



Extremvärdesteoretisk riskbedömning av svenska aktieindex

Applicerbara modeller inom riskanalys

Extreme value theory for risk assessment of Swedish stock indexes

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Jens Adolfsson

Filip Hjærtström

Simon Jonängen

Tobias Karlsson

Filip Pålsson

Extremvärdesteoretisk riskbedömning av svenska aktieindex

Applicerbara modeller inom riskanalys

Kandidatarbete i matematik inom civilingenjörsprogrammet Industriell ekonomi vid Chalmers



Jens Adolfsson Filip Hjærtström Simon Jonängen Tobias Karlsson
Filip Pålsson

Handledare: Holger Rootzén Matematiska vetenskaper

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2023

Förord

Kandidatarbetet är skrivet på institutionen Matematiska vetenskaper på Chalmers tekniska högskola. Vi vill börja med att tacka vår handledare Holger Rootzén för en professionell handledning och gjort att vi kunnat slutföra arbetet. Vi skulle även vilja tacka examinatorerna och koordinatorena Ulla Dinger och Maria Roginskaya och alla andra som på något sätt bidragit med utformningen och författandet av detta arbete. Under arbetets gång har en loggbok förts över våra enskilda medverkandes prestationer. Loggboken är inte inkluderad i denna rapport. Gruppen är överens om att alla personer i gruppen bidragit likvärdigt till arbetets slutliga presentation. Trots avsaknaden av loggbok, återfinns nedan en bidragsrapport i tabellform som återger vem/vilka som innehar huvudansvaret till vilken del av den skriftliga rapporten.

Bidragsrapport		
§	Rubrik	Huvudförfattare
-	Populärvetenskaplig presentation	Tobias
-	Sammanfattning och Abstract	Tobias & Simon
1	Bakgrund	
1.1	Kort om aktiemarknaden	Tobias
1.2	Matematiska modeller	Tobias
1.3	Indexinnehåll	Tobias
2	Inledning	
2.1	Syfte	Samtliga
2.2	Frågeställningar	Samtliga
2.3	Avgränsningar	Samtliga
2.4	Samhälleliga och etiska aspekter	Tobias
3	Teori	
3.1	Klassisk extremvärdesteori	Filip H & Simon
3.2	Block Maxima-metoden	Filip H & Simon
3.3	Peaks over Threshold-metoden	Filip H & Simon
3.4	Trender i parametrar	Filip H & Simon
3.5	Maximum Likelihood-estimat	Filip P
3.6	Anderson-Darling-test	Filip P & Jens
3.7	Riskmått	Filip P & Jens
3.8	Beroenden inom stationära serier	Filip H & Simon
3.9	Autokorrelation och korskorrelation	Filip P & Jens
3.10	Svart svan	Tobias
4	Metod	
4.1	Datainsamling	Tobias
4.2	Block Maxima-metoden	Filip H & Simon
4.3	Peaks over Threshold-metoden	Filip H, Simon & Jens
4.4	Parameterskattning	Samtliga
4.5	Godhetstest	Filip P & Jens
4.6	Beräkning av nutida VaR och ES	Filip H & Simon
4.7	VaR för trender i parametrar	Filip H & Simon
4.8	VaR över tid och volatilitet	Samtliga
4.9	Korskorrelation mellan index, VaR och avkastning	Filip H & Simon
4.10	Backtesting	Filip P & Jens
5	Resultat	Samtliga
6	Diskussion	Samtliga
-	Appendix	
A	Figurer	Filip H
B	Tabeller	Samtliga
C	Kod i 	Filip H & Simon
D	Kod i 	Filip P & Jens

Populärvetenskaplig presentation

Människor har i alla tider fascinerats av sannolikheter, slumpen och sammanträffanden. Aristoteles var tidigt ute med påståendet att "It is a part of probability that many improbable things will happen." och Neil deGrasse Tyson understryker vår förtjusning i det osannolika med "When a coincidence seems amazing, that's because the human mind isn't wired to naturally comprehend probability and statistics." Den av psykologi intresserade läsaren kanske känner igen Daniel Kahnemans "Tänka snabbt, och långsamt" eller den något mer matematiskt bevandrare Nassim Nicholas Talebs "Black Swan, the impact of the highly improbable". De båda bästsäljarna är bara några ur raden av de otaliga spaltmeter som osökt leder oss in på begreppet risk, något som investerare och aktörer på den finansiella marknaden är mycket intresserade av att undersöka, eller framförallt undvika.

För att utgå från en gemensam, matematisk, definition av begreppet "risk", anses det i det här läget motsvara den maximala, procentuella förlust som en investerare har en viss sannolikhet att förlora. Riskmåttet kallas för VaR (eng. *Value at Risk*) och exempelvis innebär ett 95%-igt VaR på 20% att en investerare står i begrepp att förlora mer än 20% av sitt kapital vid 5% av gångerna. Omvänt kan värdet tolkas som att investeraren med 95% sannolikhet inte kommer förlora mer än 20% av sin portfölj. Sannolikheten kopplad till värdet på VaR benämns som konfidensgrad.

När VaR som riskmått är etablerat och bestämt kvarstår frågan om hur värdet ska räknas fram. Genom att titta på hur ett index rör sig i prisnivå från en dag till en annan skapas en dataserie som inkluderar många små, och vissa stora, förluster. Dessa förluster förekommer naturligtvis med olika frekvenser, och därmed så finns förutsättningar för att koppla samman en förlusts storlek med dess sannolikhet. Redan här har ett VaR räknats fram, exempelvis kanske förluster på minst 2% inom en dag skedde vid 10% av fallen. Kvaliteten på värdet vid denna sorts "beräkning" är minst sagt bristande. Till exempel tas ingen hänsyn alls till möjligheten att lida en ännu större förlust än den historiskt största. Första världskriget kallades ju som sagt för "The great war" innan det var dags för krig nummer två att äntra scenen.

En något mer sofistikerad metod är att anpassa den empiriska datan till normalfördelningen. Den introducerar ett mer relativt mått och kan fortfarande ge VaR för olika konfidensgrader, vanligen någonstans kring 95% eller högre. Emellertid tenderar normalfördelningen att kraftigt underskatta risken för mycket osannolika händelser. Denna bristande förmåga är dessvärre ett avsevärt problem för investerare som vill tillämpa riskmåttet, eftersom man särskilt vill kunna förutspå de värsta krasherna, som tenderar att inträffa mycket sällan.

Det är här extremvärdesteorin kommer in i bilden. Genom att använda sig av andra fördelningar som lägger större vikt vid de stora förlusterna i datan kan bättre och mer exakta värden räknas fram. Inom extremvärdesteori finns det två konventionella metoder med var sin fördelning som används beroende på vilken typ av tidsperiod VaR ska åsyfta.

För VaR under en tidsperiod på säg en månad, används en metod som kallas *Block Maxima*. Den delar upp dataserien i mindre portioner och plockar ut det maximala värdet inom varje block, för att sedan modellera den till en *Generalized Extreme Value*-fördelning. För VaR som avser nästkommande dag, används istället *Peaks over Threshold*-metoden, som plockar ut alla datapunkter som överstiger ett visst värde och modellerar dessa efter en *Generalized Pareto*-fördelning. Båda fördelningarna uppvisar, för vissa parametrar, större storlek på "svansen", alltså de händelser som ligger långt från mitten på fördelningen och inträffar mindre ofta. Detta gör att de är mycket fördelaktiga när det kommer till att beräkna risken för stora, osannolika förluster.

Extremvärdesteorin är i sig själv rent matematisk och kan tillämpas på data som representerar vad som helst. Ett användningsområde utöver banker är till exempel försäkringsbolag som bedömer risken för väderkatastrofer. I den här rapporten riktas fokus emellertid ännu en gång mot de finansiella marknaderna. VaR och ett tillhörande riskmått, motsvarande den förväntade förlusten i de fall VaR överskrids, beräknas och analyseras för tre svenska aktieindex.

Sammanfattning





Arbetet grundar sig i den klassiska extremvärdesteorin för att beräkna riskmåttet Value at Risk och Expected Shortfall. För att beräkna dessa riskmått behöver några väl vedertagna metoder användas, vilka är Block maxima-, Peaks over Threshold- respektive normalfördelningsmetoden. För att modellera data till respektive metod, behöver den passas till en fördelningsfunktion för respektive metod, vilka är Generalized Extreme Value-, Generalized Pareto- och Normalfördelningen. Riskmått som undersöks är baserade på data från 2008-2023 för aktieindexen OMXSPI, SX20PI och SX3010PI. Hur väl data passas till respektive fördelning för respektive metod görs genom jämförande mellan densitets-grafer, QQ-grafer och Anderson-Darling-test. Studien undersöker hur riskmåtten Value at Risk ändrats över tid, vilken nivå det har idag och vilken trend det kan hålla i framtiden. Det senare nämnda har gjorts genom antagandet av icke-stationära parametrar för de extremvärdesteoretiska fördelningarna. I studiens senare skede undersöks om det finns någon korrelation mellan ett index riskmått med dess pris och om det finns någon korrelation mellan de tre indexen, för olika tidsfördröjningar. Resultatet av studien visar att SX3010PI regelbundet uppvisar de högst beräknade riskmått, både över tid och idag, SX20PI har det lägst beräknade riskmått, både över tid och idag, och OMXSPI ligger däremellan. Genom en förenande bild av dåtida Value at Risk och framtida, är det möjligt att undersöka om det finns en upp och nedgående trend. Resultatet blev att korrelationen mellan Value at Risk och indexets pris visade sig vara mycket liten medan korrelationen mellan indexens pris var korrelerade för ingen tidsfördröjning.

Abstract

The thesis originates in classic extreme value theory and aims to calculate the Value at Risk and Expected Shortfall risk measures. In order to do this, some commonly used methods are the Block Maxima, Peaks over Thresholds, and normal distribution-method. Each method has a corresponding distribution, to which data needs to be modeled. These are Generalized Extreme Value-, Generalized Pareto-, and the Normal distribution. The risk measures are based on data from stock indices OMXSPI, SX20PI, and SX3010PI in the time period of 2008-2023. The extent to which the data fits each of the distributions are evaluated using density graphs, QQ-plots, and Anderson-Darling tests. The study examines how Value at Risk changes over time, what levels it's at today, and possible trends regarding the future. The latter is done through the assumption of non-stationary parameters of the theoretical distributions. Further, the study explores possible correlation between the indices with different time delays. The result showed that SX2010PI alone has the highest risk, presently and over time, with SX20PI exhibiting the lowest risk today and over time. OMXSPI places right inbetween. Through an intertwined graph displaying both past and present Value at Risk there is a possible up- or downwards trend. No correlation between the Value at Risk and stock price was observed, nor between the indices, apart from the ones without time delay.

Innehåll

1	Inledning	1
1.1	Bakgrund	1
1.1.1	Kort om aktiemarknaden	1
1.1.2	Matematiska modeller	1
1.1.3	Indexinnehåll	2
1.2	Syfte	2
1.3	Frågeställningar	2
1.4	Avgränsningar	2
1.5	Samhälleliga och etiska aspekter	3
2	Teori	3
2.1	Klassisk extremvärdesteori	3
2.2	Block Maxima-metoden	3
2.3	Peaks over Threshold-metoden	4
2.4	Trender i parametrar	4
2.5	Normalfördelningsmetoden	5
2.6	Maximum Likelihood-estimat	5
2.7	Anderson-Darling-test	5
2.8	Riskmått	6
2.8.1	Value at Risk	6
2.8.2	Expected Shortfall	6
2.9	Beroenden inom stationära serier	6
2.9.1	Block-runs-metoden	7
2.9.2	Extremindex	7
2.10	Autokorrelation och korskorrelation	7
2.11	Svart svan	8
3	Metod	8
3.1	Datainsamling	8
3.2	Block Maxima-metoden	9
3.2.1	Val av blockbredd	9
3.2.2	Riskmåttskonvertering med extremindex	9
3.3	Peaks over Threshold-metoden	10
3.3.1	Val av tröskel	10
3.3.2	Avklustring	10
3.4	Parameterskattning	11
3.5	Godhetstest	11
3.6	Beräkning av nutida VaR och ES	13
3.7	VaR för trender i parametrar	13
3.8	VaR över tid	13
3.9	Korskorrelation mellan index, VaR och avkastning	14
3.10	Backtesting	14
4	Resultat	15
4.1	Nutida VaR och ES	15
4.2	VaR för trender i parametrarna	15
4.3	VaR över tid och volatilitet	16
4.4	Korskorrelation mellan indexen samt VaR och avkastning	17
4.5	Backtesting	17
5	Diskussion	18
5.1	Metod	18
5.2	Resultat	18
5.3	Slutsats	20

Referenser	21
Appendix	23
A Figurer	24
A.1 Val av tröskel	24
A.2 Densitets- QQ-graf	24
B Tabeller	27
B.1 Parameterskattningar från  och 	27
B.2 Konfidensintervall för parameterskattningar	27
B.3 Anderson-Darling-test	28
C Kod i 	29
C.1 Kod för nutida VaR och ES med tillhörande grafer	29
C.2 Kod för VaR över tid	32
C.3 Kod för icke-stationära parametrar	35
C.4 Kod för korskorrelation	36
C.4.1 Mellan index	36
C.4.2 Mellan VaR och pris	38
D Kod i 	41
D.1 Huvudfil	41
D.2 Funktionsfiler	44
D.2.1 Avklustring	44
D.2.2 Histogramskapare	44
D.2.3 Peaks over Threshold	45
D.2.4 Backtesting normalfördelning	45
D.2.5 Backtesting ej avklustrad PoT	45
D.2.6 Backtesting avklustrad PoT	46
D.2.7 Backtesting BM	46

1 Inledning

I detta avsnitt presenteras projektets syfte, frågeställningar, avgränsningar och samhällliga och etiska aspekter med dess ingående delar.

1.1 Bakgrund

I följande delavsnitt presenteras aktiemarknadens karaktär och hur dess rörelser kan modelleras samt de tre index som utgör studiens underlag.

1.1.1 Kort om aktiemarknaden

De finansiella marknaderna präglas på kort sikt av både positiva och negativa rörelser som vid en första anblick kan framstå som omotiverade och slumpvisa. Bara under 2022 föll OMXSPI över 4% under en och samma dag vid tre olika tillfällen [1]. Påverkan som dessa nedgångar har på individuella aktörer ska inte förringas, likväl förefaller de synnerligen försumbara i jämförelse med större krascher. Dessa ”finansiella stormar” har mycket gemensamt med exempelvis extremväder i det att de båda motsvarigheterna inträffar mer eller mindre regelbundet och generellt orsakar förödande katastrofer. Som ett exempel blev en direkt följd av finanskrisen i USA 2008 att stockholmsbörsen förlorade nästan 60% av sitt värde, vilket innebar att tusentals miljarder kronor gick om intet [1].

Det råder delade meningar kring krisens ursprung, och detsamma gäller frågan kring huruvida den hade kunnat förutspås. Om konsensus skulle motsäga det senare, benämns händelsen som en svart svan. Begreppet innebär förenklat att en händelse med allvarliga konsekvenser inte har kunnat föreställas innan den inträffat och först efteråt kunnat förklaras [2].

1.1.2 Matematiska modeller

Både med och utan krascher uppvisar finansindex betydande stokastiska karaktärsdrag. För en rationell investerare förefaller möjligheten, att på statistisk väg beräkna riskmått på stora förluster och deras respektive sannolikheter, synnerligen intressant. Ytterligare relevans återfinns i att i ett större perspektiv undersöka hur matematisk extremvärdesstatistik kan tillämpas på index över olika branscher.

Två riskmått är av särskilt intresse för studien. Initialt hur stor andel av portföljens värde en investerare står i begrepp att förlora givet en konfidensgrad och tidsram, och därefter vad investeraren förväntas förlora om resultatet utfaller lika- eller än mer negativt [3].

Vid konventionell modellering inom extremvärdesstatistik är det speciellt två metoder som ständigt återkommer och dessa utgår från täthetsfunktioner med större emfas på osannolika utfall, vilket är önskvärt vid modellering av extremvärden [3]. Fördelningarna benämns *Generalized Extreme Value* (GEV) och *Generalized Pareto* (GP). Den datamässiga utgångspunkten för de båda fördelningarna ligger i vad som kallas *Block Maxima* (BM) respektive *Peaks over Threshold* (PoT) [4]. Den förstnämnda utgår från den största förlusten inom varje tidsmässigt segment, och den efterföljande från samtliga förluster som överstiger en bestämd tröskel.

De finansiella marknadernas natur och matematiken inom extremvärdesteorin bjuder in till diskussion kring utsträckningen med vilka dessa kan tillämpas med syfte att skapa underlag för investeringsbeslut eller riskhantering. Riskmått och statistiska värden bör i nästa skede ställas i relation till yttre faktorer såsom trender och omvärldsfaktorer. Ämnet och studiens område utgörs därmed i huvudsak av kvantitativa analyser, som kompletteras av kvalitativa sådana, vilket är en förutsättning för att skapa en nyanserad och väl underbyggd diskussion kring matematikens och statistikens användningsområde i kombination med risk- och kapitalhantering på de finansiella marknaderna.

1.1.3 Indexinnehåll

Samtliga 394 aktier noterade på stockholmsbörsen utgör tillsammans indexet OMXSPI. Till följd av sin storlek, och det faktum att det ofta anses representera den svenska ekonomin i stort, jämför sig ofta investerare mot indexet. Särskilt intressant blir det att undersöka risknivån, då indexet uppvisat periodvis hög volatilitet och påverkats av flertalet krascher [1]. Ett mer snävt urval på endast aktier tillhörande hälsosektorn resulterar i SX20PI, som skiljer sig gentemot OMXSPI i det att det uppvisar lägre volatilitet över tid. Beteendet omkullkastades dock vid sommaren 2021, då indexet upplevde stora svängningar. Slutligen studeras indexet SX3010PI, som består av tio bankaktier på Stockholmsbörsen, vilket har uppvisat högre grad av volatilitet jämfört med de tidigare två indexen, vilket gör det till en intressant kandidat att utvärdera. Inget av de tre indexen justerar kursen för utdelningar.

1.2 Syfte

Syftet med arbetet är att med matematiska metoder beräkna extremvärdesteoretiska riskmått och analysera riskmåttstrender för olika svenska aktieindex. Studien ämnar att undersöka och redogöra för hur statistiska metoder kan tillämpas för att avgöra investeringsrisken för olika index, samt hur trenderna i dessa förhåller sig till varandra. Resultaten används för att avgöra i vilket av de olika aktieindexen den största investeringsrisken finns. Studien undersöker om någon korrelation erhålls med tid som variabel.

1.3 Frågeställningar

De frågor och problem som undersöks är:

- Hur stora förluster finns det risk för vid investering i var och ett av de valda indexen?
- Är rörelserna i de olika indexens värde korrelerade med varandra?
- Är det möjligt att identifiera tydliga riskmåttstrender för respektive index?
- I vilken utsträckning kan resultaten härledas till omvärldsfaktorer och betraktas som svarta svanar?

1.4 Avgränsningar

Studien avgränsas till att inkludera ett fåtal aktieindex på den svenska börsmarknaden, vilka baseras på branschrepresentation, storlek och sammanhang. Att inte fler aktieindex analyseras är en följd av arbetets begränsade omfattning och inbjuder till en mer djupgående analys och diskussion.

Ytterligare en avgränsning är att endast en tidsperiod för respektive index analyseras. Längden på vald tidsperiod väljs för att medvetet inkludera extrema värden från historiska börskrascher, som kan påvisa önskade effekter.

I studien tillämpas två metoder för extremvärdesteoretisk riskbedömning, vilka är BM och PoT. Metoderna är de vanligaste och högst ansedda vid extremvärdesteoretisk analys [4]. Dessutom prövas antagande om att de dagliga förlusterna är normalfördelade, vilket är en metod som kan passa breda index, vilka tenderar att uppvisa normalfördelade förluster. Vidare estimeras samtliga parametrar endast med hjälp av *Maximum Likelihood Estimation* (MLE)-metoden, då den är konsekvent använd inom extremvärdesteorin och anses fullgod med hänsyn till studiens syfte [3].

Då både tid och omfång är begränsad för studien kommer avgränsningen att inte undersöka ett teoretiskt tillvägagångssätt för att bestämma blockbredd att göras.

1.5 Samhälleliga och etiska aspekter

En diskussion kring arbetets etik och samhälleliga aspekter är relevant då finansiella tillämpningar i stor utsträckning kan komma att ligga till grund för investeringsbeslut eller spekulation på aktiemarknaden, vilket inte är studiens syfte. Istället ämnar studien undersöka och inbjuda till diskussion kring statistisk extremvärdesteori i ett teoretiskt sammanhang. Läsare som agerar på strategier med grund i riskmått presenterade i studien riskerar att förlora investerat kapital.

Matematisk teori, metodik och resultat presenterade i studien ska inte tolkas som investeringsunderlag. För att ytterligare understryka omfånget av studien ställs statistisk tolkning i ett sammanhang där omvärldens påverkan utvärderas. Genomgående i studien presenteras även ekonomiska samband i syfte att uppvisa en nyanserad bild av ämnet.

Studiens resultat avser endast ett på förhand utvalt urval av tidsperioder och behöver inte nödvändigtvis vara av signifikans för andra dataserier. Läsare avråds från att dra slutsatser kring de olika aktieindexens karaktär eller riskprofil utöver de resultat presenterade i arbetet.

2 Teori

I följda avsnitt kommer relevant teori förklaras för att ge läsaren en inblick i hur och varför metoderna och modellerna appliceras.

2.1 Klassisk extremvärdesteori

Extremvärdesteori är en del av statistiken som beaktar extrema händelser för en slumpvariabel X . Med hjälp av denna teori kan estimat göras för att se sannolikheten för att en extrem händelse kommer att inträffa, samt bestämma med vilken frekvens sådana händelser inträffar [4]. Grunden till extremvärdesteori är GEV-fördelningen.

Proposition 2.1 [4] *Låt X_1, \dots, X_n vara en sekvens av oberoende likafördelade (o.l.f.) slumpvariabler med fördelningsfunktionen F och $M_n = \max\{X_1, \dots, X_n\}$. Om n observationer erhålls under en given tidsperiod, kommer*

$$\mathbb{P}(M_n \leq x) = F(x)^n,$$

vara fördelningen för maximum¹.

2.2 Block Maxima-metoden

BM-metoden är en återkommande metod då extrema händelser ska analyseras i ett dataset [3]. Tidsperioden som skall analyseras delas in i icke-överlappande block där det största värdet av datasetet i just det blocket svarar för den extrema händelsen. De olika extremvärdena i varje block antas vara oberoende av varandra och modelleras med en GEV-fördelning [5]. Ju större blockbredd som väljs desto färre maxima kommer modelleras som innebär större variation gällande estimaten. Medan en allt för liten blockbredd motverkar de egenskaper som gör GEV-fördelningen lämplig vid modellering av extrema händelser.

Om X_1, \dots, X_n är o.l.f. slumpvariabler och M_n definieras enligt proposition 2.1 kan följande sats härledas från gränsvärdessatsen för maximum.

Sats 2.2 [4] *Om det existerar en sekvens av konstanter $a_n > 0$ och b_n som uppfyller*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) = \lim_{n \rightarrow \infty} F^n(a_n z + b_n) = G(z) \quad (1)$$

¹Eftersom fördelningen för minimum kan erhållas genom en negation av datasetet, kommer endast maximum att beaktas.

har G en GEV-fördelning, med fördelningsfunktion

$$G(z; \mu, \sigma, \gamma) = \exp \left\{ - \left[1 + \gamma \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\gamma} \right\} \quad (2)$$

där $1 + \gamma \left(\frac{z - \mu}{\sigma} \right) > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ och $-\infty < \gamma < \infty$.

2.3 Peaks over Threshold-metoden

Ytterligare en metod som används vid modellering av extrema händelser är PoT. Generellt kan BM-metoden anses vara icke-tillräcklig vid extremvärdesmodellering där för få extrema händelser inträffar i ett block. Om mer data finns tillgänglig kan relevant data bli försummad och resultatet mindre precist [4]. Istället för att dela in händelserna i block och se det största värdet i varje block som en extrem händelse, anses alla värden över en preciserad tröskel vara extrema i PoT-metoden. Ett överskridande beräknas som maxvärdet subtraherat med tröskelvärdet. Det betyder att utfallet X_i av en slumpvariabel X , anses vara ett extremvärde om det överstiger en tröskel u . Det betyder att sannolikheten för att den beskrivna händelsen inträffar kan skrivas som

$$\mathbb{P}(X > u + y \mid X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0, \quad (3)$$

där F är fördelningen för X_1, \dots, X_n .

Sats 2.3 [4] Låt X_1, \dots, X_n vara en följd av o.l.f. slumpvariabler med fördelningsfunktion F , och M_n definierat som i proposition 2.1. Vidare antas F uppfylla sats 2.2. Om sedan u väljs tillräckligt stort kommer ekvation (3), med substitution av ekvation (2), approximativt ha fördelningsfunktionen

$$H(y; \tilde{\sigma}, \gamma) = 1 - \left(1 + \gamma \frac{y}{\tilde{\sigma}} \right)^{-1/\gamma}, \quad (4)$$

där H är definierat för $y > 0$ och $1 + \gamma \frac{y}{\tilde{\sigma}} > 0$, och $\tilde{\sigma} = \sigma + \gamma(u - \mu)$. H benämns som GP-fördelningen.

För estimering av svansen för F så gäller följande:

$$\bar{F}(y) = 1 - F(y) = \bar{F}(u) \bar{F}_u(y - u)$$

där $\bar{F}_u(y - u)$ är den betingade fördelningsfunktionen för tröskelöverskridanden. Det följer då även att

$$\hat{\bar{F}}(y) = \hat{p}_u \hat{\bar{F}}_u(y - u). \quad (5)$$

p_u är sannolikheten för överskridande och skattas som $\frac{\text{antalet överskridanden}}{\text{antalet insamlade datapunkter}}$, och betecknas som \hat{p}_u . $\hat{\bar{F}}$ är estimatet av \bar{F} . Från ekvation (5) kan följande motiveras för en GP-fördelning

$$\hat{\bar{F}}(y) = \hat{p}_u \left(1 + \frac{\gamma}{\sigma} (y - u) \right)^{-\frac{1}{\gamma}}. \quad (6)$$

2.4 Trender i parametrar

Vid modellering av extrema fördelningar kan det finnas skäl att tro att parametrarna inte är stationära, det vill säga att de beror på tiden t [7]. I sådana fall är det möjligt att inkludera detta vid modellerandet av parametrarna för fördelningen. Detta är speciellt intressant då deterministiska datamängder eller trender studeras. För att behandla beroendet av t för parametrarna för extremvärdesteoretiska fördelningar, kan μ och σ från GEV- respektive GP-distributionen till exempel modelleras enligt ekvation (7) och ekvation (8).

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2. \quad (7)$$

På grund av att σ endast kan anta positiva värden, modelleras den vanligtvis enligt

$$\ln \sigma(t) = \phi_0 + \phi_1 t + \phi_2 t^2. \quad (8)$$

2.5 Normalfördelningsmetoden

Det finns andra fördelningar än de tidigare nämnda extremvärdesfördelningarna som kan användas vid riskbedömningsssammanhang, där en av dessa är normalfördelningen². Normalfördelningen är särskilt tillämpbar i riskbedömningsssammanhang när bredare aktieindex undersöks [3].

Definition 2.4 Låt X_1, \dots, X_n vara en sekvens av oberoende slumpvariabler med fördelningsfunktion Φ om fördelningen för dessa slumpvariabler följer

$$\Phi(x; \mu, \sigma) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz \quad (9)$$

sågs $X \sim \mathcal{N}(\mu, \sigma^2)$.

Normalfördelningsmetoden innebär att alla värden som samlas in för en dataserie antas vara normalfördelade.

2.6 Maximum Likelihood-estimat

Givet en mängd parametervärden, kommer den motsvarande täthetsfunktionen visa att vissa mät-punkter är mer sannolika än andra. I praktiken är ofta problemet av motsatt karaktär, i och med att händelsekedjan redan observerats och mät-punkter samlats in. Maximum Likelihood-metoden används för att med hjälp av historiska mätvärden bestämma den täthetsfunktion som med störst sannolikhet har genererat det observerade resultatet [8].

Låt $x = \{x_1, x_2, \dots, x_n\}$ vara observerade värden av o.l.f. slumpvariabler X_1, X_2, \dots, X_n , och $\vartheta = \{\vartheta_1, \vartheta_2, \dots, \vartheta_m\}$ vara m parametrar i täthetsfunktionen $f(x; \vartheta)$. Eftersom X_1, X_2, \dots, X_n är o.l.f. ges den simultana täthetsfunktionen $f(x_1, x_2, \dots, x_n; \vartheta)$ av

$$f(x_1, x_2, \dots, x_n; \vartheta) = \prod_{i=1}^n f(x_i; \vartheta),$$

vilken också brukar benämnas likelihood-funktionen [4]. Vid analytisk skattning av parametrar är det ofta enklare att logaritmera likelihood-funktionen $L(\vartheta)$, och därmed arbeta med log-likelihood-funktionen $l(\vartheta)$. Eftersom logaritmfunktionen är en monotont växande funktion kommer det värde på ϑ som maximerar $L(\vartheta)$ också att maximera $l(\vartheta)$ [9].

Definition 2.5 [4] Om x_1, \dots, x_n är oberoende observationer av en stokastisk variabel X med täthetsfunktion $f(x; \vartheta_0)$, definieras likelihood-funktionen

$$L(\vartheta) = \prod_{i=1}^n f(x_i; \vartheta)$$

och log-likelihood-funktionen definieras av

$$l(\vartheta) = \log(L(\vartheta)) = \sum_{i=1}^n \log(f(x_i; \vartheta)).$$

MLE är parameterskattningen $\hat{\vartheta}$ på parametern ϑ som maximerar³ likelihood-funktionen [4].

2.7 Anderson-Darling-test

Anderson-Darling-testet är ett godhetstest (eng. goodness of fit) som mäter hur mycket insamlad data avviker från en specificerad fördelning. Testet är icke-parametriskt, det vill säga oberoende av fördelning och kräver inte att data är normalfördelad, vilket innebär att också andra fördelningar än normalfördelningen kan undersökas. [10].

²Noterbart är att normalfördelningen är inte likt de andra extremvärdesdistributionerna. Istället för att sortera ut extremvärdena för en tidsserie, används all data som samlats in för tidsserien.

³Ibland går det inte att göra analytiskt och då måste det göras numeriskt. Ibland finns flera eller inget maximum.

Definition 2.6 [11] *Teststorheten, A_n^2 , för ett Anderson-Darling-test definieras som*

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F^0(x))^2}{F^0(x)(1 - F^0(x))} dF^0(x), \quad (10)$$

där n är storleken på stickprovet, x_1, x_2, \dots, x_n är stickprovets observerade mätpunkter, $F_n(x)$ är den empiriska fördelningsfunktionen och $F^0(x)$ är den teoretiska fördelningsfunktion som datan jämförs med. Den empiriska fördelningsfunktionen ges av

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}},$$

där $\mathbb{1}_{\{x_i \leq x\}}$ är indikatorfunktionen.

Sats 2.7 [11] *Ekvation (10) kan skrivas som*

$$A_n^2 = -n - \frac{1}{n} \sum_{j=1}^n (2j-1)(\log(u_{(j)}) + \log(1 - u_{(n-j+1)})), \quad (11)$$

där $u_{(j)} = F^0(x_{(j)})$ och $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ är den observerade datan sorterad i stigande ordning⁴.

2.8 Riskmått

I detta avsnitt presenteras de två mest vedertagna riskmåten inom extremvärdesteori.

2.8.1 Value at Risk

VaR är det mest använda riskmålet för kvantitativa bedömningar av finansiell risk [12], och markerar det värde som en investerare, givet konfidsensgrad $1 - \alpha$, inte riskerar att förlora under en specificerad tidsperiod. Riskmålet motsvaras av den övre α -kvantilen i en fördelning som beskriver finansiella förluster [13].

Definition 2.8 [13] *Om L är en stokastisk variabel⁵ som representerar förlustfördelningen för en given portfölj definieras VaR enligt*

$$\text{VaR}_\alpha(L) = \sup\{x \in \mathbb{R} \mid \mathbb{P}(L \geq x) > \alpha\},$$

där $\sup\{x \in \mathbb{R} \mid \mathbb{P}(L \geq x) > \alpha\}$ representerar den övre α -kvantilen av förlustfördelningen.

2.8.2 Expected Shortfall

$\text{ES}_\alpha(L)$ definieras som den förväntade förlusten, givet att den överskrider $\text{VaR}_\alpha(L)$ [13]. Ibland kallas ES även för betingat VaR [12].

Definition 2.9 [13] *Om L är en stokastisk variabel som representerar förlustfördelningen, definieras $\text{ES}_\alpha(L)$ enligt*

$$\text{ES}_\alpha(L) = \mathbb{E}(L \mid L \geq \text{VaR}_\alpha(L)) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_y(L) dy.$$

2.9 Beroenden inom stationära serier

En tidsserie vars egenskaper inte beror på tidpunkten då den observeras benämns som stationär. För sådana tidsserier förekommer ofta ett beroende mellan extrema händelser och om de inträffar tätt inpå varandra, benämns det som ett kluster [14]. En förutsättning för PoT-är att data är oberoende och därför måste dessa kluster identifieras och hanteras.

⁴Teorin bakom (10) och hur övergången mellan (10) och (11) sker kommer ej behandlas, då det ligger utanför syftet med denna rapport.

⁵Endast definitionen för det kontinuerliga fallet kommer betraktas. Detta eftersom samtliga fördelningar som behandlas i denna rapport är kontinuerliga.

2.9.1 Block-runs-metoden

För att bestämma antalet kluster samt avklustra datavärden för en tidsserie och göra dem oberoende av varandra kan Block-runs-metoden tillämpas. Metoden går ut på att det första klustret initieras vid första överskridandet av tröskeln u och innehåller alla överskridanden inom en räckvidd r [14]. Nästa kluster initieras sedan vid nästa överskridande och proceduren repeteras sedan för alla överskridanden, vilket innebär att den avklustrade data endast innehåller det största överkridande datavärdet i varje kluster.

2.9.2 Extremindex

Extremindex, $\theta \in [0, 1]$ är en parameter som beskriver till vilken grad extrema värden samlas i kluster [15]. Det vill säga, θ beskriver intensiteten av beroendet mellan extrema händelser.

Sats 2.10 [4] *För stationära tidsserier kommer extrema händelser och dess fördelning att, under lämpliga villkor, följa*

$$\mathbb{P}(M_n \leq x) \approx F(x)^{\theta n}.$$

Där $M_n = \max\{X_1, \dots, X_n\}$ för en sekvens av oberoende likafördelade (o.l.f.) slumpvariabler X_1, \dots, X_n med fördelningsfunktionen F .

I BM-metoden resulterar tillräckligt breda block i att de extrema händelserna i urvalet kan anses vara oberoende [14]. Vid användning av PoT-metoden däremot, när samtliga extrema värden över tröskeln inkluderas blir klusterintensiteten högst relevant och måste tas i beaktande.

Sats 2.11 [14] θ kan skattas som

$$\hat{\theta} = \frac{\text{antal kluster}}{\text{antal överskridanden}}.$$

2.10 Autokorrelation och korskorrelation

Korrelation är ett mått på sambandet mellan två eller fler variabler. Definitionen av korrelation är

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y},$$

där X och Y är slumpvariabler med väntevärde μ_X respektive μ_Y och standardavvikelse σ_X respektive σ_Y , och $\text{Cov}(X, Y)$ är kovariansen mellan X och Y .

Autokorrelation är ett beroendemått mellan stokastiska variabler som ingår i en och samma stokastiska process, eller mellan datapunkter i samma datamängd [16]. För stationära tidsserier är autokorrelationen som störst när förskjutningen är 0, och vanligen avtar den ju större förskjutningen är, vilket innebär att beroendet avtar i samband med att det tidsmässiga⁶ avståndet ökar [17]. Autokorrelationen mellan två tidpunkter, r och s , i en tidsserie, X_t , definieras som

$$\text{Corr}(X_r, X_s) = \frac{\text{Cov}(X_r, X_s)}{\sigma_{X_r} \sigma_{X_s}}.$$

Givet att en observerad datamängd är realiserade värden ur en stationär tidsserie kan autokorrelationsfunktionen skattas genom stickprovsautokorrelationsfunktionen [18]. Om x_1, \dots, x_n är n observationer av en stationär tidsserie, och h är tidsförskjutningen, kan stickprovsmedelvärdet definieras som

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t. \tag{12}$$

⁶Förskjutningen kommer i denna rapport endast bestå av kovariatet tid.

Stickprovsautokovariansen beräknas som

$$\hat{\gamma}_x(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad |h| < n. \quad (13)$$

Stickprovsautokorrelationen beräknas som

$$\hat{\rho}_x(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}, \quad |h| < n. \quad (14)$$

Korskorrelation definieras som autokorrelationen mellan två eller fler olika tidsserier [16]. I praktiken innebär detta att rörelsen för den ena datamängden är dynamisk medan den andra datamängden är fixerad över tid, och mäter graden av korrelation mellan två datamängder. Korskorrelationen mellan två tidsserier, X_t och Y_t , för två tidpunkter, r och s , definieras som

$$\text{Corr}(X_r, Y_s) = \frac{\text{Cov}(X_r, Y_s)}{\sigma_{X_r} \sigma_{Y_s}}.$$

I det generella fallet är inte $\hat{\rho}_{xy}(h) = \hat{\rho}_{xy}(-h)$, eftersom korskorrelationsfunktionen är asymmetrisk kring $h = 0$. En följd av detta är att stickprovskorskovariansen för en observerad datamängd definieras på olika vis beroende på om tidsförskjutningen är positiv eller negativ [19]. Om $(x_1, y_1), \dots, (x_n, y_n)$ ⁷ är n parvisa observationer av två stationära tidsserier, X_t och Y_t , h är tidsförskjutningen samt \bar{x} respektive \bar{y} är stickprovsmedelvärdet definierat enligt ekvation (12), beräknas stickprovskorskovariansen som

$$\hat{\gamma}_{xy}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (x_t - \bar{x})(y_{t+h} - \bar{y}), & 0 \leq h < n \\ \frac{1}{n} \sum_{t=1}^{n+h} (y_t - \bar{y})(x_{t-h} - \bar{x}), & -n < h < 0 \end{cases} \quad (15)$$

och stickprovskorskorrelationen beräknas som

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_{xx}(0)\hat{\gamma}_{yy}(0)}}, \quad |h| < n. \quad (16)$$

2.11 Svart svan

Ovanstående extremvärdesteori delar fortfarande den inneboende svagheten att endast titta på tidigare data som underlag till modellerna, vilket benämns som "induktionsproblemet" [2]. Teorin kring svarta svanar och deras förekomst motsätter sig genom sin definition att statistiska riskmått, som utgår från historiskt data, konsekvent kan dra helt korrekta slutsatser om framtiden. Snarare finns det tillfällen då inte heller extremvärdesteori kan förutspå osannolika händelser. Det introduceras ännu en faktor, utliggare (eng. outliers), som talar för detta, vilka inte är lika känsliga för underskattning som de är för felskattning. Ju lägre sannolikhet att en händelse inträffar, desto svårare blir det att med god säkerhet förutspå dess effekter.

3 Metod

I detta avsnitt presenteras det tillvägagångssätt som använts och legat till grund för de senare resultaten.

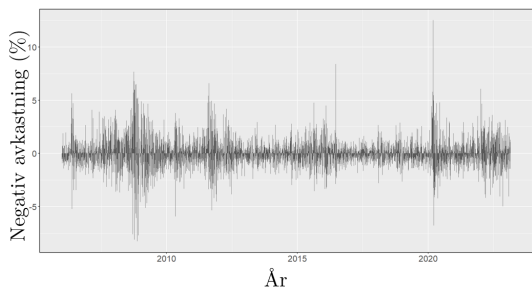
3.1 Datainsamling

Datainsamlingen genomförs identiskt för projektets alla delar som behandlas; nutida VaR och ES, VaR över tid, icke-stationära parametrar och korskorrelation mellan index. Uteslutande hämtas data från Nasdaq OMX Nordic där stängningspriset p samlas in för varje börsdag under tidsperioden 2 januari 2006 till 27 februari 2023 genom en CSV-fil med respektive datum.

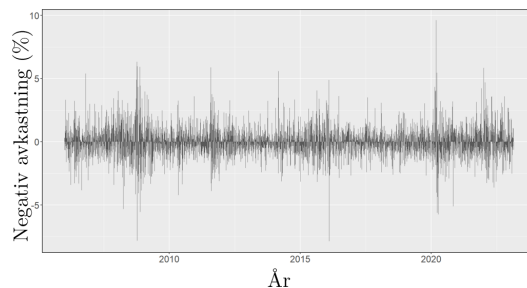
⁷Här betecknas stickprovsautokovariansen $\hat{\gamma}_{xx}(h)$ respektive $\hat{\gamma}_{yy}(h)$, för att förtydliga att det endast rör sig om en tidsserie, till skillnad från stickprovskorskorrelationen.

För att beräkna riskmått VaR och ES behöver data för en överskådlig dåtid för de olika tids-serierna samlas in. Data samlas in i form av stängningspriset p för OMXSPI, SX20PI respektive SX3010PI för varje börsdag under tidsperioden nämnd ovan. Här jämförs p med föregående börsdags p för att ta ut det procentuella förlusten. Här benämns p_i som p idag och p_{i-1} som p igår för $i \geq 2$. Den negativa procentuella avkastningen visualiseras i figur 1-3 och beräknas enligt

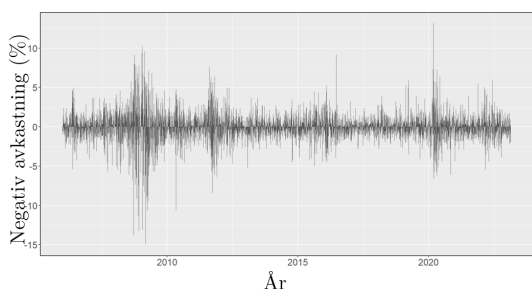
$$\left(\text{Negativ procentuell avkastning (\%)} \right)_i = -100 \cdot \frac{p_i - p_{i-1}}{p_{i-1}}.$$



Figur 1: Negativ procentuell avkastning OMX-SPI (2006-2023)



Figur 2: Negativ procentuell avkastning för SX20PI (2006-2023)



Figur 3: Negativ procentuell avkastning för SX3010PI (2006-2023)

Vektorerna som består av den negativa procentuella avkastningen, för respektive index, är de vektorer som används vid vidare beräkningar för projektets samtliga ändamål.

3.2 Block Maxima-metoden

I detta avsnitt presenteras tillvägagångssättet vid modellering av BM-metoden.

3.2.1 Val av blockbredd

Genom den teori som presenteras i avsnitt 2.2 kan datamängden ofta antas vara oberoende om tillräckligt stor blockbredd väljs. Med hänsyn till ovanstående information väljs blockbredden till en börsmånad, då detta intervall genererar tillräckligt många datapunkter för att datamängden ska kunna approximeras till en GEV-fördelning. Vidare är en börsmånad en naturlig blockbredd sett till de riskmått som räknas ut, då månatliga riskmått är intuitiva att förstå, till skillnad från ett godtyckligt antal dagar. Enligt studiens avgränsningar undersöks inte val av blockbredd ytterligare.

3.2.2 Riskmåttskonvertering med extremindex

I och med att en blockbredd på en börsmånad valts, medför det att ett månatligt VaR beräknas. Med slutändamål att beräkna ett dagligt VaR för respektive metod och index, måste därför månatligt VaR konverteras till dagligt VaR för BM-metoden. Denna konverteringen kan göras med hjälp av extremindex enligt sats 2.10, där extremindex är definierat enligt sats 2.11. Antalet kluster

identifieras via block runs-metoden enligt avsnitt 2.9.1, där räckvidden r väljs till en börsvecka, då detta intuitivt motiveras med att extrema förluster med mer än en börsveckas intervall kan betraktas som oberoende.

3.3 Peaks over Threshold-metoden

I detta avsnitt presenteras tillvägagångssättet vid modellering av PoT-metoden.

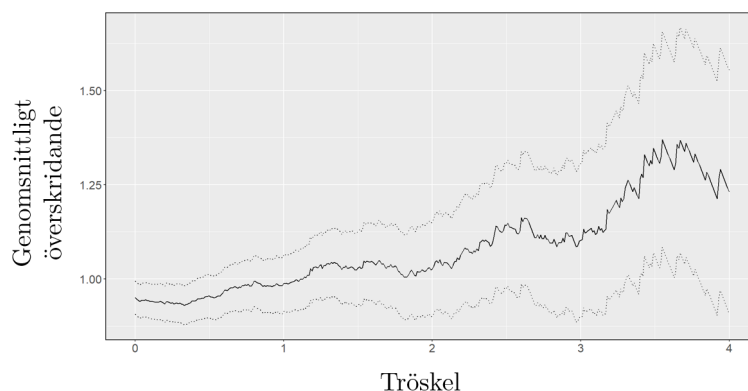
3.3.1 Val av tröskel

Valet av tröskel i PoT-metoden är en avgörande aspekt för hur väl GP-fördelningen representerar datamängden. Att bestämma tröskelvärdet är ingen exakt vetenskap, men det finns både analytiska och grafiska metoder som kan tillämpas som beslutsunderlag. Med bakgrund av detta används två metoder för att bestämma tröskelvärdet i denna studie, där den ena baserar sig på datamängdens percentil och den andra är ett grafiskt tillvägagångssätt.

Det är ofta av intresse att modellera ett förbestämt antal observationer. Detta görs genom att sortera samtliga mätvärden i storleksordning, för att sedan välja den percentil som anses vara lämplig att modellera med. Ofta föredras den 90:e percentilen [20], vilket innebär att de 10 procent mest extrema datapunkterna ligger som grund för metoden.

Ytterligare en metod för att välja tröskel är att studera *Mean Residual Life* (MRL)-graferna [4]. I detta tillvägagångssätt jämförs det genomsnittliga tröskelöverskridandet, mätt i procentenheter, för olika tröskelnivåer. Ett lämpligt tröskelvärde placeras där MRL-grafen börjar uppvisa linjära tendenser. Att bestämma tröskelvärdet genom grafiska metoder är ofta tidskrävande, eftersom dessa kräver en subjektiv bedömning.

I denna studie bestäms tröskelvärdet genom att först ta fram den 90:e percentilen för den dagliga negativa avkastningen i respektive index. Värdet valideras sedan med MRL-grafen, förutom när VaR över tid och backtestingen utförs, vilka presenteras i avsnitt 3.8 respektive 3.10, i och med att koden itereras flertalet gånger och möjligheten att subjektivt bedöma samtliga MRL-grafer inte är möjlig. För indexet OMXSPI, se figur 4, och indexet SX20PI är det svårt att identifiera ihållande linjära tendenser, vilket innebär att den 90:e percentilen varit ensamt beslutsunderlag i dessa fall. För SX3010PI uppvisas linjära tendenser vid $u = 1, 8$, vilket stämmer väl överens med den 90:e percentilen. I figur 4 ses MRL-grafen för OMXSPI. Motsvarande grafer för SX20PI och SX3010PI återfinns i figur 20 och 21 i appendix A.1. Tröskelvärdet för OMXSPI, SX20PI och SX3010PI beräknas till 1,41, 1,30 respektive 1,82.



Figur 4: MRL-graf OMXSPI med tillhörande konfidensintervall



3.3.2 Avklustring

För att kunna använda de extremvärdeteoretiska modellerna BM och PoT måste datavärdena för den negativa procentuella avkastningen vara oberoende av varandra. Detta sker som en direkt



konsekvens för BM då datavärdena delas in i stora block. Däremot måste datavärdena avklustras innan de kan passas till en GP-fördelning. Avklustringen görs med block runs-metoden i samband med beräkningen av extremindex från avsnitt 3.2.2.

3.4 Parameterskattning

För BM plockas maximumvärdena från listan av negativ procentuella avkastningen ut för respektive block och passas till GEV-fördelningen. För PoT plockas de avklustrade värdena från listan av negativ procentuella avkastningen som överskrider den beräknade tröskeln ut och passas till GP-fördelningen. För normalfördelningsmetoden behöver endast parametrar skattas för alla datapunkter som samlats in, i och med att normalfördelningen tar hänsyn till all data. Ovanstående procedur görs för varje index. Resultatet av dessa skattningar ses i tabell 1 för OMXSPI och i tabell 8 och 9 i appendix B.1 för SX20PI och SX3010PI.

Tabell 1: Parameterskattningar för GEV, GP och normalfördelningen med  och  för OMXSPI för perioden 2006-2023

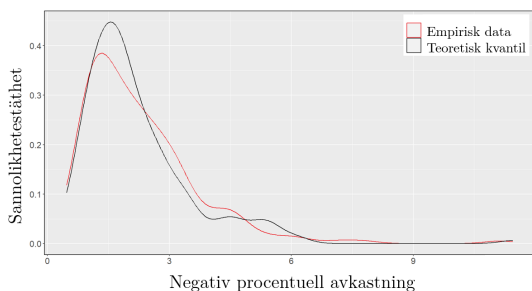
Parameterskattningar OMXSPI			
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	1,60	0,83	0,22
GP		1,26	-0,0011
Normal	-0,033	1,32	

För att säkerställa att de parameterskattningarna som gjordes i  gav ett korrekt resultat, användes  för att jämföra parameterskattningarna för OMXSPI, dessa parametrar blev samma för båda språken och återfinns i tabell 7 i appendix B.1.

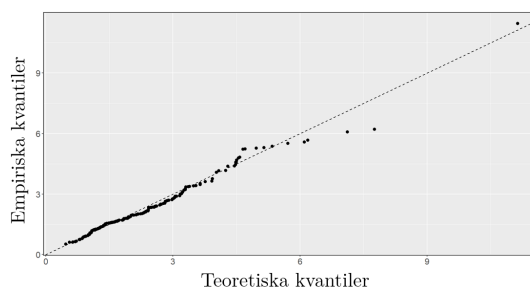
3.5 Godhetstest

För att säkerställa att de parametrar som skattas till respektive fördelning faktiskt representerar verkligheten, behöver godhetstest genomföras. Där ställs funktionsvärdena för respektive teoretiska fördelning gentemot den motsvarande empiriska. Detta görs genom att slumpmässigt generera fördelningsspecifika funktionsvärden med de tidigare skattade parametrarna och sedan grafiskt illustrera dessa täthetsfunktioner tillsammans med de empiriska datavärdena.

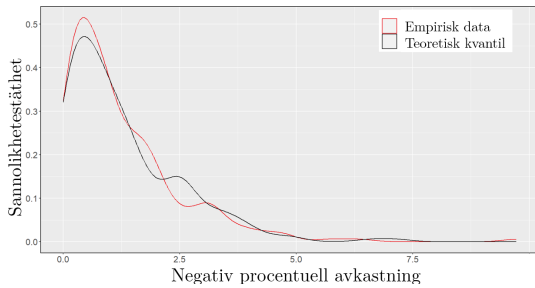
En annan metod för att bedöma hur väl den empiriska data stämmer överens med den teoretiska fördelningen visualiseras i en QQ-graf. I QQ-graferna illustreras de teoretiska kvantilerna på x-axeln, och de empiriska kvantilerna på y-axeln. Ju bättre datapunkterna följer linjen, desto bättre anses fördelningen följa den empiriska datan. Den visar hur nära de teoretiska kvantilerna ligger de empiriska. Testet är relativt informellt och subjektivt, men har fördelen att vara lätt att tolka [21]. I figur 5-10 illustreras teoretisk mot empirisk data för OMXSPI. För att se samma grafer för indexen SX20PI och SX3010PI återfinns dessa i figur 25-36 i appendix A.2.



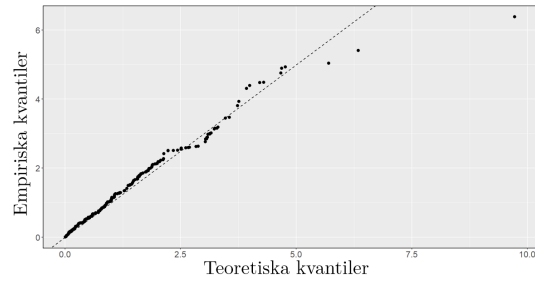
Figur 5: GEV täthetsjämförelse OMXSPI



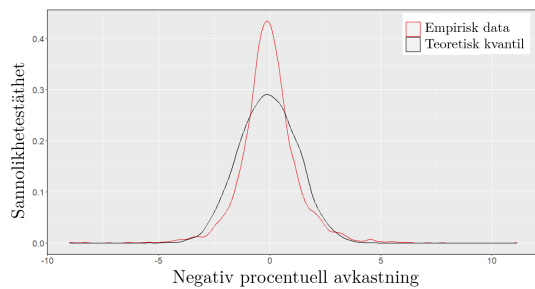
Figur 6: QQ-graf GEV OMXSPI



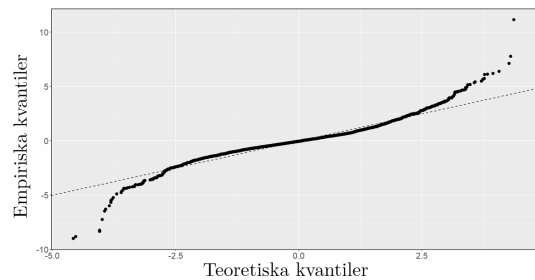
Figur 7: GP tåthetsjämförelse OMXSPI



Figur 8: QQ-graf GP OMXSPI




Figur 9: Normal tåthetsjämförelse OMXSPI



Figur 10: QQ-graf normal OMXSPI

Som nämndes ovan är dessa test förhållandevis informella och subjektiva. Därför genomförs också ett Anderson-Darling-test för att bedöma hur väl den empiriska data passar till respektive fördelning. Anledningen till att just Anderson-Darling-testet väljes som godhetstest, är att detta bedöms vara känsligt för förändringar i fördelningarnas svansar [10]. Testerna görs i enlighet med avsnitt 2.7, med en signifikansnivå $\alpha = 0,05$. I samtliga test är nollhypotesen H_0 att den empiriska data kommer från testad fördelning.

För att genomföra testet används parametrarna som skattats från den insamlade data. Därefter sorteras datapunkterna i stigande ordning för att hitta $u_{(j)}$ och A_n^2 beräknas med hjälp av ekvation (11) [22]. För att beräkna testets p -värde används formler och tabeller framtagna av D'Agostino & Stephens [23]. Dessa ekvationer beskrivs vara oberoende av testad fördelning, men stickprovets storlek har betydelse. Testets kritiska värde avläses i fördelningsspecifika tabeller, eller bestäms via Monte-Carlo-Simulering.

Om A_n^2 överstiger det kritiska värdet kan nollhypotesen förkastas [24], alltså kan det med given konfidsgrad konstateras att den empiriska datan ej bör modelleras med testad fördelning. Ibland förekommer det dock att A_n^2 inte konvergerar och vissa programvaror, däribland  som används i denna del av studien, förkastar istället H_0 om p -värdet understiger given signifikansnivå α [25]. Här är p -värdet sannolikheten att upptäcka ett provutfall minst lika extremt som observerade, under förhållandet att H_0 är sann.

Tabell 2: Resultat från Anderson-Darling-test med signifikansnivå $\alpha = 0,05$, vilket visar beslut om förkastelse (1=förkasta, 0=förkasta ej), p -värde, teststorhet samt kritiskt värde

Resultat från Anderson-Darling-test				
Fördelning	Beslut	p -värde	Teststorhet ⁸	Kritiskt värde
GEV	0	0,68	0,57	2,49
GP	0	0,96	0,26	2,49
Normal	1	$< 5 \cdot 10^{-4}$	∞	0,75

Resultatet av Anderson-Darling-testet indikerar att den empiriska datan inte kommer från en normalfördelning, vilket innebär att H_0 i detta fall kan förkastas. Eftersom att provutfallet var mindre än det kritiska värdet för både GEV- och GP-fördelningarna kan inte H_0 förkastas i något av fallen.

3.6 Beräkning av nutida VaR och ES

För nutida VaR beräknas dagligt sådant. För att bestämma VaR och ES används den metod som tillämpar klassisk extremvärdesteori, det vill säga med BM- och PoT-metoden respektive antagande av normalfördelad data. Data för negativ procentuell avkastning passas till GEV-fördelning och GP-fördelning och normalfördelning. Därefter bestäms VaR genom att beräkna kvantilerna i respektive fördelning för önskad konfidensgrad $1 - \alpha = (0,95; 0,99; 0,999)$, enligt definition 2.8. Dagligt VaR för BM kommer således från ekvation (2) och definition 2.8 beräknas som

$$\text{VaR}_\alpha(L) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}} \left(1 - (-\log(\alpha^{\hat{\theta}n}))^{-\hat{\gamma}} \right) \quad (17)$$

där $\hat{\theta}$ är extremindex och n antalet börsdagar som blocket innehåller, 21 stycken.

VaR för PoT beräknas också enligt definition 2.8. I och med att PoT tar hänsyn till dagliga datapunkter vid beräkning av VaR behöver ingen översättning med hjälp av extremindex göras. VaR beräknas från ekvation (6) och definition 2.8 som

$$\text{VaR}_\alpha(L) = u + \frac{\hat{\sigma}}{\hat{\gamma}} \left(\left(\frac{1 - \alpha}{\hat{p}_u} \right)^{-\hat{\gamma}} - 1 \right) \quad (18)$$

För normalfördelningen används samtliga datapunkter i intervallet för att skatta fördelningens parametrar, och VaR beräknas sedan enligt

$$\text{VaR}_\alpha(L) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(1 - \alpha). \quad (19)$$

I och med att inversen till normalfördelningen, $\Phi^{-1}(x)$, inte går att uttrycka med en elementär funktion beräknas denna med hjälp av R.

Beräkning av ES sker enligt definition 2.9 för ekvation (17), (18) och (19).

3.7 VaR för trender i parametrar

Vid modellering av trender i parametrar används liknande tillvägagångssätt som vid beräkning av nutida VaR. Det som skiljer tillvägagångssätten åt är att inte passa data till en fixerad parameter, utan låta parametrarna vara funktioner med tid t i enlighet med avsnitt 2.4. Genom att kombinera ekvation (7) och (17) kan icke-stationärt VaR för BM-metoden beskrivas enligt

$$\text{VaR}_\alpha(L, t) = \hat{\mu}_0 + \hat{\mu}_1 t + \hat{\mu}_2 t^2 - \frac{\hat{\sigma}}{\hat{\gamma}} \left(1 - (-\log(\alpha^{\hat{\theta}n}))^{-\hat{\gamma}} \right), \quad t \geq 1. \quad (20)$$

Motsvarande för PoT-metoden, används ekvation (8) i kombination med ekvation (18) enligt följande,

$$\text{VaR}_\alpha(L, t) = u + \frac{\exp\{\hat{\phi}_0 + \hat{\phi}_1 t + \hat{\phi}_2 t^2\}}{\hat{\gamma}} \left(\left(\frac{1 - \alpha}{\hat{p}_u} \right)^{-\hat{\gamma}} - 1 \right), \quad t \geq 1. \quad (21)$$

3.8 VaR över tid

När VaR beräknas över tid tillämpas endast konfidensgraden $1 - \alpha = 0,95$. För att beräkna VaR över tid delas datapunkterna från 2006-2023 in i sexåriga intervall. Start- och slutpunkten för påföljande intervall skiljer sig med en börsmånad och de enskilda intervallen utgör datamängden för

⁸Att teststorheten för normalfördelningen närmar sig oändligheten kan förklaras med att värdet på $u_{(j)}$ eller $u_{(n-j+1)}$ i ekvation (2.6) i vissa fall antar värden nära 0 eller 1, vilket gör att någon av logaritmfunktionerna närmar sig $-\infty$.

beräkning av det dagliga VaR direkt vid intervallets slut. Detta innebär att första datapunkten som samlas in är 2 januari 2006 och den sista datapunkten för det intervallet är 1512 (21·12·6) börsdagar senare. Vid den dagen beräknas dagligt VaR ut för respektive metod. Detta görs genom ekvation (17), (18) och $\Phi^{-1}(x)$. Nästa VaR som räknas ut utgår från data från 1 februari 2006 och den sista datapunkten i intervallet är 1512 börsdagar senare. Denna procedur upprepas tills dagen 27 februari 2023 är nådd.

Även den historiska volatiliteten i aktieindexen kommer att presenteras för att sätta VaR över tid i relation till ett annat riskmått som används av investerare för att bedöma osäkerheten på finansiella marknader [26]. Den historiska volatiliteten är ett mått på prisörligheten hos finansiella tillgångar, och beräknas baserat på den procentuella förändringen i indexpriset. Daglig volatilitet, σ , är standardavvikelsen för ett antal observationer, n , som i detta fall är 100, och skattas enligt

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

där x_i är observation i och \bar{x} är observationernas medelvärde.

3.9 Korskorrelation mellan index, VaR och avkastning

Jämförelse mellan rörelserna för OMXSPI, SX20PI och SX3010PI görs genom att beräkna likheten eller skillnaden, det vill säga positiv eller negativ korrelation, vid olika tidsfördröjningar. Beräkning av korskorrelation görs enligt ekvationer (15) och (16) med tidsfördröjningskonstanten $d = \pm 0-10$. Korrelationsvärdena representeras av staplar i en figur, där korrelationsgraden ses på y -axeln och tidsfördröjningen ses på x -axeln. Statistisk signifikans kan approximativt beräknas som

$$I_{korr} = \pm \frac{\Phi^{-1}(1 - \alpha)}{\sqrt{n}} \quad (22)$$

och illustreras i samma figur som återfinns i resultatavsnittet, där $1 - \alpha$ valts till 95%. Både OMXSPI, SX20PI och SX3010PI jämförs med varandra.

Vidare studeras korskorrelation mellan VaR och avkastning genom att beräkna månatligt VaR samt respektive månads avkastning. Det månatliga VaR fixeras över tid och korreleras med avkastningen på en tidsfördröjning på 0-20 dagar. Detta för att se om VaR kan förutse framtida avkastningar för de olika indexen. Den statistiska signifikansen beräknas genom ekvation (22).

3.10 Backtesting

I denna studie har VaR beräknats med hjälp av tre olika metoder. För att bedöma hur väl dessa tycks prestera med avseende på studiens syfte har backtesting på historisk data tillämpats.

Detta görs genom att beräkna dagligt VaR för en sexårsperiod, motsvarande 1512 datapunkter, och därefter jämföra med den negativa procentuella avkastningen för efterföljande dag. Om den negativa procentuella avkastningen är större än VaR för given konfidensgrad, $1 - \alpha$, ses detta som en överträdelse. Därefter flyttas intervallet ett steg framåt, och proceduren upprepas tills samtliga datapunkter behandlats.

Antalet faktiska överträdelser jämförs sedan med antalet förväntade överträdelser, vilken beräknas genom att multiplicera given konfidensgrad, $1 - \alpha = \{0, 95; 0, 99; 0, 999\}$, med antalet datapunkter i det intervall som utgörs av samtliga observationer exkluderat en sexårsperiod. Detta eftersom första VaR beräknas med data från en sexårsperiod. För ett antal observationer, n , beräknas det förväntade antalet överträdelser, x , för given konfidensgrad enligt $x = n\alpha$.

VaR för BM-, PoT-metoden och normalfördelningen bestäms enligt ekvation (17), (18) respektive (19). För BM beräknas ett nytt extremindex för varje intervall, samma sak gäller för tröskelvärde

i PoT, då dessa ej är konstanta. PoT-metoden backtestas med både avklustrad och icke-avklustrad data för att studera hur resultaten förhåller sig till varandra. Resultatet för backtesting presenteras i avsnitt 4.5.

4 Resultat

I detta avsnitt presenteras resultaten som erhållits under studien. Detta inkluderar nutida VaR, VaR över tid och vilka korskorrelationer som erhållits.

4.1 Nutida VaR och ES

Riskerna associerade till de olika indexen är genomgående högre för SX3010PI enligt tabell 5, följt av OMXSPI enligt tabell 3 och lägst för SX20PI enligt tabell 4. För OMXSPI blir BM-VaR cirka fyra gånger större när konfidensgraden går från 95% till 99,9%, samtidigt som SX3010PI och SX20PI uppvisar en tredubbel ökning av BM-VaR för samma konfidensgrader.

Tabell 3: VaR och ES beräknat för de olika metoderna med olika konfidensgrader för perioden 2006-2023

Riskmåtsresultat OMXSPI						
Metod	VaR (95%; 99%; 99,9%)			ES (95%; 99%; 99,9%)		
BM	2,03	3,84	7,84	3,23	5,55	10,7
PoT	1,58	3,60	6,50	2,84	4,86	7,75
Normal	2,13	3,03	4,04	2,68	3,48	4,40

Tabell 4: VaR och ES beräknat för de olika metoderna med olika konfidensgrader för perioden 2006-2023

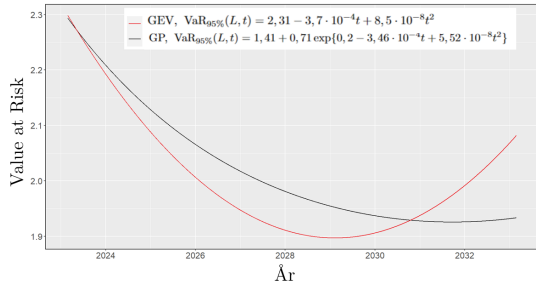
Riskmåtsresultat SX20PI						
Metod	VaR (95%; 99%; 99,9%)			ES (95%; 99%; 99,9%)		
BM	2,02	3,54	6,15	2,98	4,66	7,57
PoT	1,49	3,18	5,59	2,54	4,23	6,63
Normal	1,88	2,67	3,56	2,37	3,07	3,88

Tabell 5: VaR och ES beräknat för de olika metoderna med olika konfidensgrader för perioden 2006-2023

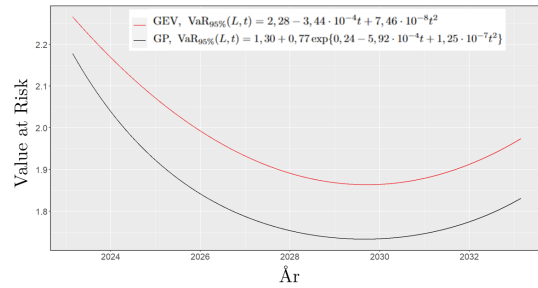
Riskmåtsresultat SX3010PI						
Metod	VaR (95%; 99%; 99,9%)			ES (95%; 99%; 99,9%)		
BM	2,66	4,99	10,3	4,22	7,26	14,2
PoT	2,06	4,76	8,83	3,75	6,52	10,7
Normal	2,95	4,18	5,56	3,70	4,79	6,06

4.2 VaR för trender i parametrarna

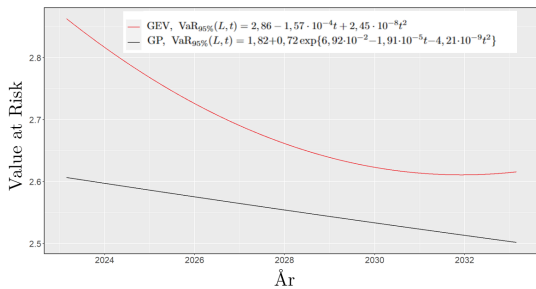
I figur 11-13 ses hur VaR förändras för perioden 2023-2032 när antagandet att parametrarna varierar över tid.



Figur 11: Trend på VaR för OMXSPI med icke stationärt μ för BM och σ för PoT



Figur 12: Trend på VaR för SX20PI med icke stationärt μ för BM och σ för PoT

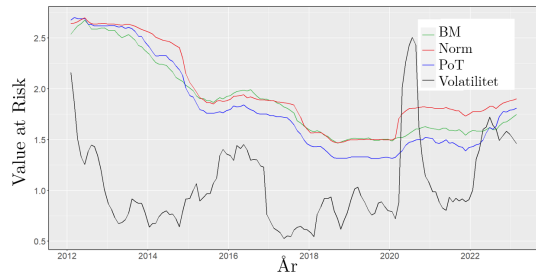


Figur 13: Trend på VaR för SX3010PI med icke stationärt μ för BM och σ för PoT

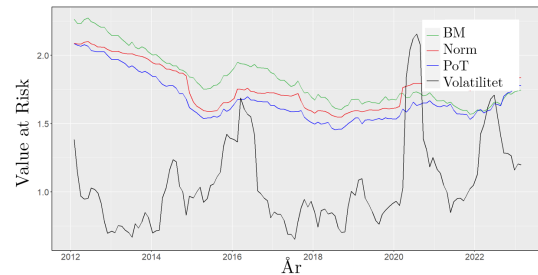
Trenden för OMXSPI och SX20PI är tydligare än för SX3010PI.

4.3 VaR över tid och volatilitet

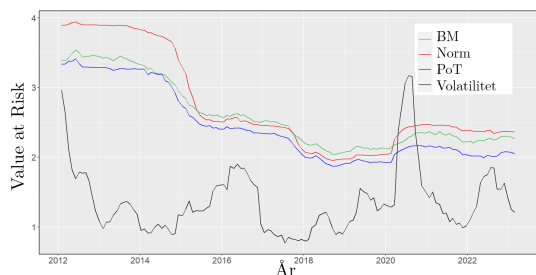
Nedan visas $\text{VaR}_{95\%}(L)$ över tid med trendlinje och trendlinjens tillhörande konfidensintervall på 95% för perioden 2012-2023. VaR och volatilitet baseras rullande på 6 år respektive 100 dagar.



Figur 14: VaR över tid och volatilitet för OMXSPI



Figur 15: VaR över tid och volatilitet för SX20PI

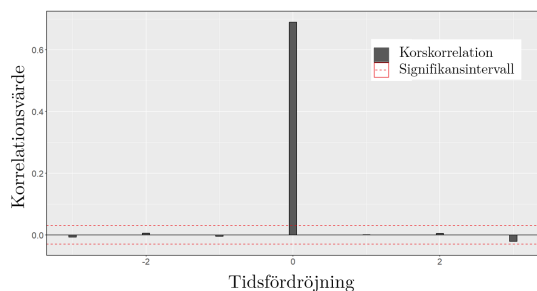


Figur 16: VaR över tid och volatilitet för SX3010PI

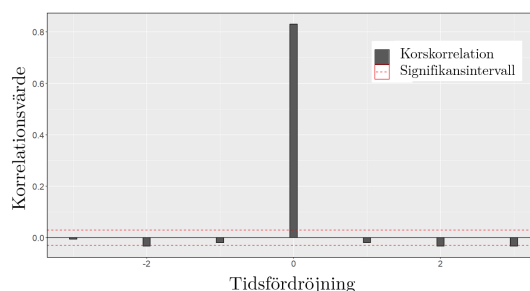
I figur 14-16 ses hur historiskt VaR över perioden 2012 till 2023. Här ses även hur volatiliteten ändras under samma tidsperiod. Volatiliteten ses reagera snabbare, medan VaR gör det något långsammare. Men överlag med liknande beteendemönster.

4.4 Korskorrelation mellan indexen samt VaR och avkastning

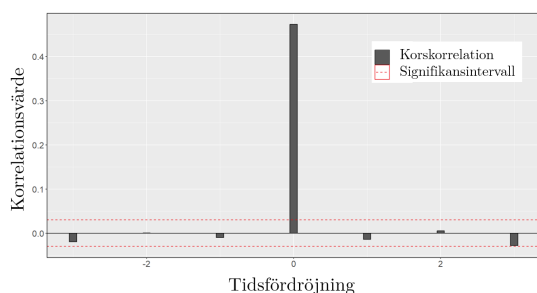
Uppvisade grader av korrelation mellan de olika indexen under perioden 2006-2023 vid olika tidsfördröjningar.



Figur 17: Korskorrelation mellan OMXSPI och SX20PI



Figur 18: Korskorrelation mellan OMXSPI och SX3010PI



Figur 19: Korskorrelation mellan SX20PI och SX3010PI

I figur 17-19 observeras en förväntad hög korrelation mellan de olika indexen vid tidsfördröjningen 0.

Angående resultatet för korskorrelationen mellan månatligt VaR och avkastningen återfanns ingen korrelation som var av statistisk signifikans.

4.5 Backtesting

I nedanstående tabell presenteras de resultat som erhålls vid backtestningen där data för rullande 6 år har använts. Förväntat antal överskridanden avser de tillfällen vid vilka förluster borde överskridit VaR med hänsyn till konfidensgraden.

Tabell 6: Antal överskridanden av dagligt VaR för de olika metoderna, respektive förväntat antal överskridanden. Data framtagen i enighet med avsnitt (3.10)

Index	Backtestingresultat								
	OMXSPI			SX20PI			SX3010PI		
Metod	95%	99%	99,9%	95%	99%	99,9%	95%	99%	99,9%
BM	147	24	2	124	28	1	100	24	3
PoT (avklustrad)	218	29	3	224	34	3	173	28	4
PoT (ej avklustrad)	140	21	3	144	31	2	107	25	4
Normal	137	50	20	134	60	22	94	37	21
Förväntat	140,2	28,0	2,8	140,2	28,0	2,8	140,2	28,0	2,8

5 Diskussion

I nedanstående avsnitt kommer diskussion kring metod- och resultat-delen av arbetet att diskuteras. Detta inkluderar de antaganden som gjorts under projektets gång och vad riskmåtsresultaten säger i relation till omvärldsfaktorer.

5.1 Metod

När data skulle avklustras gjordes detta genom Block-runs-metoden där räckvidden r valdes till fem dagar, det vill säga en börsvecka. För att validera att r valts korrekt utfördes ett test med hjälp av histogram, vilka visualiseras i figur 22-24 i appendix A.1. Av dessa figurer framgår att majoriteten av klustrerna har ett påföljande kluster inom en börsvecka, där de flesta bara ligger dagen efter. Med anledning av detta undersöktes möjligheten att sätta r till en börsmånad, 21 datapunkter. Utfallet blev återigen korta avstånd mellan kluster, och detta i kombination med att antalet överskridande datapunkter halverades innebar att r till slut valdes till fem dagar. Vid de största rörelserna är det alltså svårt att skapa ett oberoende i datan, även om r ökas.

Av figur 5-10 framgår att den empiriska datan tycks passa betydligt bättre till GEV-fördelningen samt GP-fördelningen, än till normalfördelningen. En nackdel med normalfördelningen är att den misslyckas med att modellera feta svansar och att de mest extrema händelserna således underskattas. Detta fenomen illustreras tydligt i qq-graferna, till de olika indexen, där de empiriska datapunkterna ständigt överskrider den skattade linjen i den högra svansen. Antagandet att normalfördelningen underskattar risken stärks också av tabell 3-5, där det framgår att både VaR och ES är betydligt lägre för denna fördelning än övriga två. Jämförs den empiriska datan mot fördelningarnas täthetsfunktioner, tycks både GEV- och GP-fördelningen vara bättre approximationer. Detta stärks också av de Anderson-Darling-test som gjordes, vilkas resultat visas i tabell 3.5. I dessa test förkastades normalfördelningen som en potentiell kandidat som fördelning, medan GEV- och GP-fördelningarna inte kunde förkastas för signifikansnivån $\alpha = 0,05$. Viktigt att belysa är att p -värdet, för en viss signifikansnivå, endast kan användas för att förkasta H_0 och inte bekräfta H_0 .

I tabell 4.5 framgår att normalfördelningen tycks prestera lika bra som de extremvärdesteoretiska fördelningarna på en 95%-ig konfidensnivå. Därav kan frågan lyftas om något som inträffar 5 gånger på 100 verkligen ska anses vara extremt, och således om ett 95%-igt VaR verkligen är ett lämpligt riskmått. Intressant är dock att normalfördelningen presterar bra fram till denna konfidensnivå, för att sedan kraftigt misslyckas med modelleringen. Därmed kan det antas att normalfördelningen inte lämpas att modellera stora risker.

5.2 Resultat

Målet vid val av olika index var att index med olika beteende för volatilitet och mönster skulle erhållas, för att riskminimera en investeringsportfölj. Med resultatet VaR över tid som presenteras i tabell 3-5 och graferna som illustrerar VaR över tid som ses i figur 14-16 följer de valda indexen de förväntningarna som ställdes. SX20PI, som är ett hälsovårdsindex, har lägst volatilitet och lägst nutida VaR och VaR över tid för de tre indexen. SX3010PI, som är ett bankindex, uppnår högst nivå av volatilitet av de tre valda indexen. Samma index hade högst nutida VaR och VaR över tid. OMXSPI inkluderar alla aktier på stockholmsbörsen och dess riskmått hamnade som förväntat mellan de andra två indexen.

Av tabell 3-5 framgår det att BM-metoden tycks generera ett betydligt högre värde på VaR och ES än PoT-metoden. Anledningen till detta kan vara att extremindexet θ , som används vid konvertering från block-VaR till dagligt VaR, överestimeras. I praktiken innebär det att estimatet av intensiteten mellan extrema händelser är kan ha varit för hög vilket överestimerar VaR. Av ekvation (17) att döma så kan VaR få stora svängningar med en liten förändring i θ .

Givet att förlusten överstiger VaR för given konfidensgrad är det intressant att undersöka hur stor förlusten kan förväntas bli, vilket beskrivs av ES. Av tabell 3-5 framgår att ES tycks vara

betydligt större för BM för samtliga index, jämfört med PoT och normalfördelningen. Detta är rimligt eftersom samtliga VaR för BM även är större än för resterande metoder. Det framgår också, i procentenheter mätt, att ökningen av ES mellan olika konfidsgrader är större för BM än de övriga två. En orsak till de kan vara att fördelningen av dagliga ES för BM uppvisar tunnare svansar.

För den som är intresserad över hur omvärldsfaktorer påverkar VaR är graferna över historisk VaR av stort intresse. Graferna 14-16 indikerar att riskmättet var som högst i början av det studerade tidsintervallet, då finanskrisen nyligen hade knuffat omkull hela världens börser. Riskens minskade sedan successivt hela vägen fram till tidigt 2020 då Covid-19 pandemin återigen skapade stora rörelser på aktiemarknaden, varpå VaR steg. Om Covid-kraschen anses uppfylla kraven för att definieras som en svart svan, inträffade denna utan att VaR uppvisade några indikationer på förhöjda risknivåer. Med detta perspektiv på krisen kan riskmättets fullständighet ifrågasättas.

En anledning till att VaR successivt avtog från 2014 fram till Covid-19 pandemin beror på att datainsamlingen för beräknade av historiskt VaR endast bestod av sex års data. Vid beräkandet av VaR 2014 och framåt minskade antalet datapunkter från den volatila perioden under finanskrisen, vilket i förlängningen medförde ett avtagande VaR. Liknande argument kan förvisso göras åt andra hållet, där en alltför omfattande datasamling kan leda till oönskad effekt från tidigare extremvärden och resultera i ett för högt VaR. På så sätt skulle finanskrisen förhöja VaR långt efter dess slut och resultera i att riskmättet svänger långsamt. Den påtagliga effekten som urvalet av data har på det slutgiltiga resultatet kräver konstanta avvägningar och utgör en svårighet vid användning av VaR.

Backtesting av VaR återfinns i tabell 4.5 och genererade flera resultat värda att lyfta fram. Om till en början endast OMXSPI betraktas, hamnade antalet överträdelser för den lägre konfidsgraden, $1 - \alpha = 0,95$, kring det förväntade antalet för samtliga fördelningar utom PoT med avklustrad data. Ju längre ut i svansen backtestingen utfördes, desto tydligare blev det att normalfördelningen tenderar att underskatta risken. Resterande fördelningar genererade precisa värden hela vägen ut till $1 - \alpha = 0,999$, vilket var den högsta konfidsgrad som testades i studien. Noterbart är även att avklustrad data jämfördes med icke-avklustrad data vid backtestingen och här konstaterades det att PoT-metoden presterade bättre med den icke-avklustrade datan vid den lägre konfidsgraden och att varianterna därefter genererar likvärdiga resultat för de högre konfidsgraderna.

Jämförs resultatet från backtestingen för de olika indexen, noteras att de olika metoderna tycks prestera olika bra beroende på vilket index som studeras. PoT med avklustrad data tycks, åtminstone för $1 - \alpha = 0,95$ och $1 - \alpha = 0,99$, prestera bättre på SX3010PI, där risken under studerad period varit högre. Detta i kontrast med övriga metoder som presterar sämre på detta index än övriga. På samma sätt verkar PoT utan avklustrad data prestera bättre än övriga metoder på SX20PI, vilket är det index som präglats av lägst risk.

En anledning till att antalet överträdelser blir så pass få för SX3010PI kan vara att de observerade värdena som modellen baserar sig på kommer från tidsperioden 2006-2012 där banksektorn kraschat. En följd av detta är att VaR för denna sektor är högre än för resterande index, och eftersom en överträdelse förekommer då den dagliga negativa procentuella utvecklingen överstiger dagligt VaR. Alltså överskattas risken för detta index. Åtminstone till en början där all data baseras på en osäker finansiell tid, där bankerna drabbades hårdast.

Resultaten presenterade i figur 11-13 och 17-18 har överraskande likheter. VaR över tid beräknar historiskt VaR för tidsperioden 2012-2023 och VaR för trender i parametrar använder icke-stationära parametrar för att beräkna VaR som en funktion av tid mellan 2023 och 2033. Dessa riskmåtsberäkningar visar en tydlig trend likt karakteristiken för en andragsgradsfunktion, med undantag för PoT-metoden för SX3010PI i figur 13. Även om trenderna i VaR över tid och VaR för trender i parametrarna är snarlika, så är det viktigt att poängtera att de sträcker sig över olika tidsperioder. VaR över tid säger att från och med 2020 kommer VaR att öka och det finns ingen tendens i att den skulle göra det motsatta. Däremot säger VaR för trender i parametrar att VaR

från och med 2023 kommer att minska ända fram till 2030, vilket motsäger trenden för VaR över tid. Denna motsägelse kan bero på glappet i data. Om en tidsseries historiska beteende förväntas upprepas, stärker resultat för framtida VaR denna teori.

Sätts resultaten från VaR över tid i relation till volatiliteten, ses tendenser till likartade rörelser. Att tendenserna inte alltid överensstämmer beror delvis på att VaR i detta fall baseras på 1512 datapunkter medan volatiliteten endast beror på 100, vilket kan tyckas motsägelsefullt när de ska jämföras. Anledningen är att volatiliteten är ett riskmått som ska beskriva en mer aktuell risk, och om fler datapunkter hade använts hade syftet med riskmättet urholkats. Detta medför också att volatiliteten reagerar snabbare på extrema händelser, vilket ses i början av år 2020 i figur 14-16, även om tendenserna är likartade, vilket också är det som syftas att visa. Vidare tar volatiliteten negativa såväl som positiva rörelser i beaktande, i kontrast till VaR. I tider av ihållande stora positiva kursrörelser, kommer volatiliteten öka medan VaR sannolikt inte påverkas av dessa rörelser. En ytterligare anledning till att antalet datapunkter inte är samma för de två riskmåten, är att MLE inte konvergerade för GEV-fördelningen då för få datapunkter användes.

Alla tre korskorrelationsfigurer 17-19 uppvisar korrelation vid tidsfördröjningen 0, vilket är rimligt då de påverkas av samma omvärldsfaktorer. Dessutom innehåller indexet OMXSPI alla aktier i både SX20PI och SX3010PI vilket bidrar till ökad tänkbarhet att de skulle vara korrelerade. Indexen SX20PI och SX3010PI korrelerar endast med 0,4 med tidsfördröjning 0. Detta innebär att sjukvård- och bank-branscherna inte korrelerar i den grad som OMXSPI gör med de andra två indexen, och tyder inte bara på att deras respektive rörelser är ett resultat av skilda omvärldsfaktorer utan också att deras reaktioner på en gemensam faktor uppvisar en något mer begränsad samstämmighet. Trots att indexen inte uppvisar något signifikant linjärt beroende vid tidsfördröjning skilt från 0, utesluter detta ej att det finns icke-linjära samband.

5.3 Slutsats

Sammantaget uppvisar studiens resultat att nuvarande VaR och ES indikerar att risken fortsatt är störst i banksektorn, medan OMXSPI och SX20PI uppvisar liknande värde på riskmåten där healthcare-indexet tenderar att vara något lägre. Historiska värden på VaR för de tre indexen var som högst i början av studiens intervall, i kölvattnet av finanskrisen, och föll sedan nedåt för att slutligen höjas vid Covid-kraschen till de nivåer som rädde vid studiens slut.

I mångt och mycket föll sig resultaten väl med den presenterade teorin. BM och PoT uppvisar genomgående en klart bättre förmåga än normalfördelningen att modellera extremvärden och generera empiriskt välanpassade riskmått. Skillnaderna påvisas i studien genom Anderson Darling-test, QQ-grafer och backtesting. Emellertid diskuteras även brister och begränsningar med riskmättet VaR, där bland annat urvalet av den insamlade datan har en direkt och stor påverkan på det slutgiltiga resultatet. Utöver dataurvalet har flertalet faktorer i metoden också en inverkan på det slutgiltiga resultatet. Val av tröskelvärde och blockbredd är två sådana faktorer, där studien avgränsades till att använda sig av mer subjektiva tillvägagångssätt och intuitivt sätta den senare till att motsvara en börsmanad.



Studien undersöker även korskorrelation mellan indexen, vilket inte gav några omvälvande resultat. En relativt låg korrelation mellan OMXSPI och de andra två indexen kunde påvisas, vilken möjligtvis förklaras genom deras gemensamma innehåll. Ingen korrelation mellan VaR och något av indexen kunde fastslås med statistisk signifikans.

Referenser

- [1] Nasdaq OMX Nordic. OMXSPI [Internet]. Webb sida: Nasdaq OMX Nordic; 2023 [uppdaterad 26/1-2023]. Hämtad från: https://www.nasdaqomxnordic.com/index/index_info?Instrument=SE0000744195
- [2] Taleb N. N. The Black Swan: The Impact of the Highly Improbable. Första upplagen: Random House; 2007
- [3] Rootzén H. Handledningsmöte. [Personlig intervju, 16/1] Göteborg; 2023 (opublicerad).
- [4] Coles S. An Introduction to Statistical Modelling of Extreme Values. Första upplagan. University of Bristol: Springer Verlag; 2001.
- [5] Dey D, Yan J. Extreme Value Modeling and Risk Analysis. Första upplagan. University of Connecticut: Taylor & Francis Group LLC; 2015.
- [6] Arnold BC. Pareto and generalized Pareto distributions. Modeling income distributions and Lorenz curves. 2008;119-45.
- [7] Gilleland E, Katz RW. extRemes 2.0: an extreme value analysis package in R. Journal of Statistical Software. 2016 Aug 30;72:1-39.
- [8] Myung J. Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology. 2003;47(1):90-100.
- [9] Blom G, Enger J, Englund G, Grandell J, Holst L. Sannolikhetsteori och statistiksteori med tillämpningar. 7:e uppl. Lund: Studentlitteratur; 2017
- [10] Engmann S, Cousineau D. Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. Journal of applied quantitative methods. 2011;6(3):1-17
- [11] Anderson TW. Anderson-Darling Tests of Goodness-of-Fit. | Miodrag L, redaktör. International Encyclopedia of Statistical Science. Berlin: Springer Berlin, Heidelberg; 2010. 52-4
- [12] Lauridsen S. Estimation of Value at Risk by Extreme Value Methods. Extremes. 2000;3(2):107-44.
- [13] Yamai Y, Yoshihara T. Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking & Finance. 2005;29(4):997-1015
- [14] Rootzén H. *Lecture 5*. [PowerPoint presentation]. Financial Risk, Chalmers University of Technology. [4/4-22; 7/2-23] Tillgänglig från: <https://chalmers.instructure.com/courses/18342>
- [15] Smith R, Weissman I. Estimating the Extremal Index. Journal of the Royal Statistical Society. 1994;56(3):512-28.
- [16] HSU HP. Probability, Random Variables, and Random Processes. 4:e uppl. New York: McGraw-Hill Education; 2019.
- [17] Broersen PMT. Automatic Autocorrelation and Spectral Analysis [Internet]. London: Springer London; 2006 [citerad 9 mars 2023]. Tillgänglig vid <https://link-springer-com.proxy.lib.chalmers.se/book/10.1007/1-84628-329-9>
- [18] Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. 3:e uppl. New York: Springer New York; 2016 [citerad 4 april 2023]. Hämtad från: https://link.springer.com/chapter/10.1007/0-387-21657-X_8
- [19] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 5:e uppl. New Jersey: John Wiley & Sons; 2015.

- [20] DuMouchel WH. Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique. *The Annals of Statistics*. 1983;11(4):1019-31.
- [21] Das BR, Resnick SI. QQ plots, random sets and data from a heavy tailed distribution. *Stochastic Models*. 2008 Feb 1;24(1):103-32.
- [22] Laio F. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*. 2004;40(9). doi:10.1029/2004WR003204
- [23] Jäntschi L, Bolboacă SD. Computation of Probability Associated with Anderson-Darling statistic. *Mathematics*. 2018;6(6):88. doi:10.3390/math6060088
- [24] D'Agostino RB, Stephens MA. *Goodness-of-Fit Techniques*. New York: Marcel-Dekker; 1986
- [25] MathWorks. adtest: Anderson-Darling test [Internet].
- [26] Calogero S. *A first course in Options Pricing Theory*. Göteborg: Chalmers University of Technology; 2021.

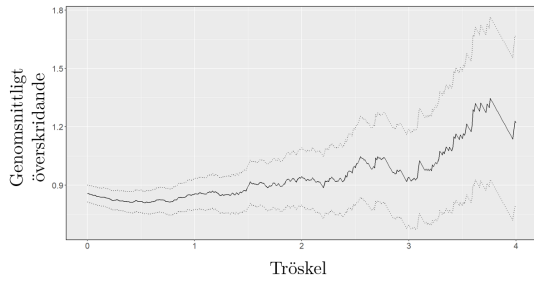
Appendix

Appendix A - Figurer
Appendix B - Tabeller
Appendix C - Kod i 
Appendix D - Kod i  MATLAB

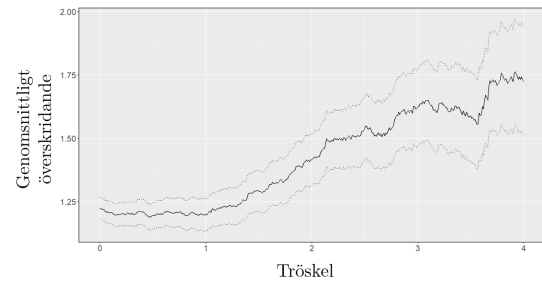
A Figurer

Här presenteras figurer som kompletterar huvuddelen av texten.

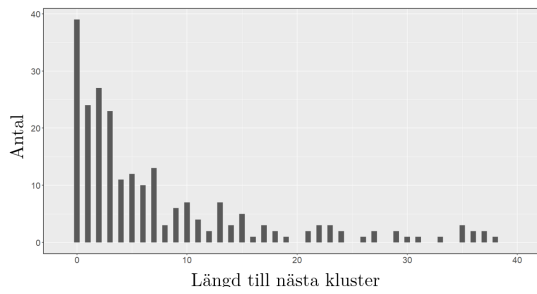
A.1 Val av tröskel



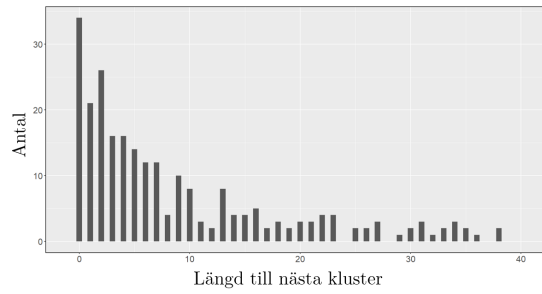
Figur 20: Mean residual life graf SX20PI med tillhörande konfidensintervall



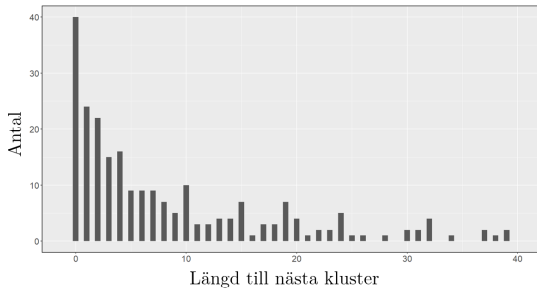
Figur 21: Mean residual life graf SX3010PI med tillhörande konfidensintervall



Figur 22: Histogram kluster OMXSPI

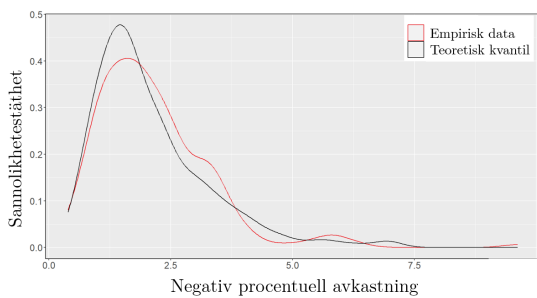


Figur 23: Histogram kluster SX20PI

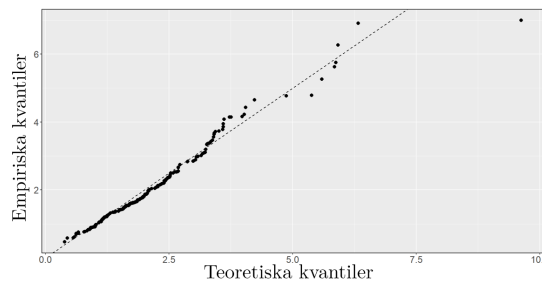


Figur 24: Histogram kluster SX3010PI

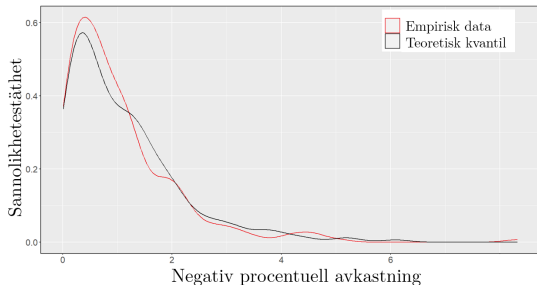
A.2 Densitets- QQ-graf



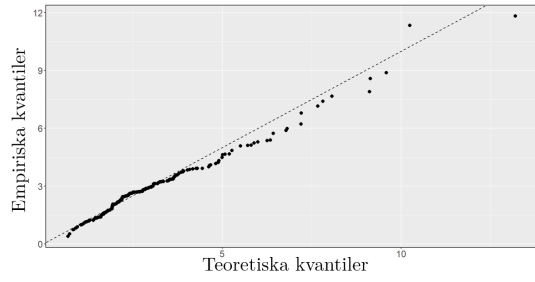
Figur 25: GEV täthetsjämförelse SX20PI



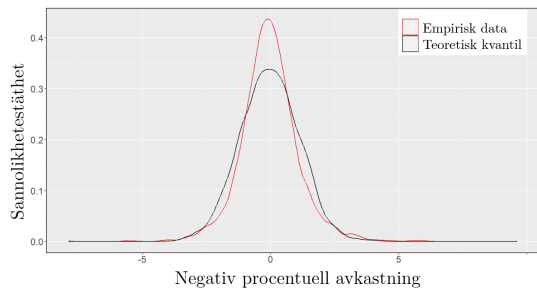
Figur 26: QQ-graf GEV SX20PI



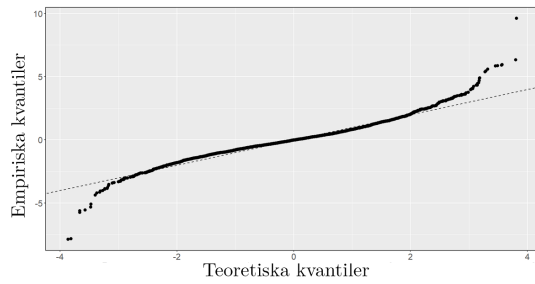
Figur 27: GP täthetsjämförelse SX20PI



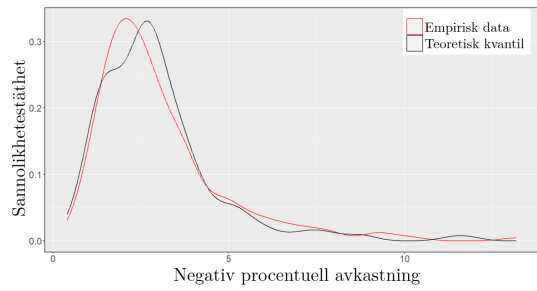
Figur 28: QQ-graf GP SX20PI



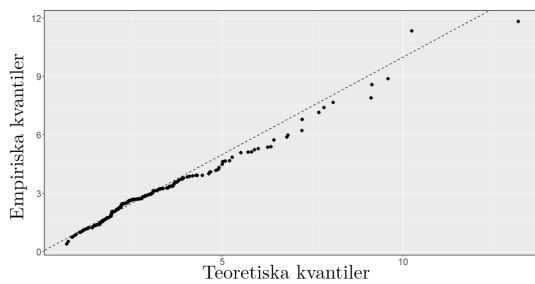
Figur 29: Normal täthetsjämförelse SX20PI



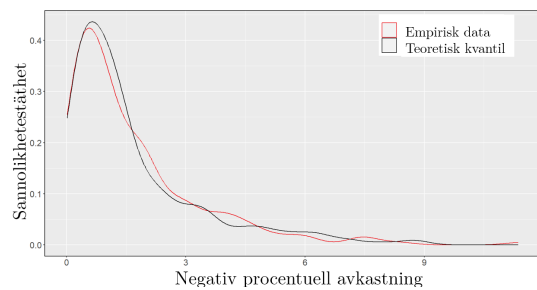
Figur 30: QQ-graf normal SX20PI



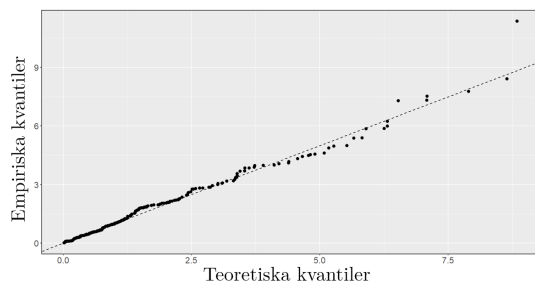
Figur 31: GEV täthetsjämförelse SX3010PI



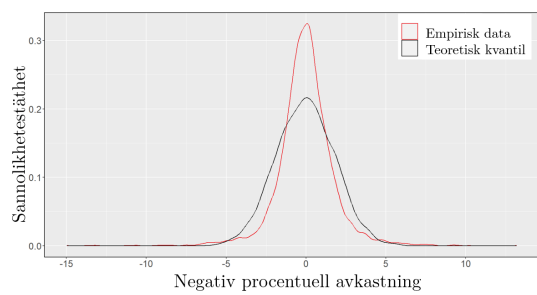
Figur 32: QQ-graf GEV SX3010PI



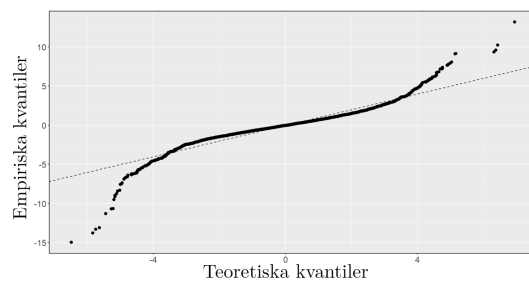
Figur 33: GP täthetsjämförelse SX3010PI



Figur 34: QQ-graf GP SX3010PI



Figur 35: Normal täthetsjämförelse SX3010PI







Figur 36: QQ-graf normal SX3010PI



B Tabeller



Här presenteras tabeller som kompletterar huvuddelen av texten.



B.1 Parameterskattningar från och



Tabell 7: Parameterskattningar för GEV-, GP- och normalfördelningen med  och  för OMXSPI

Parameterskattningar OMXSPI						
Programvara						
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	1,60	0,83	0,22	1,60	0,83	0,22
GP		1,26	-0,0011		1,26	-0,0011
Normal	-0,033	1,32		-0,033	1,32	

Tabell 8: Parameterskattningar för GEV-, GP- och normalfördelningen med  och  för SX20PI

Parameterskattningar SX20PI						
Programvara						
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	1,65	0,82	0,10	1,65	0,82	0,10
GP		1,06	-0,0040		1,06	-0,0040
Normal	-0,032	1,16		-0,032	1,16	

Tabell 9: Parameterskattningar för GEV-, GP- och normalfördelningen med  och  för SX3010PI

Parameterskattningar SX3010PI						
Programvara						
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	2,14	1,04	0,24	2,14	1,04	0,24
GP		1,64	0,026		1,64	0,026
Normal	-0,028	1,81		-0,028	1,81	

B.2 Konfidensintervall för parameterskattningar

Tabell 10: 95% konfidensintervall för parametrar i GEV, GP och normalfördelningen för OMXSPI

Konfidensintervall OMXSPI			
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	(1,47; 1,73)	(0,73; 0,94)	(0,091; 0,35)
GP		(1,06; 1,49)	(-0,11; 0,11)
Normal	(-0,072; 0,0066)	(1,29; 1,35)	

Tabell 11: 95% konfidensintervall för parametrar i GEV, GP och normalfördelningen för SX20PI

Konfidensintervall SX20PI			
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	(1,52; 1,78)	(0,73; 0,92)	(-0,0024; 0,20)
GP		(0,90; 1,24)	(-0,11; 0,10)
Normal	(-0,067; 0,0026)	(1,14; 1,19)	

Tabell 12: 95% konfidensintervall för parametrar i GEV, GP och normalfördelningen för SX3010PI

Konfidensintervall SX3010PI			
Fördelning	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
GEV	(1,97; 2,30)	(0,92; 1,19)	(0,11; 0,36)
GP		(1,37; 1,96)	(-0,10; 0,16)
Normal	(-0,082; 0,026)	(1,77; 1,85)	

B.3 Anderson-Darling-test

Tabell 13: Resultat från Anderson-Darling-test med signifikansnivå $\alpha = 0,05$, för SX3010PI

Resultat från Anderson-Darling-test				
Fördelning	Beslut	p -värde	Teststorhet	Kritiskt värde
GEV	0	0,99	0,19	2,5
GP	0	0,75	0,50	2,5
Normal	1	$< 5 \cdot 10^{-4}$	∞	0,75

Tabell 14: Resultat från Anderson-Darling-test med signifikansnivå $\alpha = 0,05$, för SX20PI

Resultat från Anderson-Darling-test				
Fördelning	Beslut	p -värde	Teststorhet	Kritiskt värde
GEV	0	0,98	0,24	2,5
GP	0	0,50	0,77	2,5
Normal	1	$< 5 \cdot 10^{-4}$	∞	0,75

C Kod i

Här presenteras koden som skrivits under arbetets gång och som ligger till grund för allt resultat. Notera att samma kod används till samma index med enda undantag att olika csv-filer läses. För VaR över tid återfinns endast koden för PoT-metoden då den är snarlik för resterande metoder.

C.1 Kod för nutida VaR och ES med tillhörande grafer

```
library("extRemes")
library("fExtremes")
library("ismev")
library("EnvStats")
library("ggplot2")
library("evd")
library("reshape2")
library("rlist")

#Decluster function
decluster_func <- function(x) {
  u = quantile(x, 0.9)
  r = 5
  decEx = list()
  i = 1
  while(i <= length(x)) {
    if (x[i] > u) {
      decEx = append(decEx, x[i])
      i=i+r
    } else {
      i = i+1
    }
  }
  return(decEx)
}

# Import .csv file
data <- read.csv("OMXSPI.csv", header = TRUE, sep = ";")
Close = data$Close
Date = data$Date
percentreturn <- -100*diff(Close)/Close[1:length(Close)-1]

mrl.plot(percentreturn, umin=0)

#Find monthly maxima
max.v <- blockMaxima(percentreturn, 21)

#Plot percentreturn
df_index = data.frame(
  date = as.Date(Date[1:length(Date)-1]),
  value = percentreturn
)

ggplot(df_index, aes(x=date, y=value))+
  geom_col(width =2.5) +
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))
```

```

#-----DECLUSTER OCH EXCEEDANCES-----

#Decluster
thr = as.numeric(quantile(percentreturn, probs = 0.9))

decEx= as.numeric(decluster_func(percentreturn))

#Find pu
exceedances = length(percentreturn[percentreturn>thr])
pu = exceedances/length(percentreturn)

#Approximate extremal index
numClust <- length(decEx)
theta = numClust/exceedances

#-----GEV FIT AND PLOT-----

#Fit to GEV
GEV <- gev.fit(max.v)
probs <- c(0.95, 0.99, 0.999)
VaR_gev_block <- GEV$mle[1]-GEV$mle[2]/GEV$mle[3] *
  (1-(-log(probs))^-GEV$mle[3])
VaR_gev_daily <- GEV$mle[1]-GEV$mle[2]/GEV$mle[3] *
  (1-(-log(probs^(theta*21))^-GEV$mle[3]))

#Plotting GEV
dens_GEV = rgev(length(max.v), loc = GEV$mle[1], scale = GEV$mle[2],
  shape = GEV$mle[3])
df_GEV <- data.frame(max.v, dens_GEV)
df_melted_GEV = melt(df_GEV)

#GEV qq plot
ggplot(mapping = aes(x = sort(max.v), y = sort(dens_GEV))) +
  geom_point(size = 2.5) +
  geom_abline(aes(slope = 1, intercept = 0), linetype = 2) +
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
    fill = NA,
    size = 0.8))

ggplot(df_melted_GEV, aes(value, color = variable)) +
  geom_density()+
  scale_color_manual(values = c("#eb0c13", "#000000"))+
  theme(legend.title = element_blank()+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
    fill = NA,
    size = 0.8))+
  theme(legend.position = c(0.8, 0.9), legend.key.size = unit(1, "cm"))

#-----GP FIT AND PLOT-----

#Fits the data to GP-distribution

```

```

GP <- gpd.fit(decEx, thr)
Var_gp <- thr + (GP$mle[1]/GP$mle[2])*(((1-probs)/pu)^(-GP$mle[2])-1)

#Plotting GP
dens_GP = thr + rgpd(length(decEx), scale = GP$mle[1], shape = GP$mle[2])

df_GP <- data.frame(decEx, dens_GP)
df_melted_GP = melt(df_GP)

#QQ-Plot
ggplot(mapping = aes(x = sort(decEx), y = sort(dens_GP))) +
  geom_point(size = 2.5) +
  geom_abline(aes(slope = 1, intercept = 0), linetype = 2) +
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))

#Density plot
ggplot(df_melted_GP, aes(value, color = variable)) +
  geom_density()+
  scale_color_manual(values = c("#eb0c13", "#000000"))+
  theme(legend.title = element_blank()+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+
  theme(legend.position = c(0.8, 0.9), legend.key.size = unit(1, "cm"))

#-----NORM FIT AND PLOT-----
sd <- sd(percentreturn)
mean <- mean(percentreturn)

#Plotting Gaussian
dens_norm = rnorm(length(percentreturn), mean = mean, sd = sd)
df_norm <- data.frame(percentreturn, dens_norm)

#Gaussian qq plot
ggplot(df_norm, aes(sample = percentreturn)) +
  geom_qq() +
  geom_qq_line()

ggplot(mapping = aes(x = sort(percentreturn), y = sort(dens_norm))) +
  geom_point(size = 2.5) +
  geom_abline(aes(slope = 1, intercept = 0), linetype = 2) +
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,

```

```

size = 0.8))

#Gaussian density
df_melted_norm = melt(df_norm)
p = ggplot(df_melted_norm, aes(x = value, color = variable)) +
  geom_density()+
  scale_color_manual(values = c("#eb0c13", "#000000"))+
  theme(legend.title = element_blank()+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+

  theme(legend.position = c(0.8, 0.9), legend.key.size = unit(1, "cm"))

p

#VaR
VaR_gs <- qnorm(probs, mean = mean, sd = sd)

#-----EXPECTED SHORTFALL-----

#Calculate ES for BM
GEV_integrand <- function(p) { # Create own function
  GEV$mle[1]-GEV$mle[2]/GEV$mle[3]*(1-(-log(p^(theta*21)))^(-GEV$mle[3]))
}

ES_gev <- numeric(length = length(probs))
for (i in 1:3) {
  ES_gev[i] = integrate(GEV_integrand, probs[i], 1)$value/(1-probs[i])
}

#Calculate ES for Peaks over Threshold
GP_integrand <- function(p) { # Create own function
  thr + (GP$mle[1]/GP$mle[2])*(((1-p)/pu)^(-GP$mle[2])-1)
}

ES_gp <- numeric(length = length(probs))
for (i in 1:3) {
  ES_gp[i] = integrate(GP_integrand, probs[i], 1)$value/(1-probs[i])
}

#Calculate ES for Gaussian
ES_norm = numeric(length=3)
for(i in 1:3){
  ES_norm[i] = integrate(qnorm, mean=mean, sd=sd, lower=probs[i],
                        upper=1)$value/(1-probs[i])
}

#-----

```

C.2 Kod för VaR över tid

```
library("extRemes")
```

```

library("fExtremes")
library("ismev")
library("EnvStats")
library("ggplot2")
library("evd")
library("reshape2")
library("ggpmisc")

decluster_func <- function(x) {
  u = quantile(x, 0.9)
  r = 5
  decEx = list()
  i = 1
  while(i <= length(x)) {
    if (x[i] > u) {
      decEx = append(decEx, x[i])
      i=i+r
    } else {
      i = i+1
    }
  }
  return(decEx)
}

#----- Function for calculating VaR with PoT -----
VaR_function_PoT <- function(x) {
  thr <- quantile(x,0.9)

  #Decluster
  dc <- as.numeric(decluster_func(x))

  #Fits the data to GP-distribution
  GP <- gpd.fit(dc, threshold = thr)

  #.dc = declustered. Function c creates a vector
  exceedances.dc <- c(dc[dc>thr])

  # Determine pu = N/n
  exceedances <- length(x[x>thr])
  pu <- exceedances/length(x)

  # Calculates Value at Risk
  VaR <- thr + (GP$mle[1]/GP$mle[2])*(((1-0.95)/pu)^(-GP$mle[2])-1)

  return(VaR)
}

# ----- Begin code -----

# Import .csv file
data_omx <- read.csv("OMXSPI.csv", header = TRUE, sep = ";")
data_sx20 <- read.csv("SX20PI.csv", header = TRUE, sep = ";")
data_sx3010 <- read.csv("SX3010PI.csv", header = TRUE, sep = ";")

Date <- data_omx$Date
Close_omx <- data_omx$Close
Close_sx20 <- data_sx20$Close
Close_sx3010 <- data_sx3010$Close
percentreturn_omx <- -100*diff(Close_omx)/Close_omx[1:(length(Close_omx)-1)]
percentreturn_sx20 <- -100*diff(Close_sx20)/Close_sx20[1:(length(Close_sx20)-1)]
percentreturn_sx3010 <- -100*diff(Close_sx3010)/Close_sx3010[1:

```

```

                                                                    (length(Close_sx3010)-1)]

percentreturn_new_omx <- list()
percentreturn_new_sx20 <- list()
percentreturn_new_sx3010 <- list()

datapoints <- 1512
step <- 21
iterations <- floor((length(percentreturn_omx)-datapoints)/step)
valueatrisk_PoT_omx <- numeric(iterations)
valueatrisk_PoT_sx20 <- numeric(iterations)
valueatrisk_PoT_sx3010 <- numeric(iterations)

for (j in 1:iterations) {
  percentreturn_new_omx <- percentreturn_omx[(j*step-(step-1)):
                                              (datapoints+j*step)]
  valueatrisk_PoT_omx[j] = VaR_function_PoT(percentreturn_new_omx)
}
for (j in 1:iterations) {
  percentreturn_new_sx20 <- percentreturn_sx20[(j*step-(step-1)):
                                                (datapoints+j*step)]
  valueatrisk_PoT_sx20[j] = VaR_function_PoT(percentreturn_new_sx20)
}
for (j in 1:iterations) {
  percentreturn_new_sx3010 <- percentreturn_sx3010[(j*step-(step-1)):
                                                    (datapoints+j*step)]
  valueatrisk_PoT_sx3010[j] = VaR_function_PoT(percentreturn_new_sx3010)
}

vardates <- character()
for (j in 1:iterations) {
  vardates[j] <- Date[datapoints+j*step]
}

trendlinje=numeric(iterations)
for (i in 1:iterations) {
  trendlinje[i] = (valueatrisk_PoT_omx[i] + valueatrisk_PoT_sx20[i] +
                  valueatrisk_PoT_sx3010[i])/3
}

df_PoT <- data.frame(
  value_omx = valueatrisk_PoT_omx,
  value_sx20 = valueatrisk_PoT_sx20,
  value_sx3010 = valueatrisk_PoT_sx3010,
  trendlinje = trendlinje,
  date = as.Date(vardates)
)

ggplot(df_PoT, aes(x=date))+
  geom_line(aes(x=date, y = value_omx, color = "omx")) +
  geom_line(aes(x=date, y = value_sx20, color = "sx20")) +
  geom_line(aes(x=date, y = value_sx3010, color = "sx3010")) +
  geom_smooth(aes(x=date, y = trendlinje, color = "trendlinje")) +
  scale_color_manual(values = c("#3CB548", "#eb0c13", "blue", "black"))+
  theme(panel.border = element_rect(color = "black",
                                    fill = NA,

```

```

                                size = 0.8))+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  theme(legend.title = element_blank()+
  theme(legend.position = c(0.8, 0.8), legend.key.size = unit(1, "cm"))

```

C.3 Kod för icke-stationära parametrar

```

library("extRemes")
library("fExtremes")
library("ismev")
library("EnvStats")
library("ggplot2")
library("evd")
library("reshape2")
library("rlist")

#Decluster function
decluster_func <- function(x) {
  u = quantile(x, 0.9)
  r = 5
  decEx = list()
  i = 1
  while(i <= length(x)) {
    if (x[i] > u) {
      decEx = append(decEx, x[i])
      i=i+r
    } else {
      i = i+1
    }
  }
  return(decEx)
}

# Import .csv file
data <- read.csv("SX3010PI.csv", header = TRUE, sep = ";")
Close = data$Close
Date = data$Date
percentreturn <- -100*diff(Close)/Close[1:length(Close)-1]

max.v <- blockMaxima(percentreturn, 21)
thr = as.numeric(quantile(percentreturn, probs = 0.9))
decEx= as.numeric(decluster_func(percentreturn))
exceedances = length(percentreturn[percentreturn>thr])
pu = exceedances/length(percentreturn)
numClust <- length(decEx)
theta = numClust/exceedances

df_gev = data.frame(val = max.v, index = seq(1,4305,21),
                    index2 = seq(1,4305,21)^2)
par_gev = fevd(x=val, data=df_gev, type = "GEV",
               location.fun = ~(index+index2))$results$par

a=floor(4305/length(decEx))*length(decEx)
b=floor(4305/length(decEx))

df_gp = data.frame(val = decEx, index = seq(1,a,b))

```

```

par_gp = fevd(x=val, data=df_gp, threshold = thr, type = "GP",
             scale.fun = ~index, use.phi = TRUE)$results$par

VaR_gev_daily <- function(mu0,mu1,mu2,sigma,xi,t) {
  return((mu0+mu1*t+mu2*t^2)-sigma/xi*(1-(-log(0.95^(theta*21)))^(-xi)))
}

VaR_gp_daily <- function(phi0,phi1,gamma,t) {
  return(thr + ((exp(phi0+phi1*t))/gamma)*(((1-0.95)/pu)^(-gamma)-1))
}

t=3650

date = seq(as.Date("2023/02/27"), by = "day", length.out = t)

y_gp = (VaR_gp_daily(par_gp[1], par_gp[2], par_gp[3], 1:t))
y_gev = VaR_gev_daily(par_gev[1], par_gev[2], par_gev[3], par_gev[4],
                     par_gev[5], 1:t)

df_gp_plot = data.frame(y=y_gp, x=date)
df_gev_plot = data.frame(y=y_gev, x=date)
df_both = data.frame(y1=y_gp,y2=y_gev, x=date)

ggplot(df_gp_plot, aes(x=x, y=y)) +
  geom_line()+
  theme(axis.title.y = element_blank()+
        theme(axis.title.x = element_blank()+
              theme(text = element_text(size = 20))+
              theme(panel.border = element_rect(color = "black",
                                                fill = NA,
                                                size = 0.8))
ggplot(df_gev_plot, aes(x=x, y=y)) +
  geom_line()+
  theme(axis.title.y = element_blank()+
        theme(axis.title.x = element_blank()+
              theme(text = element_text(size = 20))+
              theme(panel.border = element_rect(color = "black",
                                                fill = NA,
                                                size = 0.8))

```

C.4 Kod för korskorrelation

C.4.1 Mellan index

```

library("extRemes")
library("fExtremes")
library("ismev")
library("EnvStats")
library("ggplot2")
library("evd")
library("reshape2")

VaR_function_PoT <- function(x) {

```

```

thr <- quantile(x,0.9)

#Decluster
dc <- decluster(x,threshold = thr)

#Fits the data to GP-distribution
GP <- gpd.fit(dc, threshold = thr)

#.dc = declustered. Function c creates a vector
exceedances.dc <- c(dc[dc>thr])

# Determine pu = N/n
exceedances <- length(x[x>thr])
pu <- exceedances/length(x)

# Calculates Value at Risk
VaR <- thr + (GP$mle[1]/GP$mle[2])*(((1-0.95)/pu)^(-GP$mle[2])-1)

return(VaR)
}

# Import .csv file
data_omx <- read.csv("OMXSPI.csv", header = TRUE, sep = ";")
data_sx20 <- read.csv("SX20PI.csv", header = TRUE, sep = ";")
data_sx3010 <- read.csv("SX3010PI.csv", header = TRUE, sep = ";")

Date_omx <- data_omx$Date
Date_sx20 <- data_sx20$Date
Date_sx3010 <- data_sx3010$Date
Close_omx <- data_omx$Close
Close_sx20 <- data_sx20$Close
Close_sx3010 <- data_sx3010$Close
percentreturn_omx <- -100*diff(Close_omx)/Close_omx[1:(length(Close_omx)-1)]
percentreturn_sx20 <- -100*diff(Close_sx20)/Close_sx20[1:(length(Close_sx20)-1)]
percentreturn_sx3010 <- -100*diff(Close_sx3010)/Close_sx3010[1:
                                (length(Close_sx3010)-1)]

percentreturn_new_omx <- list()
percentreturn_new_sx20 <- list()
percentreturn_new_sx3010 <- list()

crosscorr_omx20 = ccf(percentreturn_omx, percentreturn_sx20, lag.max = 10)
lag_omx20 = as.numeric(crosscorr_omx20$lag)
ccf_omx20 = as.numeric(crosscorr_omx20$acf)
?ccf
ccf(percentreturn_omx, percentreturn_omx, lag.max = 10)

crosscorr_omx3010 = ccf(percentreturn_omx, percentreturn_sx3010, lag.max = 10)
lag_omx3010 = as.numeric(crosscorr_omx3010$lag)
ccf_omx3010 = as.numeric(crosscorr_omx3010$acf)

crosscorr_sx2030 = ccf(percentreturn_sx20, percentreturn_sx3010, lag.max = 10)
lag_sx2030 = as.numeric(crosscorr_sx2030$lag)
ccf_sx2030 = as.numeric(crosscorr_sx2030$acf)

sig=1.96/sqrt(length(percentreturn_omx))

df_omx20 = data.frame(lag_omx20, ccf_omx20)
df_omx3010 = data.frame(lag_omx3010, ccf_omx3010)

```

```

df_sx2030 = data.frame(lag_sx2030, ccf_sx2030)

ggplot(df_omx20, aes(x=lag_omx20, y=ccf_omx20))+
  geom_col(width = 0.2, aes(color = "corr"))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+

  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  geom_hline(aes(yintercept = sig, color = "sig"), linetype = "dashed")+
  geom_hline(aes(yintercept = -sig, color = "sig"), linetype = "dashed")+
  geom_hline(yintercept = 0)+
  scale_color_manual(values = c("black", "#eb0c13"))+
  theme(legend.title = element_blank()+
  theme(legend.position = c(0.7, 0.8), legend.key.size = unit(1, "cm"))+
  guides(color = guide_legend(override.aes = list(fill = "white")))

ggplot(df_omx3010, aes(x=lag_omx3010, y=ccf_omx3010))+
  geom_col(width = 0.2, aes(color = "corr"))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+

  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  geom_hline(aes(yintercept = sig, color = "sig"), linetype = "dashed")+
  geom_hline(aes(yintercept = -sig, color = "sig"), linetype = "dashed")+
  geom_hline(yintercept = 0)+
  scale_color_manual(values = c("black", "#eb0c13"))+
  theme(legend.title = element_blank()+
  theme(legend.position = c(0.7, 0.8), legend.key.size = unit(1, "cm"))+
  guides(color = guide_legend(override.aes = list(fill = "white")))

ggplot(df_sx2030, aes(x=lag_sx2030, y=ccf_sx2030))+
  geom_col(width = 0.2, aes(color = "corr"))+
  theme(panel.border = element_rect(color = "black",
                                     fill = NA,
                                     size = 0.8))+

  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(text = element_text(size = 20))+
  geom_hline(aes(yintercept = sig, color = "sig"), linetype = "dashed")+
  geom_hline(aes(yintercept = -sig, color = "sig"), linetype = "dashed")+
  geom_hline(yintercept = 0)+
  scale_color_manual(values = c("black", "#eb0c13"))+
  theme(legend.title = element_blank()+
  theme(legend.position = c(0.7, 0.8), legend.key.size = unit(1, "cm"))+
  guides(color = guide_legend(override.aes = list(fill = "white")))

```

C.4.2 Mellan VaR och pris

```

library("extRemes")
library("fExtremes")
library("ismev")
library("EnvStats")
library("ggplot2")
library("evd")
library("reshape2")
library("ggpmisc")

```

```

#----- Function for calculating VaR with PoT -----

VaR_function_BM <- function(x) {
  max.v <- blockMaxima(x,21)

  GEV <- gev.fit(max.v)

  # daily VaR
  VaR <- GEV$mle[1]-GEV$mle[2]/GEV$mle[3]*(1-(-log(0.95))^( -GEV$mle[3]))

  return(VaR)
}

# ----- Begin code -----

# Import .csv file
data_omx <- read.csv("OMXSPI.csv", header = TRUE, sep = ";")
data_sx20 <- read.csv("SX20PI.csv", header = TRUE, sep = ";")
data_sx3010 <- read.csv("SX3010PI.csv", header = TRUE, sep = ";")

Date <- data_omx$Date
Close_omx <- data_omx$Close
Close_sx20 <- data_sx20$Close
Close_sx3010 <- data_sx3010$Close
percentreturn_omx <- -100*diff(Close_omx)/Close_omx[1:(length(Close_omx)-1)]
percentreturn_sx20 <- -100*diff(Close_sx20)/Close_sx20[1:(length(Close_sx20)-1)]
percentreturn_sx3010 <- -100*diff(Close_sx3010)/Close_sx3010[1:
                                     (length(Close_sx3010)-1)]

percentreturn_new_omx <- list()
percentreturn_new_sx20 <- list()
percentreturn_new_sx3010 <- list()

datapoints <- 1512
step <- 21
iterations <- floor((length(percentreturn_omx)-datapoints)/step)
valueatrisk_BM_omx <- numeric(iterations)
valueatrisk_BM_sx20 <- numeric(iterations)
valueatrisk_BM_sx3010 <- numeric(iterations)

for (j in 1:iterations) {
  percentreturn_new_omx <- percentreturn_omx[(j*step-(step-1)):
                                             (datapoints+j*step)]
  valueatrisk_BM_omx[j] = VaR_function_BM(percentreturn_new_omx)
}
for (j in 1:iterations) {
  percentreturn_new_sx20 <- percentreturn_sx20[(j*step-(step-1)):
                                                (datapoints+j*step)]
  valueatrisk_BM_sx20[j] = VaR_function_BM(percentreturn_new_sx20)
}
for (j in 1:iterations) {
  percentreturn_new_sx3010 <- percentreturn_sx3010[(j*step-(step-1)):
                                                    (datapoints+j*step)]
  valueatrisk_BM_sx3010[j] = VaR_function_BM(percentreturn_new_sx3010)
}



```

```
percentreturn_ccf_omx <- numeric(iterations)

for (j in 1:iterations) {
  percentreturn_ccf_omx[j] = (percentreturn_omx[(datapoints+j*step - step)]-
    percentreturn_omx[(datapoints+j*step)]) / percentreturn_omx[(datapoints+j
      *step - step)]
}

ccf( valueatrisk_BM_omx, percentreturn_ccf_omx)
```

D Kod i MATLAB

Här presenteras koden som skrivits under arbetets gång för att säkerställa att de parametrar som skattats i  också blev detsamma i . Här presenteras även koden som skrivits för att ta fram ett histogram för överskridanden gentemot antal dagar vid block-runs metoden, samt kod för att backtesta de olika metoderna och Anderson-Darling test. I följande kod behandlas endast OMXSPI, men kod för hantering av SX3010PI och SX20PI är identisk med undantag för den CSV-fil som används.

D.1 Huvudfil

Här presenteras den fil som utgör grunden för -koden.

```
% Importera data:
load('DatesAndPricesOMXSPI.csv');
load('DatesAndPricesSX20PI.csv');
load('DatesAndPricesSX3010PI.csv');

% Spara data i variabler:
indexDate = DatesAndPricesOMXSPI(:,1:3); % Datum i matrisformat.
%indexDate = DatesAndPricesSX20PI(:,1:3);
%indexDate = DatesAndPricesSX3010PI(:,1:3);
indexPrice = DatesAndPricesOMXSPI(:,4); % Indexpris i vektorformat.
%indexPrice = DatesAndPricesSX20PI(:,4);
%indexPrice = DatesAndPricesSX3010PI(:,4);

% Skapa graf över indexets prisutveckling:
xData = linspace(0,length(indexPrice)-1,length(indexPrice));
figure
plot(xData,indexPrice)
title('Historiskt pris i indexet')
xlabel('Borsdagar sedan 2006-01-01')
ylabel('Indexpris')

% Beräkna och skapa graf över dagliga formluster i indexet:
returns = zeros((length(indexPrice)),1);
for i=1:length(indexPrice)-1
    returns(i) = -(indexPrice(i+1)-indexPrice(i))/indexPrice(i)*100;
    % Beräknar de dagliga formlusterna uttryckt i %.
end
xData = linspace(0,length(indexPrice)-1,length(indexPrice));
figure
plot(xData,returns)
title('Dagliga formluster')
xlabel('Borsdagar sedan 2006-01-01')
ylabel('Formluster')
hold on
plot([0 length(indexPrice)],[0 0],'k') % Skapa en stodlinje vid 0.
hold off

% Avklustring:
u = quantile(returns,0.9); % Troskelvärdet placerat vid den
% 90:e percentilen.
r = 5; % Langden av det glidande blocket i block-runs metoden.
[decReturns,nrClusters] = Declustering(returns,u,r); % Information
% i funktionsfilen "Declustering".

% Histogram över langden mellan kluster:
HistogramMaker(returns,u,r); % Information i funktionsfilen
% "HistogramMaker".
```

```

% Normalfordelning:
[Normmu, Normsigma, NormmuCI, NormsigmaCI] = normfit(returns, 0.05);
% Parameterskattning till normalfordelningen, inklusive konfidensintervall.
alpha = [0.05 0.01 0.001];
NormVaR = [];
NormES = [];
VaR1 = @(x) Normmu+Normsigma.*norminv(x); % Anonym funktion
for att bestamma VaR for normalfordelningen.
for j=1:3
    NormVaR(j) = Normmu+Normsigma.*norminv(1-alpha(j)); % Bestammer VaR
    % for 3 konfidensgrader.
    NormES(j) = 1./alpha(j).*integral(VaR1, 1-alpha(j), 1); % Bestammer
    % ES for 3 konfidensgrader.
end
qqplot(returns, makedist('Normal', 'mu', Normmu, 'sigma', Normsigma)) % qq-plot
% for godhetstest.

% Peaks over Threshold:
figure
plot(0:1:length(decReturns)-1, decReturns, 'k.', "MarkerSize", 3)
hold on;
plot([0 length(0:1:length(decReturns)-1)], [u u], 'r')
excesses = PeaksOverThreshold(decReturns, u); % Information i
% funktionsfilen "PeaksOverThreshold".
excessesNew = [];
counter = 0;
for i = 1:length(excesses)
    if excesses(i) > u
        plot(i, excesses(i), 'b.', "MarkerSize", 5)
        excessesNew(end+1) = excesses(i)-u;
        counter = counter + 1;
    elseif excesses(i) == u
        plot(i, excesses(i), 'g.', "MarkerSize", 5)
    end
end
end
hold off;
[parmhatPoT, parmciPoT] = gpfit(excessesNew); % Parameterskattning till
% GP-fordelningen, inklusive konfidensintervall.
PoTgamma = parmhatPoT(1);
PoTgammaCI = parmciPoT(:, 1);
PoTsigma = parmhatPoT(2);
PoTsigmaCI = parmciPoT(:, 2);
pu = counter/length(returns); % Tar fram pu-koefficienten, alltsa antalet
% overskridande delat pa antalet matpunkter.
alpha = [0.05 0.01 0.001];
PoTVaR = [];
PoTES = [];
VaR2 = @(x) u+PoTsigma./PoTgamma.*(((1-x)/pu).^(-PoTgamma)-1);
% Anonym funktion for att bestamma VaR for GP-fordelningen.
for j = 1:3
    PoTVaR(j) = u+PoTsigma./PoTgamma*((alpha(j)/pu).^(-PoTgamma)-1);
    % Bestammer VaR for 3 konfidensgrader.
    PoTES(j) = 1./alpha(j).*integral(VaR2, 1-alpha(j), 1); % Bestammer ES
    % for 3 konfidensgrader.
end
end
qqplot(excessesNew,
makedist('GeneralizedPareto', 'k', PoTgamma, 'sigma', PoTsigma, 'theta', 0))
% qq-plot for godhetstest.

% Block Maxima:
figure

```

```

plot(xData,returns,'.k')
hold on;
b = 21; % Blocklangd.
blockValues = [];
maxValues = [];
maxIndex = [];
for i=1:b:length(returns)-b
    plot([i+b-1 i+b-1],[-15 15],'b')
    for j = 1:1:b
        blockValues(j) = returns(i+j-1);
    end
    [maxValues(end+1),maxIndex(end+1)] = max(blockValues); % Maxvardet i
    % varje block laggs in i listan.
    xvals = i+maxIndex(end);
    plot(xvals,maxValues(end),'rs')
end
hold off;
[parmhatBM,parmciBM] = gevfit(maxValues); % Parameterskattning till
% GEV-fordelningen, inklusive konfidensintervall.
BMgamma = parmhatBM(1);
BMgammaCI = parmciBM(:,1);
BMsigma = parmhatBM(2);
BMsigmaCI = parmciBM(:,2);
BMmu = parmhatBM(3);
BMmuCI = parmciBM(:,3);
theta = nrClusters/(length(returns)*0.1); % Bestammer theta-varDET, alltsa
% antalet kluster delat pa antalet overskridande.
alpha = [0.05 0.01 0.001];
BMVaR = [];
BMES = [];
VaR3 = @(x) BMmu-BMsigma./BMgamma.*(1-(-log(x.^(theta*b))).^(-BMgamma));
% Anonym funktion for att bestamma VaR for GEV-fordelningen.
for j = 1:3
    BMVaR(j) = BMmu-BMsigma./BMgamma*
    (1-(-log((1-alpha(j)).^(theta*b))).^(-BMgamma));
    % Bestammer VaR for 3 konfidensgrader.
    BMES(j) = 1./alpha(j).*integral(VaR3,1-alpha(j),1); % Bestammer ES
    % for 3 konfidensgrader.
end
qqplot(maxValues,
makedist('GeneralizedExtremeValue','k',BMgamma,'sigma',BMsigma,'mu',BMmu))
% qq-plot for godhetstest.

% Anderson-Darling test:
% Normality:
[hNorm,pNorm,adstatNorm,cvNorm] = adtest(returns);
% Peaks over Threshold:
dist2 = makedist('GeneralizedPareto',
'k',PoTgamma,'sigma',PoTsigma,'theta',0);
[hPoT,pPoT,adstatPoT,cvPoT] = adtest(excessesNew,'Distribution',dist2);
% Block Maxima:
dist3 = makedist('GeneralizedExtremeValue',
'k',BMgamma,'sigma',BMsigma,'mu',BMmu);
[hBM,pBM,adstatBM,cvBM] = adtest(maxValues,'Distribution',dist3);

% Backtesting:
startpoint = 1500;
alpha = [0.05 0.01 0.001];
% Normality:
[NormViolations,expectedViolations] =
BacktestingNormality(startpoint,alpha,returns);

```

```

% Information i funktionsfilen "BacktestingNormality".
% PoT:
PoTBacktest = BacktestingPoT(startpoint,alpha,returns); % Information i
% funktionsfilen "BacktestingPoT".
PoTBacktestDecluster = BacktestingPoTDecluster(startpoint,alpha,returns,r);
% Information i funktionsfilen "BacktestingPoTDecluster".
% BM:
BMViolations = BacktestingBM(startpoint,alpha,returns,b,r); % Information i
% funktionsfilen "BacktestingBM".

```

D.2 Funktionsfiler

Här presenteras funktionsfilerna som skrivits separat och genomgående använts under arbetets gång.

D.2.1 Avklustring

```

% Denna funtionsfil avklustrar data enligt block-runs metoden.
function [decReturns,nrClusters] = Declustering(returns,u,r)
tempList = returns;
decReturns = [];
nrClusters = 0;
i = 1;
while i <= length(tempList)
    if tempList(i) > u
        nrClusters = nrClusters+1;
        clustervals = zeros(1,min(r,max(1,length(tempList)-i)));
        for j = 1:min(r,length(tempList)-i)
            clustervals(j) = tempList(i+j-1);
        end
        clustermax = max(clustervals);
        for k = 1:min(r,max(1,length(tempList)-i))
            if tempList(i+k-1) ~= clustermax
                decReturns(i+k-1) = u;
            else
                decReturns(i+k-1) = clustermax;
            end
        end
        i = i+min(r,max(1,length(tempList)-i));
    else
        decReturns(i) = tempList(i);
        i = i+1;
    end
end
end

```

D.2.2 Histogramskapare

```

% Denna funtionsfil skapar ett histogram dar staplarna representerar
% langden fran slutet av ett kluster till borjan av nasta.
function [distances,staples,counter] = HistogramMaker(returns,u,r)
distances = [];
i = 1;
counter = 0;
while i < length(returns)
    if returns(i) > u
        j = i+r;
        while returns(j) <= u && j < length(returns)
            j = j+1;
        end
        counter = counter+1;
        distances(counter) = returns(i);
        staples(counter) = returns(j);
        i = j;
    end
end

```

```

        distances(end+1) = j-(i+r)+1;
        i = i+r;
    else
        i = i+1;
    end
end
end
staples = 0.5:max(distances)-0.5;
histogram(distances,staples)

```

D.2.3 Peaks over Threshold

```

% Denna funtionsfil identifierar forluster som overskrider den bestamda
% troskeln, u, och laggar dessa i en vektor.
function excesses = PeaksOverThreshold(returns,u)
excesses = zeros(1,length(returns));
for i=1:length(returns)
    if returns(i) > u
        excesses(i) = returns(i);
    end
end
end

```

D.2.4 Backtesting normalfördelning

```

% Denna funtionsfil används for att backtesta normalfordelningsmetoden,
% samt for att bestamma det forvantade antalet overtradelser.
function [NormViolations,expectedViolations,VaR] =
BacktestingNormality(startpoint,alpha,returns)
VaR = zeros(1,length(returns)-startpoint);
NormViolations = [];
expectedViolations = [];
for j = 1:3
    expectedViolations(end+1) =
round(alpha(j)*(length(returns)-startpoint),1);
violations = 0;
for i = 1:length(returns)-startpoint
    temp = zeros(0);
    temp(1:startpoint) = returns(i:startpoint+i-1);
    [NormmuB,NormalsigmaB] = normfit(temp);
    VaR(i) = NormmuB+NormalsigmaB*norminv(1-alpha(j));
    if VaR(i) < returns(i+startpoint)
        violations = violations + 1;
    end
end
end
NormViolations(end+1) = violations;
end

```

D.2.5 Backtesting ej avklustrad PoT

```

% Denna funtionsfil används for att backtesta PoT-metoden, utan att datan
% ar avklustrad
function [PoTViolations,VaR] = BacktestingPoT(startpoint,alpha,returns)
VaR = zeros(1,length(returns)-startpoint);
PoTViolations = [];
for j = 1:3
    violations = 0;
    for i = 1:length(returns)-startpoint
        u = quantile(returns(i:i+startpoint-1),0.9);
        temp = zeros(0);

```

```

temp = returns(i:i+startpoint-1);
newtemp = zeros(0);
counter = 0;
for k = 1:length(temp)
    if temp(k) > u
        newtemp(end+1) = temp(k)-u;
        counter = counter + 1;
    end
end
parmhatPoTB = gpfitt(newtemp);
PoTgammaB = parmhatPoTB(1);
PoTsigmaB = parmhatPoTB(2);
pu = counter/startpoint;
VaR(i) = u+PoTsigmaB/PoTgammaB*((alpha(j)/pu)^(-PoTgammaB)-1);
if VaR(i) < returns(i+startpoint)
    violations = violations + 1;
end
end
PoTViolations(end+1) = violations;
end

```

D.2.6 Backtesting avklustrad PoT

```

% Denna funtionsfil används för att backtesta PoT-metoden, med avklustrad
% data.
function [PoTViolations, VaR] =
BacktestingPoTDecluster(startpoint, alpha, returns, r)
VaR = zeros(1, length(returns)-startpoint);
PoTViolations = [];
for j = 1:3
    violations = 0;
    for i = 1:length(returns)-startpoint
        u = quantile(returns(i:i+startpoint-1), 0.9);
        [decReturnsB, ~] = Declustering(returns(i:i+startpoint-1), u, r);
        temp = zeros(1, startpoint);
        temp = decReturnsB(1:startpoint);
        newtemp = zeros(0);
        counter = 0;
        for k = 1:length(temp)
            if temp(k) > u
                newtemp(end+1) = temp(k)-u;
                counter = counter + 1;
            end
        end
        parmhatPoTB = gpfitt(newtemp);
        PoTgammaB = parmhatPoTB(1);
        PoTsigmaB = parmhatPoTB(2);
        pu = counter/startpoint;
        VaR(i) = u+PoTsigmaB/PoTgammaB*((alpha(j)/pu)^(-PoTgammaB)-1);
        if VaR(i) < returns(i+startpoint)
            violations = violations + 1;
        end
    end
    PoTViolations(end+1) = violations;
end
end

```

D.2.7 Backtesting BM

```

% Denna funtionsfil används för att backtesta BM-metoden.

```

```

function [BMViolations, VaR] = BacktestingBM(startpoint, alpha, returns, b, r)
VaR = zeros(1, length(returns)-startpoint);
BMViolations = [];
theta = [];
for j = 1:3
    violations = 0;
    for i = 1:length(returns)-startpoint
        maxValues1 = zeros(0);
        for k = i:b:i+startpoint-b
            blockValues1 = zeros(0);
            for m = 1:b
                blockValues1(end+1) = returns(k+m-1);
            end
            maxValues1(end+1) = max(blockValues1);
        end
        u = quantile(returns(i:startpoint+i-1), 0.9);
        [~, nrClusters] = Declustering(returns(i:startpoint+i-1), u, r);
        theta(i) = nrClusters / (startpoint * 0.1);
        [parmhatBM, ~] = gevfit(maxValues1);
        BMgammaB = parmhatBM(1);
        BMsigmaB = parmhatBM(2);
        BMmuB = parmhatBM(3);
        VaR(i) = BMmuB - BMsigmaB / BMgammaB *
            (1 - (-log((1 - alpha(j))^(theta(i) * b))))^(-BMgammaB));
        if VaR(i) < returns(i+startpoint)
            violations = violations + 1;
        end
    end
    BMViolations(end+1) = violations;
end
end

```