VaR Estimation for Crude Oil Data via Different Approaches: Historical Simulations, EVT Model, and ACER Method

Trang Nguyen

December 2018

Master Thesis
Department of Mathematical Sciences
Norwegian University of Science and Technology

Supervisor: **Arvid Næss**
Department of Mathematical Sciences, NTNU

2nd advisor: **Sjur Westgaard**
Department of Industrial Economics and Technology Management, NTNU

Abstract

This thesis implements different approaches to predict the one-day ahead Value at Risk (VaR) of crude oil return data. The Historical Simulation (HS) approach, a non-parametric model, randomly resamples past observations with replacement to estimate the next day quantile. The Filtered HS (FHS) approach, a semi-parametric model, uses the same methodology but attempts to capture the volatility dynamics. The Conditional Extreme Value Theory (EVT) approach, a parametric model with asymptotic limits of the tail data, uses a combination of the Peaks-Over-Threshold (POT) method and the conditional variance model to extract extreme data and estimate the conditional error variance in order to compute the VaR of the next day. The Average Conditional Exceedance Rate (ACER) method, a parametric model targeting subasymptotic tail data, takes statistical dependence between the data points into account in an effort to accurately predict the extreme value distribution, i.e., the next day's VaR.

The data was retrieved from the Quandl database of crude oil continuous futures contracts traded on NYMEX WTI from April 1985 to December 2015. By dividing the data set into in-sample and out-of-sample periods, we evaluate the VaR estimates from the above approaches and assess the VaR violations based on the actual returns of the next day. From these VaR violations, we backtest the VaR estimates from these approaches via three tests. First, the unconditional coverage test checks whether the proportion of the violations is statistically different from a predetermined probability. Second, the independence test checks the clustering of these violations. The final test — a combination of the two previous tests — checks the accuracy as well as the independence of the results.

The thesis concludes that the conditional EVT approach performs best among the tested approaches. We also learn that the approaches capturing the heteroscedastic features in the data generally perform better.

Acknowledgment

First, I would like to thank my thesis supervisor, Arvid Næss. Arvid suggested the field of extreme value prediction to me and is the author of the ACER method, which not only is a compelling and integral part of this thesis but also evoked a great interest in me. Also, the book he recommended me to read, Analysis of Financial Time Series by Tsay, has been a significant resource for this thesis. Besides giving answers to my questions, Arvid has been caring to have written emails as checkups with me due to the long distance between us, and he was very understanding when I went through a difficult time. He also introduced Sjur to me, a dedicated professor in the field of risk modelling, the area that this thesis is focused on.

I would also like to thank Sjur Westgaard. After learning of my interest in actuarial science, Sjur and Arvid suggested risk management to be the topic of my master thesis. I am happy to say that it is a fascinating field that combines both mathematics and finance and has a highly practical value. Attending Sjur's lectures helped me understand more about risk management, and these lectures and his general input brought back my interest in finance, one of my bachelor's degrees, which I had mistakenly thought to be wasted. I really hope to continue working in this field in the future because I very much enjoyed the process of researching for this thesis, which was appreciably inspired by his lectures and many insightful references that Sjur kindly sent me over time. He is also the person that provided the data from Quandl for me to work with in this thesis and helped me get access to Montel database.

Next, I would like to express my profound appreciation for Erlend Magnus Viggen. Erlend has given me many great tips for presenting a research paper thanks to his expertise in academic writing. I would also like to acknowl-

edge him for having done a splendid job setting the layout of this thesis which I would otherwise have spent many hours fixing, and for consistently giving me valuable comments throughout the last stage of this thesis. Erlend has been incredible company and my biggest motivator during the last few weeks of my work, I truly appreciate his contribution to this thesis and appreciate him as a dear and important person in my life. Thank you, Erlend.

Last but most importantly, I must express my utmost sincere gratitude to my family. Thank you, mẹ Dương Thị Hòa and bố Nguyễn Bá Ngọc — my mom and dad, for having always unconditionally supported and encouraged my educational pursuit both in the U.S. and in Norway, and my other pursuits in life. Thank you, chị Nguyễn Thu Trang — my beloved sister, for always being available for me and making my time away from home never seem so challenging. My family is undoubtedly my solid emotional support that I can always lean on. They are the ones that I would like to be most proud of me. One time, my dad said that he was so proud that he couldn't sleep the night before the day we together attended a ceremony, in which I received some awards for my undergraduate achievements. I remember I had never felt happier. Without my family, I would not have been able to finish my study or to accomplish what I have today. Bố, mẹ, chị Trang, con yêu và cảm ơn nhà mình nhiều lắm!

From the bottom of my heart,
Thank You!

Sincerely yours,
Trang Nguyen

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature

1.1 Crude Oil Prices

In economics, commodities are basic goods or services that are uniform — i.e., little to no differentiation in quality — among producers of the same kind, which typically include: corn, gold, copper, coffee beans, or crude oil. Crude oil is, in fact, one of the most critical and actively traded commodities in the world [Füss, 2009]. As of June 28, 2017, nearly 1.2 million contracts were traded daily (where one contract is equivalent to 1,000 barrels) on NYMEX WTI exchange alone [CME Group, 2017]. This large amount is not very surprising because many important refined products come from crude oil, such as diesel, gasoline and a great quantity of petrochemicals. They contribute to a multitude of manufactures that affect almost every corner of consumer goods, production, and transportation of the world economy [Pines, n.d.].

Despite crude oil's great importance in the commodity exchanges, its markets are known to be highly volatile, and this makes price risk management extremely critical. Therefore, it is a vital task for risk managers to understand the factors that drive oil prices in order to manage and measure potential price risks.

There are many factors affecting the instability in crude oil prices, and *supply and demand* — a key settlement for an economic equilibrium for price and quantity in goods and services in a competitive market — needless to say, contributes towards the determination of prices in crude oil markets. In addition to these two typical factors, *financial markets* is also a major price determinant [Murphy, 2009]. Financial speculators can make a surprisingly strong impact on crude oil prices, especially because crude oil is traded at an enormous scale. To illustrate this, let's a assume a scenario where a number

Figure 1.1: Historical chart of crude oil inflation-adjusted prices from 1970–2015 [Macrotrends, 2018]

of speculators buy oil futures at a strike price that is substantially lower than the market price. This kind of event can easily provoke oil producers into expanding their current oil supply, which helps them avoid selling it later at the lower strike prices. Such a situation undesirably but essentially results in dropping both present and future prices.

Furthermore, *political situations and government policies* is certainly a key determinant in controlling oil prices and might actually wield most of the power. This is reasonably easy to understand since these events play an important and direct role in manipulating the oil supply and demand in such a macroscopic scale that can lead to drastic price changes. These extreme price fluctuations can be seen very clearly in Figure 1.1, showing how closely the crude oil volatility and the world's political setting at the time synchronized. Let's discuss some well-known events and compare them with how the oil prices changed during those periods.

- The "first oil shock" in 1973 happened when OPEC proclaimed oil embargo to punish countries supporting Israel during the Yom Kippur War, which was reflected clearly in a nearly triple price jump as shown

in Figure 1.1, the first price jump in this figure.

- In 1979, "the second oil shock" occurred as a consequence of decreasing oil output due to the Iranian Revolution and worldwide panic, and the oil prices were clearly driven up to more than double.

- Starting from 1980, some complications began within the oil supply competition. Oil production increased in many non-OPEC countries and gradually surpassed OPEC's attempt to keep prices high by limiting the supply output. The Saudis were discontent with many OPEC members who were manipulating the oil supply, they then started production at full capacity [Koepp, 1986]. All of this excessive oil supply caused oil prices to fall to an all-time low point (plunging more than half) in 1986 and to continue staying around this low level until the next oil price shock in 1990 following the Gulf War as shown in Figure 1.1.

- In the period from 1999 to mid-2008, oil prices had a fairly steady growth and reached a record peak in July 2008 due to the Middle East tension and the significant increase in oil demand in China and India [Mouwad, 2007].

- This spike was then followed by a dramatic collapse in late 2008 to early 2009 due to the global recession in 2007–2009. And once again, the oil prices peaked due to political crisis in early 2011 and remained high until the big price fall in 2014 following the diminishing demand in emerging countries and a substantial increase in the US oil production [Friedman, 2014].

- In 2015, oil prices fell considerably due to China suffering a slow economic growth while oil supply consistently had a huge surplus, and this price trend continued into 2016.

We can conclude rather surely that political setting and policies have a direct and powerful impact on oil prices, and risk managers in this field should keep in mind that the price fluctuations respond relatively quickly to these events.

Since crude oil markets are shown to be extremely volatile, it is a crucial advantage for financial participants to understand the oil market when trading in futures contracts. It does not matter whether it is from a standpoint

of hedgers, who try to lessen the effect of possible adverse price moves on oil-related assets, or speculators, who seek profits from their conjectures about oil prices' movements; benefits are in the hands of those who can estimate potential market risks, especially in extreme cases.

In this thesis, we would like to attempt using statistical methods and models to assess such extreme price risks in the crude oil market in the period from 1983 to 2015. One of the typical risk measurements for investments' loss is Value at Risk, which is given in the next section.

1.2 VaR in Risk Management

Value at Risk (VaR) is mostly concerned with, but not limited to, market risk, one of the major risks in financial markets. In a given period, the amount of risk of a specific position's devaluation can be estimated by VaR. Because VaR gauges the amount of necessary asset an institution should possess in a potential catastrophe under normal market conditions, it is one of the most widely used benchmarks to evaluate extreme risks in order to ensure the survival of financial institutions after such disastrous events. From this point of view, VaR can be interpreted as "the maximal loss of a financial position during a given time period for a given probability" [Tsay, 2010*a*, p. 326]. The literature of VaR is substantial due to its significant role in assessing financial risks. Some suggested sources for more information are Engle and Manganelli [1999], Pelletier and Christoffersen [2004], McNeil and Frey [2005, Chap. 2], and Dowd [2005].

In statistics, VaR is simply the estimate of a quantile of a random variable with a given probability distribution, or of a sample of observations. Following notations from Tsay [2010*a*], we let t be the time index of a financial position of interest for the next l periods, and $L(l)$ be the loss function of the underlying assets from time t to $t+l$. Then over the given period l with tail probability p, we define VaR as

$$p = P(L(l) \geq \text{VaR}) = 1 - P(L(l) < \text{VaR}). \tag{1.1}$$

Hence, we can interpret that over the given time horizon l, the chance of the interested position experiencing a loss greater than or equal to VaR is

$p \cdot 100\%$. Now let, $F_l(x)$ be the cumulative distribution function (CDF) of $L(l)$, and q be a probability $(0 \leq q \leq 1)$, then:

$$x_q = \inf\{x | F_l(x) \geq q\}, \quad \text{or} \quad q = \Pr(L(l) \leq x_q). \tag{1.2}$$

Here, x_q is the q^{th} quantile of $F_l(x)$, where inf denotes the smallest real number x satisfying $F_l(x) \geq q$. From (1.1) and (1.2), we know that VaR is the $(1-p)^{\text{th}}$ quantile or q^{th} quantile of the CDF of the loss function $L(l)$ for VaR $= x_{1-p} = x_q$. Thus, VaR, computed by a predictive distribution of a series of returns, estimates the potential loss of a portfolio given time horizon.

There are a few drawbacks to VaR, despite its simple concept and useful contribution. First, while VaR is a prediction and "should be computed using the predictive distribution of future returns", the majority of existing methods determining VaR disregards the reliability of parameter estimation, since the predictive distribution, which accounts for the parameter uncertainty, is difficult to implement [Tsay, 2010*a*]. However, there exist several widely used procedures to validate a set of VaR forecasts, which will be given in more details in section 2.5. Another downside to VaR is that VaR cannot provide a bigger picture of the upper tail behavior of the loss function due to being only a quantile (a single number). One of the solutions for this is the use of expected shortfall (ES), which will not be discussed in any depth in this thesis due to time constraints. ES, also called *conditional VaR*, is more sensitive to the general shape of the tail behavior of the loss function, which then provides a better sense of the tail distribution [Tsay, 2010*a*, Chap. 7].

In practice, the CDF of the loss function is, however, unknown. Therefore, calculating VaR for a financial position is essentially fitting and estimating an appropriate econometric modeling for the distribution of the observed data. An appropriate approach to assess the tail behavior of these distributions, which we essentially would like to achieve, is to apply extreme value theory. The next section will cover the literature of this methodology, whose applications can be used in predicting the probability distribution of day-to-day market risk. This methodology is also the base for two of the approaches that we use in this thesis.

1.3 Extreme Value Theory (EVT)

The field of extreme value theory (EVT) was first introduced by Leonard Tippett (1902–1985). EVT is a branch of statistics that has its chief concern of handling the most extreme deviations of a random variable given its distribution, in order to produce asymptotic models for the distribution of these deviations. The inital work was done by Fisher and H.C. Tippett [1928], and a lot of critical contributions have been made following it. Some typical ones include the extreme value theorem and family limiting distribution by Gnedenko [1943], the use of threshold in extreme value models by Balkema and De Haan [1974] and Pickands [1975], and the introduction of the generalized extreme value by Jenkinson [1955]. Because of its unique features, EVT provides important and meaningful assessments that are applied in many areas.

In the scope of finance, EVT can aim at the characteristics of min/max return of an asset over an observed period, at which "properties of the minimum return can be obtained from those of the maximum by a simple sign change" [Tsay, 2010*a*, p. 342]. For a *long financial position* (one that buys an asset with the anticipation of growing value in the future), a risk emerges when the value of the asset drops, so the minimum return (left tail of the distribution) is a big concern. Whereas the maximum return (right tail) is important to a *short financial position* (one that sells an asset they borrowed and purchases it later to deliver back to the lender, therefore, they experience a loss when the price increases). Risk managers should always be prepared for unforeseen events of extreme price changes, and it is essential for them to pay attention on modeling the tails of the returns distribution explicitly, where the EVT specializes in.

Note: *Throughout this thesis, we define equations for VaR from the perspective of a long position (hence, the interest lies in the left tail of the return distribution) with p being the tail probability.*

Tsay [2010*a*] reviewed EVT in the following manner: Let r_t be serially independent with a common CDF $F(x)$ for $(l \leq t \leq u)$ and $r_{(n)}$ be the maximum order statistic $(r_{(n)} = \max_{1 \leq j \leq n} \{r_j\})$, then the CDF of $r_{(n)}$, $F_{n,n}(x)$,

is given by

$$\begin{aligned} F_{n,n}(x) &= P\left(r_{(n)} \leq x\right) \\ &= P\left(r_{(1)} \leq x, r_{(2)} \leq x, \ldots, r_{(n)} \leq x\right) \\ &= \prod_{j=1}^{n} P\left(r_{(j)} \leq x\right) \qquad \text{(by independence)} \\ &= \prod_{j=1}^{n} F(x) = [F(x)]^n . \end{aligned} \tag{1.3}$$

This is a degenerated CDF, since as $n \to \infty$, $F_{n,n}(x) \to 0$ if $x < u$, and $F_{n,n}(x) \to 1$ if $x \geq u$. To avoid this, EVT is concerned with $\{\beta_n\}$ (the location series) and $\{\alpha_n\}$ (scaling factors series) ($\{\alpha_n\} > 0$) such that the distribution of normalized maximum $r_{(n*)} \equiv (r_{(n)} - \beta_n)/\alpha_n$ is nondegenerate.

Assuming that $r_{(n*)}$ is independent, the limiting cumulative distribution becomes

$$F_*(x) = \begin{cases} \exp\left[-(1+\xi x)^{-1/\xi}\right] & \text{if } \xi \neq 0, \\ \exp\left[-\exp(-x)\right] & \text{if } \xi = 0, \end{cases} \tag{1.4}$$

for

$$\begin{cases} x < -\frac{1}{\xi} & \text{if } \xi < 0, \\ x > -\frac{1}{\xi} & \text{if } \xi > 0, \end{cases}$$

where the subscript $*$ signifies the maximum. The result of $F_*(x)$ in the case where $\xi = 0$ is derived from the limit when $\xi \to 0$. The parameter ξ is referred to as the *shape parameter*, which regulates the tail behavior of the limiting distribution.

Jenkinson [1955] introduced the *Generalized Extreme Value* (GEV) (the limiting distribution in (1.4)) for the maximum statistic, which includes the three types of limiting distribution of Gnedenko [1943]: the Gumbel family ($\xi = 0$), the Fréchet family ($\xi > 0$), and the Weibull family ($\xi < 0$). The density functions of these families can be seen in Figure 1.2. Among these families, the Fréchet family is often used in risk management as it depicts the properties of return distribution better, which typically has a heavy (fat) tail.

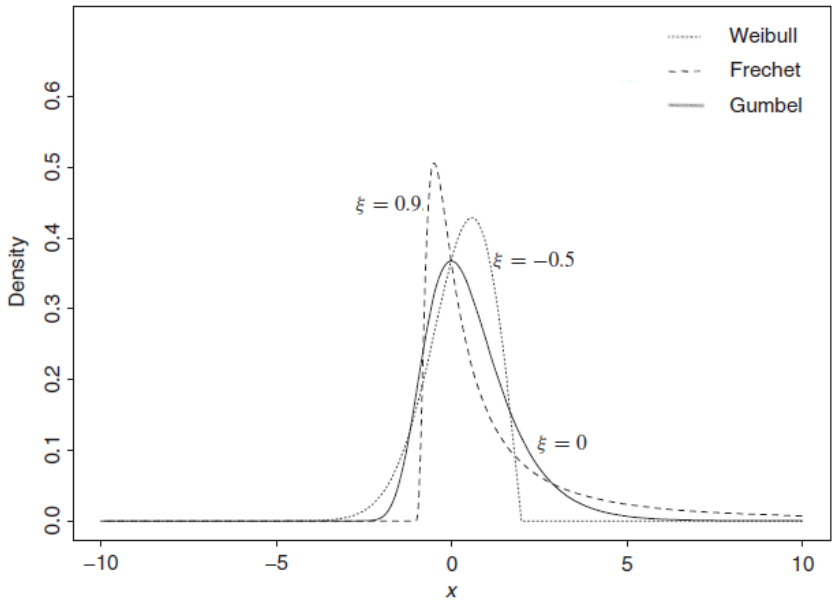

Figure 1.2: Probability density functions of extreme value distributions for maximum

There are two remarks about EVT according to Tsay [2010*a*]. First, the limiting distribution $F_*(x)$ is driven only by the tail behavior of $F(x)$, thus, EVT is largely suitable for the return r_t and is not restricted to a distribution. (Note that $F(x)$ might nevertheless decide $\{\beta_n\}$ and $\{\alpha_n\}$). Second, the tail index (ξ) is time-invariant, which makes VaR calculation much simpler.

This thesis is organized as follows: Chapter 2 gives the statistical theories and methodologies to apply in the field of finance in terms of extreme value modelings, Chapter 3 describes the crude oil data obtained from NYMEX WTI, and Chapter 4 analyzes and presents the results from the application of mentioned methods. Chapter 5 concludes the thesis.

Chapter 2

Theory and Methodology

2.1 GARCH for Volatility

Regression models are undoubtedly the most common statistical tool for time series analysis. However, most of these models in theories have a set of classical assumptions that are usually inapplicable to real-life time series data [Wei, 2006].

In standard regression analysis, we have a model where a dependent variable is described by a set of independent variables and an error term, where the error term is often assumed to be independent and identically distributed (i.i.d.) following the normal distribution with mean 0 and a constant variance. Unfortunately, when the error terms are autocorrelated over time, the common ordinary least squares (OLS) method to estimate parameters does not hold. Normally, Autoregressive-Moving Average (ARMA) models can be applied to adjust such an error structure where the error term is stationary with a constant variance. However, this assumption is most likely violated in practice, since volatility over time in financial markets is certainly seldom constant. Such a model with varying error variance is regarded to as a heteroscedasticity model and can be captured by a GARCH model.

2.1.1 ARCH Model

Following Wei [2006], let's consider the regression

$$Y_t = X_t'\beta + \varepsilon_t,$$

where t is the time index $(t = 1, 2, \ldots, n)$, Y_t is the dependent variable, X_t' is the transpose of the set of independent variables, β is the regression coefficients, and ε_t is the uncorrelated error term.

Let $\varepsilon_t = n_t$ for n_t that are uncorrelated but have time-dependent variances. In accordance with Engle [1982], assume

$$n_t = \sigma_t e_t, \tag{2.1}$$

where e_t are i.i.d. variables with mean 0 and variance 1, and

$$\sigma_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \ldots + \theta_s n_{t-s}^2. \tag{2.2}$$

Hence, the conditional variance of n_t given all information up to time $(t-1)$ is

$$\begin{aligned} \mathrm{Var}_{t-1}(n_t) &= \mathrm{E}_{t-1}\left(n_t^2\right) = \mathrm{E}\left(n_t^2 \middle| n_{t-1}, n_{t-2}, \ldots\right) \\ &= \sigma_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \ldots + \theta_s n_{t-s}^2. \end{aligned} \tag{2.3}$$

This error conditional variance depends on the squares of past errors and is time-dependent. Clearly, one large past error leads to another large error, making the variance larger. In finance, this is referred to as *volatility clustering*, which depicts a circumstance where "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" [Mandelbrot, 1963].

To have a better understanding of *volatility clustering*, let's look at Figure 2.1. Here, we have chosen a specific period in our crude oil return data (described in Chapter 3) that consists of 180 observations from 26/12/2002 to 12/09/2003. We choose a lower threshold and an upper threshold that correspond to 15^{th} and 85^{th} percentiles of the sample, respectively. Now, let us identify a *cluster* as a set of *consecutive* returns that are either altogether less than the lower threshold or altogether greater than the upper threshold, and we will name these returns *deviations*. Figure 2.1 consists of four plots, where the clusters in each plot (in reading order from the top left plot) require a different number of minimum consecutive observations (2, 3, 4, and 5, respectively). The deviation set that exceeds the upper threshold is located in between the dashed lines, while the deviation set smaller than the lower threshold is in between the solid lines. From the figure, we can see that only the plot requiring two minimum consecutive returns has "upper clusters", and other plots requiring more continuous deviations only have "lower clusters". The bottom-right plot indicates that there were at least

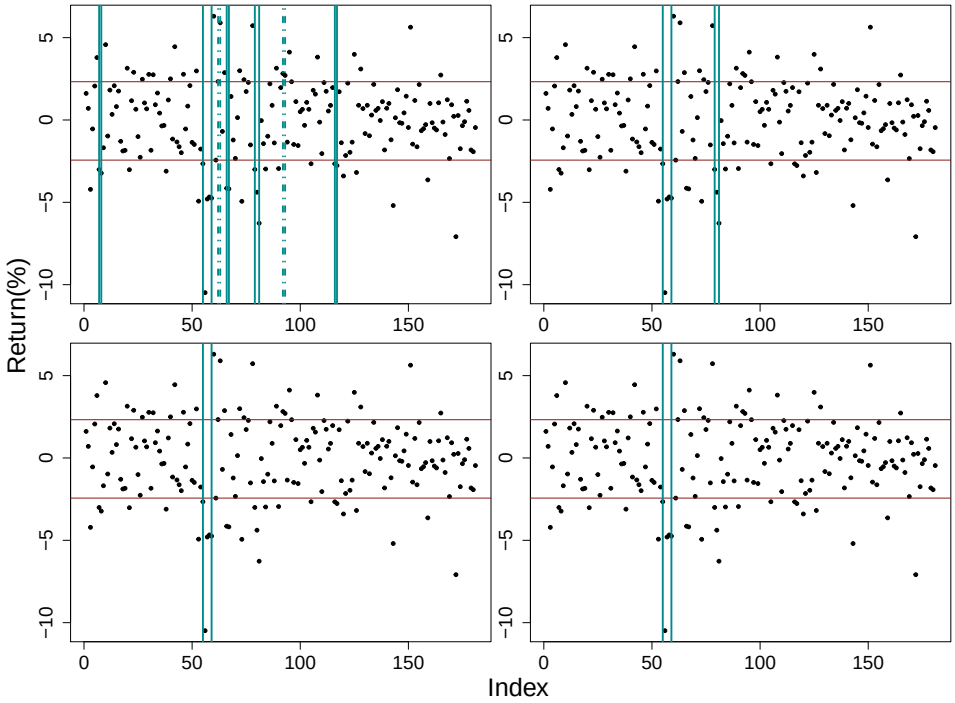

Figure 2.1: A section of the crude oil daily returns with upper and lower thresholds (red lines) corresponding to the 15th and 85th percentiles. Clusters are bordered by vertical blue lines. The returns that fall in clusters localized between dashed lines exceed the upper threshold, and the returns in solid-line clusters are smaller than the lower threshold. The different plots use different minimum numbers of consecutive returns to define a cluster. In reading order, the minimum numbers are 2, 3, 4, and 5.

five consecutive daily returns that are less than the lower threshold, so this period should be of risk managers' particular interest for a crude oil long position (since the lower the return, the bigger the risk for a long position).

Now, assuming n_t^2 follows an autoregressive model of order s, AR(s), then

$$n_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \ldots + \theta_s n_{t-s}^2 + a_t, \tag{2.4}$$

where a_t is a white noise process $\sim N(0, \sigma_a^2)$. Engle [1982] called such a model, with the error term n_t carrying a variance model as in (2.2), and the error term square following an AR process as in (2.4), the Autoregressive Conditional Heteroscedasticity (ARCH) model of order s — ARCH(s).

According to Wei [2006], the procedure testing for ARCH includes:

1. Fit the OLS regression for $(t = 1, 2, \ldots, n)$ and compute the residuals $\widehat{\varepsilon}_t = \widehat{n}_t$.
2. Gather a set of series $\left\{\widehat{n_t^2}\right\}$ and check if it follows an AR process.

2.1.2 GARCH Model

The ARCH model can be generalized so that the conditional variance of the error series depends on both the squares of past errors and the past conditional variances. Again, following the notation of Wei [2006], we have $n_t = \sigma_t e_t$, where e_t are i.i.d. variables with mean 0 and variance 1 and are independent of n_{t-i}. Then,

$$\sigma_t^2 = \theta_0 + \varphi_1 \sigma_{t-1}^2 + \ldots + \varphi_r \sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \ldots + \theta_s n_{t-s}^2 \quad (2.5)$$

such that the roots of $(1 - \varphi_1 B - \ldots - \varphi_r B^r) = 0$ lie outside of the unit circle (where $x_{t-k} = B^k x_t$, the backshift operator). The assumption for $\theta_0 > 0$ and θ_j, φ_j being nonnegative is needed to ensure a meaningful σ_t^2 where $\sigma_t^2 > 0$.

Such a model with the given error term n_t with the conditional variance specified in (2.5) is called the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model of order (r, s) —- GARCH(r, s), which was introduced by Bollerslev [1986]. Note, we can see that ARCH is simply a specific model reduced from GARCH. From the structure of the GARCH model, we can interpret the conditional variance of the error term as "a weighted function of a long-term average value (θ_0, in our notion), information about volatility during previous periods ($\theta_i n_{t-i}^2$) and the fitted variance from the model during the previous periods ($\varphi_j \sigma_{t-j}^2$)" [Brooks, 2008].

We should pay particular attention to a remark that Wei [2006] makes in his book about an easily misleading interpretation of the order (r, s) of the GARCH model. The model in (2.5) is not an ARMA(r, s) process because there is a white noise process error term in such a process, unlike σ_t^2 and n_t^2, which do not have this property. Let $a_t = (n_t^2 - \sigma_t^2)$ so that $\sigma_t^2 = (n_t^2 - a_t)$. Then (2.5) can be rewritten as

$$(1 - \alpha_1 B - \ldots - \alpha_m B^m)\, n_t^2 = \theta_0 + (1 - \phi_1 B - \ldots - \phi_r B^r)\, a_t, \quad (2.6)$$

where $m = \max(r, s)$, $\phi_i = 0$ for $i > r$, $\theta_r = 0$ for $i > s$,

$$\alpha_i = (\phi_i + \theta_i), \quad (2.7)$$

If we assume that a_t is the associated white noise process for the n_t^2 process, then (2.6) is a proper ARMA model, since $\mathrm{E}_{t-1}(n_t^2) = \sigma_t^2$, σ_t^2 is the one-step ahead forecast of n_t^2, and a_t is the corresponding one-step ahead forecast error. Thus, the GARCH(r, s) model in (2.1) and (2.5) implies that n_t^2 follows an ARMA(m, r) model in (2.6) with the AR order being $m = \max(r, s)$.

In general, the regression model with autocorrelated error can be combined with the conditional heteroscedasticity model

$$Y_t = X_t'\beta + \varepsilon_t, \tag{2.8}$$

where

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \ldots + \varphi_p\varepsilon_{t-p} + n_t, \tag{2.9}$$

$$n_t = \sigma_t e_t, \tag{2.10}$$

$$\sigma_t^2 = \theta_0 + \phi_1\sigma_{t-1}^2 + \ldots + \phi_r\sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \ldots + \theta_s n_{t-s}^2, \tag{2.11}$$

and e_t is i.i.d. with mean 0 and variance 1 and is independent of past realizations of n_{t-i}. We can test for heteroscedasticity in the error variance as the following [Wei, 2006]:

1. Calculate OLS residuals $\hat{\varepsilon}_t$ from the OLS fitting of (2.8).
2. Fit an AR(p) model (2.9) to the $\hat{\varepsilon}_t$.
3. Obtain the residuals $\hat{n}_t$ from the AR fitting in (2.9).
4. Form the series $\hat{n}_t^2$ and compute its sample autocorrelation function (ACF)

$$\hat{\rho}_i\left(\hat{n}_t^2\right) = \frac{\sum_{t=1}^{n-t}\left(\hat{n}_t^2 - \hat{\sigma}^2\right)\left(\hat{n}_{t+i}^2 - \hat{\sigma}^2\right)}{\sum_{t=1}^{n}\left(\hat{n}_t^2 - \hat{\sigma}^2\right)^2},$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{n}_t^2.$$

The partial autocorrelation function (PACF) is computed in a similar way. A pattern of ACF and PACF will indicate ARCH/GARCH errors and form a good basis for their order specification. As shown in (2.6), a general

GARCH(r,s) model for σ_t^2 corresponds to an ARMA(m,r) model for n_t^2 with $m = \max(r,s)$. Then, ACF and PACF of $\hat{n}_t^2$ will show patterns of an exponential decay.

Wei [2006] also pointed out that the GARCH$(1,1)$ model is very parsimonious, shown by the following:

1. Take the conditional variance equation in the GARCH $(1,1)$ model with different time subscripts:

$$\begin{aligned} \sigma_t^2 &= \theta_0 + \varphi_1\sigma_{t-1}^2 + \theta_1 n_{t-1}^2, \\ \text{or } \sigma_{t-1}^2 &= \theta_0 + \varphi_1\sigma_{t-2}^2 + \theta_1 n_{t-2}^2, \\ \text{or } \sigma_{t-2}^2 &= \theta_0 + \varphi_1\sigma_{t-3}^2 + \theta_1 n_{t-3}^2. \end{aligned}$$

2. Rewrite the conditional variance from the above equations:

$$\begin{aligned} \sigma_t^2 &= \theta_0 + \varphi_1(\theta_0 + \varphi_1\sigma_{t-2}^2 + \theta_1 n_{t-2}^2) + \theta_1 n_{t-1}^2 \\ &= \theta_0 + \theta_0\varphi_1 + \theta_1\varphi_1 n_{t-2}^2 + \theta_1 n_{t-1}^2 + \varphi_1^2\sigma_{t-2}^2 \\ &= \theta_0 + \theta_0\varphi_1 + \theta_1\varphi_1 n_{t-2}^2 + \theta_1 n_{t-1}^2 + \varphi_1^2\theta_0 + \varphi_1\sigma_{t-3}^2 + \theta_1 n_{t-3}^2) \\ &= \theta_0(1 + \varphi_1 + \varphi_1^2) + \theta_1 n_{t-1}^2(1 + \varphi_1 B + \varphi_1^2 B^2) + \varphi_1^3\sigma_{t-3}^2. \end{aligned}$$

3. Follow the recursive relationship, we eventually have:

$$\sigma_t^2 = \theta_0(1 + \varphi_1 + \varphi_1^2 + \ldots) + \theta_1 n_{t-1}^2(1 + \varphi_1 B + \varphi_1^2 B^2 + \ldots) + B^\infty\sigma_0^2,$$

where $\theta_0(1 + \varphi_1 + \varphi_1^2 + \ldots)$ is a constant, denoted as γ_0, and as the number of observations approaches infinity, $B^\infty \xrightarrow{\text{yields}} 0$.

Hence, the GARCH $(1,1)$ model can be rewritten as:

$$\sigma_t^2 = \gamma_0 + \gamma_1 n_{t-1}^2 + \gamma_2 n_{t-2}^2 + \ldots . \tag{2.12}$$

Equation (2.12) is a restricted infinite order ARCH model. Thus, the GARCH$(1,1)$ model is parsimonious because the conditional variance is determined by an infinite number of past squared errors despite the fact that the original model only includes three parameters [Brooks, 2008]. According to Bollerslev et al. [1994], in the scientific publications of finance, the volatility clustering in the data is often sufficiently represented by a GARCH$(1,1)$ model [Brooks, 2008]. For more literature on GARCH models, see Bollerslev et al. [1992] in addition to Bollerslev et al. [1994] and related references [Tsay, 2010*b*].

2.1.3 Estimation of GARCH Models

In order to estimate the parameters of a chosen model, the maximum likelihood estimation (MLE) method can be used quite straightforward. Again, keep in mind that we are considering the general regression model given in (2.8)–(2.11), where the error term is autocorrelated and the error conditional variance is heteroscedastic (GARCH structure). We continue following the demonstration from Wei [2006]:

Rewrite the regression as:

$$n_t = (1 - \varphi_1 B - \ldots - \varphi_p B^p)(Y_t - X_t'\beta). \tag{2.13}$$

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$, $\mathbf{X} = (X_1, \ldots, X_n)$, and Y_0 and X_0 be appropriate initial values to compute n_t. Hence, by maximizing the conditional likelihood function or the log-likelihood function under a normality assumption for $n_t \sim N(0, \sigma_t^2)$, the parameters' MLE are calculated.

Since $n_t \sim N(0, \sigma_t^2)$, then $Y_t \sim N(X_t'\beta + \varphi_1\varepsilon_{t-1} + \ldots + \varphi_p\varepsilon_{t-p}, \sigma_t^2)$, we have

$$f(Y_t|X_t'\beta, \sigma_t^2) = \frac{1}{\sigma_t^2\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y_t - X_t'\beta - \varphi_1\varepsilon_{t-1} - \ldots - \varphi_p\varepsilon_{t-p})^2}{\sigma_t^2}\right\}.$$

So, the joint distribution is

$$\begin{aligned} &f\left(Y_1, \ldots, Y_n \middle| X_1'\beta, \ldots, X_2'\beta, \sigma_1^2, \ldots, \sigma_n^2\right) \\ &\quad = \prod_{t=1}^{n} f(Y_t|X_t'\beta, \sigma_t^2) \\ &\quad = \prod_{t=1}^{n} \left(\frac{1}{2\pi\sigma_t^2}\right)^{1/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\frac{(1 - \varphi_1 B - \ldots - \varphi_p B^p)(Y_t - X_t'\beta)^2}{\sigma_t^2}\right\}, \end{aligned}$$

and the conditional likelihood function becomes

$$L(\beta, \varphi, \theta, \phi|\mathbf{Y}, \mathbf{X}, Y_0, X_0) = \prod_{t=1}^{n} \left(\frac{1}{2\pi\sigma_t^2}\right)^{1/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\frac{n_t^2}{\sigma_t^2}\right\}.$$

Accordingly, the log-likelihood function is

$$\ln L(\beta, \varphi, \theta, \phi|\mathbf{Y}, \mathbf{X}, Y_o, X_o) = \frac{1}{2}\sum_{t=1}^{n}\left(-\ln(2\pi) - \ln(\sigma_t^2) - \frac{n_t^2}{\sigma_t^2}\right).$$

where σ_t^2 is obtained from equation (2.11) and n_t from equation (2.13).

2.2 Historical Simulation (HS) Approach and Filtered HS Approach

2.2.1 Historical Simulation Approach

Historical simulation (HS) is a simple nonparametric method to estimate VaR. By assuming that the empirical distribution of past returns can well represent the future returns, or in other words, that the historical distribution remains over the next periods, HS simulates future scenarios based on what already happened in the past.

The methodology of this approach is bootstrapping, that is, the practice of estimating by random sampling with replacement. From recursively updating the return series and repeating the simulation for some N times, we can take the average of simulated returns at each time point to get a representing simulated return for that day.

Therefore, the VaR in the next period given probability p is simply the $100p$-th percentile of the set of simulated returns from historical data, $\{X_t\}$. So, assume the window rolls back to n observations, we can generate VaR using HS as

$$\text{VaR}_{p,t+1} = \text{Quantile}_p \left\{X_t\right\}_{t=1}^{n}.$$

Because of its simplicity in implementation and its model-free nature (which can eliminate misleading assumptions about the true return distribution), HS is a popular method used in practice to estimate VaR. However, this approach is also considered to be a naïve benchmark due to its many downsides.

The choice of the sample size n can be a serious drawback [Christoffersen, 2012]. If the chosen sample size n is too large, the weight of the most recent observations (which presumably are the most relevant to the likely future returns) is too little, and the potential of high variance is quite considerable. On the other hand, if the sample size is too small, there is not enough data to be relied upon to generate an accurate estimation for VaR, especially for extreme values like $\text{VaR}_{0.001}$. So the choice of sample size n alone can create a bias-variance dilemma. Additionally, the principal disadvantage of HS method is that it "ignores the potentially useful information in the volatility dynamics" [Marimoutou et al., 2009]. Since there's no distinction in terms of probability weight of the past returns (assigning equal weight to each

day's return, combined with the fact that the choice for n is ambiguous), the HS method is unrealistic. Volatility in practice tends to change through time and cluster together, which cannot be captured by this approach. A better approach should be chosen that can utilize the volatility dynamics in the observed data without making assumptions about the true return distribution, which brings us to the next part — the filtered HS approach.

2.2.2 Filtered HS Approach

Another version that can solve some of the major flaws from the original HS approach is the filtered historical simulation (FHS) approach. This approach was introduced by Hull and White [1998] and Barone-Adesi et al. [1999]. Even though this approach keeps the model-free nature of the HS approach, it can also forecast VaR via a volatility model by including the GARCH model. The biggest benefit of this approach compared to the HS approach is that it can, thanks to the GARCH model, measure risks that reflect the recent situation, regardless of how extreme the confidence level might be [Marimoutou et al., 2009].

Based on the illustration by Christoffersen [2012, Chap. 6], the procedure of this approach can be summarized as the following: First, fit a GARCH model to the return series of interest, say, from day t back to day $(t+1-m)$ (meaning that we observe data of the last m days). Note that after we get the conditional variance value at day i, we can calculate the variance in the GARCH model of day $(i+1)$ if the order of the GARCH model was (1,1). From these observed returns and their calculated conditional standard deviations, the standardized returns are computed. This step is done to include information on volatility dynamics from past data into our model, which is the *"filtering"* part of this approach. Now, we are ready to proceed the same way as in the HS approach for these past standardized returns. Instead of assuming a specific distribution to generate the standardized returns for simulation, we resample these calculated standardized returns randomly with replacement. This set of standardized returns and the calculated conditional variance at day $(t+1)$ given information available up to time t will help us obtain a set of hypothetical returns from day $(t+1)$ to day $(t+K)$.

Let's write this procedure mathematically to understand it better:

1. Assume our model of returns follow a GARCH$(1,1)$ process (as we mentioned before, GARCH(1,1) is a very parsimonious model):

$$R_t = \sigma_t z_t,$$

and

$$\sigma_{t+1}^2 = \varphi + \alpha R_t^2 + \beta \sigma_t^2.$$

2. *Filtering:* Given the returns from the past m days, $\{R_{t+1-\tau}\}_{\tau=1}^{m}$, we calculate the past standardized returns:

$$\hat{z}_{t+1-\tau} = \frac{R_{t+1-\tau}}{\sigma_{t+1-\tau}}.$$

3. *HS:* From the conditional variance calculated via the variance GARCH (1,1) model in 1., σ_{t+1}^2, and from resampling with random draw with replacement from $\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}$, we can calculate hypothetical K-day returns as

$$\begin{aligned}\hat{R}_{t+1} &= \sigma_{t+1}\hat{z}_{t+1},\\ \hat{\sigma}_{t+2}^2 &= \varphi + \alpha \hat{R}_{t+1}^2 + \beta \sigma_{t+1}^2,\end{aligned}$$

Then,

$$\hat{R}_{t+2} = \hat{\sigma}_{t+2}\hat{z}_{t+2},$$

and so forth, for $k = 2, \ldots, K$:

$$\hat{R}_{t+k} = \hat{\sigma}_{t+k}\hat{z}_{t+k}.$$

Repeating this procedure for N simulations, we get $\{\hat{R}_{n,(t+k)}\}_{n=1}^{N}$. The hypothetical K-day returns become:

$$\hat{R}_{n,(t+1):(t+K)} = \sum_{k=1}^{K} \hat{R}_{n,(t+k)} \qquad \text{for } n = 1, 2, \ldots, N.$$

Now that we have a FHS set of hypothetical returns, the K-day VaR is easily calculated:

$$\text{VaR}_{p,(t+1):(t+K)} = \text{Percentile}\{\{\hat{R}_{n,(t+1):(t+K)}\}_{n=1}^{N}, 100p\}.$$

And for the next day, VaR is quite simple since we do not have to simulate hypothetical returns in the future; we only need to resample the past standardized returns in the chosen time horizon, and VaR is therefore

$$\text{VaR}_{p,t+1} = \sigma_{t+1} \,\text{Percentile}\{\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}, 100p\}.$$

Again, the FHS approach is a good choice compared to the simple HS approach as it captures the conditional variance model while making no assumptions about the tail distribution.

2.3 Conditional EVT Method (via POT Approach)

When we introduced the EVT method in Chapter 1, we did not specify which approach we will be using. The approach used for our extreme value analysis relies on taking out values that exceed a chosen threshold, which are considered *peaks*. This approach has a straightforward name — peaks-over-threshold (POT). According to Byström [2005], the POT method is one of the most widely known methods of EVT. From now on, when we apply the EVT method, it is automatically understood that we are using the POT approach.

We can now study the EVT-POT method as described by Christoffersen [2012] in his book. Note that the theory is stated in its natural form — returns beyond a certain threshold are considered as *peaks*, while our interest is in the standpoint of a long position (i.e., returns of our interest are those of smaller value than a threshold in the left tail distribution), the equation for VaR will, thus, be adjusted accordingly to fit our assumption.

2.3.1 EVT

Let u be a threshold, consider the probability of standardized returns z less u being below a value x, given that the standardized return is beyond the threshold (hence, z is the *peak*, $z-u$ is the size of the exceedance), then the conditional cumulative distribution function can be written as

$$\begin{aligned} F_u(x) &\equiv P\{z - u \leq x | z > u\}, \text{where } x > u \\ &= \frac{P\{u < z \leq x + u\}}{P\{z > u\}} = \frac{F(x+u) - F(u)}{1 - F(u)}. \end{aligned} \tag{2.14}$$

EVT states that as the threshold u gets large, $F_u(x)$ converges in distribution to the generalized Pareto (GP) distribution:

$$G(x;\xi,\beta) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1-\exp(-x/\beta) & \text{if } \xi = 0 \end{cases},$$

where $\beta > 0$ and x satisfies

$$\begin{cases} x \geq u & \text{if } \xi \geq 0 \\ u \leq x \leq u - \frac{\beta}{\xi} & \text{if } \xi < 0 \end{cases}.$$

Student's $t(d)$ distribution has a positive tail parameter ξ as it is a standard heavy tailed distribution, which is covered by the EVT result. In finance, returns tend to have fat tails so it is often assumed that the tail parameter is positive when applying a model into the return distribution.

We can now estimate the parameters of the EVT model. Let $y = x + u$, for x and u being points in the tail of the distribution of interest, then from (2.14) we have

$$F_u(x) = F_u(y-u) = \frac{F(x+u)-F(u)}{1-F(u)} = \frac{F(y)-F(u)}{1-F(u)}.$$

Hence,

$$\begin{aligned} F(y) &= F_u(y-u)\left[1-F(u)\right] + F(u) \\ &= 1-\left[1-F(u)\right]\left[1-F_u(y-u)\right]. \end{aligned}$$

Let T be the total sample size and T_u be the number of observations beyond the threshold u. Then $1-F(u)$ can be estimated simply by the proportion T_u/T. And $F_u(y-u)$ can be estimated by MLE on the standardized observations in excess of the chosen threshold $(x = y-u)$. With the assumption of $\xi \neq 0$, the distribution becomes

$$F(y) = 1 - \frac{T_u}{T}\left(1+\xi(y-u)/\beta\right)^{-1/\xi}. \tag{2.15}$$

Since financial return distributions are typically fat tailed as mentioned, we can assume that the tail parameter $\xi > 0$. Then,

$$\begin{aligned} P(z>y) = 1-F(y) &= \frac{T_u}{T}\left(1+\xi(y-u)/\beta\right)^{-1/\xi} \\ &= L(y)y^{-1/\xi} \approx cy^{-1/\xi} \quad \text{for } y > u. \end{aligned}$$

$L(y)$ is set to a constant c, as it is a slowly varying function of y. Thus, the likelihood function for all observations y_i that are larger than the threshold u, is

$$\begin{aligned} L &= \prod_{i=1}^{T_u} f(y_i)/(1-F(u)) \\ &= \prod_{i=1}^{T_u} \frac{-\frac{1}{\xi} c {y_i}^{-(1/\xi)-1}}{(cu^{-1/\xi})}, \text{ for } y_i > u. \end{aligned}$$

By solving this function, we get

$$\xi = \frac{1}{T_u} \sum_{i=1}^{T_u} \ln\left(\frac{y_i}{u}\right). \tag{2.16}$$

This simple estimator for ξ is called the Hill estimator.

Note: *As suggested by Christoffersen [2012, Chap. 6], "a good rule of thumb is to set the threshold so as to keep the largest 5% of the observations for estimating ξ".*

Set $1 - F(u) = cu^{-1/\xi}$ equal to 1 minus the proportion of the data points beyond the threshold. Then, the parameter c is estimated by:

$$c = \frac{T_u}{T} u^{1/\xi}.$$

The CDF for observations y_i larger than u is then:

$$F(y) = 1 - cy^{-1/\xi} = 1 - \frac{T_u}{T}(y/u)^{-1/\xi}. \tag{2.17}$$

2.3.2 Conditional EVT (GARCH-EVT)

A drawback of the EVT method is that the returns' i.i.d. property in practice does not actually hold, and this property is a requirement to use the limiting cumulative distribution. Therefore, in order to apply the EVT method to real-life data, we must get rid of the time-dependent variance patterns to generate independent inputs from our original data. One way to do that is to obtain standardized data — or specifically, in the case when the data is financial returns, we should obtain a set of standardized returns $\{z_t\}$. Standardized return is the result of the return at time t, R_t, minus the average return (which is essentially 0), which is then divided by the

conditional standard deviation at time t, σ_t. These z_t values can reasonably be considered as i.i.d. in most situations [Christoffersen, 2012, Chap. 6]. The extrapolated quantile of the next day, which is inherently VaR, would then be calculated by multiplying the conditional standard deviation at time $t+1$ by the quantile of standardized returns that is estimated by the EVT model.

We know from earlier section that the GARCH model can conveniently model the next day's conditional standard deviation. However, a downside of this model when it comes to extreme value analysis is that it focuses on the whole return distribution, instead of just the tail, which we are primarily interested in. Therefore, a combination of the GARCH and EVT approaches, which was suggested by McNeil and Frey [2000] and which we here call conditional EVT, is a reasonable solution to deal with these downsides — lack of i.i.d. property and explicit tail distribution [Marimoutou et al., 2009].

To apply the conditional EVT method, we first need to fit a GARCH model into the return data and obtain the conditional standard deviation for each day. We can then extract the standardized returns, z_t, and estimate the conditional variance at time $t+1$, σ_{t+1}, from the fitted model. This is the *conditional* part of this method. After doing that, we can use the EVT approach to model our data and proceed to estimate VaR. As noted from the beginning of the section, we are interested in the left tail distribution. Therefore, to calculate VaR of returns in this tail from the EVT-POT introduced above, we must make some adjustments. The details of these adjustments are discussed in the next part.

Here is a summary of the procedure for the general conditional EVT (GARCH-EVT) approach based on the EVT method discussed above (hence, without adjustments to the left tail distribution for consistency with the previous part), where the fitted model is assumed to be a GARCH(1,1) model:

1. *Conditioning*: Assuming that our returns follow a GARCH(1,1) process (which, as we mentioned before, is a very parsimonious model), the conditional variance for each day is

$$\sigma_t^2 = \varphi + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2.$$

We can then calculate the conditional standard deviation and extract

the standardized returns as

$$z_t = \frac{R_t}{\sigma_t},$$

and calculate the one-day ahead conditional variance,

$$\sigma_{t+1}^2 = \varphi + \alpha R_t^2 + \beta \sigma_t^2.$$

2. *Applying EVT*: Following the EVT method above, we can calculate the quantile Φ_p^{-1} of the *peak* standardized returns (standardized returns z's that are larger than the threshold u) by setting the CDF in (2.17) of these *peaks* to some probability p. Then we have

$$F\left(\Phi_p^{-1}\right) = 1 - \frac{T_u}{T}(z/u)^{-1/\xi} = p,$$

where ξ is the Hill estimator as calculated in (2.16).

The one-day ahead VaR is then the p^{th} quantile of our returns, which is the product of the conditional standard deviation for the next day and the p^{th} quantile of the *peaks* of the standardized returns:

$$\text{VaR}_{p,t+1} = \sigma_{t+1}\Phi_p^{-1}.$$

2.3.3 VaR from (Conditional) EVT Quantile

As we have emphasized before, the EVT-POT method gives attention to the *peak*s beyond some threshold while we are interested in the returns that are lower than a certain threshold. Therefore, in principle, we are looking at the left tail of the return distribution while the theory is stated in the interest of the right tail distribution.

In order to fit the EVT approach to our intended goal, the first necessary step is to adjust our standardized returns by flipping them over, i.e., changing their sign. We will thus use the negative standardized returns, which we name y_t, instead of the standardized returns z_t. Assuming that the original return series is a set $\{R_t\}$ and its corresponding time-dependent standard deviation is σ_t, then the negative standardized returns y_t are

$$y_t = -z_t = -\frac{R_t}{\sigma_t}.$$

Again, from a risk manager's perspective, our biggest concern lies with the returns that are less than a low quantile, say $\text{VaR}_{p=0.01}\{R_t\}$, of our original raw returns. Then from the CDF in (2.17), we would like $(1-p)\cdot 100\% = 99\%$ of our negative standardized returns to be smaller than the quantile F^{-1}_{1-p} of those negative standardized returns larger than a threshold u. Thus, letting (2.17) take the value of $(1-p)$, we have

$$\begin{aligned} F\left(F^{-1}_{1-p}\right) &= 1-p \\ &= 1-\frac{T_u}{T}\left(F^{-1}_{1-p}/u\right)^{-1/\xi}. \\ \text{So,} \qquad p &= \frac{T_u}{T}\left(F^{-1}_{1-p}/u\right)^{-1/\xi}, \\ \text{and} \quad F^{-1}_{1-p} &= u\left[p/(T_u/T)\right]^{-\xi}. \end{aligned} \tag{2.18}$$

Finally, the $\text{VaR}_{p,t}\{R_t\}$ of the original return of the next day, combined with the variance model and calculated standardized return quantile F^{-1}_{1-p}, is estimated as

$$\text{VaR}_{p,t+1} = -\sigma_{t+1}F^{-1}_{1-p}.$$

2.4 ACER

According to Næss [2010], the problem of EVT methods is that it is uncertain to what extent the asymptotic results from the EVT based on the observed data are applicable in practice. The average conditional exceedance rate (ACER) method was therefore proposed by Næss [2010] to overcome some disadvantages of EVT methods, which require i.i.d. observations and use asymptotic assumptions for the data [Dahlen et al., 2015]. The method is sectioned into 3 parts as below.

Note: *The reasonings, theories, and notations below are largely taken directly from Næss et al. [2013]).*

2.4.1 Cascade of Conditioning Approximations

Consider a stochastic process $Z(t)$ for t in $(0,T)$ and let $X_1,\ldots,X_N$ be derived from discrete times $t_1,\ldots,t_N$ in $(0,T)$. Let $M_N = \max\{X_j; j=1,\ldots,N\}$.

Then we want to have a precise estimation of $P(\eta) = P(M_N \leq \eta)$ for large η.

$$\begin{aligned} P(\eta) &= P(X_1 \leq \eta, \ldots, X_N \leq \eta) \\ &= P(X_N \leq \eta | X_{N-1} \leq \eta, \ldots, X_1 \leq \eta) P(X_{N-1} \leq \eta, \ldots, X_1 \leq \eta) \\ &= \left[\prod_{j=2}^{N} P(X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta) \right] P(X_1 \leq \eta). \end{aligned} \tag{2.19}$$

Assuming a one-step memory approximation, we condition only on the previous observation, then

$$\begin{aligned} &P\left(X_j \leq \eta \middle| X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\right) \\ &\approx P\left(X_j \leq \eta \middle| X_{j-1} \leq \eta\right) \qquad\qquad 2 \leq j \leq N. \end{aligned} \tag{2.20}$$

By conditioning on one more observation, the approximation becomes

$$\begin{aligned} &P\left(X_j \leq \eta \middle| X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\right) \\ &\approx P\left(X_j \leq \eta \middle| X_{j-2} \leq \eta, X_{j-1} \leq \eta\right) \quad 3 \leq j \leq N(3). \end{aligned} \tag{2.21}$$

Let $p_{kj}(\eta) = P(X_{j-k+1} \leq \eta, \ldots, X_j \leq \eta)$ for $j > k$ (the probability of k consecutive elements of X from X_{j-k+1} to X_j smaller or equal to η). Then from (2.19) and (2.20), $P(\eta)$ becomes

$$P(\eta) = \left[\prod_{j=2}^{N} P(X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta) \right] P(X_1 \leq \eta) \tag{2.22}$$

$$\begin{aligned} &\approx \left[\prod_{j=2}^{N} P(X_j \leq \eta | X_{j-1} \leq \eta) \right] P(X_1 \leq \eta) \\ &= \frac{\prod_{j=2}^{N} p_{2j}(\eta) P(X_1 \leq \eta)}{P(X_1 \leq \eta) P(X_2 \leq \eta) \cdots P(X_{N-1} \leq \eta)} \\ &= \frac{\prod_{j=2}^{N} p_{2j}(\eta)}{\prod_{j=2}^{N-1} p_{1j}(\eta)}. \end{aligned} \tag{2.23}$$

Note that if the values of X_j are independent, then (2.22) is the classical approximation result

$$P(\eta) \approx P_1(\eta) = \prod_{j=1}^{N} P(X_j \leq \eta). \tag{2.24}$$

Now let $\alpha_{1j}(\eta) = P(X_j > \eta) = 1 - p_{1j}$, then (2.23) becomes

$$\begin{aligned} P(\eta) &\approx \prod_{j=1}^{N} \left(1 - \alpha_{1j}(\eta)\right) \approx \prod_{j=1}^{N} \exp\left\{-\alpha_{1j}(\eta)\right\} \\ &= \exp\left\{-\sum_{j=1}^{N} \alpha_{1j}(\eta)\right\}. \qquad (\text{since } 1 - y \approx e^{-y}) \end{aligned} \tag{2.25}$$

Building from $\alpha_{1j}(\eta)$, we can decluster consecutive exceeders X_j by defining $\alpha_{kj}(\eta)$:

$$\begin{aligned} \alpha_{kj}(\eta) &= P\left(X_j > \eta \middle| X_{j-k+1} \leq \eta, \ldots, X_{j-1} \leq \eta\right) \\ &= 1 - P\left(X_j \leq \eta \middle| X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\right) \\ &= 1 - \frac{p_{kj}(\eta)}{p_{k-1,j-1}(\eta)}, \qquad j \geq k \geq 2. \end{aligned}$$

Then,

$$1 - \alpha_{2j}(\eta) = \frac{p_{2j}(\eta)}{p_{1,j-1}(\eta)},$$

so that

$$\prod_{j=2}^{N} \left(1 - \alpha_{2j}(\eta)\right) = \prod_{j=2}^{N} \frac{p_{2j}(\eta)}{p_{1,j-1}(\eta)} = \frac{p_{23}(\eta) p_{23}(\eta) \cdots p_{2N}(\eta)}{p_{11}(\eta) p_{12}(\eta) \cdots p_{1,N-1}(\eta)}.$$

Thus, (2.23), which considers two consecutive observations at each conditioning, now becomes

$$\begin{aligned} P\left(\eta\right) \approx P_2\left(\eta\right) &= \prod_{j=2}^{N} \left(1 - \alpha_{2j}(\eta)\right) p_{11}\left(\eta\right) \\ &\approx \exp\left\{-\sum_{j=2}^{N} \alpha_{2j}(\eta)\right\} \left(1 - \alpha_{11}\left(\eta\right)\right) \\ &\approx \exp\left\{-\sum_{j=2}^{N} \alpha_{2j}(\eta) - \alpha_{11}\left(\eta\right)\right\}. \end{aligned}$$

Similarly, we get

$$P\left(\eta\right) \approx P_3\left(\eta\right) \approx \exp\left\{-\sum_{j=3}^{N} \alpha_{3j}(\eta) - \alpha_{22}\left(\eta\right) - \alpha_{11}\left(\eta\right)\right\}.$$

Therefore, in general, for $j \geq 2, j > k$,

$$P(\eta) \approx P_k(\eta) \approx \exp\left\{ -\sum_{j=k}^{N} \alpha_{kj}(\eta) - \sum_{j=1}^{k-1} \alpha_{jj}(\eta) \right\}, \tag{2.26}$$

and $P_k(\eta) \to P(\eta)$ as $k \to N$ with $P_k(\eta) = P(\eta)$ as $\eta \to \infty$.

For the cascade of approximations $P_k(\eta)$ to have practical significance, an implicit assumption is made about a cut-off value k_c satisfying $k_c \ll N$ such that effectively $P_{k_c}(\eta) = P_N(\eta)$. Note that, for data where X_i and X_j are independent whenever $|j-i| > k$ (k-dependent stationary data sequences), then $P(\eta) = P_{k+1}(\eta)$ and under mild conditions on the joint distributions of the data, $\lim_{N\to\infty} P_1(\eta) = \lim_{N\to\infty} P(\eta)$. For finite values of N, this is much more complex, and purely asymptotic results should be used with caution.

With the assumed cut-off value k_c and all k-values of interest such that $k \ll N$, and with the extreme value prediction by the conditioning approach described above, $\sum_{j=1}^{k-1} \alpha_{jj}(\eta)$ becomes effectively negligible compared to $\sum_{j=k}^{N} \alpha_{kj}(\eta)$. Thus, a simpler approximation that is applicable to both stationary and nonstationary data is adopted as

$$P(\eta) \approx P_k(\eta) = \exp\left\{ -\sum_{j=k}^{N} \alpha_{kj}(\eta) \right\}, \qquad k \geq 1. \tag{2.27}$$

From the definition of $\alpha_{1j}(\eta) = P(X_j > \eta)$, it follows that $\sum_{j=1}^{N} \alpha_{1j}(\eta)$ is equal to the expected number of exceedances of the threshold η during the time interval $(0, T)$.

Assuming that the data is independent, the approximation can be written as

$$\begin{aligned} P(\eta) &= \prod_{j=2}^{N} P(X_j \leq \eta) = \prod_{j=2}^{N} \left(1 - \alpha_{1j}(\eta)\right) \\ &\approx \exp\left\{ -\sum_{j=1}^{N} \alpha_{1j}(\eta) \right\} = P_1(\eta). \end{aligned}$$

As a quick reminder, if X is a Poisson random variable with mean λ, then $P(X \leq k) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$.

Thus, $P_1(\eta)$ expresses the approximation that the stream of exceedance events constitutes a nonstationary Poisson process. This also leads to an

understanding of (2.27): $\sum_{j=k}^{N} \alpha_{kj}(\eta)$ can be interpreted as the expected effective number of independent exceedance events provided by conditioning on $(k-1)$ previous observations.

2.4.2 Empirical Estimation of the Average Conditional Exceedance Rates

Define the average conditional exceedance rate (ACER) of order k for $k = 1, 2, \ldots$ as

$$\varepsilon_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^{N} \alpha_{kj}(\eta). \tag{2.28}$$

Hence, the empirical estimation of the ACER function $\bar{\varepsilon}_k(\eta)$ can be obtained by counting the total number of favorable incidents (exceedances conditioned on a determined number of previous non-exceedances) from the total data series and divided by $N-k+1 \approx N$.

In practice, there are two typical cases for the underlying process $Z(t)$. One, the process is stationary or even ergodic. Two, the process depends on certain parameters whose variation in time may be modelled as an ergodic process. For each set of values of the parameters, the premise is that $Z(t)$ can then be modelled as an ergodic process, which can be used to model long-term statistics [Næss, 1984; Schall et al., 1991]. For both scenarios, the empirical estimation above proceeds the same way, and it can also apply for the long-term situation.

Now, for $j = k, \ldots, N$ and $k = 2, 3, \ldots, j$, we define two random functions:

$$\begin{aligned} A_{kj}(\eta) &= I\left\{X_j > \eta, X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\right\}, \\ B_{kj}(\eta) &= I\left\{X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\right\}, \end{aligned}$$

where $I\{\mathcal{A}\}$ denotes the indicator function of some event $\mathcal{A}$. Then,

$$\alpha_{kj}(\eta) = \frac{\mathbb{E}\left[A_{kj}(\eta)\right]}{\mathbb{E}\left[B_{kj}(\eta)\right]}, \quad \text{with } j = k, \ldots, N \text{ and } k = 2, 3, \ldots, j,$$

where $\mathbb{E}[\ldots]$ denotes the expectation operator.

Assuming an ergodic process, then $\varepsilon_k(\eta) = \alpha_{kk}(\eta) = \ldots = \alpha_{kN}(\eta)$, and by replacing ensemble means with corresponding time averages, it may be assumed that for the time series at hand

$$\varepsilon_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} a_{kj}(\eta)}{\sum_{j=k}^{N} b_{kj}(\eta)}, \tag{2.29}$$

where $a_{kj}(\eta)$ and $b_{kj}(\eta)$ are realizations of $A_{kj}(\eta)$ and $B_{kj}(\eta)$, respectively, for the observed time series.

As $\eta \to \infty$, clearly $\mathbb{E}\left[B_{kj}(\eta)\right] = 1$, then $\lim_{\eta\to\infty} \sum_{j=k}^{N} b_{kj}(\eta) = N - k + 1$. Now, denote $\widetilde{\varepsilon}_k(\eta)$ to be:

$$\widetilde{\varepsilon}_k(\eta) = \frac{\sum_{j=k}^{N} \mathbb{E}\left[A_{kj}(\eta)\right]}{N - k + 1}, \tag{2.30}$$

then it follows that $\lim_{\eta\to\infty} \frac{\widetilde{\varepsilon}_k(\eta)}{\varepsilon_k(\eta)} = 1$.

For $k \geq 2$, using the modified ACER function $\widetilde{\varepsilon}_k(\eta)$ is obviously easier for long-term statistics than $\varepsilon_k(\eta)$. It is also more convenient to apply for nonstationary time series because

$$\begin{aligned} P(\eta) &\approx \exp\left\{-\sum_{j=k}^{N} \alpha_{kj}(\eta)\right\} \\ &= \exp\left\{-\sum_{j=k}^{N} \frac{\mathbb{E}\left[A_{kj}(\eta)\right]}{\mathbb{E}\left[B_{kj}(\eta)\right]}\right\} \overset{\eta\to\infty}{\cong} \exp\left\{-\sum_{j=k}^{N} \mathbb{E}\left[A_{kj}(\eta)\right]\right\}. \end{aligned} \tag{2.31}$$

Let us assume that the time series can be segmented into K blocks, such that $\mathbb{E}\left[A_{kj}(\eta)\right]$ remains approximately constant within each block. Let us also assume that for a sufficient range of η-values we have $\sum_{j\in C_i} \mathbb{E}\left[A_{kj}(\eta)\right] \approx \sum_{j\in C_i} a_{kj}(\eta)$, where C_i denotes the set of indices for block no. $i = 1, \ldots, K$. Then,

$$\sum_{j=k}^{N} \mathbb{E}\left[A_{kj}(\eta)\right] \approx \sum_{j=k}^{N} a_{kj}(\eta).$$

Thus, we can now rewrite (2.31) as

$$P(\eta) \approx \exp\left\{-(N - k + 1)\hat{\varepsilon}_k(\eta)\right\}, \tag{2.32}$$

where $\hat{\varepsilon}_k(\eta) = \frac{\sum_{j=k}^{N} a_{kj}(\eta)}{N-k+1}$.

We would like to know which events are counted for the estimation of the various $\varepsilon_k(\eta)$, for $k \geq 2$. From the definition of ACER, $\varepsilon_2(\eta)(N-1)$ can be interpreted as the expected number of exceedances above threshold η, conditioning on the previous observation (an exceedance is only counted if it is immediately preceded by a non-exceedance).

Defining a clump of exceedances as a maximum number of consecutive exceedances above η, then in general, $\varepsilon_k(\eta)(N-k+1)$ is equal to the average number of clumps of exceedances above η, for the realizations considered, separated by at least $(k-1)$ non-exceedances.

Keep in mind that the ACER functions automatically account for the clumping or clustering tendency of the data and its effect on the extreme value distribution.

Now, assuming a stationary time series, we can estimate a confidence interval for $\varepsilon_k(\eta)$ with R available realizations of the requisite length of the time series, or one long realization segmented into R subseries. Define

$$\hat{\varepsilon}_k^{(r)} = \frac{\sum_{j=k}^{N} a_{kj}^{(r)}(\eta)}{N-k+1},$$

where the index r refers to the r^{th} realization. Then, for both stationary and non-stationary time series, the sample estimate of $\widetilde{\varepsilon}_k(\eta)$, $\hat{\varepsilon}_k(\eta)$, is

$$\hat{\varepsilon}_k(\eta) = \frac{1}{R}\sum_{r=1}^{R} \hat{\varepsilon}_k^{(r)}(\eta).$$

The sample variance $\hat{s}_k^2(\eta)$, thus, can be estimated as

$$\hat{s}_k^2(\eta) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\varepsilon}_k^{(r)}(\eta) - \hat{\varepsilon}_k(\eta)\right)^2.$$

If the realizations are independent for a suitable number R, say $R \geq 20$, we can approximate the 95% confidence interval CI for the value $\varepsilon_k(\eta)$, where

$$CI^{\pm}(\eta) = \hat{\varepsilon}_k(\eta) \pm \frac{1.96\hat{s}_k(\eta)}{\sqrt{R}}.$$

Now, assuming that the stream of conditional exceedances over a threshold η constitute a Poisson process, which also applies to the non-stationary case, then the variance of the estimator $\hat{E}_k(\eta)$ of $\tilde{\varepsilon}_k(\eta)$, where

$$\hat{E}_k(\eta) = \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1}$$

is $\mathrm{Var}[\hat{E}_k(\eta)] = \tilde{\varepsilon}_k(\eta)$, Hence, for high levels η, the arpproximate limits of a 95% confidence interval of $\tilde{\varepsilon}_k(\eta)$ and $\varepsilon_k(\eta)$ is

$$CI^{\pm}(\eta) = \hat{\varepsilon}_k(\eta)\left(1 \pm \frac{1.96}{\sqrt{(N-k+1)\hat{\varepsilon}_k(\eta)}}\right).$$

2.4.3 Estimation of Extremes for the General Case

This approach to extreme value prediction derives from an underlying premise concerning the relevant asymptotic extreme value distribution, which is assumed here to be of Fréchet type (typical for financial data). This premise leads to the assumption about the sampled time series to be used as a basis for prediction. For independent data, this assumption can be expressed in terms of the ACER function $\varepsilon_1(\eta)$ as

$$\varepsilon_1(\eta) \approx \left[1 + \xi\left(a(\eta - b)\right)\right]^{-1/\xi}, \text{ for } \eta \geq \eta_0, \tag{2.33}$$

for a suitable asymptotic tail marker η_0, where $a > 0$, b, $\xi > 0$ are constants. (2.33) corresponds to (1.4) in the case where $x = a(\eta - b) > \frac{-1}{\xi}$, and $\xi > 0$.

Using the asymptotic form as a guide, it is assumed that the behavior of the mean exceedance rate in the subasymptotic part of the tail will follow a function largely of the form $\left[1 + \xi\left(a(\eta - b)\right)\right]^{-1/\xi}$ $(\eta \geq \eta_1 \geq b)$, where $a > 0$, b, $c > 0$, and $\xi > 0$ are suitable constants, and η_1 is an appropriately chosen tail level. Then,

$$\varepsilon_k(\eta) \approx q_k(\eta)\left[1 + \xi_k\left(a_k(\eta - b_k)^{c_k}\right)\right]^{-1/\xi_k}, \qquad \eta \geq \eta_1, \tag{2.34}$$

where the function $q_k(\eta)$ is weakly varying compared with the function $\left[1 + \xi_k\left(a_k(\eta - b_k)^{c_k}\right)\right]^{-1/\xi_k}$, and $a_k > 0$, b_k, $c_k > 0$, and $\xi_k > 0$ are suitable constants. When $c_k = 1$ and $q_k(\eta) = 1$, it is the asymptotic limit.

It is also expedient to assume that the unknown function $q_k(\eta)$ varies sufficiently slowly in the tail region to be replaced by a constant for $\eta \geq \eta_1$. Then for simplicity of notation, we suppress the index k on the ACER functions, which can be rewritten as

$$\varepsilon(\eta) \approx q\left[1 + \tilde{a}(\eta - b)^c\right]^{-\gamma}, \quad \eta \geq \eta_1, \quad \text{where } \gamma = \frac{1}{\xi}, \quad \tilde{a} = a\xi. \tag{2.35}$$

In practice, the tail marker η_1 is provisionally identified from visual inspection of the log plot $(\eta, \ln \hat{\varepsilon}_k(\eta))$, where the chosen value corresponds to the beginning of regular tail behavior in a sense to be discussed below.

The optimization process to estimate the parameters is done relative to the log plot. The mean square error (MSE) function to be minimized is

$$F(\tilde{a}, b, c, q, \gamma) = \sum_{j=1}^{N} w_j \left| \log \hat{\varepsilon}\left(\eta_j\right) - \log q + \gamma \ln \left[1 + \tilde{a}\left(\eta_j - b\right)^c\right] \right|^2, \quad (2.36)$$

where $w_j = \left(\log CI^+(\eta_j) - \log CI^-(\eta_j)\right)^{-2}$ denotes a weight factor putting more emphasis on more reliable data points.

With $\tilde{a}$, b, and c fixed, the optimal values of γ and $\log q$ are found by using closed form weighted linear regression formulas in terms of w_j. Let $y_j = \log \hat{\varepsilon}(\eta_j)$ and $x_j = 1 + \tilde{a}(\eta_j - b)^c$, then the optimal values of γ and $\log q$ are

$$\gamma^*(\tilde{a}, b, c) = -\frac{\sum_{j=1}^{N} w_j (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{N} w_j (x_j - \bar{x})^2} \quad (2.37)$$

and

$$\log q^*(\tilde{a}, b, c) = \bar{y} + \gamma^*(\tilde{a}, b, c)\bar{x}. \quad (2.38)$$

To solve for an optimal set of parameters $\{\tilde{a}, b, c\}$, we can use the Levenberg-Marquardt method on the function $\tilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$, then following (2.37) and (2.38) will calculate the corresponding γ^* and q^*.

For estimation of the confidence interval for the predicted return value provided by the optimal curve, the empirical confidence band is reanchored to the optimal curve. The range of fitted curves that stay within the reanchored confidence band will determine an optimized confidence interval of the predicted return value. As a final point, it has been observed that the predicted return value is not very sensitive to the choice of η_1. However, this sensitivity should always be checked to verify the robustness of the obtained predictions.

Finally, we can calculate VaR for this ACER method. This is quite simple to achieve since we only need to invert the ACER function in (2.35) to solve for η. Assuming the loss will not exceed VaR with probability $1 - p$, $P(X_t < \mathrm{VaR}_{1-p}) = 1 - p$, then,

$$\mathrm{VaR}_{1-p} = b + \left(\frac{1}{\tilde{a}}\left(\frac{q}{p}\right)^{1/\gamma} - 1\right)^{1/c}.$$

Note that the VaR calculated above is for potential losses, which are the negative of the returns. The VaR_p for our original returns in the left tail is

$$\text{VaR}_{p,\text{return}} = -\,\text{VaR}_{1-p,\text{loss}} = -\left(b + \left(\frac{1}{\tilde{a}}\left(\frac{q}{p}\right)^{1/\gamma} - 1\right)^{1/c}\right).$$

2.5 Test and Evaluation

After forecasting extreme returns via estimating VaRs from different methods, we need to backtest the resulting VaR estimates by comparing them with the realized returns (out-of-sample period). There are a set of statistical techniques to evaluate aggregate risk models, where an unconditional test can be implemented for the average probability of a VaR violation, or a conditional coverage test consisted of the unconditional test and an independence test. In this section, we will follow the reasoning and notations from Christoffersen [2012, Chap. 13].

First, we shall introduce what a VaR violation is. The purpose of VaR in risk management is to gauge the maximal potential loss during a given time period, hence, if the realized return is *worse* than the predicted VaR, there is a violation (showing less reliability in our models). The definition of *worse* here is relative to each particular financial position, because a bad event considered by a long position is opposite to one by a short position. (Refer back to section 1.3 for the definition of these two financial positions.)

Let R_{PF} be the portfolio return, $\text{VaR}_{p,t+1}$ be the one-day predicted VaR. This means that we expect "the actual return will only be worse than the $\text{VaR}_{p,t+1}$ forecast $p \cdot 100\%$ of the time" [Christoffersen, 2012]. The "hit sequence" of VaR violations is defined in Christoffersen [2012, Chap. 13] and can be adjusted to our perspective of a long position as

$$I_{t+1} = \begin{cases} 1 & \text{if } R_{PF,t+1} < \text{VaR}_{p,t+1}, \\ 0 & \text{if } R_{PF,t+1} \geq \text{VaR}_{p,t+1}\,. \end{cases}$$

Hence, $\{I_{t+1}\}_{t=1}^{T}$ is a sequence showing whether past violations occurred in the period of T days.

Assuming we have an optimal VaR model, then given all the information available up to time t, we should only be able to forecast VaR but not the

VaR violations. The capability of predicting these violations means that we obtained information that could have been used for building a better model, implying that our current VaR model would not be optimal. Therefore, the hit sequence from a correctly specified risk model should look like a series of random tosses of a coin with the probability of coming up heads $p \cdot 100\%$ of the time, assuming the coverage rate of VaR is p. Thus, the hit sequence should follow the Bernoulli distribution over time. We have the following null hypothesis:

$$H_0 : I_{t+1} \sim \text{i.i.d. Bernoulli } (p).$$

We can now move on with the backtesting procedure for VaR models.

2.5.1 Unconditional Coverage Testing

The unconditional coverage hypothesis examines whether the fraction of violations of a particular risk model, π, is significantly different from the expected fraction, p.

Denote T_0 and T_1 to be the numbers of 0s and 1s in the sample, respectively, for T days, then the likelihood function of an i.i.d. Bernoulli(π) hit sequence is

$$L(\pi) = \prod_{t=1}^{T} (1-\pi)^{1-I_{t+1}} \pi^{I_{t+1}} = (1-\pi)^{T_0} \pi^{T_1}.$$

Thus, π can be easily estimated by computing the proportion of observed violations in the sequence, $\hat{\pi} = T_1/T$. The likelihood function is then optimized:

$$\begin{aligned} & L(\hat{\pi}) = (1 - T_1/T)^{T_0} (T_1/T)^{T_1}, \\ \xrightarrow{\text{under } H_0} \quad & L(p) = (1-p)^{T_0} p^{T_1}. \end{aligned}$$

Kupiec [1995] suggested to conduct this proportion of failures test (unconditional coverage test) by using a likelihood ratio test:

$$LR_{uc} = -2 \ln \left[L(p) / L(\hat{\pi}) \right].$$

As $T \to \infty$, $LR_{uc} \sim \chi^2_{(1)}$. Therefore, we get the asymptotic result

$$LR_{uc} = -2 \ln \left\{ (1-p)^{T_0} p^{T_1} / \left[(1 - T_1/T)^{T_0} (T_1/T)^{T_1} \right] \right\} \sim \chi^2_{(1)}.$$

There are two types of errors that need assessing. Type I error occurs when we reject a correct model, while Type II error occurs when we fail to reject an incorrect model. When the significance level increases, the probability of making Type I error is larger while making Type II errors is smaller, and vice versa. According to Christoffersen [2012, Chap. 13], committing Type II errors in risk management can lead to very costly damage, so an appropriate choice of significance level is recommended to be 10%.

2.5.2 Independence Testing

Assuming that we get a correct unconditional coverage rate for our estimated VaR model, our job is still not done, however. If all of these VaR violations occur around the same period, this is a cluster of violations. These violations are not representative of the whole data because they happen around the same time, so they provide a skewed picture of the total sample, creating a wrong expectation of potential risks when the violations actually spread over time. It is thus desirable to reject VaR models that imply hits clustered in time.

In order to do this, let's assume the hit sequence is dependent over time and can be described as a first-order Markov sequence. This means given the fact that there is a nonviolation today ($I_t = 0$), the probabily of tomorrow being a violation ($I_{t+1} = 1$) is π_{01}; and if today is also a violation ($I_t = 1$), then the probability becomes π_{11}. These two probabilities (π_{01}, π_{11}) can describe the whole process given our assumption that tomorrow's outcome depends only on the outcome today. Since π_{00} and π_{10} are just the other possible outcome of the next day given the information today, then

$$\pi_{00} = 1 - \pi_{01},$$
$$\pi_{10} = 1 - \pi_{11}.$$

This first-order Markov property can thus be described through the transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}.$$

Let T_{ij} be the number of observations with a j following an i, for $i, j = 0, 1$. The likelihood function of this Markov process becomes

$$L(\Pi_1) = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}}.$$

The MLE of π_{01}, π_{11} can be calculated as

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00} + T_{01}},$$
$$\hat{\pi}_{11} = \frac{T_{11}}{T_{10} + T_{11}},$$

The estimated transition probability matrix is then

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix}.$$

The probability of a violation following a violation being larger than the probability of a violation following a nonviolation is typically worrisome. This is because it indicates that our VaR estimations are frequently incorrect and they seem not to capture the volatility clusters. However, in the case where the hits are independent over time, $\pi_{01} = \pi_{11} = \pi$, the transition matrix is simply:

$$\hat{\Pi} = \begin{bmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{bmatrix}.$$

Let the null hypothesis be $H_0 : \pi_{01} = \pi_{11}$. We can then test it by using a likelihood ratio test:

$$LR_{ind} = -2\ln\left[L(\hat{\pi})/L(\hat{\Pi}_1)\right] \sim \chi^2_{(1)}.$$

($L(\hat{\pi})$ is obtained from the unconditional coverage test.)

2.5.3 Conditional Coverage Testing

The ultimate goal to backtest VaR models is testing the coverage rate of VaR while at the same time checking if the violations are independent. Hence, we can combine the two tests above for testing conditional coverage under the null hypothesis $H_0 : \pi_{01} = \pi_{11} = p$:

$$LR_{cc} = -2\ln\left[L(p)/L(\hat{\Pi}_1)\right] \sim \chi^2_{(2)}.$$

Note that the LR_{cc} test takes the null hypothesis likelihood in the unconditional coverage test and the alternative hypothesis likelihood from the

independence test. The joint test of conditional coverage can actually be computed by taking the sum of the two previous tests, since

$$\begin{aligned} LR_{cc} &= -2\ln\left[L(p)/L(\hat{I\!I}_1)\right] \\ &= -2\ln\left[\{L(p)/L(\hat{\pi})\}/\{/L(\hat{\pi})/L(\hat{I\!I}_1)\}\right] \\ &= -2\ln\left[L(p)/L(\hat{\pi})\right] - 2\ln\left[L(\hat{\pi})/L(\hat{I\!I}_1)\right] \\ &= LR_{uc} + LR_{ind}. \end{aligned}$$

The three tests mentioned above are all simple to implement as they only use information on past VaR violations. They can be used as a convenient diagnostic check on risk models.

We are now ready to move on to the next sections where we analyze data by applying these methods, calculate VaR for each method and backtest these estimates to compare different models.

Chapter 3

Data

3.1 Introduction

Before we look at the data, we should first understand what type of data was collected and why we should use it. The data set we collected was CME data retrieved from the Quandl Stevens Continuous database. CME Group Inc. (Chicago Mercantile Exchange & Chicago Board of Trade) is a financial market institution that operates derivatives exchanges, which is "evolving into an ever more sophisticated institution that plays a key role in many sorts of financing", according to The Economist [2013]. The New York Mercantile Exchange (NYMEX), owned by the CME group, has West Texas Intermediate (WTI) as the underlying commodity for oil futures contracts — "the world's most liquid and actively traded crude oil contract", claimed by the CME Group [2017].

These crude oil prices are quoted from the settlement price traded in continuous contracts. While short-term contracts have properties unsuited for long-term trend analysis, such as short duration and varying liquidity, continuous contracts can give a historical reflection of prices by chaining together these consecutive single contracts by end-to-end roll method (i.e., the last trading day as roll date). Another thing to note here is that these prices are adjusted by the calendar-weighted method, which gives a rather smooth transition between contracts because the weights of combined prices are averaged out. According to Quandl's recommendation, "calendar-weighted rolling" should be the choice when it comes to forecast and analyze regression on economic data. Hence, the data that is used in this thesis is collected from continuous futures contracts with the aforementioned methods in the period of over 30 years (April 1985–December 2015).

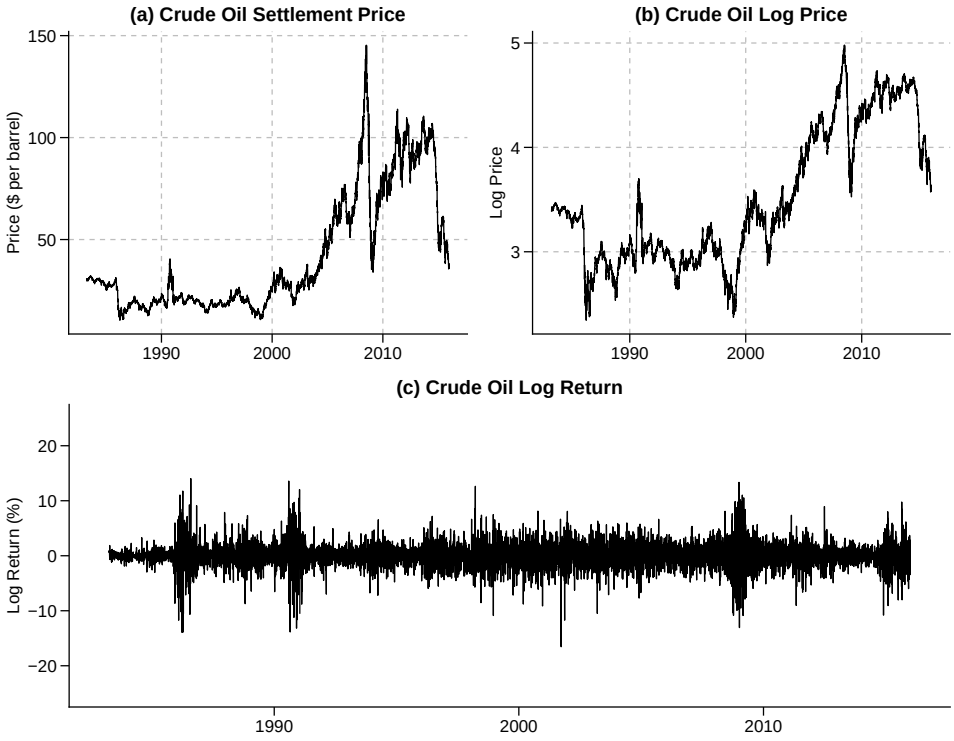

Figure 3.1: Calendar-weighted rolling crude oil prices for continuous futures contracts — from 04/04/1985 to 31/12/2015

3.2 Data

Now, we can take a look at our data through different plots in Figure 3.1. Figure 3.1(a) shows the crude oil settlement price trend in our chosen period. As we have discussed in Chapter 1, these raw prices show extreme price fluctuations (ranging from $10 to $145 per barrel), especially in the periods where there were international crises in oil supply distribution. However, to analyze our data, we will not use these raw prices directly, but instead, we will use returns — a much preferable input. By using raw prices, we build a system that is inconvenient to do cross-evaluation among variables with unequal values. Also, the numbers themselves do not have much meaning because of the lack of a setting benchmark. Therefore, we will use returns, which are measured in percentage, that way, we normalize our variable and bring meaning to each number in our metric.

Returns contain different types, and the regular simple returns might be the most well-known. A simple return at time t is the proportion of the change in value from time $(t-1)$ to time t, $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$. Nevertheless, we will be using *log returns* because of their several advantages over normal returns. Log return is continuously compounded return, calculated by taking the natural logarithm of simple gross return. Hence, let r_t be log return at time t and R_t be the simple return, and then

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}).$$

From the definition, we can see quite clearly the time-additivity property of these returns as

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) \\ &= \ln((1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})) \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \ldots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \ldots + r_{t-k+1}, \end{aligned}$$

meaning that the continuously compounded multi-period return is the sum of the continuously compounded one-period returns. This property makes log returns a lot more manageable compared to the simple returns [Tsay, 2010*b*], and it is the main reason why we use log returns for our data here in this thesis.

There are other theoretical and algorithmic benefits of log returns over simple returns, such as log-normality (convenient for many classical assumptions in statistics), mathematical ease in stochastic processes (calculus properties of exponential function e^x), numerical stability (avoiding arithmetic underflow by summation via log transformation), etc. [Quantivity, 2011]. Throughout the thesis from now on, log returns are referred to as either *log returns*, *returns*, or *our data*.

Figure 3.1(b) shows the graph of the natural log price of crude oil. Again, the fluctuations are shown clearly in this graph, but it is a bit smoother compared to the raw price due to the natural log transformation, making the range less dramatic. What we are really interested in is to look at the returns of this crude oil corresponding to the analyzed period in Figure 3.1(c). By the definition above, these returns are simply the difference of the log prices calculated from Figure 3.1(b) (in percentage). We can see that these crude

Table 3.1: Summary statistics of daily returns on crude oil from 04/04/1985 to 31/12/2015

Statistic	Value
Size	8222
Mean (%)	0.0029
Std. Dev (%)	2.2436
Min (%)	−40.8062
Max (%)	14.0326
Skewness	−0.8832
Excess Kurtosis	17.6007
Jarque-Bera	107256.25*
Ljung-Box	72.68*
$q_{0.001}$	−11.6984
$q_{0.01}$	−6.1584
$q_{0.05}$	−3.4829
$q_{0.95}$	3.3409
$q_{0.99}$	5.9621
$q_{0.999}$	10.9129

Note: Statistics marked with (*) are significantly different from 0 at the 1% significance level.

oil returns were especially volatile in the three periods where oil production increased in the 80's, around the Gulf War period (1990–1991) and the global recession (2008–2009), reflected by the extreme variations in a short time interval. The returns are shown to be stationary (mostly fluctuate around 0) and seem to have several volatility clusterings.

3.3 Description

Table 3.1 gives some descriptive statistics on the return sequence along with test values of the Jarque-Bera test for normality and Ljung-Box test for serial correlation.

We have a total of 8,222 observations in our analyzed period from April 4, 1985 to December 31, 2015 for our crude oil data. We observe that the mean is very close to 0 (as expected in financial returns), while the minimum and

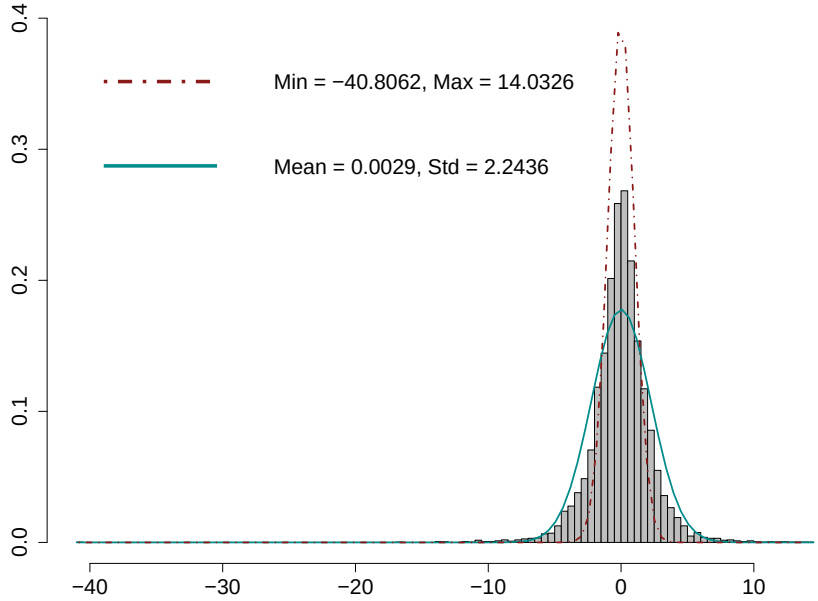

Figure 3.2: Empirical distribution of the return series with relative frequency density on the y-axis (i.e., the total area under the histogram is 1) with two fitted normal cumulative distribution functions

maximum are quite at a distance to the mean (especially the minimum), suggesting a long left tail. The negative skewness and high excess kurtosis clearly suggest that the series does not follow the normal distribution, which has the skewness of 0 and excess kurtosis of 2.96.

The Jarque-Bera statistic shows that the null hypothesis of normality is rejected at any level of significance, which is consistent with the result presented by the skewness and kurtosis discussed above. The Ljung-Box statistic is the result of testing for conditional heteroscedasticity with the null hypothesis of no autocorrelation up to the 20th lag. The p-value of this statistic is significant, 6.614×10^{-8}, suggesting that we reject the null hypothesis. Hence, we can conclude that our observations are serially autocorrelated, which indicates that we should apply a GARCH model before applying the EVT method. In other words, we shall use the conditional EVT approach when analyzing because our data is clearly not i.i.d. — a prerequisite for EVT methods. Pairs of the empirical upper and lower 0.1%, 1%, and 5% quantiles are not too far off from being symmetric to each other, which is supported by the small negative skewness.

Figure 3.2 shows the empirical distribution of our data, which has a long left tail as expected due to our extreme minimum. The histogram is slightly skewed to the left, indicating the calculated negative skewness. We also fitted two normal CDF's onto our return histogram. The red dashed line illustrates a normal curve with the same range as our data (i.e., same minimum and maximum), while the blue line represents a normal curve with mean and standard deviation the same as our empirical mean and standard deviation. The two curves are clearly ill-fitted, and we expected this result because we know from the summary statistics in Table 3.1 that our data does not fit the normal distribution.

We are now ready to analyze our return data with the different simulation approaches, EVT models and the ACER method that we studied in Chapter 2.

Chapter 4

Analysis and Results

From different modelling methods introduced in Chapter 2, we can now examine the return data of the crude oil market from April 4, 1983 to December 31, 2015 to estimate the one-day ahead VaR's and to backtest them using the unconditional coverage, independence, and conditional coverage tests. Let us first summarize some results we got from each method before evaluating our VaR estimations.

4.1 Methods Summary

4.1.1 Historical Simulation

Figure 4.1 shows the histogram of our simulations for crude oil returns by the HS approach. Here, we simulated 10,000 possible 250-day trading periods and compare the mean of our simulated returns with the true empirical mean return. As we can see from the figure, they almost hold the same value that is very close to 0. This result is quite predictable because we know that the mean return in financial data is 0 for the most part due to the competitive market conditions.

We again simulate 10,000 possible 250-day trading periods and calculate the VaR of the next day. This means that for each simulation for the returns from day t to day $(t + 249)$ (resampling with replacement), we predict $\text{VaR}_{p,t+250}$. Hence, we obtain 10,000 different results for the p quantile in each simulation, and the histogram of these VaR's is illustrated in Figure 4.2. Here, we use $p = 0.05$, and we also include the vertical red dashed line to mark the true empirical 0.05 quantile. The mean VaR of the simulated returns and the empirical quantile are shown to be quite close together.

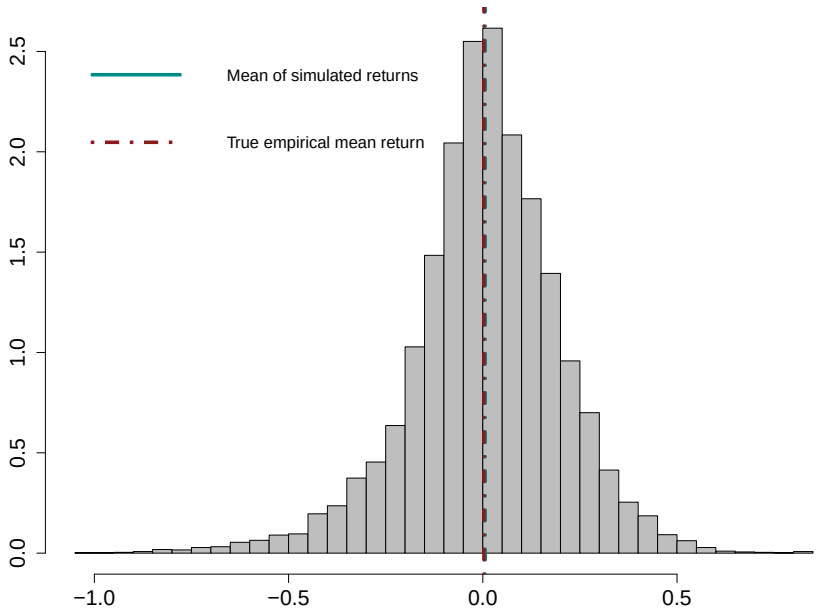

Figure 4.1: Histogram of HS returns (%) for 10,000 possible 250-day trading periods

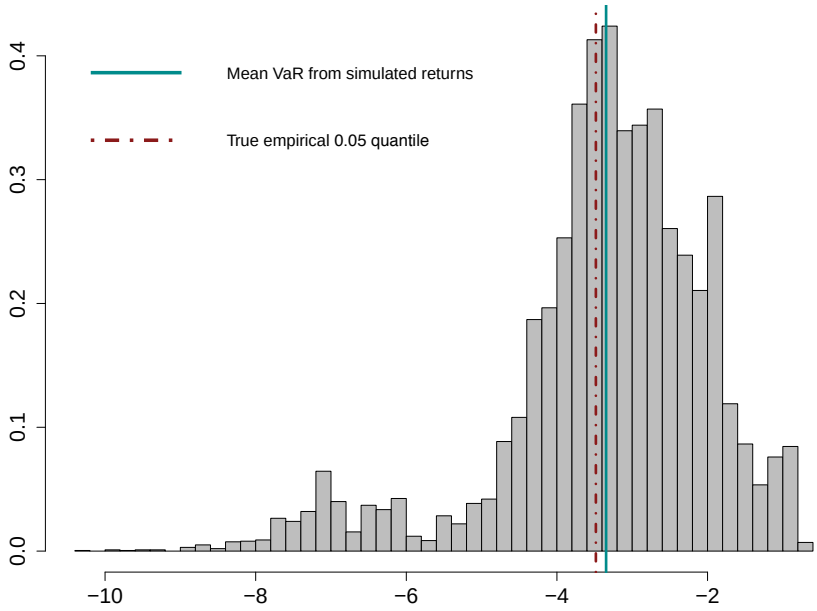

Figure 4.2: Histogram of 10,000 $VaR_{0.05}$'s for historically simulated returns (%) of 250-day trading periods

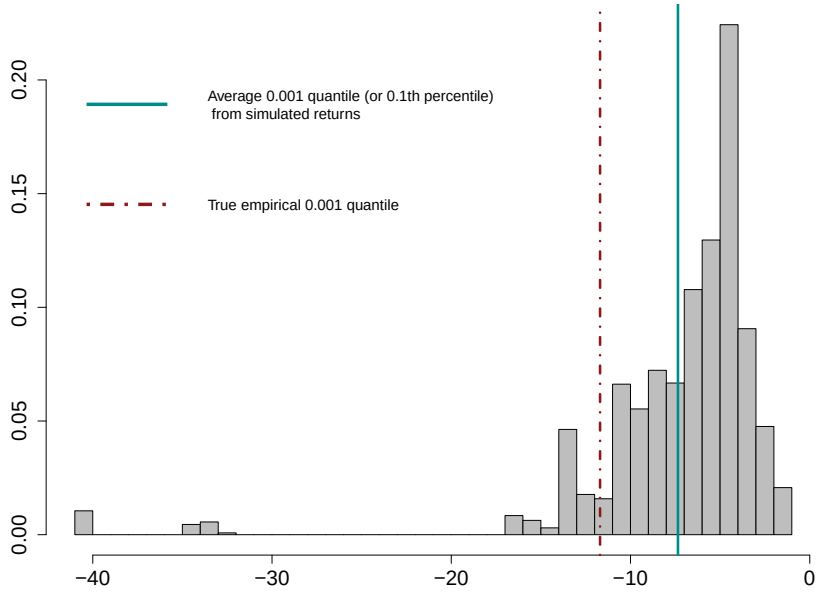

Figure 4.3: Histogram of 10,000 extreme quantiles (0.1^{th} percentile) for historically simulated returns (%) of 250-day trading periods

Similar to what we did above, we now have a histogram of simulated 10,000 VaR's of 250-day trading period but with a very small quantile, $p = 0.001$. This quantile is often used in stress testing. To learn more about stress testing, read Quagliariello [2009]. Figure 4.3 shows that the average of the calculated 0.001 quantiles is not very close to the true empirical 0.001 quantile. What generates this difference is that the empirical 0.001 quantile is of the whole data where each return enters as one entry, while the original return at each particular day t , for $1 \leq t \leq 250$, would be an input in t simulated 250-day periods, and for $251 \leq t \leq 7972$, this return is a potential input in 250 different simulations of random sampling with replacement (remember that we use a 250-day trading day for each simulation). Hence, the range of the simulated extreme quantiles is quite big (it actually covers all the negative values in the data) and contains a few results at the minimum value of our returns.

With the same simulation approach as above, Figure 4.4 plots our HS VaR's and the true 0.01 quantile in the out-of-sample period. We choose a time horizon of 250 days to obtain each quantile of the out-of-sample period, i.e., the true quantile for each day is calculated from its last 250 observations.

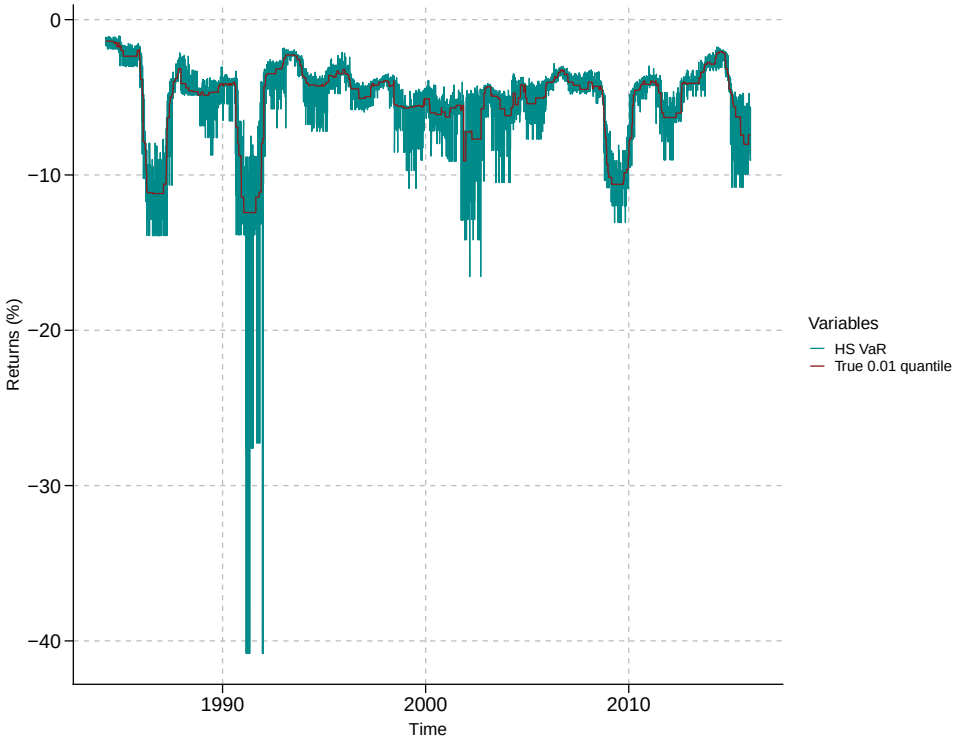

Figure 4.4: Predicted $VaR_{0.01}$ of the next day within a rolling window of 250 days using HS approach compared with the true 0.01 quantile of a 250-day horizon

The figure shows that they harmonize with each other in terms of the direction of movement, but the HS VaR's are a lot more extreme. This extremity results from the resampling method, which potentially makes those extreme returns in our original data manipulate the quantile if they appear too many times in the same simulation period.

4.1.2 Filtered HS

For the FHS method, we begin by capturing the heteroscadasticity in the error variance. The first step we can do is to look at the ACF and PACF of our return data, which is illustrated in Figure 4.5. While the ACF shows no significant lags, the PACF indicates that our return series is not stationary in variance, which is consistent with the volatility clustering and high fluctuations shown in Figure 3.1. As in time series analysis, we should

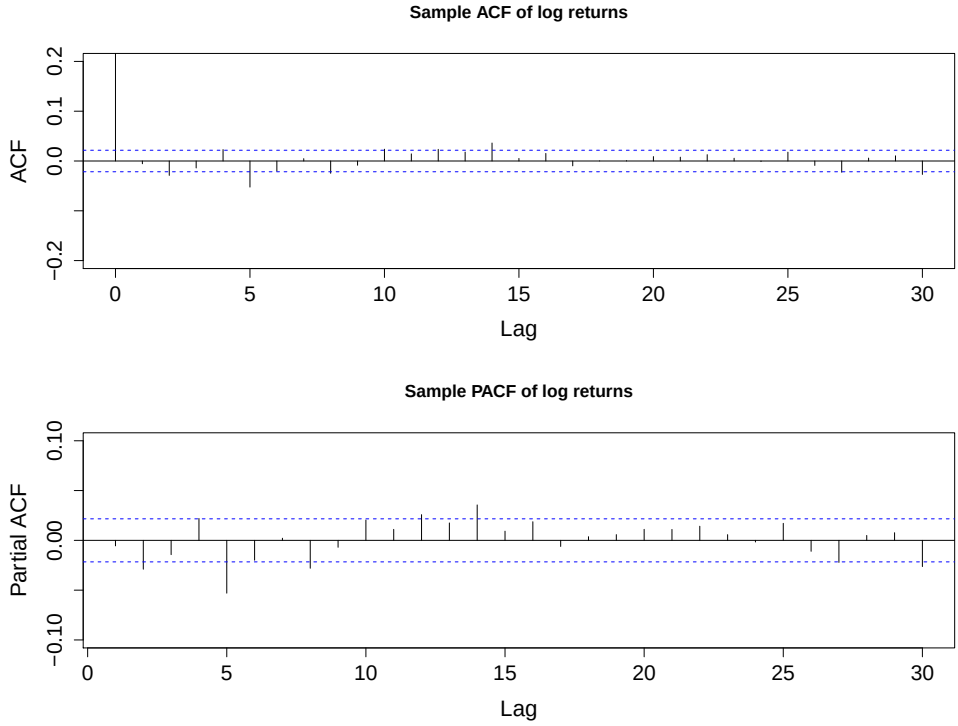

Figure 4.5: Sample ACF and PACF of crude oil returns. The dashed lines give the values beyond which the autocorrelations and partial autocorrelations are significantly different from 0.

apply an autoregressive integrated moving average (ARIMA) model to our data when nonstationarity seems to exist. Therefore, we fit a combination of ARIMA models into our data with the different possible orders of the AR and MA parts running from 1–5 (a maximum of 5 lags, which represents the last 5 days — the length of a one week trading period). The data can then be integrated by a differencing process if necessary. For more details of ARIMA model and time series analysis, we recommend Wei [2006].

The ARIMA model that gives us the best Akaike information criterion (AIC) is ARIMA(3,0,3), which corresponds to ARMA(3,3). We can now plot the ACF and PACF of the squared residuals of this fitted ARMA model onto our return data, which is shown in Figure 4.6. Here, we included 200 lags, and the figure shows that both the ACF and PACF of the squared residuals decay very slowly, suggesting a very high order of the ARMA model when fitting these squared residuals. As we know from Chapter 2, the order of

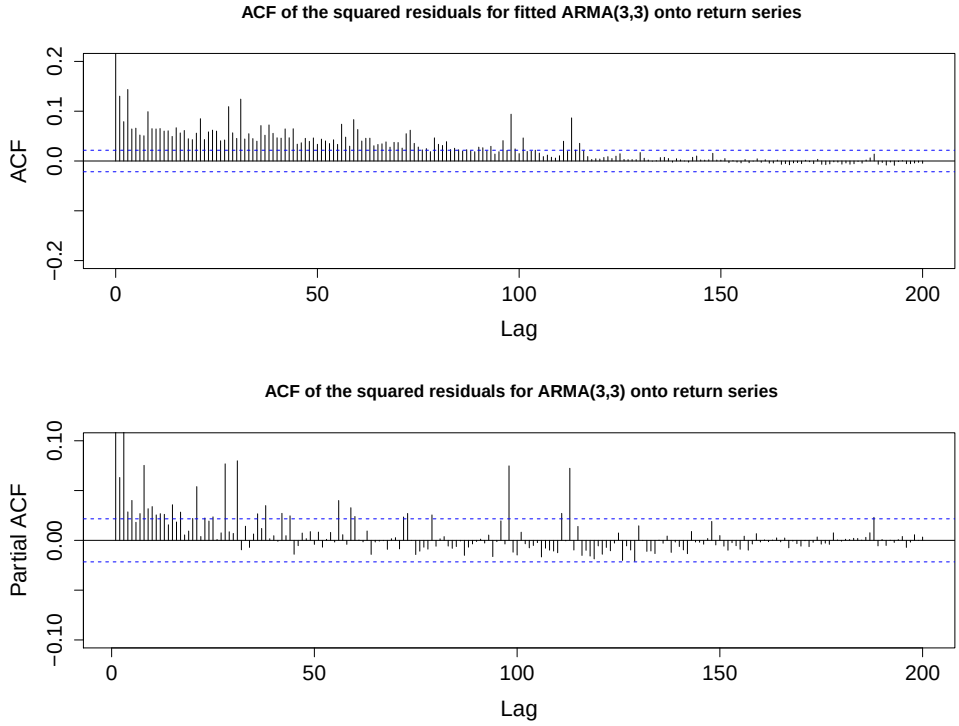

Figure 4.6: Sample ACF and PACF of the squared residuals of fitted ARMA(3,3) onto returns

the ARMA model for the squared residuals corresponds to the order of GARCH model in our return data. This would suggest that our GARCH model would have a very high order, which is quite impractical. Therefore, we will use the GARCH(1,1) model as it is parsimonious as shown before, and as Wei [2006, p. 380] suggested, "a very simple GARCH(r, s) model with $r \leq 2$ and $s \leq 2$ is sufficient to provide a significant improvement over the traditional homoscedasticity models".

Therefore, we will fit 4 different models to our data as the *filtering* step of this approach. The first two models are ARMA(3,3) - GARCH(1,1) with the innovation e_t in (2.10) following the normal and Student's t distribution. For the last two models, we fit the GARCH(1,1) model directly to the returns, with, again, normally distributed innovation in one model, and Student's t distributed innovation in the other. Figure 4.7 shows the ACF of the residuals and squared residuals for each fitted models. We can see that

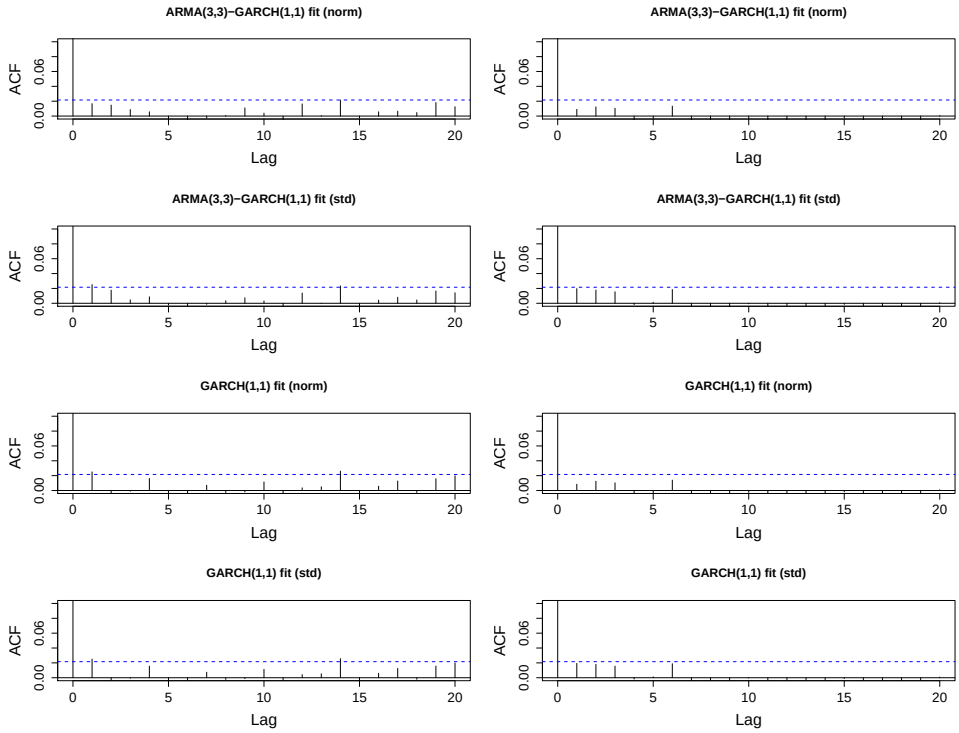

Figure 4.7: Sample ACF of the residuals and squared residuals of the fitted models. The ACF plots on the left hand side are for residuals, and the ACF plots on the right are for the squared residuals. From the top row to the bottom row, the plots obtain from the following fitted models: ARMA(3,3) - GARCH(1,1) with normal distributed innovation, ARMA(3,3) - GARCH(1,1) with student's t distributed innovation, GARCH(1,1) with normal distributed innovation, and GARCH(1,1) with student's t distributed innovation.

these plots suggest no serial correlations in the residual series (no significant lags), hence the conditional heteroscedasticity is captured. The result for the parameter estimates of the fitted models and their correspondent AIC is given in Table 4.1. The parameter "Shape" indicates the degrees of freedom for the Student's t distributed innovation e_t (with mean 0 and variance 1).

We can see from the table that the coefficients of the AR(1) and MA(1) parts of the first two models are not significant, and the AIC of the ARMA-GARCH model with the Student's t distributed e_t is slightly better. For the GARCH models, all parameters are significant at the 0.01 significance level, and GARCH (1,1) with Student's t distributed innovation has a lower AIC. Therefore, we will choose the GARCH (1,1) with Student's t distributed e_t

Table 4.1: Parameter estimates of fitted models with Information Criterion Statistic AIC

	ARMA (3,3) - GARCH (1,1) (norm)	ARMA (3,3) - GARCH (1,1) (std)	GARCH (1,1) (norm)	GARCH (1,1) (std)
μ	0.007	0.010	0.019	0.016
AR1	0.122	0.064	–	–
AR2	−0.290***	−0.362***	–	–
AR3	0.789***	0.717	–	–
MA1	−0.111	−0.063	–	–
MA2	0.269***	0.339***	–	–
MA3	−0.796***	−0.722***	–	–
ω	0.011***	0.006**	0.011***	0.006**
α_1	0.089***	0.075***	0.089***	0.075***
β_1	0.916***	0.930***	0.916***	0.930***
Shape	–	6.871	–	6.836
AIC	4.066	4.023	4.067	4.023

Note: Statistics marked with (***) and (**) are significantly different from 0 at the 0.1% and 1% confidence level, respectively. Shape is the degrees of freedom of the Student's t-distribution.

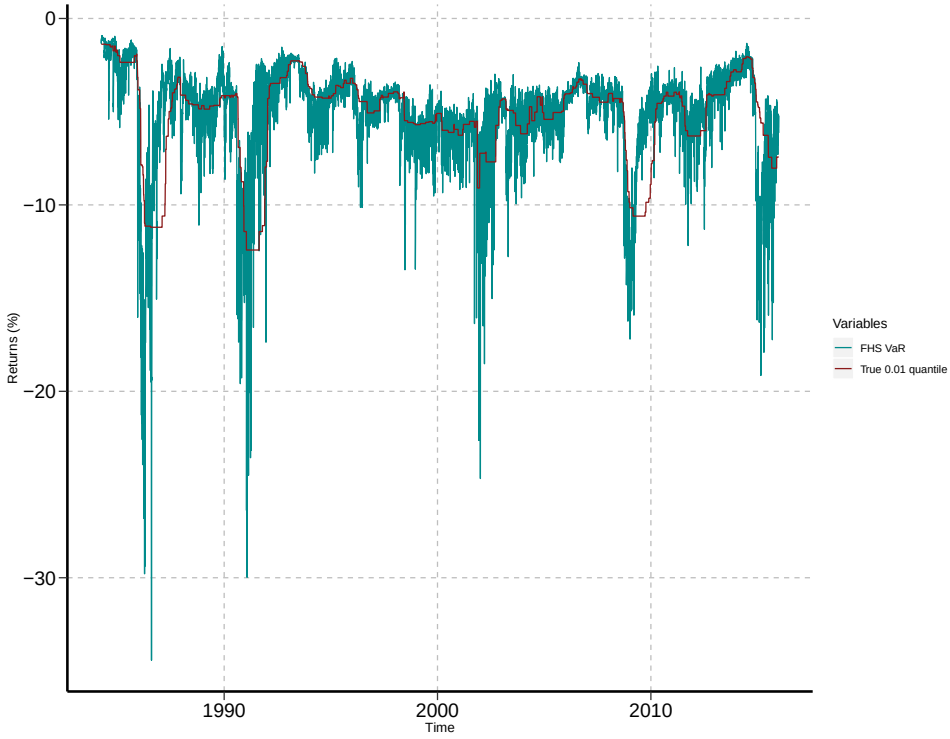

Figure 4.8: Predicted $\text{VaR}_{0.01}$ of the next day within a rolling window of 250 days using FHS approach compared with the true 0.01 quantile of a 250-day horizon.

model as our final model for the *filtering* step of the FHS approach because it has the lowest AIC among the models with all significant coefficients. GARCH(1,1) is also simpler than the ARMA-GARCH model and is confirmed in the aforementioned literature to be sufficient.

After fitting the chosen GARCH (1,1) model for each day t, we will obtain the conditional standard deviation for that day. We can then calculate the standardized return, $\hat{z}_t$, and predict the one-day ahead conditional standard deviation, σ_{t+1}, as described in 2 for the FHS approach. We complete this approach by resampling the standardized returns from the last 250-trading days with replacement to generate a new standardized return for the next day and compute our one-day ahead VaR's.

Like in the HS approach, we plot our FHS VaR's result and the true 0.01 quantile in the out-of-sample period in Figure 4.8. Our interest is to see, from our estimation of these calculated VaR's, if there is a VaR violation for

each day by comparing with the corresponding true return. This result is presented in the next section, section 4.2.

4.1.3 Conditional EVT

As we discussed in Chapter 2, we fit the GARCH model onto the crude oil return to obtain the conditional standard deviation. We then can compute the standardized returns, which are safely assumed to be independent. The GARCH model fitted here is GARCH(1,1) process with Student's t distributed innovations.

One challenging part of this method is to determine which threshold to use. The choice we make will influence the outcome heavily due to the bias and variance dilemma, because we directly decide the size of observations considered. The choice for the threshold here is in accordance with Christoffersen [2012]'s suggestion that we mentioned earlier, that the threshold is the 95% quantile of the negative standardized returns.

The Hill estimator is then computed in order for us to compute the quantile F^{-1}_{1-p} in (2.18). We then proceed to estimate the one-day ahead VaR's and plot them in Figure 4.9. In the figure, beside the difference in magnitude of the estimated VaR's and the true 0.01 quantile of the last 250 days, we can see that there seems to exist some lag between the general shape of the two series. Again, we should keep in mind that this figure only compares our VaR estimation with the quantile of a given time, it does not assess our model. Our ultimate goal is to compare our computed VaR with the return of each corresponding day to observe if there was any violation.

4.1.4 ACER

Like in the EVT approach, before applying the ACER method, we modify our data by considering the negative values of our return data. We first examine this method by looking at the k-plot in Figure 4.10.

The k-plot shows the comparison among ACER estimates conditioning on $(k-1)$ previous data points for different η values (%). The statistical dependence with different degrees of conditioning does not seem to be so strong after $\eta > 0$, and the ACER functions converge around $\eta = 12\%$. Notice that the effect of conditioning on previous data can be captured to

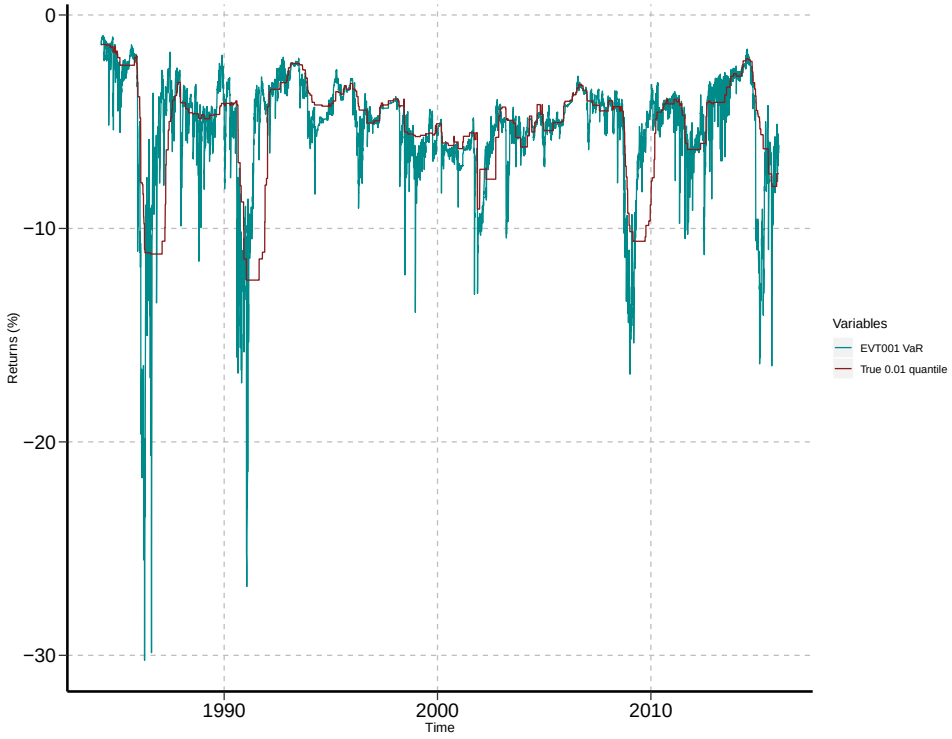

Figure 4.9: Predicted $\text{VaR}_{0.01}$ of the next day within a rolling window of 250 days using a conditional EVT approach compared with the true 0.01 quantile of a 250-day horizon.

some extent by $\hat{\varepsilon}_3(\eta)$, meaning that we only need to condition on the two previous observations. Conditioning on more data points does not seem to give any significant difference in ACER estimates for tail values roughly after $\eta = 8\%$, which is the start of the tail's regular behavior. Næss et al. [2013] uses the first ACER function for the prediction in the paper and the given reason was that all the ACER estimates appeared to merge in the far tail, so "the obvious choice is to use the first ACER function, which allows us to use all the data in its estimation and thereby increase accuracy".

For our data, we have tried various models by conditioning on different k values as well as setting different sample sizes. For each model, we calculate $\text{VaR}_{0.001}$, $\text{VaR}_{0.01}$, and $\text{VaR}_{0.05}$ for the next day. Among the models we have fitted, $k = 3$ did the best for our data in terms of the number of violations. Other than the amount of the previous observations we condition on, the sample size seems to have a certain impact on the outcome: As we condition

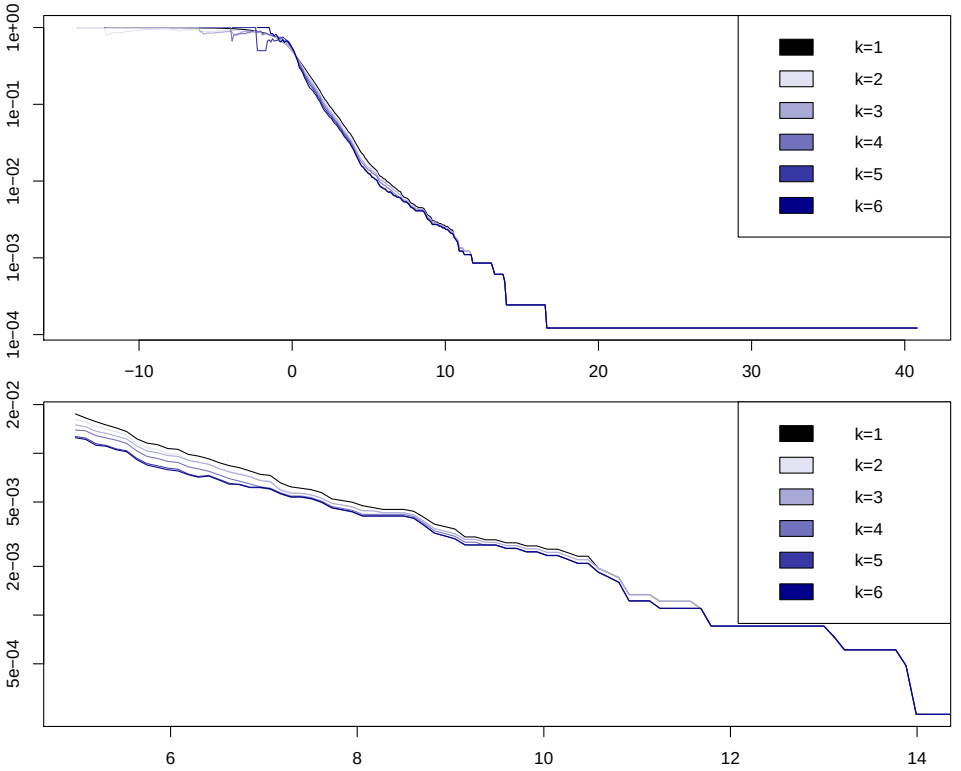

Figure 4.10: The top plot is the k-plot of $(\eta, \hat{\varepsilon}_k(\eta))$ computed from the negative of our return data for different values of k (values of η are in percentage). The bottom plot is the result of zooming in the top plot for $\eta = (5, 14)$ to identify convergence.

on more data points, we reduce the sample of exceedances significantly if clusters occur (and we already know that volatility clusters exist in our data set), so we generally want to have a larger sample size to improve precision.

The reason why we do not keep $k = 1$ as in Næss et al. [2013], although our ACER functions converge in the far tail, is that these functions do not have a steady flow until $\eta > 0$ and not really converge until $\eta > 8\%$ for our modified data. In our original data, this is equivalent to a stable result for the ACER functions when the threshold η is set to be less than 0%, and these functions converge when η is less than -8%.

However, about 50% of the crude oil return data is larger than 0% and only the minimum of our data is less than -8%. Therefore, if we only base on the k-plot when the tail seems to have regular behavior, we would use

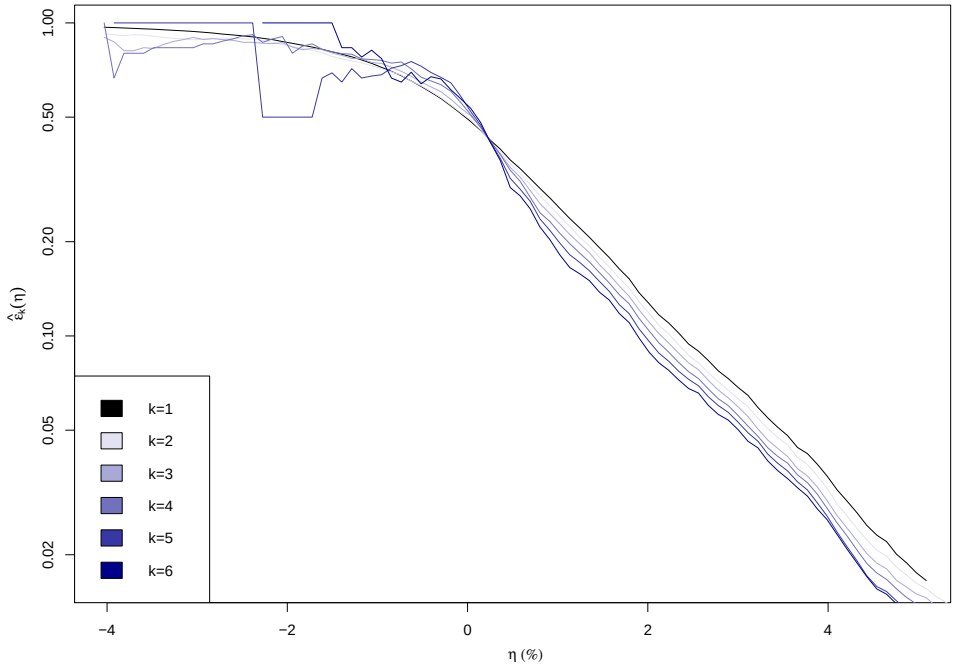

Figure 4.11: The k-plot for η in $(-4\%, 5\%)$ to identify a reasonable choice of k

very little to no data from our data set. In fact, 90% of our original data is in the range of $(-3.48\%, 3.34\%)$, so we would like to look at a more reasonable range of η. Figure 4.11 shows the $k-$plot for η in the range of $(-4\%, 5\%)$. From this figure, it seems like choosing $k = 3$, i.e., conditioning on two previous observations, is a sensible choice. As a matter of fact, we fit a variety of models and $k = 3$ always gives the best result among the same sample size.

The final model we choose to present for the ACER method to predict the VaR of the next day is from the rolling samples of 500 observations with $k = 3$. The results are shown in the next part.

4.2 Evaluating VaR estimations

Our data set of crude oil returns is divided into an in-sample period and an out-of-sample period to assess VaR performance. For each model used

Table 4.2: Out-of-sample VaR violations of different models

		HS	FHS	Cond. EVT	ACER
p = 0.001	Violations	70	64	10	124
	Proportion	0.0088	0.0080	0.0013	0.0160
	Rank	3	2	1	4
p = 0.01	Violations	156	131	101	246
	Proportion	0.0196	0.0164	0.0127	0.0318
	Rank	3	2	1	4
p = 0.05	Violations	496	346	410	698
	Proportion	0.0622	0.0434	0.0514	0.0841
	Rank	3	1	2	4

Note: A violation occurs when $R_{PF,t+1} < \text{VaR}_{p,t+1}$ at time $t+1$. The ACER method has 250 observations fewer compared to the other approaches due to the larger sample size for each fitting period.

(except for the ACER method), we use a rolling window of 250 days (approximately a typical trading year, which is 252 days) from our data to simulate daily returns. This makes the in-sample and out-of-sample period contain 7972 observations each. Hence, the first in-sample period starts at the 1st observation and the last in-sample period starts at the 7972nd observation. Accordingly, the out-of-sample period starts from the 251st and ends at the 8222nd observation.

Since we condition on $k = 3$ for the ACER method, we use a rolling window of 500 days (approximately two trading years) to fit our model in order to have enough data points in each sample. This means that the out-of-sample period for the ACER method has 250 observations fewer compared to other approaches (contains 7722 observations, from the 501st to the 8222nd observation).

The one-day ahead VaR is estimated using each method described in Chapter 2, and we count the number of the out-of-sample VaR violations for each fitted model for three quantiles of $p = 0.001$, $p = 0.01$, and $p = 0.05$. Table 4.2 summarizes the results with the proportion of violations for each approach and the rank of how well these approaches perform.

We can see that for all the given quantiles, the conditional EVT approach gives very good results and is ranked the first for quantiles 0.001 and 0.01. The proportion of VaR violations made by this approach is also very close

to our given probability to calculate VaR. In addition, the FHS approach performs very well compared to the remaining approaches, and for quantile 0.05, it even outperforms the conditional EVT approach.

The HS approach also has a good performance and provides results that are very close to the FHS approach, except for when the quantile is 0.05. The ACER method is expected to do well since it takes the statistical dependence of the data into account, however, it does not seem to perform very well in our data. This might be because the value of η chosen for each fitting sample was automatically set to when the regular behavior of the tail started. Wrong choices for η can lead to biased results, especially when the returns experience some extreme changes, which, as we know, happen a few times in our crude oil data.

The results for different backtesting methods are presented in Table 4.3. As a reminder, we have the following tests: unconditional coverage testing ($H_0 : \pi = p$), independence testing ($H_0 : \pi_{01} = \pi_{11}$), and conditional coverage testing ($H_0 : \pi = p$ and $\pi_{01} = \pi_{11}$).

The likelihood ratio statistic for each test is computed according to the methodology discussed in Chapter 2. Two significance levels are chosen here, $\alpha = 0.05$ and $\alpha = 0.1$. Like we mentioned earlier, Christoffersen [2012, Chap. 13] recommended to use $\alpha = 0.1$ (which gives a smaller critical value compared to $\alpha = 0.05$, hence, the possibility of rejecting a model is higher) to minimize the chance of failing to reject an incorrect model, which in risk management can lead to a catastrophic event.

As shown in Table 4.3, the estimation of VaR from three of our approaches (HS, FHS, and ACER) do not pass either the coverage test or the independence test, except for the FHS, which passes the independence test when $\alpha = 0.05$ for the 0.01 quantile. Nonetheless, the conditional EVT approach performs really well as it passes 13 out of 18 tests. From the results of Table 4.2, we could already expect the conditional EVT approach to pass most of the unconditional coverage tests because the proportion of the VaR violations was very close to our given probability. The approach also provides very good results for the independence test and the conditional coverage test. When the quantile is 0.01, this approach does not seem to perform quite as well — it only passes the independence test when $\alpha = 0.05$. Nevertheless, the conditional EVT approach generally gives outstanding results.

Table 4.3: Different likelihood ratio tests for the fitted models

		HS	FHS	Cond. EVT	ACER
	L_{uc}				
	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x
	L_{ind}				
p = 0.001	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x
	L_{cc}				
	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x
	L_{uc}				
	$\alpha = 0.05$	x	x	x	x
	$\alpha = 0.1$	x	x	x	x
	L_{ind}				
p = 0.01	$\alpha = 0.05$	x	✓	✓	x
	$\alpha = 0.1$	x	x	x	x
	L_{cc}				
	$\alpha = 0.05$	x	x	x	x
	$\alpha = 0.1$	x	x	x	x
	L_{uc}				
	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x
	L_{ind}				
p = 0.05	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x
	L_{cc}				
	$\alpha = 0.05$	x	x	✓	x
	$\alpha = 0.1$	x	x	✓	x

Notes: These tests are under different null hypotheses: For L_{uc}, $H_0 : \pi = p$, for L_{ind}, $H_0 : \pi_{01} = \pi_{11}$, and for L_{cc}, $H_0 : \pi = p$ and $\pi_{01} = \pi_{11}$.
The likelihood ratio statistics are calculated and compared with the critical values. At the $\alpha \cdot 100\%$ significance level, "x" means that we reject the specified VaR model, and "✓" means that we do not reject it.
Critical values: $\chi^2_{(1)}$ = 2.706 at $\alpha = 0.1$, 3.841 at $\alpha = 0.05$ (for L_{uc} and L_{ind}),
$\chi^2_{(2)}$ = 4.605 at $\alpha = 0.1$, 5.991 at $\alpha = 0.05$ (for L_{cc}).

Chapter 5

Conclusion and Future Work

This thesis studies different methods for extreme value prediction from sampled time series of crude oil prices. In particular, we estimated the value at risk (VaR) of the next day via different approaches and evaluated our results. We included some widely used approaches in risk management such as historical simulation (HS) and filtered HS (FHS) with the GARCH model (which captures the heteroscedasticity in our data). To target the extreme values better, we also used the conditional extreme value theory (EVT) method in order to model the far tail of the distribution, as well as to get rid of the heteroscedastic effect. The last approach used in this thesis is a conditional average exceedance rate (ACER) method that handles the dependence between the data points and targets the subasymptotic data.

Results of the VaR estimation for the aforementioned approaches were presented and backtested by three different tests. These tests examined whether the models we used have a statistically correct result for the coverage rate and whether there exists a problem of violation clustering in these estimations. By comparing and evaluating these estimates, we conclude that the conditional EVT approach has an outstanding performance compared to the others, in both accuracy and independence testing.

We note that the approaches used in this thesis can be grouped in three ways. The first group contains the approaches that use the bootstrapping methodology (HS and FHS), the second contains those that target the use of (sub)asymptotic data (conditional EVT and ACER), and the third group contains those that deal with the heteroscedasticity in the data (FHS and conditional EVT). We found that it is the third group that performs the best in estimating VaR values.

We should also note that even if an approach performs well in one data set, it does not mean that it will have the same performance in other sets of

data. Each data set has its unique properties depending on what category this data falls into. For example, an approach that works well for modeling the price risk of crude oil might not work as well for modeling the price risk of a soft drink. The reason is that soft drinks are a type of goods that is very elastic — i.e., the consumption has a highly sensitive response to changes in price due to many existing substitutes. Thus, the prices of a soft drink should not be nearly as volatile as the prices of crude oil. Therefore, the results from this thesis only apply to the given crude oil data set, and they do not indicate that the conditional EVT approach always performs better than, say, the ACER method.

There are several potential improvements that we can make in the future. Since VaR is only a single point estimate of the next period's quantile, it does not reflect the extent of how big a potential loss could become beyond that point, which is essential in risk management. To acquire better knowledge of what might happen, we can compute VaR values for different quantiles farther in the tail of interest. However, we would then retrieve only a set of discrete VaR values whereas we want to extract more information from our continuous loss function. What we might then be interested in is the average potential loss in the case where the actual loss goes beyond our estimation of VaR. This is where a risk measure such as expected shortfall (ES) comes in handy. Given that the loss exceeds VaR_p, ES gives us the average loss in the worst $p \cdot 100\%$ of the time. Therefore, ES should be investigated in future work in order to have a better evaluation of the tail of the loss distribution, in addition to the VaR values that we estimated.

Another possible improvement is to increase the amount of information for backtesting our VaR estimations. The tests implemented are merely based on the past VaR violations and cannot easily uncover an incorrect model. Therefore, we might want to include other information (i.e., related variables) that may have a correlation or causal effect with the violations via some regression-based approach [Christoffersen, 2012]. According to Christoffersen [2012], we can better understand "the areas in which the risk model is misspecified", a substantial task in risk modeling, by increasing the information set.

Another idea for future improvement is to find an algorithm to identify a better η for the ACER method. The choice of η in this thesis, which is generated automatically from the tail's behavior, may have strongly negatively impacted the results of this approach.

Bibliography

Balkema, A. and De Haan, L. (1974), 'Residual life time at great age', *The Annals of Probability* **2**.

Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (1999), 'VaR without correlations for nonlinear portfolios', *Journal of Futures Markets* **19**, 583–602.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, T., Chou, R. and F. Kroner, K. (1992), 'ARCH modeling in Finance: A review of the theory and empirical evidence', *Journal of Econometrics* **52**, 5–59.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994), 'ARCH models', *Handbook of Econometrics* **4**(3), 2959–3038.

Brooks, C. (2008), *Introductory Econometrics for Finance*, 2nd edn, Cambridge University Press, chapter Modelling volatility and correlation.

Byström, H. (2005), 'Extreme value theory and extremely large electricity price changes', *International Review of Economics and Finance* **14**, 41–55.

Christoffersen, P. (2012), *Elements of Financial Risk Management*, 2nd edn, Academic Press.

CME Group (2017), 'NYMEX WTI light sweet crude oil futures'. [Accessed 2018-12-08].
URL: *https://www.cmegroup.com/trading/why-futures/welcome-to-nymex-wti-light-sweet-crude-oil-futures.html*

Dahlen, K., Solibakke, P., Westgaard, S. and Næss, A. (2015), 'On the estimation of extreme values for risk assessement and management: The ACER method', *International Journal of Business* **20**(1), 33–51.

Dowd, K. (2005), *Measuring Market Risk*, 2nd edn, John Wiley & Sons.

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

Engle, R. F. and Manganelli, S. (1999), 'Caviar: Conditional autoregressive value at risk by regression quantiles', *Journal of Business and Economic Statistics* **22**.

Fisher, R. and H.C. Tippett, L. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

Friedman, N. (2014), 'U.S. oil prices fall 46% this year, steepest loss since 2008', *The Wall Street Journal* . [Published 2014-12-31, accessed 2018-12-08].
URL: *https://www.wsj.com/articles/oil-prices-hit-again-by-weak-china-data-buildup-in-u-s-stockpiles-1420019256*

Füss, R. (2009), Crude oil market, *in* G. Gregoriou, ed., 'Encyclopedia of Alternative Investments', Taylor & Francis Group, LLC, pp. 129–130.

Gnedenko, B. (1943), 'Sur la distribution limite du terme maximum d'une serie aleatoire', *Annals of Mathematics* **44**(3), 423–453.

Hull, J. and White, A. (1998), 'Incorporating volatility updating into the historical simulation method for value-at-risk', *Journal of Risk* **1**, 1–19.

Jenkinson, A. F. (1955), 'The frequency distribution of the annual maximum (or minimum) of meteorological elements', *Quarterly Journal of the Royal Meteorological Society* **81**, 158–171.

Koepp, S. (1986), 'Cheap oil!', *TIME Magazine* . Monday, Apr. 14, 1986 issue.

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The Journal of Derivatives* **3**(2), 73–84.

Macrotrends (2018), 'Crude oil prices - 70 year historical chart'. [Accessed 2018-12-08].
URL: *https://www.macrotrends.net/1369/crude-oil-price-history-chart/*

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

Marimoutou, V., Raggad, B. and Trabelsi, A. (2009), 'Extreme value theory and value at risk: Application to oil market', *Energy Economics* **31**, 519–530.

McNeil, A. and Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach', *Journal of Empirical Finance* **7**, 271–300.

McNeil, A. and Frey, Rüdigerand Embrechts, P. (2005), *Quantitative Risk Management: Concepts Techniques and Tools*, Princeton University Press.

Mouwad, J. (2007), 'Rising demand for oil provokes new energy crisis', *The New York Times* . [Published 2007-11-09, accessed 2018-12-08].
URL: *https://www.nytimes.com/2007/11/09/business/worldbusiness/09oil.html*

Murphy, C. (2009), 'Why do oil prices swing so wildly?'. [Accessed 2018-12-08].
URL: *https://www.cbsnews.com/news/why-do-oil-prices-swing-so-wildly/*

Næss, A. (1984), 'On the long-term statistics of extremes', *Applied Ocean Reseach* **6**(4), 227–228.

Næss, A. (2010), 'Estimation of extreme values of time series with heavy tails', *Preprint Statistics* **14**.

Næss, A., Gaidai, O. and Karpa, O. (2013), 'Estimation of extreme values by the average conditional exceedance rate method', *Journal of Probability and Statistics* **2013**.

Pelletier, D. and Christoffersen, P. (2004), 'Backtesting value-at-risk: A duration-based approach', *Journal of Financial Econometrics* **2**, 84–108.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **3**(1), 119–131.

Pines, L. (n.d.), 'What is oil, how is oil used & why it's critical to the world economy'. [Accessed 2018-12-08].
URL: *https://commodity.com/energy/oil/*

Quagliariello, M. (2009), 'Stress testing the banking system: Methodologies and applications'.

Quantivity (2011), 'Why log returns'. [Accessed 2018-12-08].
URL: *https://quantivity.wordpress.com/2011/02/21/why-log-returns/*

Schall, G., Faber, M. and Rackwitz, R. (1991), 'The ergodicity assumption for sea states in the reliability estimation of offshore structures', *Journal of Offshore Mechanics and Arctic Engineering-transactions of The Asme - J OFFSHORE MECH ARCTIC ENG* **113**(3), 241–246.

The Economist (2013), 'The futures of capitalism', The Economist Group Limited. [Accessed 2018-12-08].
URL: *https://www.economist.com/finance-and-economics/2013/05/11/the-futures-of-capitalism*

Tsay, R. (2010*a*), *Analysis of Financial Time Series*, 3rd edn, Wiley, chapter Extreme Values, Quantiles, and Value at Risk.

Tsay, R. (2010*b*), *Analysis of Financial Time Series*, 3rd edn, Wiley, chapter Financial Time Series and Their Characteristics.

Wei, W. (2006), *Time Series Analysis: Univariate and Mulvariate Methods*, 2nd edn, Pearson, chapter Time Series Regression and GARCH Models.