

Intro til Matematisk Statistik Projekt: Handel med ETF

August Semrau Andersen - s183918

June 2019

Contents

1	Beskrivende analyse	2
1.1	Del <i>a</i> : Beskrivelse af data	2
1.2	Del <i>b</i> : Analyse af empirisk tæthed af fire udvalgte ETF'er	2
1.3	Del <i>c</i> : Beskrivelse på baggrund af analysen	4
2	Problem 1 - Opbygning af portefølje	5
2.1	Del <i>d</i> : Covariansmatricen	5
2.2	Del <i>e</i> : Portefølje af to ETF'er med mindste varians	5
3	Problem 2 - Den bedste investering	7
3.1	Del <i>f</i> : Opstilling af modeller	7
3.2	Del <i>g</i> : Bestemmelse af konfidensintervaller	9
3.3	Del <i>h</i> : Bestemmelse af konfidensintervaller med bootsstrapping	9
3.4	Del <i>i</i> : Test af hypotese	10
4	Problem 3	11
4.1	Del <i>j</i> : Signifikant forskel mellem ETF'er	11
5	Problem 4	12
5.1	Del <i>k</i> : Korrelation mellem risikomål	12
5.2	Del <i>l</i> : Lineær regressionsmodel for risikomål	13
5.3	Del <i>m</i> : Bestemmelse af model-koefficienter	13
5.4	Del <i>n</i> : Test af modellens korrelation	14
6	Bilag	15
6.1	Bilag 1: Beskrivelser af AGG, VAW, IWN og SPY	15
6.2	Bilag 2: Varians som funktion af α - plots.	16
6.3	Bilag 3: Forventet ugentligt relativt afkast for seks porteføljer.	17
6.4	Bilag 4: Korrelationsmatrix.	18

1 Beskrivende analyse

1.1 Del a: Beskrivelse af data

Datasætter præsenteret i filen "finans1_data.csv" har nogle bestemte karakteristika som i denne første del bliver beskrevet.

Dataene viser generelt set det relative ugentlige afkast for 95 forskellige ETF'ere (kolonner) over en tidsperiode der strækker sig 454 målinger (rækker) fra den første måling taget den 5. Maj 2006 (05/05/2006) og den sidste måling der blev taget den 8. Maj 2015 (08/05/2015). Den første kolonne viser datoen for målingen, den første række angiver navnet på ETF'en og de resterende celler indeholder så det relative ugentlige afkast, altså forskellen mellem slut og start kurs af ETF'en for den pågældende uge minus 1.

For datoerne gælder det, at selvom en konkret dato er give, selve datoen ikke er forbeholdt én bestemt ugedag (f.eks. mandag), men skifter ugedag løbende. Da det er det relative ugentlige afkast vi kigger på er det dog blot hvilket uge datoen falder på der er interessant, men hertil lægges der mærke til, at der i nogle få tilfælde bliver sprunget over en hel uge og dertil de ugentligt afkast, f.eks. i række 54 for en måling af det ugentlige afkast for uge 21 i 2007 (dato. 25/05/2007), hvor den næste række data er for uge 23 (dato. 04/06/2007). I det korrekte tidsspan fra første til sidste måling er der 470 uger, altså mangler der med de kun 454 målinger data fra 16 forskellige uger.

Af de $95 \cdot 454 = 43130$ målinger er ingen uden data, men i 210 celler er målingen lig nul, hvilket betyder at der ingen ændring i kurs er sket for den pågældende ETF, hvilket virker usandsynligt.

Udover manglen på nogle bestemte uger, lader dataen til at være af en høj kvalitet.

1.2 Del b: Analyse af empirisk tæthed af fire udvalgte ETF'er

Af de 95 forskellige ETF'er er fire udvalgt, AGG, VAW, IWN og SPY, som vi nu vil kigge nærmere på. Til hver af dem følger en investerings-økonomisk beskrivelse, som kan ses i bilaget side 15, men for nu fokuseres der på at danne en statistisk beskrivelse af dem hver især.

Et hurtigt overblik kan dannes ved at kigge nærmere på de fire ETF'ers statistiske nøgletal. I tabellen nedenfor ses for de fire ETF'er antallet af observationer der opfattes som en stikprøve for ETF'ernes relative afkast, gennemsnittet \bar{x} af stikprøven, variansen s^2 samt standardafvigelsen s af stikprøven, såvel som første kvartil Q_1 , medianen Q_2 og tredje kvartil Q_3 .

ETF:	Antal obs.	Stikprøve gennemsnit	Stikprøve varians	Stikprøve standardafvigelse	Nedre kvartil	Median	Øvre kvartil
AGG	454	2.65757E-04	0.3571068E-04	59.75841E-04	-2.9733E-03	0.2374E-03	3.8932E-03
VAW	454	17.9379E-04	13.01973E-04	360.8286E-04	-16.096E-03	4.798E-03	19.685E-03
IWN	454	11.87679E-04	10.2499E-04	320.1547E-04	-14.305E-03	3.120E-03	19.056E-03
SPY	454	13.60105E-04	6.143463E-04	247.8601E-04	-11.325E-03	4.216E-03	14.498E-03

Table 1: Beskrivende statistiske nøgletal for de fire ETF'er.

I tabellen ovenfor ser vi forskellige similariteter mellem de fire ETF'er; Deres stikprøve/sample gennemsnit ligger alle kun en anelse over nul, AGG med den laveste og VAW med den højeste, hvor AGG stikker ud fra resten ved at have en markant mindre varians. For at få et bedre billede af disse, er de grafisk repræsenteret som tæthedsfunktioner nedenfor. Her ses en lodret

linje i 0, som hjælper med at forstå ETF'ens relative afkast, da en funktion med en større tæthed til højre for linjen end til venstre vil tendere mere mod et generelt positivt afkast.

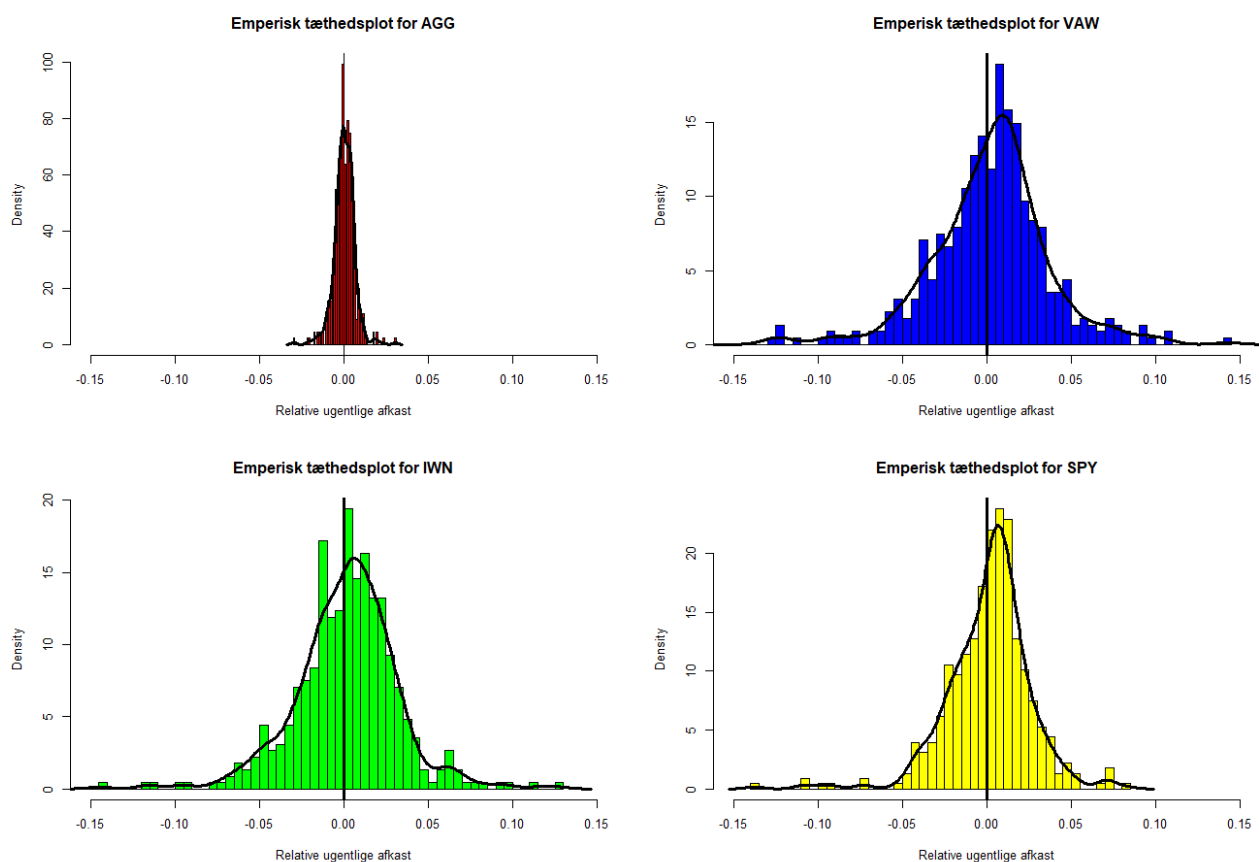


Figure 1: Tæthedsfunktioner for de fire ETF'er.

For en mere direkte sammenligning mellem de fire ETF'er kan et boxplot med fordel benyttes.

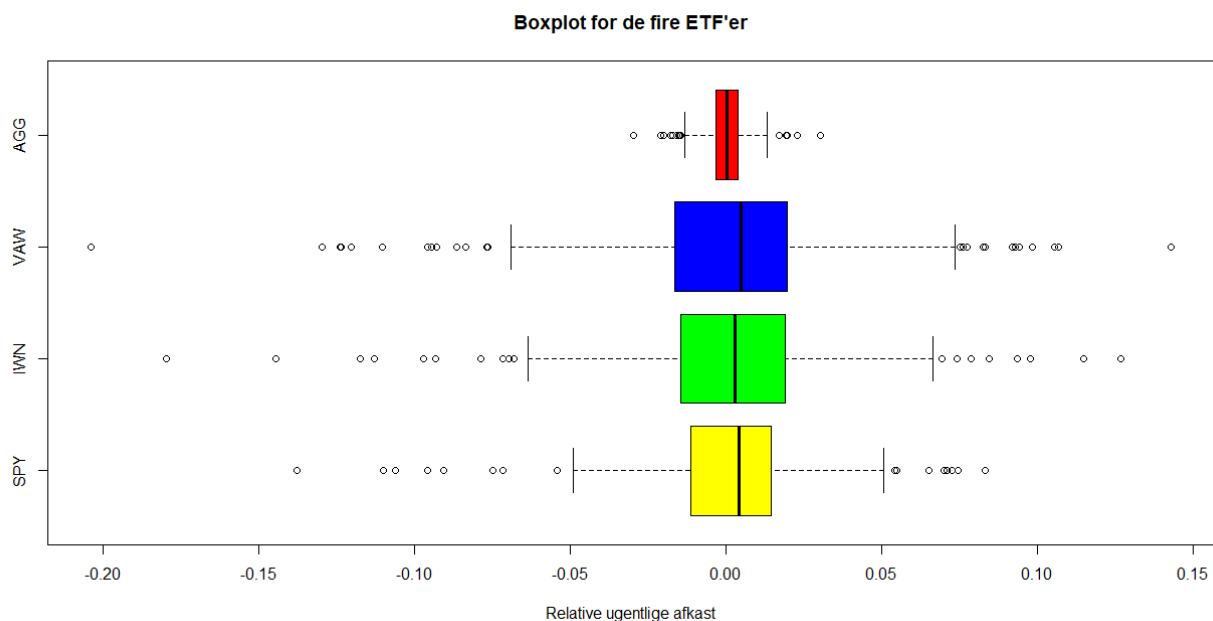


Figure 2: Boxplot over de fire ETF'er, hvori det farvede område er det inter-kvartile område hvor venstre side af det er Q1, den sorte linje i midten er medianen og højre er Q3.

På boxplottet er det tydeligt at se, at des større medianen er for ETF'en des større virker variansen også til at være, hvilket også ses på tabellen.

En vigtig note at tage med fra dataen præsenteret for ETF'erne er, at selvom de alle har positive sample-gennemsnit ligger de alle og varierer under nul, og da det desuden er det *relative* ugentlige afkast vi kigger på, betyder det at et fald og derefter en stigning på f.eks. 1% i både forlæns og baglæns rækkefølge betyder et fald i ETF'ens værdi.

1.3 Del c: Beskrivelse på baggrund af analysen

Hvis vi tager et nærmere kig på dataen af de fire ETF'er, ser vi først og fremmest at AGG er markant anderledes end de andre tre ETF'er, VAW, IWN og SPY.

AGG har en langt mindre varians og standardafvigelse sammenlignet med de andre, og det kan derfor påpeges at den rent økonomisk vil have en mindre risiko forbundet med sig, men samtidigt også et mindre potentielt udbytte.

De andre tre har langt større spredning, og samtidigt ser vi på tæthedsfunktionerne plottet på figur 1, at de alle er skæve mod positive værdier for det relative ugentlige afkast, hvilket betyder at de vil have tendens til give et positivt afkast i led med deres positive gennemsnitsværdier. AGG er mere symmetrisk fordelt.

Ser vi på boxplottet skal der også lægges mærke til de mange ekstremaer som ikke er medtaget i 'selve' boxplottet, da vi ignorerer ekstremaer som ligger over $1.5 \cdot IQR$ væk fra sample mean, hvori IQR er den interkvartile range $Q3 - Q1$. Især VAW, som både har højest mean og varians blandt ETF'erne, viser uger hvor det relative ugentlige afkast har være omkring 15%, altså meget positivt, men også et ekstremumspunkt på -20%, hvilket tyder på en dårlig uge for investorene. Sådanne ekstramaer, som vi også ser for IWN og SPY til en vis grad, opstår i sjældne tilfælde, da vi for hver af dem ikke ser mere end omkring 20 ekstremaer ud af de i alt 454 datapunkter.

Statistisk analyse I

2 Problem 1 - Opbygning af portefølje

2.1 Del d: Covariansmatricen

Her undersøges seks ETF'er (AGG, VAW, IWN, SPY, EWG og EWW) for deres tilbøjelighed til at variere med hinanden, ved at udregne deres indbyrdes Covarians opstillet i en matrix. Covariansen kan beregnes således:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Her ses i diagonalen sample variansen s^2 for hver af ETF'erne, og i de andre celler ses covariancen $Cov(X, Y)$ mellem dem indbyrdes, som altså spejler sig langs diagonalen.

	AGG	VAW	IWN	SPY	EWG	EWW
AGG	3,571E-05	-4,260E-05	-2,588E-05	-3,240E-05	-5,084E-05	-3,711E-05
VAW	-4,260E-05	1,302E-03	9,838E-04	7,927E-04	1,110E-03	1,185E-03
IWN	-2,588E-05	9,838E-04	1,025E-03	7,222E-04	9,502E-04	1,010E-03
SPY	-3,240E-05	7,927E-04	7,222E-04	6,143E-04	8,046E-04	8,153E-04
EWG	-5,084E-05	1,110E-03	9,502E-04	8,046E-04	1,444E-03	1,180E-03
EWW	-3,711E-05	1,185E-03	1,010E-03	8,153E-04	1,180E-03	1,659E-03

Figure 3: Covarians-Matrix for de seks ETF'er.

2.2 Del e: Portefølje af to ETF'er med mindste varians

En stor varians for en ETF betyder, at den er mere risikabel at investere i, selvom den potentielt set kan give et større afkast. Det gælder dog for alle økonomiske investeringer, at man helst vil minimere risikoen mens man maksimerer afkastet, og dette kan man lettere gøre ved at sprede sine penge ud over flere investeringer på samme tid.

Her vil vi altså forsøge at finde det sæt af udvalgte ETF'er der sammen har den mindste varians, altså den mindste risiko, og dette gør vi ved at samle dem til én stokastisk variabel:

$$P = \alpha \cdot X_{ETF1} + (1 - \alpha) \cdot X_{ETF2}$$

For denne stokastiske variabel P ses det, at der indgår en variabel α , der bestemmer hvor meget af den samlede investering skal investeres i henholdsvis $ETF1$ og $ETF2$. Her vil $\alpha = 1$ altså betyde at hele investeringen var i $ETF1$ og omvendt $\alpha = 0$ i $ETF2$, men *netop* disse to tilfælde vil dog ikke medtages i undersøgelsen, da vi kun er interesserede i en portefølje bestående af parvise ETF'er.

De seks porteføljer vi vil kigge på er følgende:

$$\begin{aligned} P_1 &= \alpha_1 \cdot X_{EWG} + (1 - \alpha_1) \cdot X_{EWW} \\ P_2 &= \alpha_2 \cdot X_{AGG} + (1 - \alpha_2) \cdot X_{SPY} \\ P_3 &= \alpha_3 \cdot X_{VAW} + (1 - \alpha_3) \cdot X_{IWN} \\ P_4 &= \alpha_4 \cdot X_{VAW} + (1 - \alpha_4) \cdot X_{EWG} \\ P_5 &= \alpha_5 \cdot X_{VAW} + (1 - \alpha_5) \cdot X_{EWW} \\ P_6 &= \alpha_6 \cdot X_{IWN} + (1 - \alpha_6) \cdot X_{EWG} \end{aligned}$$

For den stokastiske variable P har vi, at variansen kan skrives op som en funktion af α :

$$V_P(\alpha) = \alpha^2 \cdot V(X_{ETF1}) + (1 - \alpha)^2 \cdot V(X_{ETF2}) + 2 \cdot \alpha \cdot (1 - \alpha) \cdot Cov(X_{ETF1}, X_{ETF2})$$

Som eksempel kan variansen som funktion af α beregnes for den stokastiske variabel bestående af de to ETF'er EWG og EWW:

$$V_{P1}(\alpha) = \alpha^2 \cdot V(X_{EWG}) + (1 - \alpha)^2 \cdot V(X_{EWW}) + 2 \cdot \alpha \cdot (1 - \alpha) \cdot Cov(X_{EWG}, X_{EWW})$$

$$\Leftrightarrow V_{P1}(\alpha) = 0.000744194 \cdot \alpha^2 - 0.000959294 \cdot \alpha + 0.001659257$$

Da variansen kan opskrives som funktion af denne værdi α , faktisk et andengradspolynomie, kan sammensætningen af porteføljerne, altså hvor meget der er af den ene kontra den anden ETF, bestemmes ift. hvornår der er mindst varians. Dette kan med fordel visualiseres som de plttede polynomier i intervallet $\alpha \in [0, 1]$, hvilket nedenfor kan ses for $P1$, og for resten kan ses i Bilag 2 side 16.

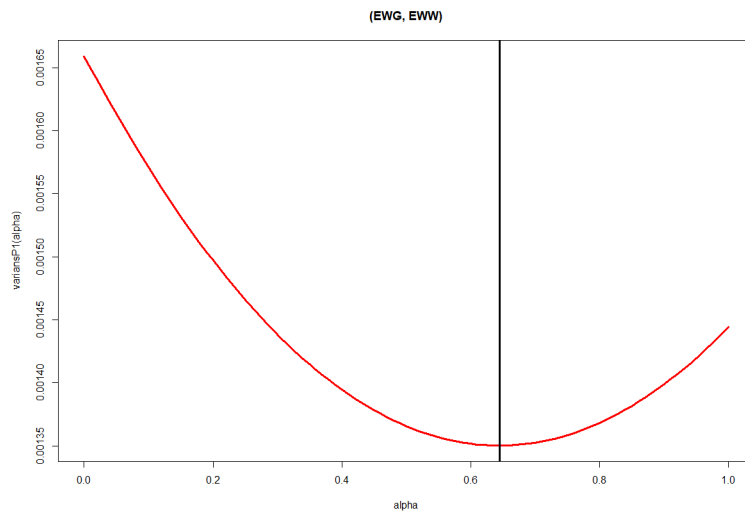


Figure 4: Variansen af P1: (EWG, EWW) porteføljen som funktion af α med minimum.

Ved at opstille de stokastiske variable P1-P6 ud fra de fundne α -værdier, findes porteføljen med mindst varians til at være $P2 : (AGG, SPY)$. Da denne indeholder ETF'en AGG, som vi tidligere har set til at have en anderledes smal fordeling ift. andre ETF'er, kan det også være interessant at se at variansen blandt de resterende porteføljer ligger markant tættere, hvori variansen for $P6 : (IWN, EWG)$ er mindst. Statistiske nøgletal for porteføljerne (som stokastiske variable) ses i tabellen nedenfor.

Portefølje	α	Varians s^2	Std. afvigelse s	Forventet ugentligt relativt afkast X
P1: EWG, EWW	0.6445187	1.350115E-03	0.03674392	0.001402737
P2: AGG, SPY	0.9047261	0.029221E-03	0.005405731	0.0003700198
P3: VAW, IWN	0.114569	1.020274E-03	0.03194172	0.001257121
P4: VAW, EWG	0.6350648	1.231874E-03	0.03509807	0.001590483
P5: VAW, EWW	0.8019989	1.278782E-03	0.03576007	0.001775972
P6: IWN, EWG	0.868502	1.015156E-03	0.03186151	0.001194124

Table 2: Tabel over α_m værdier og den dertilhørende varians, standard afvigelse og gennemsnit.

Hvis vi skulle vælge den bedste portefølje at investere i, ville $P2$ være den bedste baseret på denne opgaves idé om at den mindste varians og derved den sikreste investering er bedst.

Dog er det forventede afkast også markant mindre end de andre porteføljer, for hvilke $P5 : (Vaw, EWW)$ har det største forventede relative ugentlige afkast (fundet som sample mean af den stokastiske variabel). Ved at kigge på tæthedskurver for de seks porteføljer, som alle kan ses i bilag 3 på side 17, ses det også at der er en positiv skævhed for $P5$, så på trods af den store varians, vurderes denne $P5(Vaw, EWW)$ portefølje til at være den bedste investering.

3 Problem 2 - Den bedste investering

3.1 Del f: Opstilling af modeller

Nu da vi har vurderet både enkelte ETF'er såvel som porteføljer bestående af flere, på baggrund af deres stikprøve-værdier, ville det være interessant for os at kunne kreere en model for ETF'erne så vi kan tilgå statistiske nøgletal uden brug af stikprøve-værdierne. For at validere om det er muligt at modellere fordelingen af de fire ETF'er vi undersøgte i opgave b; AGG, VAW, IWN og SPY, ud fra normalfordelingen, vil vi her plotte datapunkterne (stikprøven) som et såkaldt QQ-plot. QQ-plottet er en grafisk metode hvormed vi kan sammenligne to sandsynlighedsfordelingers kvartiler, i dette tilfælde ETF'ernes data mod en teoretisk normalfordeling, og det gælder at ETF'ens sandsynlighedsfordeling er normalfordelt hvis QQ-plottet af ETF'ernes datapunkter og den teoretiske normalfordeling på det samlede plot nedenfor følges ad.

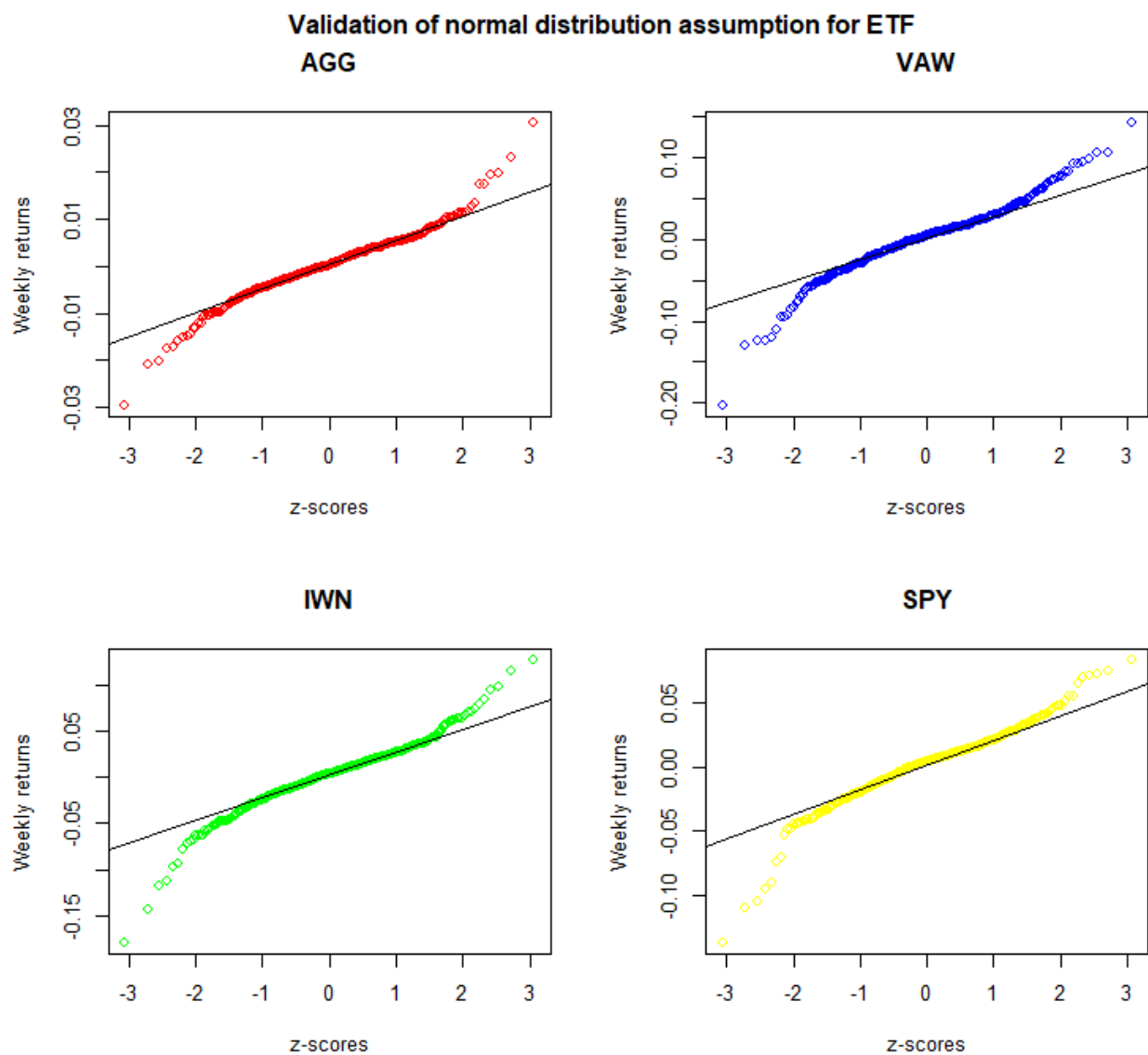


Figure 5: QQ-plottet og den teoretiske normalfordeling.

På plottet ser vi, at der er en stor overensstemmelse mellem den teoretiske normalfordeling og QQ-plottet for alle fire ETF'er i z-score intervallerne fra omkring -2 til 2, og vi vurderer dem derfor til at være normalfordelte.

3.2 Del g: Bestemmelse af konfidensintervaller

I denne delopgave skal et 95% konfidensinterval bestemmes for det gennemsnitlige ugentlige afkast af de fire ETF'er som blev fundet til at være normalfordelte i opgaven ovenfor, f . Dette finder vi ved formlen:

$$\bar{x} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

Heri forstås $t_{1-\frac{\alpha}{2}}$ som værdien for den $1 - \frac{\alpha}{2}$ 'te fraktil af en normaliseret t-fordeling, som er en normalfordeling divideret med en Chi i anden χ^2 fordeling, der ligner normalfordelingen mere og mere jo større n bliver.

For et 95% konfidensinterval er $\alpha = 0.05$, og derved findes fraktilen vha. R's qt-funktion:

$$t_{0.975} = 1.965215$$

Konfidensintervallet for variansen findes ved den lidt anderledes formel:

$$\overline{s^2} \pm \left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \right]$$

Tabellen nedenfor viser konfidensintervaller (KI) og KI for varians for de fire ETF'er, samt bredden af disse intervaller.

ETF:	KI nedre grænse	KI øvre grænse	Bredde af KI	KI varians nedre	KI varians øvre	Bredde af varians KI
AGG	-2.854073E-04	8.169213E-04	1.102329E-03	0.314808E-04	0.408589E-04	9.37803E-06
VAW	-15.34208E-04	51.21788E-04	6.655998E-03	11.47759E-04	14.89672E-04	3.41913E-04
IWN	-17.65174E-04	41.40533E-04	5.905708E-03	9.035838E-04	11.72758E-04	2.691742E-04
SPY	-9.25960E-04	36.46171E-04	4.572132E-03	5.415792E-04	7.029136E-04	1.613344E-04

Table 3: 95% Konfidens Interval, 95% Konfidens Interval for varians samt bredden af disse intervaller.

VI ser i tabellen, at AGG som i tidligere opgaver igen står ud, men en markant mindre spredning og derved også et mindre konfidensinterval. De andre tre ETF'er følges ad meget pænt ift. bredderne på intervallerne, hvor de bredeste konfidensintervaller tilhører ETF'er med større varians som set i tidligere opgaver.

3.3 Del h: Bestemmelse af konfidensintervaller med bootsstrapping

I tilfælde af at de fire ETF'ers normalfordeling ikke havde været klar nok, ville det være en fordel at kunne finde en fordeling uden at have en forantagelse for fordelingen af dataen. Derfor benyttes her ikke-parametrisk bootsstrapping til at finde konfidens intervaller for afkastet og variansen af afkastet, da denne uden bias vil danne en fordeling hvis intervaller kan sammenlignes med resultaterne for opgave g , for at eftervise normalfordelingen. Ikke-parametrisk bootsstrapping er en metode med hvilken man ud fra et relativt småt datasæt kan skabe stor statistisk sikkerhed, ved at mange gentagne gange vælge tilfældige data ud af en stikprøve med tilbagelægning.

Tabellen nedenfor er udfyldt ved brug af R-funktionen replicate med replace = TRUE.

ETF:	BS KI nedre	BS KI øvre	BS KI bredde	BS KI var nedre	BS KI var øvre	BS KI var bredde
AGG	-2.872946E-04	0.8173335E-03	1.104628E-03	0.2843316E-04	0.4400079E-04	0.1556763E-04
VAW	-15.57933E-04	5.083007E-03	6.64094E-03	10.35784E-04	16.09184E-04	5.734E-04
IWN	-17.92994E-04	4.099481E-03	5.892475E-03	8.070048E-04	12.749193E-04	4.679145E-04
SPY	-9.461376E-04	3.6178889E-03	0.004564027	4.82482E-04	7.65535E-04	2.83053E-04

Table 4: Bootstrapped 95% Konfidens Interval, 95% KI for varians samt bredden af disse intervaller.

Kigger vi tilbage på tabellen med konfidensintervallerne fra forrige opgave ser vi, at de bootstrappede værdier ligger rigtigt tæt på de beregnede værdier, hvori især KI for afkastet ser meget fornuftigt ud, mens KI for variansen afviger en smule mere.

3.4 Del i: Test af hypotese

For at vurdere om det bedre kan betale sig at gemme sine penge derhjemme, end at investere dem i en af ETF'erne, opstilles hypotesen der siger, at der ingen signifikant forskel er mellem disse to muligheder. Sagt med andre ord, 'påstår' nulhypotesen at det forventede relative afkast for en ETF ikke er signifikant forskelligt fra nul, altså har vi at:

$$H_0 : \mu_{ETF} = 0$$

$$H_1 : \mu_{ETF} \neq 0$$

Her gælder det, at vi ved brug af R-funktionen t.test finder p-værdien for hver ETF, med hvilken vi accepterer eller afviser nulhypotesen. I dette tilfælde med 95% konfidens interval, skal p-værdien være mindre end signifikans-niveau værdien $\alpha = 0.05$ for at afvise H_0 .

For de fire ETF'er finder vi, at p-værdierne er markant over α , og for dem alle accepteres nulhypotesen, hvilket ses i tabellen for neden.

ETF:	p-værdi	H_0	H_1
AGG	0.3430	Accepteret	Afvist
VAW	0.2901	Accepteret	Afvist
IWN	0.4297	Accepteret	Afvist
SPY	0.2429	Accepteret	Afvist

Table 5: Tabel over p-værdier for de fire ETF'er

Altså ses det, at det statistisk set ikke kan betale sig at investere i nogle af de fire ETF'er, AGG, VAW, IWN og SPY, sammenlignet med at gemme sine penge derhjemme.

4 Problem 3

4.1 Del j: Signifikant forskel mellem ETF'er

I opgaven forinden i, blev de fire ETF'er vurderet til ikke at være signifikant bedre at investere i, sammenlignet med ingen investering overhovedet. På trods af dette, er det stadig interessant at kigge på, om der indbyrdes blandt dem kunne findes en signifikant forskel mellem den ETF med størst og den med laveste forventede månedlige relative afkast.

Altså, hvis vi *skulle* investere i én af de fire ETF'er, ville der så være signifikant forskel mellem den umiddelbart bedste og værste eller er det ligegyldigt hvilken vi investerer i?

Dette kan undersøges på samme måde som i forrige opgave, da vi ved at finde en p-værdi større end signifikans-niveauet $\alpha = 0.05$ kan acceptere en nulhypotese:

$$H_0 : \mu_{ETF(minst)} = \mu_{ETF(størst)}$$

VAW er ETF'en med det største forventede ugentlige relative afkast (μ_{VAW}), og AGG er den med den mindste (μ_{AGG}). For disse findes p-værdien til at være $p = 0.3738$, og H_0 accepteres herved, hvorved det viser sig at der ikke er nogen statistisk signifikant forskel på om man investerer sine penge i VAW eller AGG.

Statistisk analyse II

5 Problem 4

5.1 Del *k*: Korrelation mellem risikomål

Korrelation mellem forskellige stokastiske variable beskriver sammenhængen mellem de variable, altså hvor meget den ene afhænger af den anden. Korrelation er et tal mellem -1 og 1, hvor -1 er fuld negativ korrelation, 0 betyder ingen korrelation/sammenhæng overhovedet, og 1 er en fuld positiv korrelation.

I det andet datasæt, "data/finans2_data.csv", finder vi 8 variable for hver ETF, som hver beskriver bestemte risikomål for ETF'erne.

Nedenfor er korrelationen mellem fem af disse plottet, hvor det altså er muligt at se hvilke risikomål korrelerer med andre.

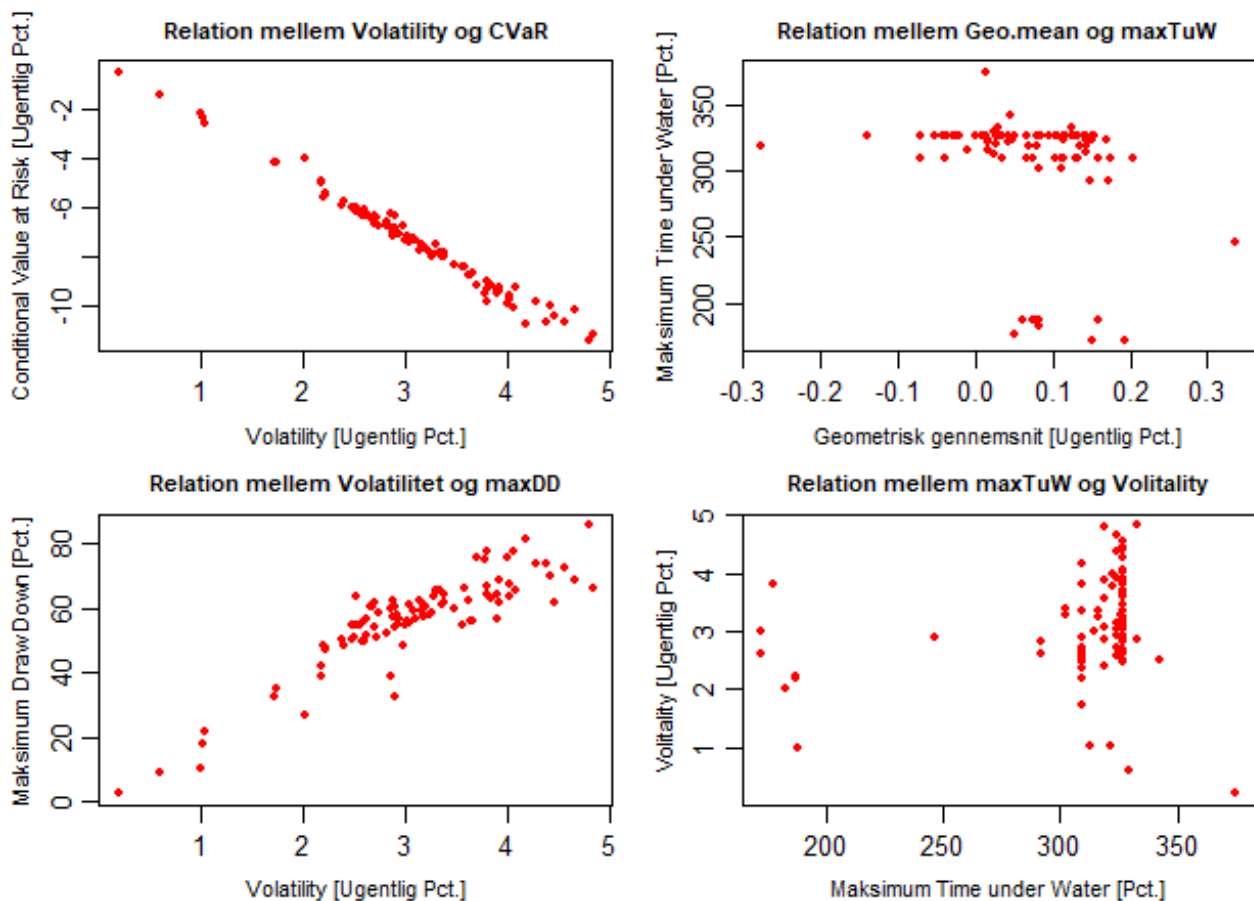


Figure 6: Korrelationen i fire forskellige tilfælde.

Vi ser at der er en stor negativ korrelation, faktisk på $r = -0.9922225$, mellem Volatilitet og Conditional-Value-at-Risk CVaR. Altså bliver CVaR større, jo mindre volatilitet der er for ETF'en.

Der er ikke nogen reel korrelation mellem Geometrisk gennemsnits relative ugentlige afkast Geo.mean og Maksimale Time under Water maxTuW der er $r = -0.2176515$.

Mellem Volatilitet og Maksimal Draw Down er der dog igen en høj korrelation, denne gang positiv $r = 0.8797983$, og det lader derfor til at volatilitet korrelerer meget med disse to risikomål. På det sidste plot ser vi dog at volatilitet ikke korrelerer med alle risikomål, da korrelationen med Maksimal Time under Water ikke er stor, med $r = 0.2505564$

Korrelation mellem to stokastiske variable kan findes ved brug af formlen, der simpelt siger af korrelationen er covariansen divideret med den ene standardafvigelse gange den anden:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) = \frac{Cov(x, y)}{s_x \cdot s_y}$$

Herved kan korrelationen mellem geo.mean og maxTuW findes:

$$\frac{Cov(Geo.mean, maxTuW)}{s_{Geo.mean} \cdot s_{maxTuW}} = -0.2176515$$

Dette er ikke en specielt entydig korrelation.

5.2 Del l: Lineær regressionsmodel for risikomål

I denne opgave skal en lineær regressionsmodel med Geo.mean som responsvariabel Y_i bestemmes, og til dette formål skal en forklarende variabel x_i i form af én af de andre risikomål benyttes. Denne forklarende variabel vælges ud fra hvilket risikomål der korrelerer mest med Geo.mean, hvilket er det ugentlige Value-at-Risk VaR med en korrelation på $r = 0.4110733$. Correlationsmatrix ses i Bilag 4 på side 18.

Herved kan jeg opstille regressionsmodellen, for hvilken vi forudsætter at der er en lineær sammenhæng til trods for den lidt lave korrelation. Heri er responsvariabelen Y_i Geo.mean og den forklarende variabel x_i er VaR:

$$Y_i = \beta_0 + \beta_1 \cdot x_i$$

5.3 Del m: Bestemmelse af model-koefficienter

Ud fra lm-funktionen i R kan model-koefficienterne β_0 og β_1 bestemmes:

$$Y_i = 0.186419 + 0.023529 \cdot x_i$$

Disse kan dog også findes manuelt, ved formlerne forneden, hvori $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 187.6348$ er summen af den kvadrerede fejl. Udregningen er lavet i R:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{S_{xx}} = 0.02352877$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = 0.07690416 - 0.02352877 \cdot (-4.65452) = 0.1864193$$

Her ser vi, på samme måde som en hvilken som helst anden lineær regression, at hældningen $\beta_1 = 0.02352877$ fortæller noget om at når Geo.mean bliver større bliver VaR også større. $\beta_0 = 0.1864193$, der er skæringen med y-aksen, fortæller hvad værdien af VaR er når Geo.mean

er nul.

Modellens varians er også interessant at kende, og denne kan beregnes ved formelen:

$$\hat{\sigma}^2 = \frac{RSS(\hat{\beta}_0, \hat{\beta}_1)}{n - 2}$$

Heri har vi at summen af de kvadrerede afvigelse, $RSS(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$.
Vha. formlerne udregnes modellens varians, hvor vi har fundet at $RSS(\beta_0, \beta_1) = 0.510839$.

$$\hat{\sigma}^2 = \frac{RSS(\hat{\beta}_0, \hat{\beta}_1)}{n - 2} = \frac{0.510839}{95 - 2} = 0.005492893$$

Disse værdier ser ud til at være korrekte, ift hvis de holdes op imod samme værdier beregnet med R-funktionen *lm*.

Denne funktion outputter endvidere en række andre tal, som kan bruges til at fortolke modellen, bl.a. p-værdier som bruges i den næste (og sidste) opgave.

5.4 Del *n*: Test af modellens korrelation

I fortsættelse af ovenstående, ser vi at R-funktionen *lm* også outputter en p-værdi for modellens koefficienter β_1 og β_0 , som er henholdsvis $p_{\beta_1} = 3.50\text{E-}05$ og $p_{\beta_0} = 2.61\text{E-}10$. Disse p-værdier er begge markant mindre en 0.05, hvilket vi fra tidligere ved betyder, at en nulhypoteses bliver forkastet. Altså, viser p-værdierne at der *er* en statistisk signifikant lineær sammenhæng (korrelation) mellem Geo.mean og VaR.

6 Bilag

6.1 Bilag 1: Beskrivelser af AGG, VAW, IWN og SPY

AGG: iShares Core Total US Bond Market ETF, formerly iShares Lehman Aggregate Bond Fund (the Fund) seeks investment results that correspond generally to the price and yield performance of the total United States investment-grade bond market as defined by the Lehman Brothers U.S. Aggregate Index (the Index). The Index measures the performance of the United States investment-grade bond market, which includes investment-grade United States Treasury bonds, government-related bonds, investment-grade corporate bonds, mortgage pass-through securities, commercial mortgage-backed securities and assetbacked securities that are publicly offered for sale in the United States. The securities in the Index must have at least one year remaining to maturity. In addition, the securities must be denominated in United States dollars, and must be fixed rate, non-convertible and taxable. The Index is market capitalization weighted. The Fund uses a representative sampling strategy to track the Index. The Fund's investment advisor is Barclays Global Fund Advisors (BGFA).

VAW: Vanguard Materials ETF (the Fund), formerly known as Vanguard Materials VIPERs, is an exchange-traded share class of Vanguard Materials Index Fund, which employs a passive management or indexing investment approach designed to track the performance of the Morgan Stanley Capital International (MSCI) US Investable Market Materials Index (the Index). The Index is an index of stocks of large, medium and small United States companies in the materials sector, as classified under the Global Industry Classification Standard (GICS). This GICS sector is made up of companies in a range of commodity-related manufacturing industries. Included within this sector are companies that manufacture chemicals, construction materials, glass, paper, forest products and related packaging products, as well as metals, minerals and mining companies, including producers of steel. The Fund attempts to replicate the Index by investing all, or substantially all, of its assets in the stocks that make up the Index, holding each stock in approximately the same proportion as its weighting in the Index. The Fund also may sample its target index by holding stocks that, in the aggregate, are intended to approximate the Index in terms of key characteristics, such as price/earnings ratio, earnings growth and dividend yield.

IWN: iShares Russell 2000 Value Index Fund (the Fund) seeks investment results that correspond generally to the price and yield performance of the Russell 2000 Value Index (the Index). The Index measures the performance of the small-capitalization value sector of the United States equity market. It is a subset of the Russell 2000 Index. The Index is a capitalization-weighted index and consists of those companies or portion of a company, with lower price-to-book ratios and lower forecasted growth within the Russell 2000 Index. The Index represents approximately 50% of the total market capitalization of the Russell 2000 Index. The Fund invests in a representative sample of securities included in the Index that collectively has an investment profile similar to the Index. iShares Russell 2000 Value Index Fund's investment advisor is Barclays Global Fund Advisors.

SPY: SPDR Trust, Series 1 (the Trust) is a unit investment trust. The Trust is an exchange-traded fund created to provide investors with the opportunity to purchase a security representing a proportionate undivided interest in a portfolio of securities consisting of substantially all of the common stocks, in substantially the same weighting, which comprise the Standard and

Poor's 500 Composite Price Index (the SP Index). Each unit of fractional undivided interest in the Trust is referred to as a Standard and Poor's Depositary Receipt (SPDR). The Trust utilizes a full replication approach. With this approach, all 500 securities of the Index are owned by the Trust in their approximate market capitalization weight.

6.2 Bilag 2: Varians som funktion af α - plots.

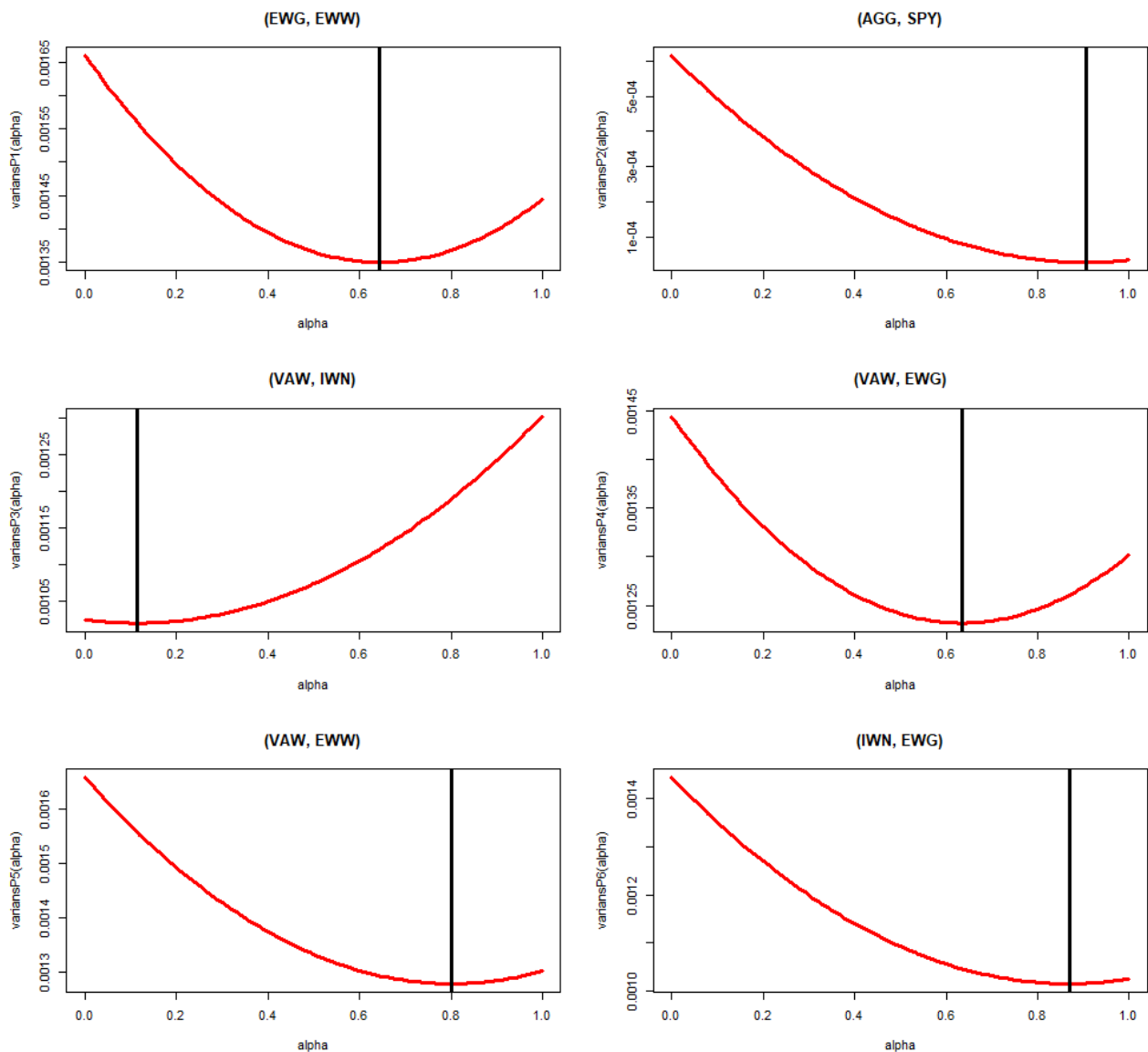


Figure 7: Variansen for de seks porteføljer som funktion af α -værdien

6.3 Bilag 3: Forventet ugentligt relativt afkast for seks porteføljer.

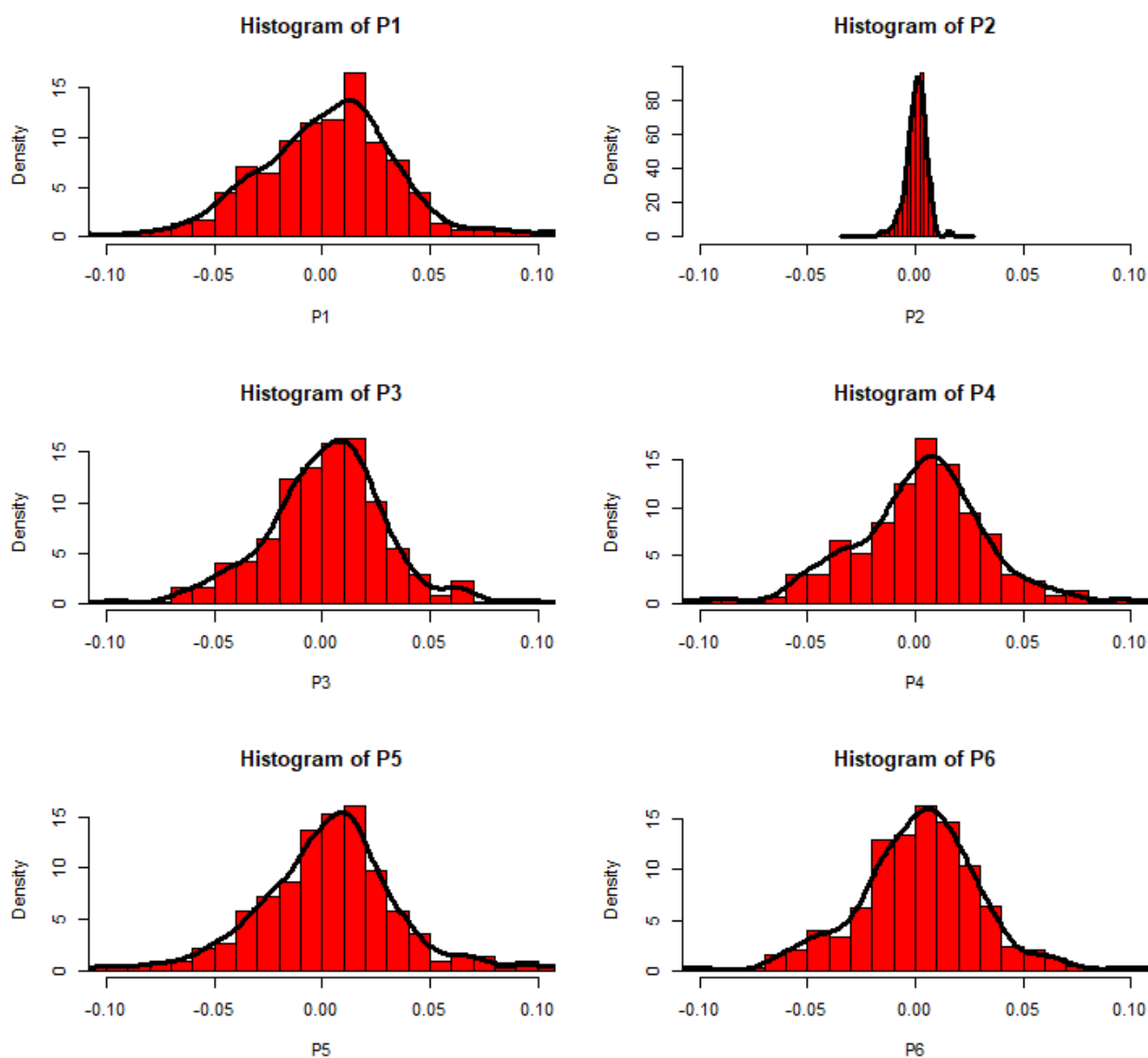


Figure 8: Tæthedsfunktioner for de seks porteføljer ift. deres relative ugentlige afkast.

6.4 Bilag 4: Korrelationsmatrix.

	Geo.mean	Volatility	maxDD	maxTuW	VaR	CVaR
Geo.mean	1	-0,35729	-0,39682	-0,21765	0,411073	0,384272
Volatility	-0,35729	1	0,879798	0,250556	-0,97349	-0,9922
maxDD	-0,39682	0,879798	1	0,281172	-0,86289	-0,90868
maxTuW	-0,21765	0,250556	0,281172	1	-0,26697	-0,26406
VaR	0,411073	-0,97349	-0,86289	-0,26697	1	0,967631
CVaR	0,384272	-0,9922	-0,90868	-0,26406	0,967631	1

Figure 9: Matrix med korrelasjoner mellom risikomålene.