



Bachelorprojekt

William B. W. Lin (lgt226)

Risikostyring: Backtesting expected shortfall

Vejleder: Nina Lange

Institut for Matematiske Fag. 6. juni 2025

Resumé

Denne afhandling analyserer forskellige teststørrelser til backtesting af risikomålet expected shortfall (ES), som er blevet standard af Basel-komiteén for finansielle institutioner. ES er et afledt mål for det andet risikomål VaR og besidder pænere matematiske egenskaber. Vores mål er, at analysere ES teststørrelsernes robusthed og anvendelighed i praksis. Vi gennemgår først den grundlæggende teori for risikomål og derefter de modeller og teststørrelser, der indgår vores analyser. Vi undersøger teststørrelsernes styrke-egenskaber under iid såvel som under mere realistiske antagelser gennem Monte Carlo simulationer. Derefter undersøger vi teststørrelsernes evne til at afvise ikke-parametriske modeller. Resultaterne viser, at ES teststørrelserne sagtens kan sammenlignes med Basel VaR testen og kan effektivt fange, når ES er stærkt misspecifieret, både når vi anvender parametriske og ikke-parametriske prædiktive modeller. En særlig opmærksomhed gives til CC-testen, som er den eneste med en asymptotisk fordeling. Dette gør den potentiel mere anvendelig i praksis end de andre teststørrelser. Vi ser dog, at dette i praksis ikke er tilfældet, da konvergensen mod den asymptotiske fordeling er meget langsom.

Indhold

Resumé	2
1 Introduktion	4
2 Grundlæggende om risikomål	5
2.1 Egenskaber ved risikomål	5
2.2 Value-at-risk og expected shortfall	6
3 Tidsserieanalyser: GARCH processer	8
4 Setup for backtesting af ES	11
4.1 Teststørrelse 1	12
4.2 Teststørrelse 2	13
4.3 Teststørrelse 3	13
4.4 Teststørrelse 4	14
5 Analyse af styrke under urealistiske antagelser	16
5.1 Estimering af styrke	17
5.2 Resultater med iid setup	18
6 Analyse af styrke under mere realistiske antagelser	22
6.1 Mere realistisk model	22
6.2 Resultater med GARCH setup	23
7 Afvisningsrater med ikke-parametriske modeller	26
7.1 Historisk simulation	27
7.2 Vægtet historisk simulation	27
7.3 Resultater med ikke-parametriske modeller	28
8 Inspektion af CC-testen	29
8.1 Empirisk applikation	30
8.2 Stikprøvestørrelsen	31
9 Diskussion	32
10 Konklusion	33
Appendix	35
Litteratur	37

1 Introduktion

Riskostyring udgør en central del af arbejdet i mange finansielle institutioner, som typisk handler en bred vifte af finansielle instrumenter og påtager sig dermed i nogle tilfælde betydelig risiko. Denne risiko skal håndteres, både for institutionens egne interesser, men også ud fra regulatoriske hensyn, bl.a. for at undgå insolvens. Derfor er det påkrævet for mange finansielle institutioner at tilbageholde kapital, således at det kan dække ekstreme kapitalkrav f.eks. fra store markedsudsving [1]. For at vurdere hvor meget kapital, der skal tilbageholdes, er det brugbart at have en metode til at kvantificere risiko. Til dette formål bruger man bekvemt inden for risikostyring såkaldte risikomål.

Vi vil i denne afhandling hovedsageligt fokusere på *expected shortfall* (ES), som er et risikomål afledt af det andet udbredte risikomål *value-at-risk* (VaR). Disse to mål er blandt de hyppigst anvendte inden for risikostyring. ES er særligt interessant, da det på flere måder anses for at være et bedre risikomål end VaR. Dette skyldes, at ES besidder pænere matematiske egenskaber og giver en bedre indikation på størrelsen af ekstreme tab. Til gengæld er der en del tvivl, om hvorvidt ES kan blive backtestet (Vi vil uddybe alle detaljerne senere).

På trods af tvivlen er der alligevel foreslået teststørrelser i litteraturen til backtesting af ES. For at få en ide om disse teststørrelsers anvendelighed og brugbarhed i praksis, ønsker vi specifikt at analysere, hvor følsomme og pålidelige disse tests er til at opdage, prædiktive modellers ES og VaR forudsigelser ikke stemmer overens med data. Dette vil vi gøre med forskellige setups, både hvor vi bruger parametriske og ikke-parametriske metoder til estimering. For de parametriske setups vil vi først undersøge et simpelt iid tilfælde. Vi undersøger derefter, om testene klarer sig ligeså godt, når vi benytter *GARCH modeller*, som er velegnet til modellering af tidsafhængig data. For ikke-parametriske modeller vil vi undersøge to varianter af *historisk simulation* (HS), nemlig den klassiske og *vægtet historisk simulation*. Dette er især relevant siden den klassiske HS er den mest anvendte metode hos finansielle institutioner, mens vægtet HS kan ses som en udvidet version af den klassiske.

En af de teststørrelser vi vælger at undersøge er den simple CC-test introduceret af Nolde og Ziegel (2017), som har en asymptotisk fordeling. Denne egenskab gør teststørrelsen umiddelbart meget lettere at bruge i praksis end de andre tests. Siden dette betyder meget for anvendeligheden af denne test (og er umiddelbart dens store fordel over de andre tests), vil vi også undersøge dette nærmere.

Vores kode med alle implementeringer er tilgængelig på Github: <https://github.com/OiaaOiaaOiaa/Bachelor-projekt>.

2 Grundlæggende om risikomål

Vi vil i det følgende undersøge, hvad et risikomål er samt motivere introduktionen af risikomålet expected shortfall og undersøge, hvorfor den typisk er præfereret over value-at-risk. Der er ikke en streng definition af, hvad et risikomål er, men i denne afhandling vil vi tænke det som en afbildung fra en stokastisk variabel til de reelle tal. Mere præcist, hvis vi lader \mathcal{M} betegne mængden af alle stokastiske variable, kan vi definere et risikomål ρ som værende en afbildung $\rho : \mathcal{M} \rightarrow \mathbb{R}$. Vi vil betragte $X \in \mathcal{M}$ som værende en stokastisk variabel for profit & loss (P&L), og vi benytter derfor konventionen, at positive værdier er profit, og negative værdier er tab. Vi vil derudover også benytte konventionen, at store værdier for ρ indikerer høj risiko, mens lave værdier indikerer lav risiko. Et af de tidligste og mest naturlige risikomål i brug var variansen. Nogle af dens tydeligste svagheder er dog, at den kun giver mening for fordelinger med endelig varians samt fordelinger, der nogenlunde er symmetriske, siden variansen ikke fanger om afvigelser fra middelværdien er positive eller negative. Sidenhen er der blevet udviklet andre risikomål såsom VaR og ES ([2], s. 1-3).

2.1 Egenskaber ved risikomål

Definitionen af et risikomål er meget bred, så en måde at indsnævre de interessante risikomål er at undersøge, hvilke egenskaber er ønskværdige for et sådant mål. Dette bliver typisk vurderet baseret på en samling af pæne matematiske egenskaber kaldet *koherens* introduceret af Artzner et al. (1998) ([2], s. 3). Vi definerer koherens i det følgende.

Definition 2.1. (Artzner et al. [3]) *Lad X være en stokastisk variabel. Vi siger risikomålet ρ er koherent, hvis den opfylder følgende betingelser:*

- *Positiv homogenitet: For alle X og $a \geq 0$ gælder, at*

$$\rho(aX) = a\rho(X).$$

- *Sub-additivitet: For alle stokastiske variable X_1 og X_2 gælder, at*

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2).$$

- *Monotonicitet: For alle stokastiske variable X_1 og X_2 gælder, at*

$$X_1 \leq X_2 \implies \rho(X_2) \leq \rho(X_1).$$

- *Translation-invariant:* For alle stokastiske variable X og $a \in \mathbb{R}$ gælder, at

$$\rho(X + a) = \rho(X) - a.$$

Disse påne matematiske egenskaber ser umiddelbart tekniske ud, men de har intutive fortolkninger. Sub-additivitet kan fortolkes som, at øget diversifikation reducerer risiko. Homogenitet falder naturligt ud som følge af sub-additivitet, da for $n \in \mathbb{N}$ er $\rho(nX) \leq n\rho(X)$, og siden der ikke er noget diversifikation, er det naturligt at antage lighed. Monotonicitet kan fortolkes som, at hvis et aktiv X_1 klarer sig værre end et andet aktiv X_2 i alle tilfælde, er X_1 mere risikabelt end X_2 . Og endeligt, translation-invarians kan tænkes som, at hvis man tilføjer et risikofrit aktiv $a \in \mathbb{R}$, vil kapitalkrav og dermed risiko sænkes med netop den mængde. Alle disse egenskaber passer altså nogenlunde med intuitionen ([4], s. 72-73).

Til sidst vil vi nævne egenskaben *elicitabilitet*. Denne egenskab har spillet en bemærkelsesværdig rolle i kritikken mod ES, som vi vil se senere.

Definition 2.2. (*Acerbi og Szekely [5]*) Risikomålet $\rho(X)$ for en stokastisk variabel X siges at være *elicitable*, hvis den minimerer forventningen af en scoringsfunktion $S(x, y)$:

$$\rho = \arg \min_x E[S(x, X)].$$

Elicitabilitet er et brugbart egenskab at besidde, siden den kan benyttes til at vurdere, om en prædiktiv model er god. For at illustrere dette, lad x_t, \dots, x_T være realisationer for en stokastisk proces $(X_t)_{t=1, \dots, T}$, som vi er interesseret i. Lad ρ_t, \dots, ρ_T være forecasts fra en prædiktiv model for et risikomål ρ . Har vi en scoringsfunktion S til risikomålet ρ , kan vi evaluere hvor god vores prædiktion var ved at se om $\bar{S} = \frac{1}{T} \sum_{t=1}^T S(\rho_t, x_t)$ er så lav som mulig [5].

2.2 Value-at-risk og expected shortfall

Som nævnt er nogle af de mest anvendte risikomål value-at-risk og expected shortfall. Vi vil definere disse i det følgende. Lad F betegne fordelingsfunktionen for X . Vi starter med at betragte VaR, som vi definerer som en negeret fraktil i en P&L fordeling.

Definition 2.3. (*Bayer og Dimitriadi [6]*) Value-at-Risk, $VaR_\beta(X)$, med niveau $\beta \in (0, 1)$ og fordelingsfunktion F , defineres ved:

$$VaR_\beta(X) = -\inf \{y \in \mathbb{R} : F(y) \geq \beta\} = -F^{-1}(\beta), \quad (2.1)$$

hvor F^{-1} betegner fraktilfunktionen.

Value-at-risk blev introduceret af RiskMetrics i 1996, og kort tid efter gjorde Basel-komiteén det til standard at benytte det i den finansielle sektor. Men allerede i 1998 blev akademiske artikler udgivet, hvor risikomålet modtog en del kritik. Det kan nemlig vises, at risikomålet ikke besidder sub-additivitet-egenskaben, og det er således ikke et koherent risikomål. Risikomålet kan altså i nogle tilfælde indikere, at øget diversifikation øger risikoen [5].

Derudover, value-at-risk med niveau β angiver blot det tab, som bliver overskredet ved fordelingens hale med sandsynlighed β , men giver til gengæld ingen indikation på størrelsen af potentielle tab. Disse er en væsentlige svagheder ved VaR som et risikomål, og har derfor motiveret introduktionen af det afledte mål ES, som modsat VaR, besidder disse egenskaber [2].

Definition 2.4. (*Acerbi og Szekely [5]*) *Expected shortfall, $ES_\alpha(X)$, med niveau $\alpha \in (0, 1)$ og fordelingsfunktion F , defineres ved*

$$ES_\alpha(X) = -\frac{1}{\alpha} \int_0^\alpha F^{-1}(q)dq, \quad (2.2)$$

hvor F^{-1} er fraktilfunktionen.

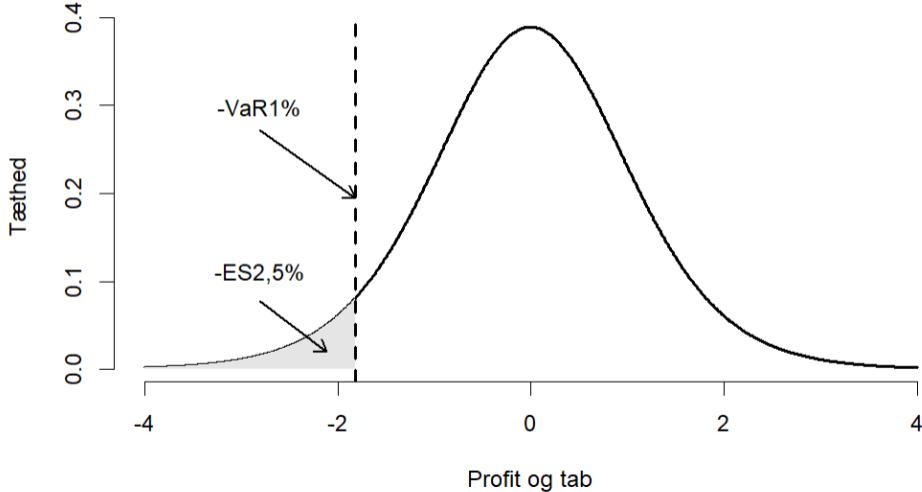
Basel-komiteén foreslog i første omgang allerede i 2013 om at gøre ES til standard i den finansielle sektor, siden den netop bedre kan fange størrelsen af tabene ved fordelingers haler. Foreslaget blev dog mødt med kritik, da det kan vises, at risikomålet ikke er elicitable [5], altså at der ikke findes en scoringsfunktion, som opfylder Definition 1.2. for ES. Dette betyder umiddelbart, at ES prædiktioner ikke direkte kan evalueres ved at udnytte elicitabilitet. VaR lider ikke af dette problem, hvor det kan vises, at den har en scoringsfunktion på formen $S(x, y) = (1(x > y) - \alpha)(x - y)$.

Manglen af elicitabilitet har været problematisk, da det har ført til diskussioner, om hvorvidt ES så også kan blive backtestet [5], hvilket ville være en stor svaghed, hvis det ikke var tilfældet. På trods af denne tvivl er ES sidenhen alligevel blevet gjort til standard af Basel-komiteén [7].

Vi vil ikke behandle elicitabilitet yderligere eller vurdere, hvorvidt det faktisk har indflydelse på et risikomåls evne til at blive backtestet i denne afhandling. På trods af ikke at være elicitable vil vi se i de følgende afsnit, at der alligevel er udviklet ES teststørrelser til backtesting. Dette rejser spørgsmålet, om hvor effektive og robuste disse teststørrelser så er i praksis, og hvordan de klarer sig i forhold til en VaR test, siden der ikke er tvivl, om VaR kan blive backtestet. Dette motiverer vores undersøgelser i resten af afhandlingen.

Vi kan visualisere disse to risikomål i figur 1 i en generel P&L fordeling, hvor den stiplede linje angiver VaR ved $\beta = 1\%$ og det skraverede område er ES ved

$\alpha = 2,5\%$. Disse niveauer for ES og VaR er gjort til standard af Basel komitéen, da for normalfordelingen er $VaR_{1\%}$ netop lig med $ES_{2.5\%}$ [5]. Vi vil også benytte disse niveauer i de efterfølgende afsnit, så vi vil fremover lade $\alpha = 2,5\%$ og $\beta = 1\%$.



Figur 1: Visualisering af ES og VaR.

Vi vil afslutningsvis præsentere alternative og brugbare udtryk for ES i det tilfælde, hvor den underliggende stokastiske variabel har en kontinuert og strengt voksende fordelingsfunktion.

Proposition 2.5. (*McNeil et al. [4]*) *Lad X have en kontinuert og strengt voksende fordelingsfunktion F . Da kan vi udtrykke expected shortfall ES med niveau $\alpha \in (0, 1)$ ved*

$$ES_\alpha(X) = -E \left[\frac{X I}{\alpha} \right] = -E [X | X + VaR_\alpha < 0], \quad (2.3)$$

hvor $I = 1(X + VaR_\alpha < 0)$ er en indikatorfunktion.

3 Tidsserieanalyser: GARCH processer

Vi vil i det følgende afsnit give en kort introduktion til de væsentligste koncepter indenfor finansiel tidsserieanalyse, der er relevant for denne afhandling. Vores formål er at introducere en særlig simpel model, som er meget brugbar til at beskrive og replike reelle finansielle tidsseriedata såsom aktier. Vi taler her om de såkaldte GARCH modeller (*generalized autoregressive conditionally heteroscedastic*), som benyttes til modellering af volatilitetsprocesser. Her refererer generalized autoregressive til, at afkast og volatilitet afhænger af både tidligere volatilitet og afkast. Conditionally heteroscedastic betyder blot, at volatiliteten givet tidligere information ændrer sig over tid ([4], s. 113-118).

En tidsserie definerer vi som en diskret stokastisk proces $(X_t)_{t \in \mathbb{Z}}$, dvs. vi betragter det som en familie af stokastiske variable $X_t \in \mathbb{R}$ indekseret med de hele tal. Hvis de eksisterer kan vi definere middelværdifunktionen $\mu(t)$ og autokovariansfunktionen $\gamma(t, s)$ for $(X_t)_{t \in \mathbb{Z}}$ ved henholdsvis:

$$\begin{aligned}\mu(t) &= E[X_t], \quad t \in \mathbb{Z}, \\ \gamma(t, s) &= E[(X_t - \mu(t))(X_s - \mu(s))], \quad t, s \in \mathbb{Z}.\end{aligned}$$

Det skal bemærkes, at der gælder $\gamma(s, t) = \gamma(t, s)$ samt at variansen af X_t er $\gamma(t, t) = \text{var}(X_t)$ ([4], s. 98-99). Vi kan nu give en teknisk definition af *stationaritet*, som fanger en ide om, at uanset hvilket tidsvindue af en tidsserie man betragter, er der nogle bestemte egenskaber, der går igen.

Definition 3.1. (*McNeil et al. [4]*) *Vi siger en tidsserie $(X_t)_{t \in \mathbb{Z}}$ er stationær, hvis den opfylder en eller begge af følgende typer af stationaritet:*

- *Streng stationaritet: Tidsserien $(X_t)_{t \in \mathbb{Z}}$ er streng stationær, hvis $(X_{t_1}, \dots, X_{t_n})$ har samme fordeling som $(X_{t_1+k}, \dots, X_{t_n+k})$ for alle $t_1, \dots, t_n, k \in \mathbb{Z}$ og for alle $n \in \mathbb{N}$.*
- *Kovarians stationaritet: Tidsrækken $(X_t)_{t \in \mathbb{Z}}$ siges at være kovarians stationær hvis første og andet moment eksisterer og*

$$\begin{aligned}\mu(t) &= \mu, \quad t \in \mathbb{Z}, \\ \gamma(t, s) &= \gamma(t + k, s + k), \quad t, s, k \in \mathbb{Z}.\end{aligned}$$

Streng stationaritet beskriver derfor med andre ord ideen om, at forskellige tidsvinduer i en tidsserie har nøjagtig de samme egenskaber uafhængig af, hvor i tidsserien de forskellige tidsvinduer kommer fra. Kovarians stationaritet (nogle gange kaldet svag stationaritet) derimod er en svagere form for stationaritet, som kun kræver uniform opførsel af middelværdien, varians og kovarians ([8], s. 307-308).

I praksis kan antagelsen om stationaritet for data være en lidt streng antagelse. Det er således almindeligt at benytte *differencing* en eller flere gange til at gøre data stationært. Har vi en tidsserie proces $(Y_t)_{t \in \mathbb{Z}}$ og udfører differencing på det, betragter vi i stedet tidsserieprocessen $(Y_t - Y_{t-1})_{t \in \mathbb{Z}}$. Dette giver ingen problemer i praksis fortolkningsmæssigt, siden det er almindeligt at betragte log-afkast i stedet for aktiepriser. Det vil sige, givet vi har aktiepriser $(Y_t)_{t \in \mathbb{Z}}$ og dermed log-priser $(\log(Y_t))_{t \in \mathbb{Z}}$, er det mere bekvemt at vi i stedet betragter tidsserien $(X_t)_{t \in \mathbb{Z}} = (\log Y_t - \log Y_{t-1})_{t \in \mathbb{Z}}$, der som bekendt approksimerer de relative afkastrater $(Y_t - Y_{t-1})/Y_{t-1}$ ([4], s. 105).

Et andet nyttigt begreb indenfor finansiell tidsserieanalyse er de såkaldte *strenghvid støj* processer, som vi definerer i det følgende.

Definition 3.2. (*McNeil et al. [4]*) Vi siger $(X_t)_{t \in \mathbb{Z}}$ er en *strenghvid støj proces*, hvis den er en serie af iid stokastiske variable med endelig varians.

Vi vil betegne strenghvid støj processer med middelværdi 0 og varians σ^2 ved $SWN(0, \sigma^2)$ ligesom McNeil et al (2015). Vi kan nu definere GARCH modellen.

Definition 3.3. (*McNeil et al. [4]*) Lad $(Z_t)_{t \in \mathbb{Z}}$ være en $SWN(0, 1)$. Vi siger en proces $(X_t)_{t \in \mathbb{Z}}$ er en *GARCH(p, q)* proces, hvis den er streng stationær og opfylder følgende ligninger for alle $t \in \mathbb{Z}$:

$$X_t = \sigma_t Z_t, \quad (3.1)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (3.2)$$

hvor $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, \dots, p$ og $\beta_j \geq 0$ for $j = 1, \dots, q$.

I praksis benyttes typisk en GARCH model af lav orden, og vi vil derfor i denne afhandling blot fokusere på det særlige tilfælde hvor $p = q = 1$, altså GARCH(1,1) modeller. Når $q = 0$ vil σ_t ikke afhænge af σ_{t-1} , hvorved modellen blot reduceres til en såkaldt ARCH model. α_1 kaldes derfor ARCH-parameteren.

En stor grund til, hvorfor vi er lige præcis interesseret i GARCH modellen, er, at der er mange typiske egenskaber ved tidsserier af finansielle afkast, som GARCH modellen fanger godt. Siden afkast og volatilitet afhænger af både tidligere afkast og volatilitet, vil stød i denne model udvise persistens og har altså tendens til at blive efterfulgt af hinanden. Derudover, siden den betinget volatilitet varierer over tid, og siden en smule serial korrelation er til stede, er afkast ikke uafhængige og identisk fordelte (iid) i denne model. Til sidst ses nemt, at middelværdien for GARCH processen er nul, og med det rigtige valg af fordelingen af innovationerne, vil fordelingen af afkast også være tunghalet. Alle disse egenskaber er meget typiske for tidsserier af finansielle afkast ifølge McNeil et al. ([4], s. 79-80).

Vi vil afslutningsvis bemærke, at siden GARCH processer kun forudsætter streng stationaritet, kunne man komme ud i, at modellen har uendelig varians og er dermed ikke kovarians stationær. Følgende proposition giver en nødvendig og tilstrækkelig betingelse for, at modellen har endelig varians.

Proposition 3.4. (*McNeil et al. [4]*) *GARCH(1,1) processen er kovarians stationær hvis og kun hvis $\alpha_1 + \beta_1 < 1$. I så fald er den ubetinget varians $E[\sigma_t^2] = \omega / (1 - \alpha_1 - \beta_1)$. Der gælder særligt $\alpha_1 + \beta_1$ er persistensen for processen.*

4 Setup for backtesting af ES

Backtesting kan betragtes som en måde at evaluere, om en metode til at lave out-of-sample estimerer af et risikomål overhovedet passer med de faktiske realisationer. Til dette formål kan det være nyttigt at bruge forskellige teststørrelser. Vi vil i dette afsnit præsentere fire forskellige teststørrelser til backtesting af ES fra forskellige steder i litteraturen. For at undersøge nærmere egenskaberne ved dem benytter vi et hypothesetest framework. Dette framework vil vi bruge til vores analyser i de efterfølgende afsnit, hvor vi undersøger teststørrelsernes præstation under forsimplede antagelser, og senere under mere realistiske antagelser.

Vi betragter en finansiel tidsserie $\vec{X} = (X_t)_{t=1,\dots,T}$ i perioderne $t = 1, \dots, T$, hvor X_t er en stokastisk variabel for profit og tab i periode t , som har en ukendt og sand fordeling F_t . Vi definerer derudover filtreringen $(\mathcal{F}_t)_{t=1,\dots,T}$, hvor $\mathcal{F}_t = \sigma(X_s, s \leq t)$ er den tidlige tilgængelige information om vores finansielle tidsserie til og med tid t . For at lemppe notationen vil vi benytte $E_t[\cdot]$ for middelværdien til tid t givet tidligere information \mathcal{F}_{t-1} .

Vi konstruerer en prædiktiv model P_t baseret på tidlige tilgængelig information, som benyttes til at give et one-step-ahead estimat for VaR til tid t med niveau α :

$$VaR_{\alpha,t} = -P_t^{-1}(\alpha). \quad (4.1)$$

One-step-ahead estimerer for ES til tid t og niveau α beregnes ved:

$$ES_{\alpha,t} = -\frac{1}{\alpha} \int_0^\alpha P_t^{-1}(q) dq. \quad (4.2)$$

Lad da $\overrightarrow{ES}_\alpha = (ES_{\alpha,t})_{t=1,\dots,T}$ og $\overrightarrow{VaR}_\alpha = (VaR_{\alpha,t})_{t=1,\dots,T}$. De sande VaR og ES betegner vi med $VaR_{\alpha,t}^F$ hhv. $ES_{\alpha,t}^F$ [5].

Vi ønsker at konstruere teststørrelser med henblik på backtesting af ES, som skal indikere fejl, når vores prædiktive modeller laver forkerte estimerer for ES (og VaR). Vi vil kun undersøge ensidet tests, så nulhypotesen H_0 vil generelt være antagelsen, om at vores forudsigelser er korrekte, eller at vi har overestimeret risikoen. Den alternative hypotese H_1 er, at vi har underestimeret risikoen. Dette setup giver mening fra regulatorens synspunkt, siden der ikke er noget galt i, at et finansielt institution holder mere kapital tilbage end nødvendigt [5, 1]. Vi vil nu introducere de fire teststørrelser for ES, som vi vil benytte i de senere analyser.

4.1 Teststørrelse 1

Den første teststørrelse vi vil betragte kaldes ER (exceedance residual) testen og tager udgangspunkt i, hvor stor forskellen mellem realisationer og vores ES forecasts er, givet VaR er overskredet. Lad \vec{X} vere en finansiel tidsserie. Det ses nemt ved brug af proposition 2.5. at med de sande VaR og ES værdier, $VaR_{\alpha,t}^F$ og $ES_{\alpha,t}^F$, er:

$$E_t[X_t + ES_{\alpha,t}^F \mid X_t + VaR_{\alpha,t}^F < 0] = -ES_{\alpha,t}^F + ES_{\alpha,t}^F = 0.$$

Inspireret af dette, med ES og VaR estimatorer $\overrightarrow{ES}_\alpha$ og $\overrightarrow{VaR}_\alpha$, kan vi nu definere de rå exceedance residualer som værende forskellen mellem ES og observerede værdier:

$$ER_t = (X_t + ES_{\alpha,t})1(X_t + VaR_{\alpha,t} < 0),$$

for $t = 1, \dots, T$. Bemærk, at hvis ES er underestimeret vil ER_t estimererne være mindre end de sande, hvorfor vi i så fald vil få en lavere middelværdi for vores estimeret ER_t . Det omvendte gælder for overestimering af ES.

Vi kan nu definere teststørrelsen:

$$Z_1(\vec{X}, \overrightarrow{ES}_\alpha, \overrightarrow{VaR}_\alpha) = \frac{1}{\sum_{t=1}^T I_t} \sum_{t=1}^T ER_t, \quad (4.3)$$

hvor $I_t = 1(X_t + VaR_{\alpha,t} < 0)$ er en indikatorfunktion.

Lad $E_t[ER_t \mid X_t + VaR_{\alpha,t} < 0] = \mu$. Vores H_0 hhv. H_1 er derfor:

$$\begin{aligned} H_0 : \mu &\geq 0, \\ H_1 : \mu &< 0. \end{aligned}$$

H_0 er ækvivalent med at teste for overestimering af ES, og under H_1 underestimeres ES [1]. Bemærk, at ligesom den første teststørrelse foreslægt af Acerbi og Szekely (2014) vil denne teststørrelse kun reagere på forskelle ved fordelingens hale.

Det er klart fra (4.3), at denne test ikke tager hensyn til, hvor mange overskridelser af VaR, der bør være ved niveau α . Derfor vil denne teststørrelse være rimelig afhængig af, at VaR nogenlunde er estimeret korrekt ([5], s. 4). Vi bemærker til sidst, at siden denne test måler forskellen på middelværdien af X_t og ES givet VaR overskridelser, indikerer denne test særligt fejl, når forholdet mellem ES og VaR estimererne er meget forkerte. Dette betyder også, at testen vil have det svært med at opdage simultane misspecifikationer af ES og VaR forudsigelser kommende fra bl.a. misspecifikation af volatiliteten i location scale modeller [6].

4.2 Teststørrelse 2

Fra proposition 2.5. vides, at ES kan beskrives ved følgende forventning:

$$ES_{\alpha,t}^F = -E_t \left[\frac{X_t I_t}{\alpha} \right].$$

Givet tidsserien \vec{X} og estimatorer $\overrightarrow{ES}_\alpha$ og $\overrightarrow{VaR}_\alpha$ kunne man derfor håbe på, at i gennemsnit ligger de observerede $-X_t I_t / \alpha$ rimelig tæt på de tilsvarende ES estimatorer således at $ES_{\alpha,t} + X_t I_t / \alpha$ er omrent lig 0. Inspireret af dette definerer vi teststørrelsen:

$$Z_2(\vec{X}, \overrightarrow{ES}_\alpha, \overrightarrow{VaR}_\alpha) = \sum_{t=1}^T \frac{X_t I_t}{T \alpha ES_{\alpha,t}} + 1 \quad (4.4)$$

Vores nulhypotese hhv. alternative hypotese:

$$\begin{aligned} H_0 : & ES_{\alpha,t}^F \leq ES_{\alpha,t}, \text{ for alle } t, \\ & VaR_{\alpha,t}^F \leq VaR_{\alpha,t}, \text{ for alle } t. \\ H_1 : & ES_{\alpha,t}^F \geq ES_{\alpha,t}, \text{ for alle } t \text{ og } > \text{ for nogle } t, \\ & VaR_{\alpha,t}^F \geq VaR_{\alpha,t}, \text{ for alle } t. \end{aligned}$$

Det ses nemt, at under H_1 vil summen i Z_2 blive mindre end -1 . I så fald bliver Z_2 negativ, og testen vil da indikere fejl. Denne teststørrelse vil dog også opfange afvigelser i VaR, siden underestimering af VaR betyder flere overskridelser, hvilket også gør summen lavere end -1 . De omvendte hændelser sker under H_0 ved overestimering af ES hhv. VaR. Bemærk desuden, at ligesom Z_1 reagerer denne test kun ved ændringer i fordelingens hale [5].

4.3 Teststørrelse 3

Givet tidsserien \vec{X} og prædiktive modeller P_1, \dots, P_T , vides fra basal sandsynlighedsregning, at hvis vores prædiktive modeller er korrekte vil de observerede rang $U_t = P_t(X_t)$ være iid uniform fordelt $U(0, 1)$. Lad $\vec{U} = (U_t)_{t=1, \dots, T}$ være vektoren af alle U_t . Vi definerer:

$$\widehat{ES}_\alpha^{(N)}(\vec{Y}) = -\frac{1}{\lfloor N\alpha \rfloor} \sum_i^{\lfloor N\alpha \rfloor} Y_{i:N}, \quad (4.5)$$

som en estimator for ES for N iid udtræk fra $\vec{Y} = (Y_i)_i$, hvor $Y_{i:N}$ betegner den i 'te ordnede observation blandt de N udtræk, og $\lfloor \cdot \rfloor$ er floor funktionen. Vi definerer nu

vores tredje teststørrelse ved:

$$Z_3(\vec{X}) = -\frac{1}{T} \sum_{t=1}^T \frac{\widehat{ES}_\alpha^{(T)}(P_t^{-1}(\vec{U}))}{E_V [\widehat{ES}_\alpha^{(T)}(P_t^{-1}(\vec{V}))]} + 1, \quad (4.6)$$

hvor \vec{V} også er en vektor af stokastiske variable, som også er iid $U(0, 1)$. Med denne test forsøger vi at estimere ES til tid t ved brug af de observerede rang \vec{U} , siden hvis P_t er korrekt og U_t -erne er uniform fordelte, vil $P_t^{-1}(\vec{U})$ nogenlunde simulere X_t . I nævneren har vi et estimat for middelværdien af ES estimatet i tælleren, hvis rolle er at normalisere testen og kompensere for bias for ES estimatoren i tælleren. Med disse estimatorer fås T kvotienter, som vi kan tage gennemsnit over. Nævnerne i (4.6) har følgende analytiske form:

$$E_V [\widehat{ES}_\alpha^{(T)}(P_t^{-1}(\vec{V}))] = -\frac{T}{[T\alpha]} \int_0^1 I_{1-p}(T - [T\alpha], [T\alpha]) P_t^{-1}(p) dp, \quad (4.7)$$

hvor $I_x(a, b)$ er en regularized betafunktion (dette bevises af Acerbi og Szekely 2014, [5]). Der gælder også for denne teststørrelse, at korrekt estimering af risikoen vil gøre middelværdien af Z_3 lig nul, mens overestimering af risikoen vil øge middelværdien. Underestimering vil gøre middelværdien mindre end nul, hvilket indikerer fejl. Vores nulhypotese og alternative hypotese er:

$$H_0 : F_t \succsim P_t, \text{ for alle } t.$$

$$H_1 : F_t \precsim P_t, \text{ for alle } t \text{ og } \succ \text{ for nogle } t.$$

Her betegner \succsim og \succ svag hhv. streng første ordens stokastisk dominans. Under H_0 betyder dette blot, at $F_t(x) \leq P_t(x)$. Tilsvarende under H_1 er $P_t(x) \leq F_t(x)$, hvor der skal gælde $<$ for nogle t [9]. Det ses også fra hypoteserne, at modsat Z_1 og Z_2 er Z_3 en test om hele fordelingen passer, ikke blot halen [5].

4.4 Teststørrelse 4

Den sidste test vi vil betragte er den såkaldte simple CC-test (conditional calibration test) introduceret af Nolde og Ziegel (2017) [10]. Der gælder, at parret (VaR_α, ES_α) har en identifikationsfunktion V :

$$V(X, VaR_\alpha, ES_\alpha) = \begin{pmatrix} \alpha - 1(X + VaR_\alpha < 0) \\ VaR_\alpha - ES_\alpha - \frac{VaR_\alpha + X}{\alpha} 1(X + VaR_\alpha < 0) \end{pmatrix}.$$

Givet en identifikationsfunktion siger vi, at serien af forudsigelser $(VaR_{\alpha,t}, ES_{\alpha,t})_{t \in \mathbb{N}}$ er betinget kalibreret for (VaR, ES) hvis (bemærk betinget fordi det er betinget på tidligere information):

$$E[V(X_t, VaR_{\alpha,t}, ES_{\alpha,t}) | \mathcal{F}_{t-1}] = 0,$$

for alle $t \in \mathbb{N}$. Vi siger den er betinget super-kalibreret for (VaR, ES) , hvis vi i stedet har ≥ 0 og betinget sub-kalibreret hvis ≤ 0 . Nolde og Ziegel har vist, at med den givne identifikationsfunktion, svarer en test om korrekt eller overestimering af ES til en test om sub-kalibrering ([10], s. 1845). Sub-kalibrering vil derfor være vores nulhypotese. Derfor, givet tidsserien \vec{X} og forudsigelser \vec{ES}_α og \vec{VaR}_α , er vores nulhypotese hhv. alternative hypotese:

$$\begin{aligned} H_0 : E[V(X_t, VaR_{\alpha,t}, ES_{\alpha,t}) | \mathcal{F}_{t-1}] &\leq 0, \text{ for alle } t = 1, \dots, T \\ H_1 : E[V(X_t, VaR_{\alpha,t}, ES_{\alpha,t}) | \mathcal{F}_{t-1}] &> 0, \text{ for alle } t = 1, \dots, T. \end{aligned}$$

Bemærk hypoteserne gælder komponentvis for hver af de to komponenter af identifikationsfunktionen. I praksis vil vi benytte en anden variant af disse hypoteser. Definer \mathcal{F}_{t-1} -målbare funktioner $h_{1,t}$ og $h_{2,t}$ (vi kalder dem testfunktioner), som er q_1 - hhv. q_2 -vektorer. Det ses pr. definition af den betinget middelværdi, at nulhypotesen svarer til at teste følgende for hver komponent af identifikationsfunktionen:

$$\begin{aligned} H_0 : E[h_{1,t}^T V_1(X_t, VaR_{\alpha,t}, ES_{\alpha,t})] &\leq 0, \text{ for alle } t = 1, \dots, T, \\ E[h_{2,t}^T V_2(X_t, VaR_{\alpha,t}, ES_{\alpha,t})] &\leq 0, \text{ for alle } t = 1, \dots, T. \end{aligned}$$

Vi kan kombinere de to målbare funktioner til at få en samlet testfunktion, som er en $q \times 2$ matrix, hvor $q = q_1 + q_2$:

$$h_t = \begin{pmatrix} h_{1,t} & 0 \\ 0 & h_{2,t} \end{pmatrix}.$$

Vi kan da alternativt udtrykke nulhypotesen og den alternative hypotese simplere ved:

$$\begin{aligned} H_0 : E[h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t})] &\leq 0, \text{ for alle } t = 1, \dots, T, \\ H_1 : E[h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t})] &> 0, \text{ for alle } t = 1, \dots, T, \end{aligned}$$

hvor det skal bemærkes, at $h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t})$ er en q -vektor. Udtrykket for den alternative hypotese H_1 fandt vi på en tilsvarende måde som for H_0 . Det er disse udtryk for H_0 og H_1 , som vi vil anvende i praksis. Den ensidet test, som Nolde og

Ziegel foreslår, tager formen:

$$Z_4(\vec{X}, \overrightarrow{ES}_\alpha, \overrightarrow{VaR}_\alpha) = \sqrt{T}^{-1} \widehat{\Omega}_T^{-1/2} \sum_{t=1}^T h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t}) \xrightarrow{d} N(0, I_q), \quad (4.8)$$

hvor:

$$\widehat{\Omega}_T = \frac{1}{T} \sum_{t=1}^T (h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t})) (h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t}))^T \quad (4.9)$$

er en $q \times q$ konsistent estimator for kovariansmatricen for $h_t V(X_t, VaR_{\alpha,t}, ES_{\alpha,t})$. Test af samme type som Z_4 kaldes en CC-test.

Det særlige ved teststørrelsen Z_4 er, at modsat de andre teststørrelser vi introducerede tidligere, har denne teststørrelse en asymptotisk fordeling (asymptotisk standard normalfordelt for alle komponenterne af Z_4) [10]. Dette egenskab ved Z_4 gør testen umiddelbart særlig brugbar i praktiske anvendelser, særligt til at undersøge, om vores prædiktive modeller understøttes af data, når vi kun har en enkel sample, som det er tilfældet for finansielle tidsserier.

Den simpleste testfunktion, som Nolde og Ziegel foreslår er et hvor $q_1 = q_2 = 1$ og $h_{1,t} = h_{2,t} = 1$, således at $h_t = I_2$ blot er identitetsmatricen. I så fald kaldes Z_4 en simpel CC-test [10]. Det vil være denne variant vi vil fokusere på i denne afhandling. Bemærk, at med vores valg af testfunktion består Z_4 af to komponenter $Z_4 = (Z_{4,1}, Z_{4,2})$, modsat de andre tests vi har introduceret, der er en-dimensionelle.

5 Analyse af styrke under urealistiske antagelser

For at få et nemt overblik over vores metoder, og hvor robuste vores teststørrelser er, vil vi først foretage en simpel sensitivitetsanalyse. I vores simple setup vil vi derfor betragte det urealistiske men pæne scenarie, hvor vores finansielle tidsserie $\vec{X} = (X_t)_{t=1,\dots,T}$ er iid i stil med Acerbi og Szekelys (2014) setup [5]. Vi vil særligt undersøge styrken for vores ES_α teststørrelser og sammenligne dem med Basel VaR_β testen, som blot tæller antal overskridelser af VaR_β :

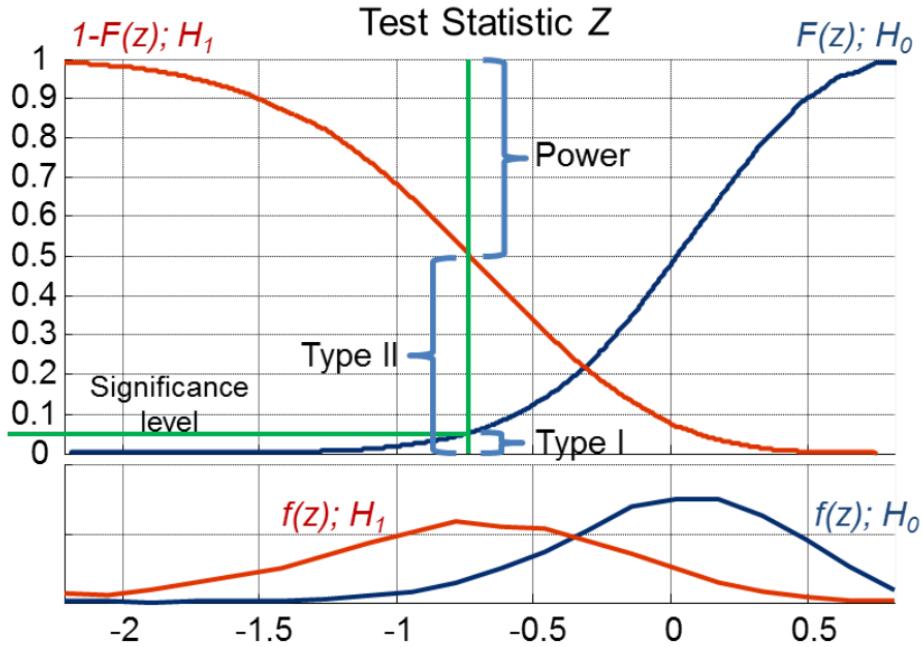
$$N(\vec{X}, \overrightarrow{VaR}_\beta) = \sum_{t=1}^T 1(X_t + VaR_{\beta,t} < 0). \quad (5.1)$$

En teststørrelsес styrke er særligt interessant i forhold til backtesting. Høj styrke indikerer, at en teststørrelse er god til at korrekt afvise forkerte nulhypoteser. Dermed undgås type 2 fejl. I finansielle sammenhænge er det ønskværdigt, hvis vores teststørrelser faktisk har en god evne til at afvise prædiktive modeller, der giver upræcise forudsigelser for ES.

5.1 Estimering af styrke

Vi estimerer styrke ved Monte Carlo simulation. Mere specifikt simulerer vi fordelingen af teststørrelsen $Z = Z_i$ for $i = 1, 2, 3$ (og basel testen) under H_0 samt H_1 , som vi betegner $F_Z^{H_0}$ og $F_Z^{H_1}$. Dette gøres ved at generere $M = 10000$ samples af størrelse $T = 250$ under H_0 hhv. H_1 . For hvert sample estimerer vi ES og VaR med ligninger (4.2) og (4.1), som bruges til at beregne M værdier for Z . Bemærk vi estimerer ES og VaR ved brug af den prædictive model både, når vi simulerer under H_0 og H_1 .

Vi ønsker egentlig at beregne styrke med signifikansniveauerne 5% og 10%, men siden Basel VaR testen er en diskret stokastisk variabel, benytter vi i stedet de diskrete niveauer, der ligger tættest på 5% og 10% for VaR testen. Dette gøres så vi kan sammenligne dens styrke med de andre tests. For alle teststørrelserne, undtagen CC-testen Z_4 , vil vi beregne styrken med signifikansniveau η ved først at finde den empiriske η -fraktil for Z 's fordeling under H_0 . Hvis vi betegner denne ved $q_\eta^{H_0}$, kan vi beregne styrken for testene ved $styrke = F_Z^{H_1}(q_\eta^{H_0})$. Denne metode er illustreret af Acerbi og Szekely (2014) og som vi viser i figur 2.



Figur 2: Visualisering af hvordan styrke beregnes for teststørrelserne [5].

Beregning af styrken for CC-testen Z_4 er lidt mere besværlig, siden den har to komponenter. Til at løse dette problem, benytter vi Hommels metode, som beskrevet af Nolde og Ziegel (2017), der er særligt egnet til teststørrelser med flere komponenter. For hver af Z_4 's to komponenter vil vi udføre en lignende metode som før, hvor vi simulerer fordelingen af de to komponenter under H_0 og H_1 . Ligesom før genererer vi $M = 10000$ samples af størrelse $T = 250$ under både H_0 og H_1 . Vi estimerer

derefter ES og VaR og beregner så en værdi for $Z_{4,1}$ og $Z_{4,2}$ for hvert sample. Vi får da simuleret fordelingerne $F_{Z_{4,1}}^{H_0}$ og $F_{Z_{4,2}}^{H_0}$. Med vores M realisationer af $Z_{4,1}$ og $Z_{4,2}$ under H_1 , $z_{(4,1),1}^{H_1}, \dots, z_{(4,1),M}^{H_1}$ og $z_{(4,2),1}^{H_1}, \dots, z_{(4,2),M}^{H_1}$, kan vi beregne M par af p-værdier

$$(\pi_{1,n}, \pi_{2,n}) = \left(1 - F_{Z_{4,1}}^{H_0}(z_{(4,1),n}^{H_1}), 1 - F_{Z_{4,2}}^{H_0}(z_{(4,2),n}^{H_1})\right),$$

for $n = 1, \dots, M$. Vi kan da benytte Hommels metode til at få M p-værdier under den globale hypotese:

$$\tilde{\pi}_n = 2C_2 \min \left\{ \pi_{(m),n}/m, m = 1, 2 \right\}, \quad C_2 = \frac{1}{1} + \frac{1}{2} = \frac{3}{2}.$$

Her betegner $\pi_{(1),n}$ og $\pi_{(2),n}$ de ordnede p-værdier for sample $n \in \{1, \dots, M\}$ [10]. Vi kan da beregne styrken ved at tælle, hvor stor andel af vores M samples fra H_1 giver en p-værdi under η og dermed afvist under H_0 . Vi tæller altså andelen af vores $\tilde{\pi}_n$ -er, hvor $\tilde{\pi}_n < \eta$. Bemærk vi ikke benytter Z_4 's asymptotiske fordeling her. Begrundelsen for dette vil vi forklare i afsnit 8.

5.2 Resultater med iid setup

Til beregning af styrken for vores teststørrelser lader vi vores model under H_0 , altså vores prædiktive model, være en misspecifieret version af den sande model (modellen under H_1). Med dette simple setup kan vi nemt se, gennem vores valg af de parametriske modeller, hvilke typer misspecifikationer, testene er mest sensitive overfor. I denne simple analyse fokuserer vi udelukkende på varianter af t-fordelingen. Vi betragter følgende valg af H_0 og H_1 :

- (a) Vi lader H_1 være en skaleret version af den prædiktive model under H_0 , dvs. $F(x) = P(x/\gamma)$, for $\gamma > 1$. Vi ønsker, at dækningen af ES ikke passer i de to modeller, så $ES_\alpha^P = ES_{\alpha'}^F$, hvor vi lader $\alpha' = 5\%, 10\%$. Vi vælger γ således at $\gamma = ES_\alpha^P/ES_{\alpha'}^F$. Med dette setup har vi tydeligt underestimeret ES med vores prædiktive model. Vi vælger at se, hvordan resultaterne ser ud når t-fordelingen har frihedsgrader $\nu = 100$ og $\nu = 5$. Vi viser resultaterne i tabel 1.
- (b) Vi lader H_1 være en t-fordeling med lavere frihedsgrader end modellen under H_0 . Det ses, at med lavere frihedsgrader stiger variansen $\sigma^2 = \nu/(\nu - 2)$, og det er udbredt viden, at t-fordelingen med lavere frihedsgrader har tungere haler, hvorfor vi med modellen under H_0 igen underestimerer ES. Vi undersøger tilfældene, hvor under H_0 er $\nu = 100$ eller $\nu = 10$, og under H_1 er $\nu = 10, 5$ eller 3.
- (c) Vi foretager den samme analyse som for (b), men hvor vi i stedet betragter den

standardiserede t-fordeling. Dermed vil forskelle ikke komme fra variansen, men udelukkende gennem forskelle i deres haleegenskaber. Vi viser resultaterne for denne analyse sammen med dem fra (b) i tabel 2.

- (d) Vi gentager (b) og (c), men hvor vi forskyder fordelingen under H_1 således at VaR_α er ens for H_0 og H_1 . På den måde kan vi se, hvordan vores tests klarer det, hvis vi kun underestimerer ES, men ikke VaR. i opsummerer resultaterne i tabel 3.

ν	Signifikans niveau (%)	Dækning α (%)	Skala γ (%)	Styrke				
				Z_1 (%)	Z_2 (%)	Z_3 (%)	Z_4 (%)	$VaR_{1\%}$ (%)
5	4,1	5	21,9	3,3	51,1	24,2	29,0	38,2
		10	53	4,6	98,5	75,9	93,8	93,6
	10,6	5	21,9	12,0	69,1	45,6	43,3	55,3
		10	53	20,6	99,6	92,4	96,9	97,2
100	4,3	5	13,7	8,0	47,5	37,7	30,9	39,2
		10	34	27,0	97,8	93,9	94,6	94,7
	11,3	5	13,7	21,6	65,2	57,5	44,3	56,7
		10	34	57,0	99,3	97,9	97,4	97,9

Tabel 1: Resultater for (a).

Det ses fra tabel 1, at Z_2 klarer sig bedst af alle teststørrelser (endda bedre end VaR testen), når den sande model blot er en skaleret version af den prædictive model. Den sande models volatilitet er $\gamma > 1$ større end den prædictive models, som pr. konstruktion gør, så vores ES og VaR er underestimeret, og hyppigheden af VaR overskridelser er større. Disse typer misspecifikationer gør Z_2 i gennemsnit mere negativ, hvorved den vil indikere fejl. Z_4 er også en test om halen passer med den sande og Z_3 tester om hele fordelingen er prædikteret korrekt. Misspecifikation af volatilitet vil derfor også let kunne fanges af disse teststørrelser. Teststørrelserne vil som forventet se øget styrke, jo større misspecifikationer er, samt når signifikansniveauet øges. Vi bemærker, at styrken for Z_1 er en del lavere end for de andre tests i alle tilfælde. Dette skyldes, at siden VaR underestimeres vil der typisk være flere overskridelser end der teoretisk bør være. Dette tager Z_1 slet ikke hensyn til, og kombineret med, at vi har simultant fejlestimeret ES og VaR gennem volatiliteten, giver det mening, hvorfor styrken er så lav.

Det ses fra tabel 2, at Z_3 har særlig høj styrke sammenlignet med de andre teststørrelser, når vi betragter ændringer i frihedsgrader for t-fordelingen. Som før nævnt indikerer Z_3 , at der er et problem, når hele fordelingen ikke passer, ikke kun i halen. Da t-fordelinger er meget forskellige ved forskellige frihedsgrader, særligt pga. forskelle

	ν	Signifikans niveau (%)	ν	Styrke				
	H_0		H_1	Z_1 (%)	Z_2 (%)	Z_3 (%)	Z_4 (%)	$VaR_{1\%}$ (%)
t-fordeling	10	3,9	5	21,8 69,1	40,3 91,6	46,5 93,4	27,5 81,2	37,4 87,4
			10,7	5 3	41,9 85,6	59,5 96,4	64,0 97,1	41,9 90,7
	100	4,3	10	27,9 97,1	39,5 99,4	52,1 99,8	29,0 96,8	37,5 98,8
			11,3	10 3	46,6 99,1	57,1 99,8	67,8 99,9	43,4 99,0
	standardiseret t-fordeling	3,9	5	18,6 45,0	6,5 7,3	17,7 30,9	4,8 7,3	8,5 7,4
			10,7	5 3	33,3 58,9	16,0 15,4	30,2 41,0	12 19,0
	100	4,3	10	21,2 70,0	8,8 12,1	21,6 51,4	6,5 11,5	10,5 13,0
			11,3	10 3	35,3 78,3	18,7 21,3	34,1 58,7	14,3 27,1

Tabel 2: Resultater for (b) og (c).

i volatiliteten og tunghalethed, giver det mening, hvorfor Z_3 har en udmærket evne til at fange misspecifikationer her. Z_2 og Z_4 ser også ud til at være sensitiv overfor misspecifikationer i frihedsgrader, særligt når de er store. Dette sker selvfølgelig som følge af, at haleegenskaberne ændrer sig meget, hvilket resulterer i meget forskellige ES og VaR, når frihedsgraderne ændres. Vi illustrerer dette i tabel 3.

	ν	$VaR_{1\%}$	$VaR_{2,5\%}$	$ES_{2,5\%}$
t-fordeling	100	2,36	1,98	2,38
	10	2,76	2,23	2,82
	5	3,36	2,57	3,52
	3	4,54	3,18	5,04
Standardiserede t-fordeling	100	2,34	1,96	2,35
	10	2,47	1,99	2,52
	5	2,60	1,99	2,72
	3	2,62	1,84	2,90

Tabel 3: VaR og ES for t-fordelingen og standard t-fordeling [5].

Når vi benytter den standardiserede t-fordeling virker alle teststørrelserne til at have lavere styrke. Dette kan forklares ved, at for den standardiseret t-fordeling vil de misspecificerede ES og VaR ligge meget tæt på de sande ES og VaR, som også illustreret i tabel 3. Derudover er variansen for fordelinger præcis det samme. Vores prædiktive model minder derfor meget om den sande model, hvilket gør det svært for teststørrelserne at fange misspecifikationerne. For både t-fordelingen og den stan-

dardiserede t-fordeling ses det, at Z_1 klarer sig bedre her end i analysen før. Siden haleegenskaberne er anderledes i den prædiktive model, er forholdet mellem ES og VaR derfor meget forkerte, hvorfor Z_1 kan opnå høj styrke.

	ν H_0	Signifikans niveau (%)	ν H_1	Styrke				
				Z_1 (%)	Z_2 (%)	Z_3 (%)	Z_4 (%)	$VaR_{1\%}$ (%)
t-fordeling	10	3,9	5	27,8	8,5	26,0	7,0	11,7
			3	72,2	26,7	62,2	20,1	25,7
		10,7	5	45,0	19,4	39,6	16,5	24,2
	100	4,3	5	82,9	40,0	71,1	37,1	42,3
			3	92,0	37,9	80,2	31,1	33,9
		11,3	10	44,8	16,8	38,3	16,1	23,2
		3	94,7	50,4	84,3	51,6	50,8	
standardiserede t-fordeling	10	3,9	5	18,5	6,5	17,7	4,8	8,6
			3	42,6	14,0	38,0	9,3	14,0
		10,7	5	33,2	16,1	30,4	12,1	18,9
	100	4,3	5	57,5	25,8	50,3	20,6	27,4
			3	70,0	20,0	59,0	15,4	20,0
		11,3	10	35,4	15,3	30,5	13,0	19,6
		3	78,7	32,0	67,2	31,3	35,2	

Tabel 4: Resultater for (d).

I tabel 4 ses det, at når vi undersøger for t-fordelingen og VaR er korrekt estimeret, er det Z_1 og Z_3 , der har højest styrke blandt ES teststørrelserne. For Z_1 skal det bemærkes, at dens styrke i denne analyse generelt ligger højt og tæt på de resultater, vi fandt i (b) og (c). Dette skyldes den samme grund som vi beskrev tidligere. Resultaterne for Z_3 er også forventeligt siden forskydningen af den sande model ift. vores prædiktive model gør, at fordelingerne under H_0 og H_1 er meget forskellige.

Når vi undersøger den standardiserede t-fordeling under med denne type misspecifikation er det igen Z_1 og Z_3 , der udviser højest styrke. Forklaringen er præcis den samme som for den ikke-standardiserede t-fordeling. Det bemærkes igen, at styrken for alle testene er lavere når vi betragter den standardiserede t-fordeling, netop fordi deres egenskaber er mere ens, end hvis de ikke var standardiserede. Dette kan også ses i halen, hvor forskellene i ES og VaR under forskellige frihedsgrader ses at variere mindre, når vi betragter den standardiserede t-fordeling. Dette illustreres i tabel 5.

Samlet set ser vi, at teststørrelserne ser ud til at være sensitive overfor misspecifikationer, i hvertfald under iid antagelser. Det ses også, at ofte vil ES testene endda have bedre styrke-egenskaber end Basel VaR testen.

	ν	$VaR_{1\%}$	$VaR_{2,5\%}$	$ES_{2,5\%}$
t-fordeling	100	2,36	1,98	2,38
	10	2,52	1,98	2,57
	5	2,78	1,98	2,93
	3	3,34	1,98	3,84
Standardiserede	100	2,34	1,96	2,35
t-fordeling	10	2,44	1,96	2,49
	5	2,58	1,96	2,70
	3	2,75	1,96	3,04

Tabel 5: VaR og ES for t-fordelingen og standard t-fordeling, hvor fordelingerne er forskudt, så VaR_α er ens under både H_0 og H_1 [5].

6 Analyse af styrke under mere realistiske antagelser

I det tidligere afsnit så vi, at vores teststørrelser for ES ser ud til at besidde gode styrke-egenskaber, særligt hvis misspecifikationerne er store. Vores resultater blev dog beregnet under antagelsen, at vores finansielle tidsserie kommer fra en iid datagenereringsproces. Dette er som sagt en urealistisk antagelse. I dette afsnit vil vi derfor undersøge, hvor sensitive vores teststørrelser er, når vi betragter et mere realistisk setup. Dette vil vi gøre ved igen at foretage en sensitivitetsanalyse og undersøge vores teststørrelsers styrke.

6.1 Mere realistisk model

Vi antager nu, at vores finansielle tidsserie $\vec{X} = (X_t)_{t=1,\dots,T}$ ikke er iid fordelt og kommer fra en GARCH(1,1) model. Vi lader følgende realistiske GARCH model med standardiserede t-fordelte innovationer være vores **prædiktive model**:

$$X_t = \sigma_t Z_t, \quad \text{hvor } Z_t \sim t_\nu, \quad (6.1)$$

$$\sigma_t = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (6.2)$$

hvor vi lader $\omega = 0,01$, $\alpha_1 = 0,1$, $\beta_1 = 0,85$ og $\nu = 10$. Ligesom før genererer vi $M = 10000$ samples fra denne model af størrelse $T = 250$, som benyttes til at simulere teststørrelsernes fordeling under H_0 . Vi gør det samme for at simulere fordelingen under H_1 , hvor den sande model vælges til den samme som den prædiktive model, bortset fra nogle af parametrene varieres. Vi kan således se, hvor gode vores teststørrelser er til at afvise vores prædiktive model ved forskellige former for misspecifikationer. Da GARCH modellen defineres rekursivt, har vi brug for en burn-in periode. Vi lader denne periode være 250 lang, hvor vi vælger initialværdierne til den ubetinget varians $\sigma_0^2 = \frac{\omega}{1-\alpha_1-\beta_1}$ og ubetinget middelværdi $X_0 = 0$. Den 250 lange burn-in periode er der,

så vores samples fra GARCH modellerne ser tilfældige ud, og er ikke i så høj grad påvirket af vores valg af initialparametre.

Vi estimerer ES og VaR ved brug af vores prædiktive model både når vi sampler under H_0 og H_1 . Ved brug af positiv homogenitet kan de estimeres på følgende måde:

$$ES_{\alpha,t} = \hat{\sigma}_t ES_\alpha(Z_t), \\ VaR_{\alpha,t} = \hat{\sigma}_t VaR_\alpha(Z_t),$$

hvor $\hat{\sigma}_t^2$ er vores estimater for volatiliteten. Siden vi med vores prædiktive model ved, hvilken fordeling innovationerne har, kan vi beregne ES for innovationerne analytisk.

Når vi simulerer under H_0 kan vi direkte bruge de $\sigma_t = \hat{\sigma}_t$, som blev brugt til generering af \vec{X} til estimering af ES og VaR. Når vi simulerer under H_1 må vi selvfølgelig beregne volatilitetsestimatorne $\hat{\sigma}_t$ som om de kom fra den prædiktive model.

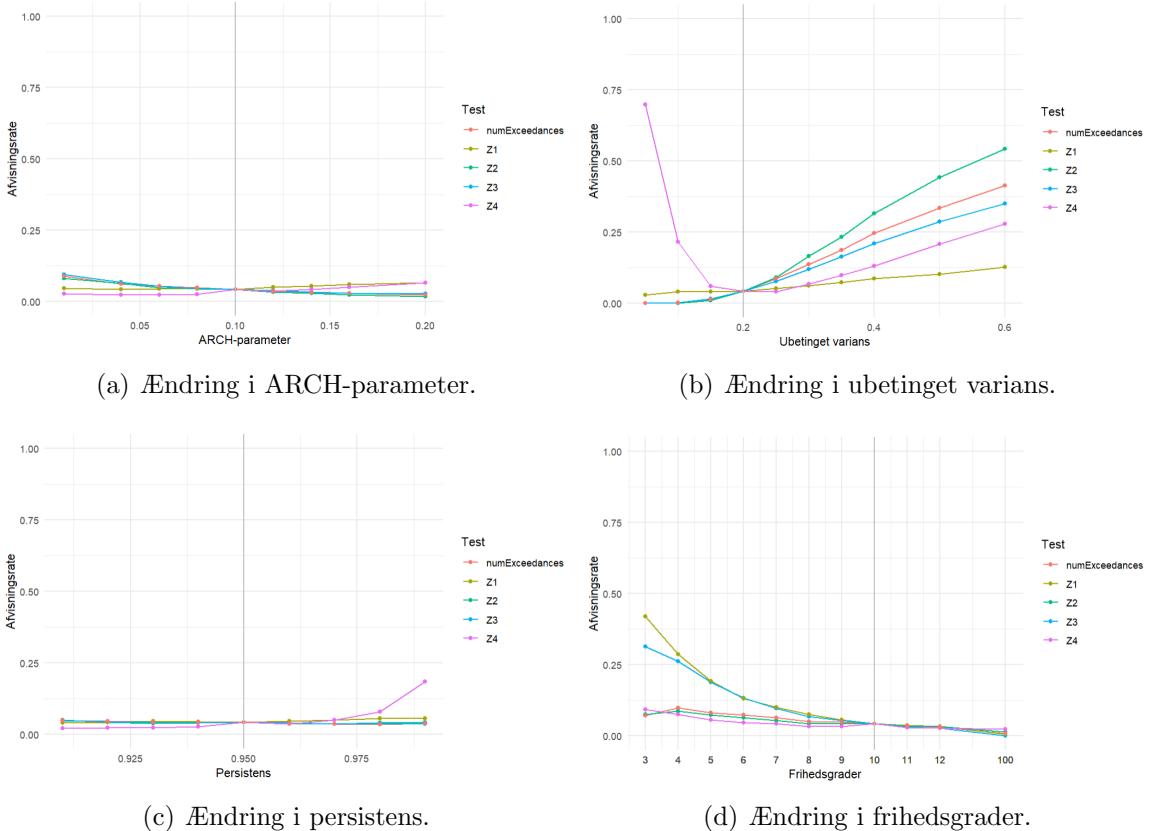
6.2 Resultater med GARCH setup

Vi undersøger sensitiviteten af vores teststørrelser ved at variere den prædiktive model på følgende måder (styrken for teststørrelserne beregnes på samme måde som tidligere):

- (a) Vi varierer ARCH-parameteren α_1 mellem 0, 01 og 0, 2 og bevarer persistensen for GARCH processen ved at justere β_1 . Det vil sige vi lader $\tilde{\alpha}_1 + \tilde{\beta}_1 = \alpha_1 + \beta_1 = 0,95$. Med denne type misspecifikation undersøger vi, hvor sensitive vores teststørrelser er, når den sande model påvirkes anderledes af stød. Hvis $\tilde{\alpha}_1 < \alpha_1$ reagerer den sande model mindre til stød og har mere konstant varians. Det er dog ikke klart, om dette vil føre til under- eller overestimering af ES og VaR.
- (b) Siden vi har valgt $\alpha_1 + \beta_1 < 1$ er den ubetinget varians endelig, og vi kan derfor undersøge teststørrelsernes sensitivitet, når den ubetinget varians $E[\sigma_t^2] = \omega/(1 - \alpha_1 - \alpha_2)$ varieres mellem 0, 05 og 0, 6. Dette gøres ved at ændre på ω og holde α_1 og β_1 konstant. Når $\tilde{\omega} > \omega$ vil den sande model have større ubetinget varians og vil derfor variere mere end den prædiktive model. Vi forventer i så fald, at vores prædiktive model underestimerer ES og VaR.
- (c) Vi varierer persistensen $\alpha_1 + \beta_1$ mellem 0, 91 og 0, 99. Dette gøres ved at lade $\tilde{\alpha}_1 = d\alpha_1$ og $\tilde{\beta}_1 = d\beta_1$, så den sande models persistens er $\tilde{\alpha}_1 + \tilde{\beta}_1 = d(\alpha_1 + \beta_1)$ for et $d > 0$. For at sørge for kun persistensen ændres, justerer vi også ω , så den ubetinget varians er konstant, dvs. $\tilde{\omega}/(1 - \tilde{\alpha}_1 - \tilde{\beta}_1) = \omega/(1 - \alpha_1 - \beta_1)$. Når persistensen for den sande model er større end vores prædiktive model vil stød påvirke i højere grad de efterfølgende afkast. Dette gælder for både positive og negative stød, hvorfor det ikke er klart, om ES og VaR under- eller overestimeres.

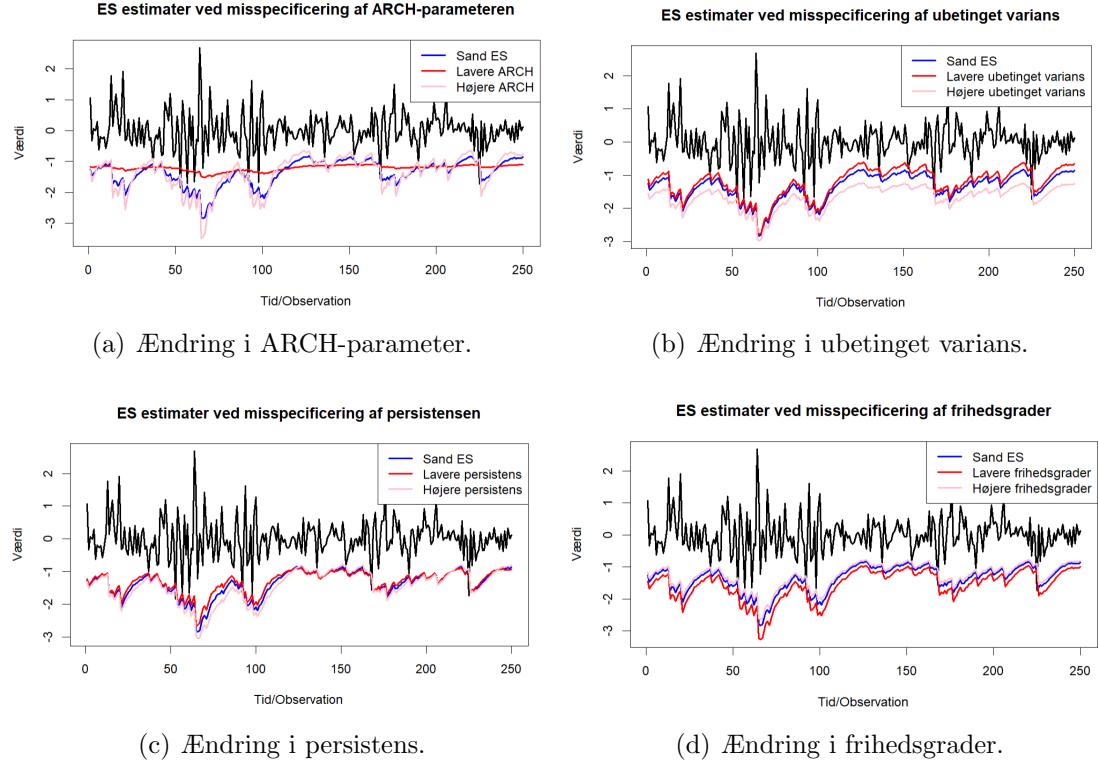
(d) Vi varierer innovationernes frihedsgrader ν mellem 3 og 100. Siden vi benytter den standardiserede t-fordeling vil andre parametre som den ubetinget varians ikke ændres, men haleegenskaberne vil. For lave frihedsgrader bliver den standardiserede t-fodeling mere tunghalet, og vi vil derfor forvente, den prædiktive model underestimerer ES og VaR.

For at få et godt overblik over styrken for testene, illustrerer vi denne gang sensitiviteten med et plot, som vist i figur 3. Den grå linje indikerer, hvilken værdi for den pågældende parameter, vores prædiktive model har.



Figur 3: Sensitivitet for ES teststørrelser.

Det ses i figur 3(a), at vores teststørrelser generelt udviser meget lav styrke, både når vi med vores prædiktive model under- og overestimerer ARCH-parametren. Dette havde vi allerede forventet, da lav $\tilde{\alpha}_1$ med høj $\tilde{\beta}_1$ betyder, at volatiliteten påvirkes i mindre grad af tidligere afkast, og mere af tidligere volatilitet. I så fald forventes afkastene i tidsserien at have mindre variation, men til gengæld er volatiliteten mere konstant, hvilket reflekteres i ES og VaR estimaterne. Samlet set giver dette både under- og overestimater for ES og VaR i begge tilfælde, hvorfor alle vores teststørrelser har lav styrke. I figur 4(a) illustrerer vi, hvordan ES estimaterne ser ud, når vi har valgt en sand model, men estimerer ES ved brug af prædiktive modeller, der hhv.



Figur 4: Påvirkningen af ES estimerer, når vi benytter en prædiktiv model, der er en misspecifieret version af den sande model. Vi betragter her både når den pågældende parameter over- og underestimeres af en prædiktiv model.

under- og overestimerer ARCH-parameteren. Dette sample stemmer overens med vores forklaring.

Figur 3(b) viser, at vores teststørrelser godt kan opnå høj styrke, særligt når vores misspecifikation er stor. Når datas sande fordeling har en større ubetinget varians, end vi forventer med vores prædiktive model, vil data variere mere end vi forventer, hvilket typisk giver anledning underestimering af VaR og ES. Det er derfor, når vi vælger F_t til at have en stor ubetinget varians, at mange af vores teststørrelser har høj styrke. Vores setup i denne undersøgelse minder meget om scenarie (a) i den simple analyse. Det er derfor ikke overraskende, at resultaterne har mange ligheder, hvor Z_2 udviser særligt høj sensitivitet efterfulgt af Z_3 og Z_4 , og Z_1 er igen mindst sensitiv. Forklaringen på dette er derfor identisk med vores forklaring for scenarie (a) fra iid-tilfældet. Vi bemærker, at CC-testen Z_4 fejlagtigt også udviser sensitivitet overfor overestimering af den ubetinget varians, og dermed ES og VaR, på trods af at være en ensidet test. Dette er formentlig forårsaget af, at når der ikke sker nogen VaR-overskridelser vil VaR delen af identifikationsfunktionen for (VaR, ES) nemt indikere fejl. I figur 4(b) illustrerer vi med en enkel sample, som umiddelbart bekræfter vores formodninger.

Det er tydeligt udfra figur 3(c), at alle vores teststørrelser heller ikke er sensitive overfor misspecifikering af persistensen. Som nævnt, hvis den prædiktive model un-

derestimerer persistensen, vil effekten af stød i den sande model aftage langsommere. Når vi overestimerer persistensen, vil effekten af stød hurtigt forsvinde. I begge tilfælde vil vi typisk hverken systematisk under- eller overestimere ES og VaR, siden disse effekter både går ud over positive og negative stød. Dette illustrerer vi i figur 4(c). Dermed vil teststørrelserne ikke se nogen høj styrke. Den eneste teststørrelse, der ser ud til at reagere på denne type misspecifikation er Z_4 , hvor den får højere styrke ved overestimering af persistensen. Ved meget høj persistens vil vedvarenhed af stødene være bemærkelsesværdig nok til, at testen i visse tilfælde kan opfange, at der er begået fejlestimering. Det samme mønster ses egentlig også for de andre tests, men i langt mindre grad.

Når vi betragter ændring i frihedsgrader i figur 3(d), ses det, at ved lavere frihedsgrader vil alle vores teststørrelser se øget styrke. Ligesom i tilfældet i den simple analyse (scenarie (b) og (c)) skyldes dette, at når den sande fordeling F_t har lavere frihedsgrader, vil vi med vores prædictive model underestimere, hvor tunghalet fordelingen i virkeligheden er. Dette giver anledning til underestimering af ES og VaR, hvilket vi også ser i figur 4(d). Siden denne analyse minder om scenarie (b) og (c) i den simple analyse, er det ikke overraskende, at Z_1 igen har højest styrke, selv når vi fjerner iid antagelsen. Begrundelsen for dette er dermed også det samme.

Vi ser samlet set, at vores teststørrelser kan være sensitive overfor misspecifikationer i den prædictive model, når vi betragter et mere realistisk tilfælde. Deres styrker er dog typisk ikke så høje, hvorfor testene ofte vil lave type 2 fejl. Til gengæld ses det igen, at testene kan udvise meget sensitivitet, når misspecifikationerne er meget store.

7 Afvisningsrater med ikke-parametriske modeller

Vi har indtil videre undersøgt sensitiviteten for vores teststørrelser, når vi benytter parametriske modeller, hvilket lader os kontrollere, hvilken type misspecifikation vi har begået med vores prædictive model. I praksis benytter finansielle institutioner ofte ikke-parametriske modeller til estimering af ES og VaR. Her er det ikke lige så oplagt, hvilken type misspecifikation der er blevet lavet. Vi vil i dette afsnit undersøge den mest udbredte metode til estimering af ES og VaR nemlig historisk simulation. Vores mål er at undersøge, om denne metode giver en gode nok prædiktioner af ES og VaR og se, om vores ES teststørrelser kan fange det. Som nævnt vil vi undersøge to varianter af historisk simulation: Den klassiske historisk simulation og vægtet historisk simulation [11].

7.1 Historisk simulation

Hovedideen bag HS er, at vi forestiller os, at fortidig data kan benyttes til simulering af, hvordan fremtidig data vil se ud. Vi kan herfra empirisk estimere ES og VaR. Den præcise metodologi for HS og vægtet HS beskriver vi i større detalje i det følgende.

Vi betragter det finansielle tidsserie for afkast X_1, \dots, X_N i perioderne $1, \dots, N$. Lad dette være det in-sample data, som vi vil benytte til at give et one-step-ahead estimat for ES og VaR. Hvis vi står i tid N , ønsker vi altså at give estimaterne $ES_{\alpha, N+1}$ og $VaR_{\alpha, N+1}$. Ideen bag den klassiske HS er, at vi forestiller os, at afkast i morgen kan modelleres med en fordeling, som kan antage værdierne X_1, \dots, X_N hver med sandsynlighed $1/N$. Lad følgende betegne afkastraterne sorteret fra mindst til størst $X_{(1)}, \dots, X_{(N)}$. For at estimere VaR_α vil vi summe $1/N$ indtil vi lige præcis rammer eller overskrider $\alpha = 2,5\%$. I tilfældet hvor vi summer n gange og når lige præcis $2,5\%$, dvs. $S_n = \sum_{i=1}^n 1/N = 2,5\%$, er vores VaR estimat $VaR_{\alpha, N+1} = -X_{(n)}$. Vi har da empirisk estimeret og negeret α -fraktilen i X_1, \dots, X_N . I praksis sker $S_n = 2,5\%$ ikke, hvis vi vælger vores in-sample størrelse til f.eks. $N = 250$. I tilfældet at S_n er lige præcis over $2,5\%$, må vi benytte lineær interpolation. Hvis $X_{(n)}$ er det afkast lige over niveau $2,5\%$, vil vores $VaR_{\alpha, N+1}$ estimat være:

$$VaR_{\alpha, N+1} = - \left(X_{(n-1)} + \frac{X_{(n)} - X_{(n-1)}}{S_n - S_{n-1}} (\alpha - S_{n-1}) \right). \quad (7.1)$$

$ES_{\alpha, N+1}$ kan da estimeres empirisk ved gennemsnittet af de værste afkast divideret med α :

$$ES_{\alpha, N+1} = - \frac{\frac{1}{N}X_{(1)} + \dots + \frac{1}{N}X_{(n-1)} + (\alpha - S_{n-1})X_{(n)}}{\alpha}. \quad (7.2)$$

Bemærk, at vi her har justeret vægten for $X_{(n)}$, så vi får et bedre estimat for ES ved niveau α [11].

7.2 Vægtet historisk simulation

Den klassiske HS bruger antagelsen, at X_{N+1} kan antage værdierne X_1, \dots, X_N med lige stor sandsynlighed $1/N$. Dette implicerer umiddelbart, at man forventer, at alle de tidligere N afkastrater er lige gode indikationer på, hvad afkast er for i morgen. Boudoukh, Richardson og Whitelaw (1998) foreslår, at de seneste data bør gives mere vægt end gammelt data, da de giver bedre indikation på, hvordan volatiliteten og den økonomiske situation er i dag. Dette tager vægtet HS hensyn til, hvor i stedet for X_{N+1} antager værdien $X_n \in \{X_1, \dots, X_N\}$ med sandsynlighed $1/N$, benyttes sandsynlighederne $\lambda^{N-n}(1-\lambda)/(1-\lambda^N)$ for $\lambda \in (0, 1)$. Her vil sandsynligheden for, at X_{N+1} antager værdien X_n være λ større end for X_{n-1} , altså de seneste værdier får

relativ større vægt sammenlignet med gamle. VaR og ES estimeres da på samme måde som for den klassiske HS med de vægtede sandsynligheder [11].

7.3 Resultater med ikke-parametriske modeller

Vi betragter følgende GRACH(1,1)-model med standard t-fordelte innovationer, som beskriver den **sande datagenereringsproces**:

$$X_t = \sigma_t Z_t, \quad \text{hvor } Z_t \sim t_{10}, \quad (7.3)$$

$$\sigma_t = 0,01 + 0,1 X_{t-1}^2 + 0,85 \sigma_{t-1}^2, \quad (7.4)$$

Vi simulerer $M = 10000$ samples fra denne model af størrelse $T = 500$. For hvert sample kan vi estimere ES og VaR ved brug af HS og vægtet HS med et rullende vindue og $N = 250$ data in-sample. Samtidig vil vi beregne de ideelle estimatorer for ES og VaR, hvor vi benytter information om den sande datagenereringsproces. Vi kan da beregne en afvisningsrate, dvs. hvor ofte HS og vægtet HS bliver afvist som prædiktiv model for data. Vi benytter $\lambda = 0,995$ for vægtet HS.

For at udfordre HS vil vi også betragte et tilfælde, hvor data i en lille periode i virkeligheden kommer fra en datagenereringsproces med større volatilitet og flere ekstreme udsving. For at modellere dette vil vi lade afkastraterne X_{351}, \dots, X_{380} komme fra følgende GARCH(1, 1) datagenereringsproces:

$$X_t = \sigma_t Z_t, \quad \text{hvor } Z_t \sim t_3, \quad (7.5)$$

$$\sigma_t = 0,05 + 0,1 X_{t-1}^2 + 0,85 \sigma_{t-1}^2, \quad (7.6)$$

hvor vi har lavere frihedsgrader på innovationerne for at få tungere haler, og volatilitten er større. Vi gentager derefter dette, men hvor vi tilføjer endnu en volatil perioden. Vi vil her også lade datagenereringsprocessen for X_{421}, \dots, X_{450} beskrives ved (7.5) og (7.6). Disse scenarier kan fortolkes som, at der i en periode sker uroligheder, som fører til nogle stød. Bemærk i begge tilfælde har vi valgt, så stødene sker i de sidste 250 dage af vores $T = 500$ sample. Vi forventer derfor, at HS og vægtet HS vil underestimere VaR og ES.

Vores resultater vises i tabel 6, hvor afvisninsraterne beregnes på samme måde som styrke i vores tidligere analyser. Bemærk vi ikke bruger Z_3 , da den kræver en parametrisk prædiktiv model.

Vi bemærker, at i alle tilfælde bliver den vægtet HS sjældnere afvist end den almindelige HS. Dette er præcis som forventet, da de seneste afkastrater er mere repræsentativ for i morgen pr. konstruktion af GARCH modellen, hvilket gør vægtet HS til

Sand model	Prædiktiv model	Signifikans niveau (%)	Afvisningsrate			
			Z_1 (%)	Z_2 (%)	Z_4 (%)	$VaR_1\%$ (%)
GARCH uden ekstra volatil periode	HS	4,0	7,9	12,9	14,8	7,1
	Vægtet HS	4,0	6,7	8,2	9,7	3,7
GARCH med ekstra volatil periode	HS	4,1	6,7	26,4	24,7	14,0
	Vægtet HS	4,1	6,1	16,5	15,4	6,4
GARCH med ekstra to volatile perioder	HS	4,0	3,3	44,2	43,0	25,7
	Vægtet HS	4,0	3,6	32,3	30,6	13,2

Tabel 6: Afvisningsrater for HS og vægtet HS.

en bedre prædiktiv model. Vi får allerede her en indikation på, at vores teststørrelser kan skelne mellem upræcise og mere præcise ikke-parametriske prædiktive modeller.

Det ses generelt, at jo flere volatile perioder vi inkluderer, jo større bliver afvisningsraterne. For at HS virker godt skal tidligere data være en god repræsentation af, hvordan fremtidig data ser ud. De ekstra volatile perioder, som vi har indskudt, gør altså data mere uforudsigtigt, hvilket forværret både den klassiske HS' og vægtet HS' præstation. Vi ser derfor igen, at store misspecifikationer er teststørrelserne bedre til at fange. Den eneste teststørrelse, der opfører sig anderledes er Z_1 , hvor begge HS metoder bliver lidt sjældnere afvist, når den sande model inkluderer ekstra volatile perioder. Dette kan skyldes numeriske forskelle samt, at teststørrelsen er afhængig af en nogenlunde korrekt estimeret VaR, hvilket slet ikke er tilfældet, når vi indskyder uforudsigtige perioder. Derfor forbliver afvisningsraten lav. Bemærk at Z_2 og Z_4 afviser de prædiktive modeller oftere end VaR testen, hvilket tyder på, at disse ES tests er bedre til at fange upræcise estimerater fra HS.

8 Inspektion af CC-testen

Vi har i tidligere analyser set, at CC-testen generelt udviser styrke-egenskaber, der er sammenlignelige med de øvrige teststørrelser. En særlig egenskab ved CC-testen er dog, at den, som nævnt i afsnit 4.4, har en asymptotisk fordeling. Det kan derfor umiddelbart virke som om, testen er særlig anvendelig i praksis, da man på enkel vis kan beregne en p-værdi uden behov for tunge simulationer. Dette er især relevant i finansiell praksis, hvor metoder som historisk simulering (HS) ofte anvendes. Da HS ikke bygger på en parametrisk model, og den sande fordeling er ukendt, er simulering typisk ikke mulig. Dette er i modsætning til vores tidligere analyser, hvor vi enten anvendte en parametrisk prædiktiv model eller kendte den sande datagenereringsproces.

8.1 Empirisk applikation

For at demonstrere den umiddelbare fordel ved CC-testen udfører vi en backtest med både HS og GARCH modeller som prædiktive modeller for S&P 500-data, hvorefter vi diskuterer vores metode. Denne tilgang er inspireret af analysen af Nolde og Ziegel (2017). Vi har denne gang valgt at bruge reel data, da det demonstrerer den reelle udfordring ved ikke at kende den sande datagenereringsproces, hvilket fremhæver CC-testens særlige egenskab. Derudover, havde vi i stedet selv genereret data fra f.eks. en GARCH model, ville en efterfølgende fit af samme type model naturligt føre til gode resultater. I så fald er det meningsløst at kigge på parametriske modeller som prædiktive modeller, medmindre vi bevidst havde misspecifieret modellen.

Vi henter justeret lukkepris for S&P 500 fra 23-02-2021 til 01-01-2025 og omregner det til log-afkast. Dette giver os et 6000 stort datasæt. Vi kan nu benytte begge varianter af HS med et rullende vindue med 1000 datapunkter in-sample og estimere ES og VaR en dag forud. Dette giver os ES og VaR estimerater for de sidste 5000 dage. Det samme setup bruges når vi estimerer ES og VaR med GARCH modeller, hvor vi for hvert vindue fitter en GARCH(1,1) model med standard t-fordelte innovationer. Dette benyttes til at give et one-step-ahead forecast for volatiliteten, som sammen med den fittede fordeling af innovationerne, kan benyttes til estimering af ES og VaR, ligesom vi gjorde i afsnit 6. Ved brug af Z_4 og dens asymptotiske fordeling får vi p-værdierne som vist i tabel 7.

Prædiktiv model	p-værdi (%)	Afvist med 5% signifikansniveau?
HS	6,8	ikke afvist
Vægtet HS	72,9	ikke afvist
GARCH fits	5,5	ikke afvist

Tabel 7: p-værdier for HS og vægtet HS.

Vi ser altså, at alle p-værdier er over 5% signifikansniveau, hvorfor vi ikke kan afvise nogen af modellerne. Det skal dog bemærkes, at ES og VaR estimeraterne kommende fra vores GARCH(1,1) fits giver en p-værdi meget tæt på 5% og dermed meget tæt på at være signifikant. Vi ville i så fald have afvist modellen. Dette kan potentielt skyldes, at vi bare har fittet en GARCH(1,1) model over hvert vindue uden at have undersøgt nærmere, om det giver en god beskrivelse af data.

Vi har her nemt opstillet en analyse, hvor vi kan afgøre, om både parametriske og ikke-parametriske metoder er gode prædiktive modeller for S&P 500 data. Den samme undersøgelse kan i princippet også udføres for de andre teststørrelser, men kun for den parametriske model. Dette kræver jo som sagt simulation under nulhypotesen, hvilket er langt mere beregningstungt.

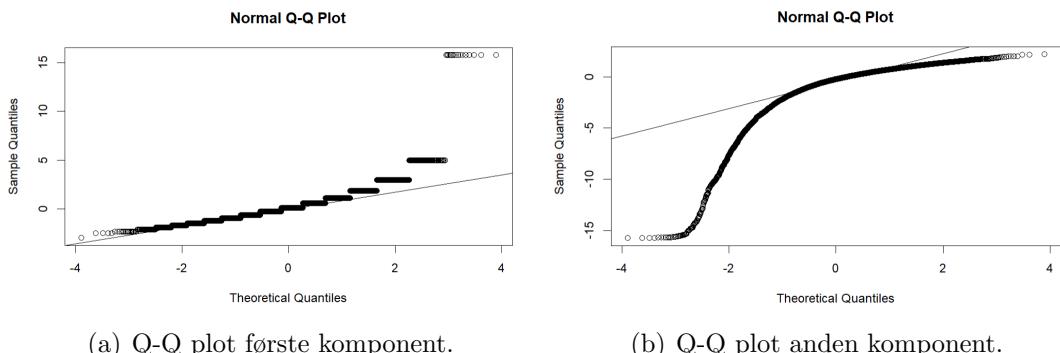
8.2 Stikprøvestørrelsen

Z_4 's asymptotiske fordeling virker umiddelbart til at være en meget stor fordel over de andre teststørrelser i praksis baseret på vores analyse i 8.1. Vi vil dog påstå, dette er til diskussion. Husk vi i 8.1 brugte en tidsserie med 6000 observationer, hvor vi udregnede ES og VaR estimatorer for de seneste 5000 dage. Nolde og Ziegel (2017) bruger også en lignende stikprøvestørrelse for deres analyse. En stor stikprøvestørrelse er nødvendig for, at asymptotiske resultater kan anvendes. Spørgsmålet er imidlertid, hvor stor er stor nok? For at undersøge dette, betragt et tilfælde, hvor bruger følgende prædiktive model til at beskrive noget data:

$$X_t = \sigma_t Z_t, \quad \text{hvor } Z_t \sim t_{10}, \quad (8.1)$$

$$\sigma_t = 0,01 + 0,1 X_{t-1}^2 + 0,85 \sigma_{t-1}^2. \quad (8.2)$$

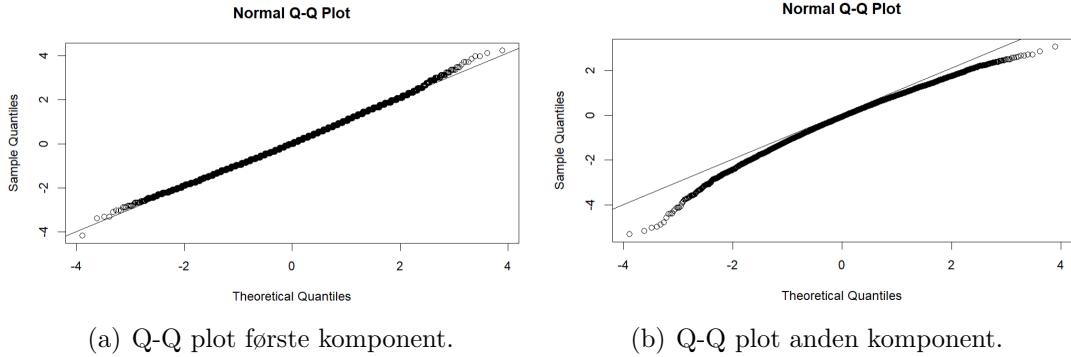
Vi gentager beregningerne fra tidligere afsnit, hvor vi sampler 10000 gange fra denne model af størrelse 250 og beregner ES og VaR ved brug af modellens parametre for hvert sample. Vi kan da simulere Z_4 's to komponenters fordelinger under den prædiktive model. Vi sammenligner hver af komponenternes fordelinger med normalfordelingen i Q-Q plots som vist i figur 5.



Figur 5: Q-Q plots for med stikprøvestørrelse 250.

Det ses tydeligt, at når vi kun benytter en stikprøvestørrelse på 250 er Z_4 's to komponenter slet ikke normalfordelte. Det var netop på grund af dette, at vi i de tidligere afsnit undlod at bruge Z_4 's asymptotiske fordeling til beregning af styrke, på trods af det var lettere. Det vil også sige, at den øvelse vi lavede i 8.1 egentlig ikke er mulig, hvis vi kun er interesseret i de seneste 250 dage svarende til ca. et år talt i handelsdage. Selv når vi lader stikprøvestørrelsen være 5000, ligesom vi gjorde i 8.1, fås Q-Q plotsene som vist i figur 6.

Det ses, at første komponent for Z_4 ser ud til at være nogenlunde standard normalfordelt, ligesom man kunne have håbet. For den anden komponent er situationen



Figur 6: Q-Q plots for med stikprøvestørrelse 5000.

imidlertid anderledes, idet fordelingen er let venstreskæv. Vores simulerede fordeling har konsekvent lavere fraktiler end de tilsvarende fraktiler i normalfordelingen. Det ser altså ud til, at ikke engang en 5000 stor stikprøvestørrelse, som Nolde og Ziegel endda selv bruger, er nok til at være sikker på, at brugen af asymptotiske resultater er forsvarligt. Kravet på en mindst lige så stor stikprøvestørrelse gør dette egenskab for CC-testen uinteressant i praksis for finansielle institutioner og regulatorer, da 5000 dage svarer til ca. 20 år. I praksis er det mere almindeligt at interessere sig for at evaluere ES og VaR estimatorer for de seneste et eller to år, dvs. de seneste ca. 250-500 dage. I så fald er simulation alligevel nødvendigt for at beregne en p-værdi for Z_4 , ligesom for de andre teststørrelser.

9 Diskussion

En stor del af vores undersøgelse om ES teststørrelsernes anvendelighed drejer sig om deres sensitivitet overfor misspecifikationer i både et iid og GARCH setup. Resultaterne minder meget om hinanden, hvilket er tydeligst, når vi sammenligner, hvad der sker, når vi ændrer på volatiliteten eller frihedsgraderne. Vi ser umiddelbart også, at vores teststørrelser sagtens kan opnå høj styrke, hvis misspecifikationerne er store i begge tilfælde. Vi betragtede GARCH modeller og t-fordelinger, da vi nemt kan undersøge egenskaber som tunghæthed og volatilitet. Til gengæld er vores resultater kun meningsfuldt for et begrænset antal modeller. F.eks. er det ikke trivielt, at vi ville have set samme resultater, hvis vi betragtede skæve t-fordelinger for at introducere noget asymmetri, ligesom Nolde og Ziegel (2017) gør. Vi kunne også have undersøgt andre typer af modelklasser, da vi betragtede ikke-iid tilfældet, f.eks. ved at benytte EGARCH, ligesom Bayer og Dimitriadis (2020) gør. På denne måde undersøger vi ES teststørrelsernes styrke-egenskaber for en bredere vifte af modeller, som finansielle tidsserier kan komme fra.

I vores empiriske analyse brugte vi log-afkast, både fordi det nemt kan fortolkes, og fordi det potentielt gør vores tidsseriedata stationært, hvilket ville gøre vores brug af GARCH modeller forsvarligt. I praksis kan en hypotesetest afgøre, om en enkel differencing var nok til at sikre stationaritet ([8], s. 338-341). Da vi ikke gjorde dette, kunne det være et af grundene til, at vores fittede GARCH modeller næsten blev afvist. Men siden vi fittede for hvert rullende vindue, ville vi skulle lave sådanne hypotesetests 5000 gange. Fravalget at gøre dette var derfor for simplicitetens skyld. Alternativt kunne man bruge de første 1000 dage til at konstruere en prædiktiv model, som bruges til ES og VaR estimering over de næste 5000 dage. Da behøves kun en enkel hypotesetest. Dette er dog uinteressant i praksis, siden det ville betyde, at en finansiel institution bruger den samme prædiktive model i ca. 20 år.

Andre måder at tilgå vores empiriske analyse på er at benytte andre teststørrelser med asymptotiske fordelinger, f.eks. nogle af Bayer og Dimitriadi's (2020) regression-baseret teststørrelser. Vi kan her undersøge, om de konvergerer mod deres asymptotiske fordeling hurtigere. Vi besluttede dog ikke at bruge dem, da de krævede meget computerkraft og lang tid at beregne.

I vores HS analyse ser vi, at teststørrelserne en gang imellem kan fange, når den sande datagenereringsproces' frihedsgrader og volatilitet ændrer sig, selv i små perioder. Vi så et lignende resultat fra vores parametriske analyse. Vi undersøgte dog ikke andre egenskaber som persistens eller ARCH-parameteren ligesom før. Vores fravælg af dette var grundet en forventning om at se lignende resultater som før, men for at være helt sikker bør en nærmere analyse foretages. Vi vil til slut bemærke, at vi benyttede $\lambda = 0,995$ for vægtet HS, da Hull (2015) også brugte samme værdi. I praksis bør en dybere undersøgelse for λ foretages for at optimere præstationen af vægtet HS. I vores analyse var $\lambda = 0,995$ nok til at demonstrere fordelen ved vægtet HS, men vi forventer dette resultat kan gøres tydeligere, hvis λ var valgt mindre arbitraert.

10 Konklusion

For at undersøge anvendeligheden af de forskellige ES teststørrelser til backtesting af ES har vi udført en sensitivitetsanalyse af dem. Vi har her både undersøgt et iid setup og et GARCH setup. Formålet med iid setuppen er blot, at få et nemt overblik over, hvordan mekanismerne fungerer, mens med GARCH setuppen ser vi om resultaterne også gælder i et mere realistisk setup. Vi ser i begge tilfælde, at teststørrelserne kan have det svært med at fange små misspecifikationer, og nogle typer af misspecifikationer fanges bedre end andre. Teststørrelserne klarer sig typisk bedst, når vores prædiktive model misspecifierer parametre som den langsigtede volatilitet eller frihedsgraden

på innovationerne. Dette sker, da disse parametre har en mere direkte påvirkning på ES og VaR estimaterne. Vi har også undersøgt ikke-parametriske prædiktive modeller, mere specifikt to varianter af historisk simulation. De to HS metoder bliver ikke så ofte afvist af teststørrelserne, når data er mere forudsigtig, hvilket tyder på, at HS somme tider kan fungere som en god prædiktiv model for ES og VaR. Når datagenereringsprocessen er meget uforudsigtig, aferiser teststørrelserne modellerne oftere. Vi ser altså, at teststørrelserne sagtens også kan backteste ikke-parametriske modeller. Samlet set ser ES teststørrelserne ud til at have rimelige styrke-egenskaber og er sammenlignelige med VaR testen. De kan altså sagtens fungere effektivt i praksis til at fange upræcise modeller.

På overfladen kan det virke som om, at Z_4 er en særlig imponerende teststørrelse, da dens asymptotiske fordeling kan gøre implementeringen af den en del lettere. Vi ser dog, at dette egentlig ikke er tilfældet grundet dens krav på en enorm stikprøvestørrelse.

Appendix

Bevis for proposition 2.5. (inspireret af McNeil et al., side 70 [4]):

Bevis. Lad F være en kontinuert og strengt voksende fordelingsfunktion. Lad $U : \Omega \rightarrow [0, 1]$ være en uniform fordelt stokastisk variabel på $[0, 1]$. Da har den stokastiske variabel $X = F^{-1}(U)$ fordelingsfunktion F . Det ses, at da F er strengt voksende, er:

$$-E\left[\frac{X I}{\alpha}\right] = \frac{-E[F^{-1}(U) 1(F^{-1}(U) - F^{-1}(\alpha) < 0)]}{\alpha} = \frac{-E[F^{-1}(U) 1(U < \alpha)]}{\alpha}.$$

Det bemærkes at $-\frac{1}{\alpha}E[F^{-1}(U) 1(U < \alpha)] = -\frac{1}{\alpha} \int_0^\alpha F^{-1}(q)dq$ og vi har nu første lighed. For at se den anden lighed, betragt partitionen af Ω rummet $\{X + VaR_\alpha < 0\}$ og $\{X + VaR_\alpha < 0\}^c$, og lad \mathcal{G} være den mindste σ -algebra genereret af partitionen. Fra defintionen af betinget forventning er:

$$E[E[X | \mathcal{G}] 1(X + VaR_\alpha < 0)] = E[X 1(X + VaR_\alpha < 0)] = E[X I].$$

Da $E[X | \mathcal{G}]$ er \mathcal{G} -målelig er den konstant for $\omega \in \{X + VaR_\alpha < 0\}$. Dermed kan venstre side omskrives til

$$E[E[X | \mathcal{G}] 1(X + VaR_\alpha < 0)] = E[X | \mathcal{G}] E[1(X + VaR_\alpha < 0)] = E[X | \mathcal{G}] \alpha,$$

for $\omega \in \{X + VaR_\alpha < 0\}$ og det ses let, at vi nu har vist den anden lighed. \square

Bevis for proposition 3.4. (inspireret af McNeil et al., side 114-119 [4]):

Bevis. Vi ønsker at vise begge implikationer.

” \implies ”: Antag, at GARCH(1,1) processen er kovarians stationær. Da er den ubetinget varians og middelværdi ens for alle X_t . Bemærk første og andet moment eksisterer og $E[X_t] = E[\sigma_t]E[Z_t] = 0$. Dermed er variansen lig det andet moment $\sigma_X^2 = E[X_t^2]$. Der gælder derudover $E[Z_t^2] = 1$, da innovationerne er en SWN(0, 1). Se desuden:

$$E[X_t^2] = E[\sigma_t^2]E[Z_t^2] = E[\sigma_t^2] = \omega + \alpha_1 E[X_{t-1}^2] + \beta_1 E[\sigma_{t-1}^2].$$

Bemærk at $\sigma_X^2 = E[X_t^2] = E[\sigma_t^2]$. Da vi har antaget stationaritet fås:

$$\sigma_X^2 = \omega + \alpha_1 \sigma_X^2 + \beta_1 \sigma_X^2.$$

Det ses nu let, at $\sigma_X^2 = \omega/(1 - \alpha_1 - \beta_1)$, og den er endelig hvis $\alpha_1 + \beta_1 < 0$.

” \Leftarrow ”: Antag, at $\alpha_1 + \beta_1 < 1$. Vi omskriver udtrykket for volatiliteten for GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha_1 Z_{t-1}^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + (\alpha_1 Z_{t-1}^2 + \beta_1) \sigma_{t-1}^2.$$

Dette kan ses som en *stokastisk rekurrens relation* af form $Y_t = A_t Y_{t-1} + B_t$. Den har en løsning, hvis $E[\max\{0, \ln |B_t|\}] < \infty$ og $E[\ln |A_t|] < 0$. I så fald er løsningen:

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}.$$

Lad $B_t = \omega$ og $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$. Det ses let, at $E[\max\{0, \ln |\omega|\}] < \infty$ samt:

$$E[\ln(\alpha_1 Z_{t-1}^2 + \beta_1)] \leq \ln(E[\alpha_1 Z_{t-1}^2 + \beta_1]) = \ln(\alpha_1 + \beta_1) < 0,$$

ved brug af Jensens ulighed og at $\alpha_1 + \beta_1 < 1$. Løsningen til rekurrens ligningen er da:

$$\sigma_t^2 = \omega + \sum_{i=1}^{\infty} \omega \prod_{j=0}^{i-1} (\alpha_1 Z_{t-j-1}^2 + \beta_1) = \omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta_1) \right).$$

Vi ser da:

$$\begin{aligned} E[X_t^2] &= E[\sigma_t^2] E[Z_t^2] \\ &= E\left[\omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta_1) \right)\right] \\ &= \omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 + \beta_1) \right) \\ &= \omega \left(1 + \sum_{i=1}^{\infty} (\alpha_1 + \beta_1)^i \right). \end{aligned}$$

Siden $\alpha_1 + \beta_1 < 1$ har denne geometriske række en endelig sum:

$$E[X_t^2] = \omega \left(1 + \frac{1}{1 - \alpha_1 - \beta_1} - 1 \right) = \frac{\omega}{1 - \alpha_1 - \beta_1},$$

for alle t . GARCH(1,1) processen er således kovarians stationær. \square

Litteratur

- [1] McNeil, A.J., Frey, R. (2000): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach" side 1-30, Journal of Empirical Finance.
- [2] Emmer, S., Kratz, M. & Tasche, D. (2015): "What is the Best Risk Measure in Practice? A comparison of standard measures" (side 1-27), SSRN.
- [3] Artzner, P., Delbaen, S.F., et al. (1998): "Coherent Measures of Risk" side 1-24, ETH Zürich.
- [4] McNeil, A.F., Frey, R. & Embrechts, P. (2015): "Quantitative Risk Management: Concepts, Techniques and Tools" kapitler 2-4, Princeton University Press.
- [5] Acerbi, C. & Szekely, B. (2014): "Backtesting Expected Shortfall" (side 1-37), MSCI.
- [6] Bayer, S., Dimitriadis, T. (2020): "Regression-Based Expected Shortfall Backtesting" side 1-41, Journal of Financial Econometrics.
- [7] Chang, C., et al. (2018): "Choosing expected shortfall over VaR in Basel III using stochastic dominance" side 95-97, International Review of Economics and Finance.
- [8] Ruppert, D., Matteson, D.S. (2015): "Statistics and Data Analysis for Financial Engineering" side 307-341, Springer Texts in Statistics.
- [9] Jiang, X. og Neapolitan, R.E. (2007): "Probabilistic Methods for Financial and Marketing Informatics" kapitel 6, Morgan Kaufmann.
- [10] Nolde, N. og Ziegel, J. (2017): "Elicitability and Backtesting: Perspectives for Banking Regulation" i The Annals of Applied Statistics side 1833-1872, Institute of Mathematical Statistics.
- [11] Hull, C.J. (2015): "Risk Management and Financial Institutions Fourth Edition" side 277-285, Wiley.



Deklaration for anvendelse af generative AI-værktøjer (studerende)

På Københavns Universitet udfører vi vores arbejde med **ansvarsfølelse og respekt for samfund, kulturarv, miljø og mennesker** omkring os.

Integritet, ærlighed og transparens er forudsætninger for akademisk arbejde. Vi forventer derfor, at eksamenspræstationer afspejler **den studerendes egen læring og selvstændige indsats**.

Akademisk arbejde baserer sig altid på andres indsigter, viden og bidrag, men altid med **grundig anerkendelse, respekt og kreditering** af dette arbejde.

Dette gælder også ved brug af generativ kunstig intelligens.

Vejledning

Brug af generativ AI ved eksamen

I henhold til KU's regler for brugen af værktøjer, der er baseret på generativ AI (GAI), skal du være transparent om din anvendelse af teknologien, fx i dit metodeafsnit og/eller ved at udfylde og vedlægge nedenstående deklarationsskabelon som bilag til skriftlige opgavebesvarelser.

Når du skriver din deklaration, er det vigtigt, at læseren får et tydeligt billede af, om og hvordan generativ AI har bidraget til det endelige produkt.

Hvis det er besluttet, at du skal bruge skabelonen i dit fag, skal du også benytte den, når du *ikke* har anvendt GAI-værktøjer som hjælpemiddel. I dette tilfælde skal du dog blot krydse af, at du *ikke* har brugt GAI, og behøver ikke at udfylde resten.

Ved at deklarere din brug af GAI-værktøjer sikrer du, at der ikke opstår udfordringer i forhold til reglerne om eksamenssnyd.

I kurser, hvor brugen af GAI er integreret i fagligheden, kan refleksion og kritisk vurdering af anvendelsen af GAI-værktøjer også indgå som en del af et metodeafsnit i din opgave. Spørg din underviser eller vejleder, hvis du er tvivl, om det er tilfældet i dit kursus.

Hvis generativ AI er objekt for din undersøgelse, vil det fremgå af dine forskningsspørgsmål, din metodebeskrivelse, din analyse og konklusion, hvilken rolle GAI spiller i din opgave. Hvis du

samtidig også bruger GAI som hjælpemiddel i processen, skal du deklarere denne anvendelse særskilt.

Opmærksomhedspunkter:

- Hvis GAI er et tilladt hjælpemiddel på dit kursus, må du anvende GAI til dialog og sparring under udarbejdelsen af din opgave, men du må **ikke** overlade udfærdigelsen af din opgavebesvarelse til GAI-værktøjer, selvom alle hjælpemidler er tilladt.
- Hvis materiale fra GAI inkluderes som kilde (direkte eller i redigeret form) i din besvarelse, gælder de samme krav om brug af citationstegn og kildehenvisning som ved alle andre kilder, da der ellers vil være tale om plagiat.
- Brug aldrig personhenførbare, ophavsretsbeskyttede eller fortrolige data i et AI-værktøj.
- Husk altid at undersøge gældende regler og retningslinjer for brug af generativ AI på KU.
- Læs kursusbeskrivelsen grundigt. Det er vigtigt at du ved, hvilke anvendelser der er tilladt i dit kursus. Der kan eventuelt være yderligere krav om dokumentation, fx at du skal beskrive dine centrale prompts og evt. kildemateriale (hvad har du givet af kontekst, hvad har du fodret værktøjet med, hvad har du bedt værktøjet om at gøre), beskrive outputtet (hvilke svar du fik af værktøjet), beskrive processen, f.eks. historik og iterationer (hvis du har skrevet frem og tilbage med værktøjet ad flere omgange for at komme frem til et brugbart svar).
- Tal med din underviser eller vejleder, hvis du er i tvivl.

Deklaration for anvendelse af generativ AI-værktøjer (studerende)

Jeg/vi har benyttet generativ AI som hjælpemiddel/værktøj (sæt kryds)

Jeg/vi har **IKKE** benyttet generativ AI som hjælpemiddel/værktøj (sæt kryds)

Hvis brug af generativ AI er tilladt til eksamen, men du ikke har benyttet det i din opgave, skal du blot krydse af, at du ikke har brugt GAI, og behøver ikke at udfylde resten.

Oplist, hvilke GAI-værktøjer der er benyttet, inkl. link til platformen (hvis muligt):

*Eksempel: [Copilot med enterprise-databeskyttelse (KU-licens),
<https://copilot.microsoft.com>]*

Svar:

*ChatGPT med privat mail, <https://chatgpt.com/>
DeepSeek med privat mail, <https://chat.deepseek.com/>*

Beskriv hvordan generativ AI er anvendt i opgaven:

- 1) Formål (hvad har du/I brugt værktøjet til)
- 2) Arbejdsfase (hvornår i arbejdsprocessen har du/I brugt GAI)
- 3) Hvad gjorde du/I med outputtet (herunder også, om du/I har redigeret outputtet og arbejdet videre med det)

Svar:

Jeg har hovedsageligt brugt GAI i løbet af min arbejdsproces som sparringspartner, hvor jeg spørger den om mine fortolkninger af mine analyser giver faktuelt mening. Outputtet bruger jeg ikke, da samtalen blot bruges til at bekraefte for mig selv, at det, jeg skriver, er nogenlunde korrekt.

Jeg har derudover undervejs brugt den til at hjælpe mig med at få ideer til, hvordan jeg kan udvide og forbedre mine analyser. Blandt andet fik jeg det forslag, at jeg kan indslyde ekstra volatile perioder i datagenereringsprocessen. Implementeringen stod jeg selv for at udføre.

Jeg har også brugt GAI til kildesøgning, på samme måde som jeg bruger Google. Dette har især hjulpet mig med at finde relevante pakker i R. Jeg har også fået hjælp til at forstå syntaksen i de indbyggede R-kommandoer, men jeg har selv sammensat alle programmer i hele opgaven.

NB. GAI-genereret indhold brugt som kilde i opgaven kræver korrekt brug af citationstegn og kildhenvisning. [Læs retningslinjer fra Københavns Universitetsbibliotek på KUnet](#).