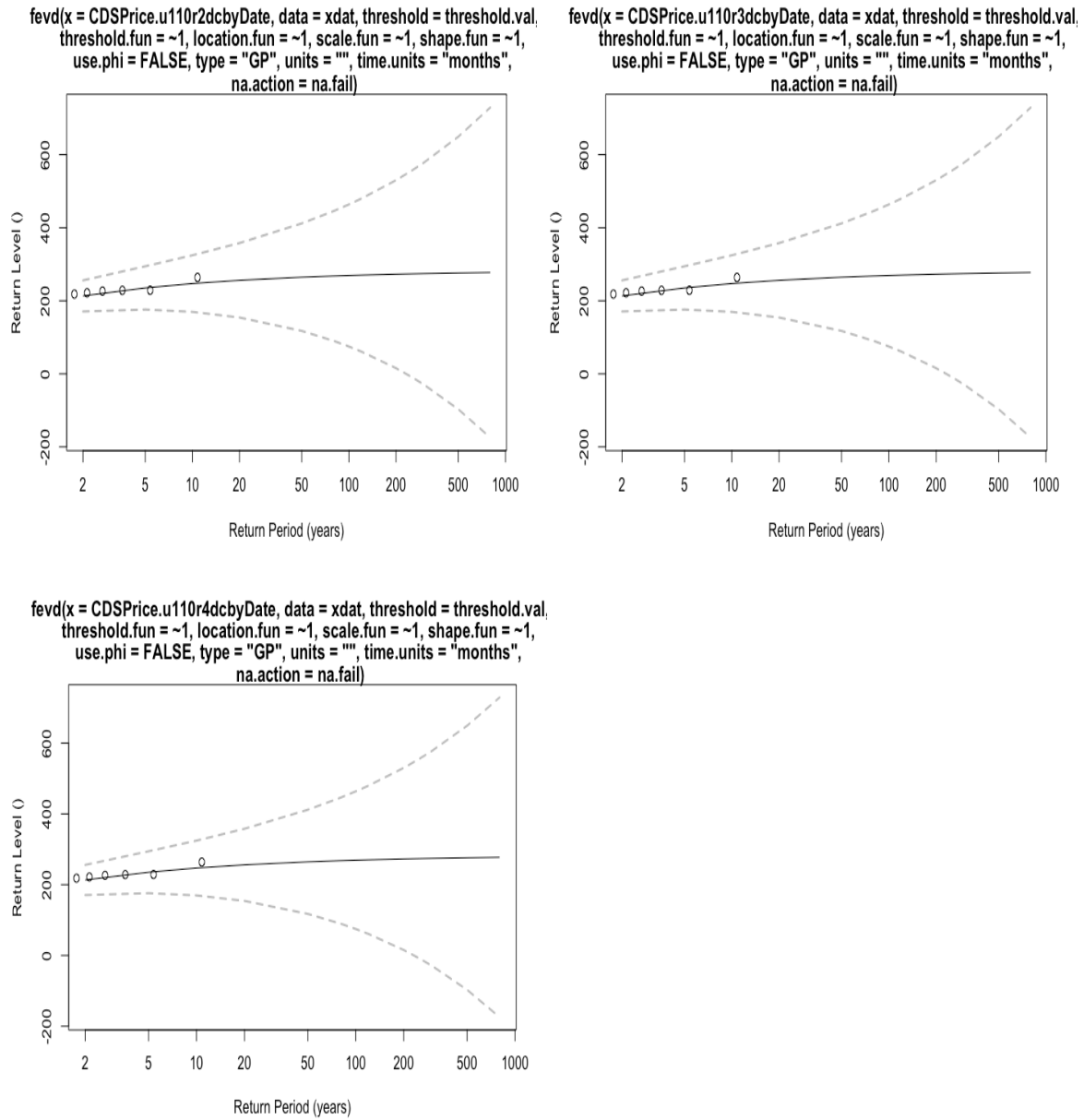


Figure 24. Declustering Return Level Plots with $u=110$, $r=2$, $r=3$ and $r=4$

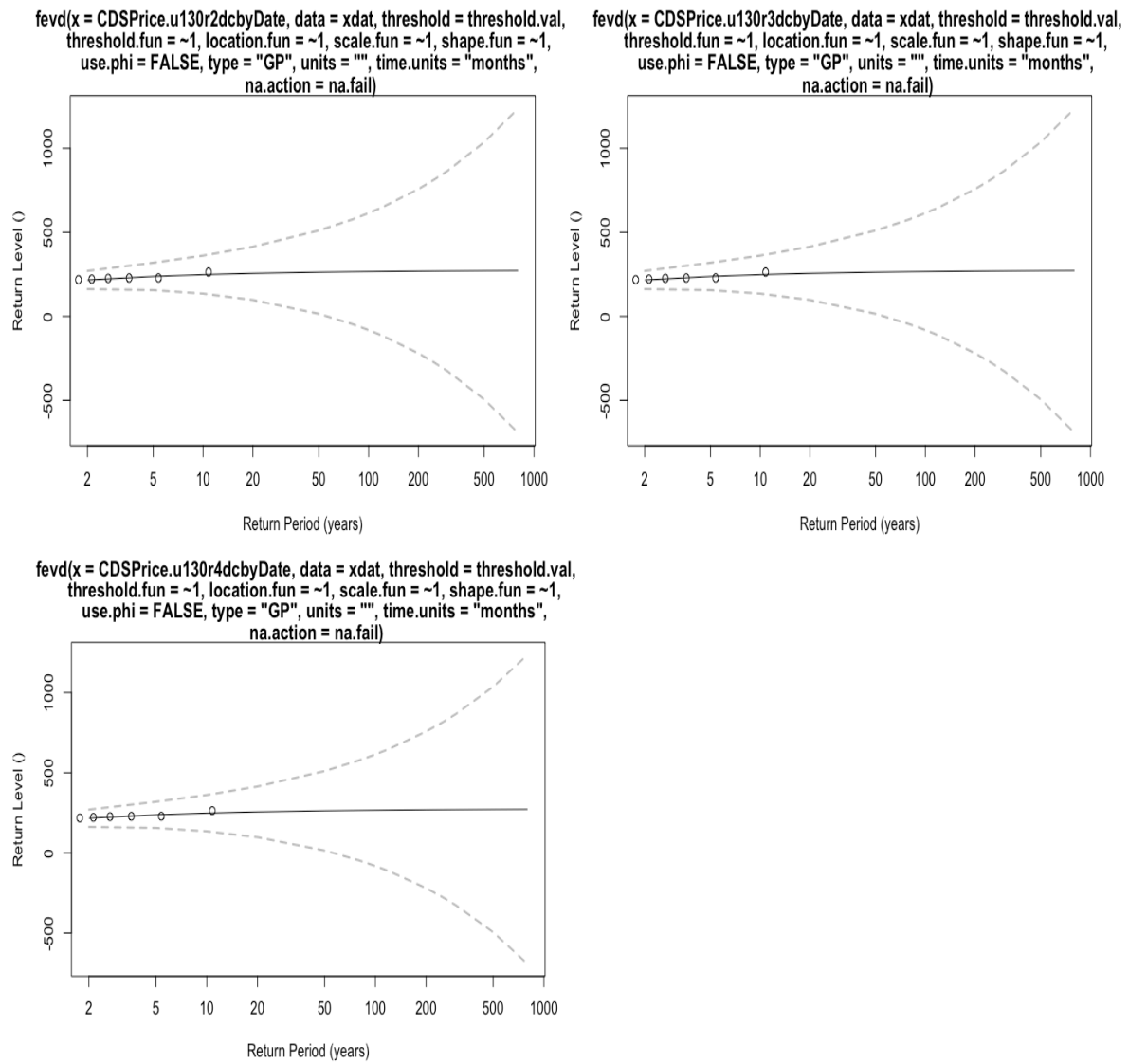


Figure 25. Declustering Return Level Plots with $u=130$, $r=2$, $r=3$ and $r=4$

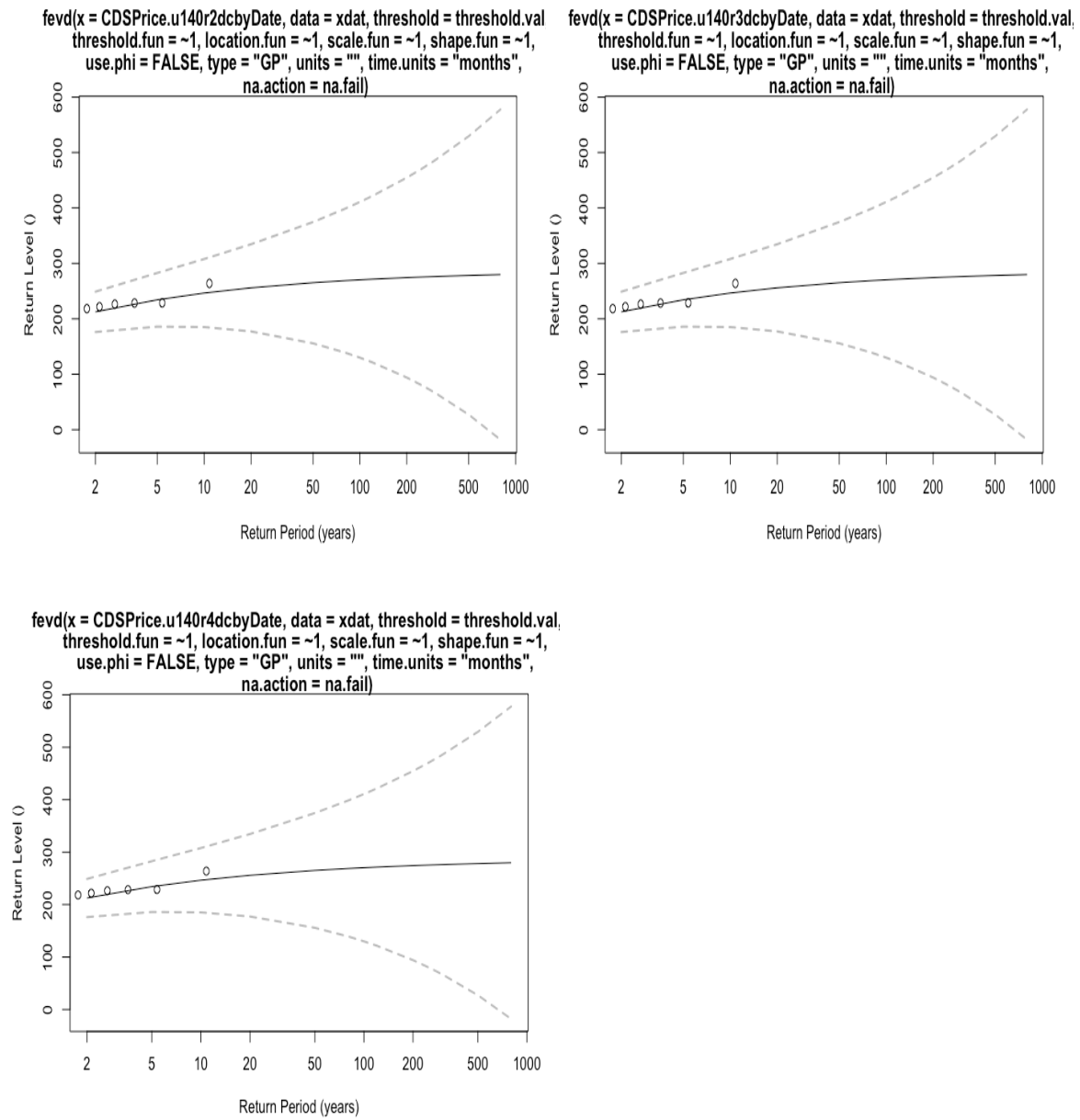


Figure 26. Declustering Return Level Plots with $u=140$, $r=2$, $r=3$ and $r=4$

Probable Maximum Loss: PML

The Probable Maximum Loss, PML, gives an estimate of the return levels with respect to different time horizons and levels of risk. It is estimated by a given worst loss with regards to chosen risk level and time horizon by calculating the upper quantile, a range of r , versus different time horizons T .

$$x_{T,p} = u + \frac{\sigma}{\xi} \left(\frac{(\lambda T)^\xi}{(-\log(1-p))^\xi} - 1 \right)$$

$$\hat{\lambda} = \frac{n_c}{n_{years}}$$

Parameter Estimates

The Probable Maximum Loss, PML, takes into consideration the parameter estimates from the declustering method with a chosen threshold of 130, a minimum gap, $r = 3$, and the parameters derived from the POT model with respect of number of cluster maxima in different time periods. As mentioned before, dependence between the different CDS prices should be handled in this model as well.

Table 30. Lambda, $\hat{\lambda}$, estimate with number of clusters, n_c , and number of years of observation, n_{years} , derived from when $u = 130$ and $r = 3$

Number of clusters, n_c	4
Number of years of observation, n_{years}	10.66
Lambda estimate, $\hat{\lambda}$	0.3750

Table 31. Scale and shape estimates derived from Peak-over-Threshold (POT)

Threshold estimate, u	130
Scale parameter, σ	71.2500
Shape parameter, ξ	-0.4911

Return Level

Table 32: Estimated quantile PML for risk levels 10% and 1% and time horizons of 2 years, 5 years, 10 years and 30 years

Risk, p	Next 2 year	Next 5 years	Next 10 years	Next 30 Years
10%	238.34	251.66	258.42	265.37
1%	263.50	267.69	269.83	272.02

5. Conclusion

To conclude, this thesis investigates different methods appropriate for describing how credit risk could be priced for different time periods by exploiting underlying CDS prices. Interestingly, when investigating what distribution is most appropriate describing the underlying CDS data, the Block Maxima approach and the POT model give rise to different outcomes.

The Block Maxima approach results in an estimated positive shape parameter for the upper bound hence the underlying distribution is assumed to belong to the Fréchet family in that region. However, the lower bound gives rise to a negative shape parameter, i.e. the underlying distribution assumes to belong the Weibull family. The above contradictory outcome gives an incentive to apply the Gumbel distribution with a shape parameter equal to zero. The likelihood ratio test is carried out and rejects the alternative model. Conclusively, when applying the Block Maxima approach, the distribution of the CDS price under and after the financial crisis is most accurate explained by a Gumbel distribution with a shape parameter equal to zero. The POT model, on the other hand, indicates a negative shape parameter for the whole confidence interval hence the distribution of the CDS price shall not be fitted to an exponential distribution. This conclusion is supported by the likelihood ratio test which rejects the null hypothesis (that the underlying data is exponential distributed) and the CDS price shall therefor be fitted to the GP distribution when applying the POT model. Conclusively, due to a negative shape parameter, an upper end-point of the return level exists when applying the POT model.

The reason for the contradictory outcomes when determining the underlying distribution's family for the models above might be due to the chosen threshold value of 130 for the POT model. Even though the threshold value is assumed to be sufficient, the number of exceedances might be too few for making correct statistical interpretations. Hence a lower threshold value might give rise to a shape parameter equal to zero, i.e. the underlying CDS data belongs to an exponential distribution, for the POT model.

As mentioned before, independency in the underlying data is vital to make accurate statistical assumptions hence the declustering method is of high importance. In line with theory, the scale parameter is to change meanwhile the shape parameter is to be stable when changing the threshold values, u . This is not the case for this thesis. The problem is, as mentioned before, due to the strong short term dependence in the data, one needs a much longer series of observations to make inference on the extremes of the future CDS prices. What happens is that it is too few cluster maxima in the dataset which make it problematic to estimate the scale and shape parameter. As a result, the likelihood function is practically flat and returns the start values of the scale and shape parameter as an outcome of optimisation. For verifications, different thresholds and minimum gaps is combined to find the

equilibrium for accurate estimations. Interestingly, higher threshold leads to higher return level until a threshold value of 130, which seems to be the maximum point of return level value. Thereafter the return level decays once again. As in line with the mean residual life plot and the model based approach, a threshold value of 130 is an appropriate threshold value, hence that the return level takes on its maximum value at that threshold value is expected. Further, as mentioned before, higher u and r contributes to less significant standard errors. Since the scale and shape parameter are fitted to GP distribution, as in the POT model, it comes as no revelation that the scale and shape parameter estimates are the same in the declustering method as for the POT model. Likewise, the 10-year return level is the same.

Not only is the credit risk with respect to different time periods of interest, but also the level of risk that is priced in. This thesis incorporates this by applying the Probable Maximum Loss method. As expected, a larger risk level and a longer time horizon give a higher uncertainty of credit event hence a higher price of the CDS. Likewise, a lower risk level and shorter time horizons, which implicates more certainty, results in a lower CDS price.

Conclusively, pricing credit risk by applying statistical extreme value models, which is what this thesis has thrived to examine, is an approach with suggestive results. However, the issue lies in the dependency in the data hence makes the mathematical models ambiguous and longer series of observations is needed to make inference on the extremes of the future CDS prices. This is a problem when pricing credit risk since CDS prices are mainly priced on a monthly basis. Favourably, one would need more frequent data points in order to make accurate estimations and results. Noteworthy, this do not imply that observations further back in time are sufficient in this context since historical data too far back in time might not be of relevance of pricing future credit risk.

Bibliography

- [1] **Coles S.** Springer Series in Statistics. *An Introduction to Statistical Modeling of Extreme Values*. 2001
- [2] **Gilleland E.** *Intro the extRemes Package*. 2014. Version 1.0-1.
URL: <http://www.assessment.ucar.edu/toolkit/>
- [3] **Gilleland E, W Katz R.** *in2extRemes: Into the R package extRemes Extreme Value Analysis for Weather and Climate Applications*. 2013. Weather and Climate Impacts Assessment Science Program, Geophysical Statistics Project. National Center for Atmospheric Research
- [4] **Gregory J.** *The XVA Challenge. Counterparty Credit Risk, Funding, Collateral and Capital*. 2015
- [5] **Lindgren G, Rootzén H, Sandsten M.** *Stationary Stochastic Processes for Scientists and Engineers*. 2014
- [6] **Rootzén H, Tajvidi N.** *Extreme Value Statistics and Wind Storm Losses: A Case Study*. *Scand. Actuarial J.* 1997; 1: 70-94
- [7] **Venables W N, Smith D M. The R Core Team.** *An Introduction to R*. 2018. Notes on R: A Programming Environment for Data Analysis and Graphics Version 3.5.1
- [8] **Weistroffer C.** *Credit default swaps. Heading towards a more stable system*. 2009. Deutsche Bank Research, Frankfurt am Main, Germany

Master's Theses in Mathematical Sciences 2018:E69
ISSN 1404-6342
LUTFMS-3358-2018
Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden
<http://www.maths.lth.se/>