

TO PROBLEMSTILLINGER RUNDT MINIMERING AV VARIANS I MAT1020

FRED ESPEN BENTH

1. BAKGRUNN: FORVENTNING, VARIANS, KOVARIANS OG KORRELASJON TIL DATA

Anta at du har en serie med n observasjoner, eller data, $x_1, x_2, \dots, x_n$. Disse dataene kan være temperaturmålinger i en by, aksjekursene til en bedrift eller hvor mye CO₂ som er sluppet ut i atmosfæren. For eksempel, i Figur 1 viser den blå

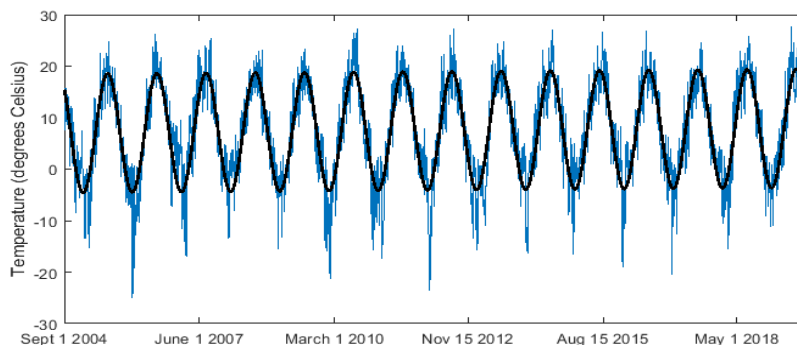


FIGURE 1. Serie av daglige temperaturer (definert som snittet av daglig maksimum og minimum, blå kurve) sammen med historisk sesongtemperatur (svart kurve). Dataene er fra Vilnius, Litauen.

kurven daglige målte temperaturer fra 1. september 2004 til 30. september 2019 mens den svarte viser den historiske sesongtemperaturen gitt ved en sinus-funksjon. Hvis vi noterer målingene som t_i , temperatur på dato i , mens middeltemperaturen noteres som s_i , kan vi regne ut data for *avvik* fra normalen. Avvik fra normalen er definert som $x_i := t_i - s_i$.

I statistikk er det to helt sentrale mål som man ofte bruker når data skal analyseres, og det er *gjennomsnitt* og *varians*. Gjennomsnittet definerer vi til å være middelverdien til dataene,

$$(1) \quad \hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n x_i$$

Vi burde strengt tatt hengt på en indeks n på $\hat{\mu}$ siden dens verdi avhenger av hvor mange data vi har, gitt ved n . For å spare oss for mye notasjon, "glemmer" vi denne indeksen i den videre analysen.

Date: July 2, 2024.

Variansen er et mål på *hvor mye dataene avviker fra sin gjennomsnittsverdi*. Datamåling x_i sitt avvik fra gjennomsnittsverdien er $x_i - \hat{\mu}_x$. Vi vil gjerne la positive og negative avvik telle like mye, så vi ser på kvadratet til avviket $(x_i - \hat{\mu}_x)^2$. Gjennomsnittsverdien til alle de kvadratiske avvikene kalles *variansen* til dataene

$$(2) \quad \hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)^2$$

Man kan regne ut at (vis dette selv!)

$$(3) \quad \hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \hat{\mu}_x^2$$

Uttrykket i (3) kan brukes som et alternativ til uttrykket i (2) når man beregner variansen til en serie med data. De gir begge samme svar.

Ofte snakker man om *standardavviket* til dataene. Standardavviket er rett og slett kvadratroten til variansen,

$$(4) \quad \hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)^2}$$

Hvis du tenker på målinger av temperaturer, så er enheten til dataene grader Celsius, $^{\circ}C$. Gjennomsnittet har samme enhet, mens variansen får enhet $(^{\circ}C)^2$. Standardavviket, derimot, vil ha samme enhet som gjennomsnittet, $^{\circ}C$. Dette er en av grunnene til at standardavviket er mye brukt som mål på variasjonen i data, eller hvor mye dataene avviker i snitt fra gjennomsnittsverdien.

Data er noe man har observert (selvsagt!), og dermed et bilde av historien. Ofte bruker man historien som utgangspunkt for å si noe om hva man tror fremtiden vil bli. For eksempel, hvis $x_1, x_2, \dots, x_{365}$ er daglige målinger av temperaturavviket fra normalen på Blindern gjennom fjoråret, så kan man tenke seg at avviket fra normalen for det kommende året blir i snitt $\hat{\mu}_x$, med en daglig variasjon rundt dette snittet gitt ved standardavviket $\hat{\sigma}_x$. Man sier at temperaturavviket fra normalen i *forventning* blir $\hat{\mu}_x$ med en *usikkerhet* gitt av variansen, evt. standardavviket. I begynnelsen av det kommende året vet vi selvsagt ikke hva som faktisk skjer, så dette er noe vi *predikerer* kommer til å skje. Vi har faktisk laget oss en enkel *statistisk modell* for hva som *kan komme til å skje med temperaturene i det kommende året!*.

Risiko er et annet begrep man bruker om noe som er usikkert. Risikoen måles med varians eller standardavvik.

Hvis vi har to dataserier, $x_1, x_2, \dots, x_n$ og $y_1, y_2, \dots, y_n$, bruker man *kovariansen* til å måle i hvor stor grad dataene fra den første serien av observasjoner samvarierer med den andre serien. Kovariansen defineres som

$$(5) \quad \hat{c}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

Merk at $\hat{c}_{xx} = \hat{\sigma}_x^2$ og $\hat{c}_{xy} = \hat{c}_{yx}$

Korrelasjonen til dataseriene er gitt ved

$$(6) \quad \hat{\rho}_{xy} = \frac{\hat{c}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$$

Vi legger merke til at korrelasjonen er uten enhet, eller med andre ord dimensjonsløs. Dessuten ser vi at $\hat{\rho}_{xy} = \hat{\rho}_{yx}$.

Teorem 1.1. *Vi har at $-1 \leq \hat{\rho}_{xy} \leq 1$ og $\hat{\rho}_{xx} = 1$.*

Proof. Vi viser at $-1 \leq \hat{\rho}_{xy} \leq 1$ ved induksjon. Først viser vi en nyttig ulikhet: Hvis a og b er to tall, så vet vi fra kvadratsetningene at $(a - b)^2 = a^2 - 2ab + b^2$. Men samtidig er også $(a - b)^2 \geq 0$, og det følger derfor at

$$(7) \quad 2ab \leq a^2 + b^2$$

Vi setter $a_i = x_i - \hat{\mu}_x$ og $b_i = y_i - \hat{\mu}_y$. Bruker vi ulikheten (7), vil

$$\begin{aligned} (a_1 b_1 + a_2 b_2)^2 &= a_1^2 b_1^2 + 2a_1 b_1 a_2 b_2 + a_2^2 b_2^2 \\ &\leq a_1^2 b_1^2 + a_1^2 b_2^2 + a_2^2 b_1^2 + a_2^2 b_2^2 \\ &= (a_1^2 + a_2^2)(b_1^2 + b_2^2) \end{aligned}$$

Anta nå at (dette er induksjonshypotesen)

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right) \left(\sum_{i=1}^n b_i^2\right)$$

For $n + 1$ finner vi,

$$\begin{aligned} \left(\sum_{i=1}^{n+1} a_i b_i\right)^2 &= \left(\sum_{i=1}^n a_i b_i + a_{n+1} b_{n+1}\right)^2 \\ &= \left(\sum_{i=1}^n a_i b_i\right)^2 + 2a_{n+1} b_{n+1} \sum_{i=1}^n a_i b_i + a_{n+1}^2 b_{n+1}^2 \\ &\leq \left(\sum_{i=1}^n a_i b_i\right)^2 + \sum_{i=1}^n a_i^2 b_{n+1}^2 + \sum_{i=1}^n a_{n+1}^2 b_i^2 + a_{n+1}^2 b_{n+1}^2 \\ &= \left(\sum_{i=1}^{n+1} a_i^2\right) \left(\sum_{i=1}^{n+1} b_i^2\right) \end{aligned}$$

der vi har brukt den nyttige ulikheten (7) til å konkludere med at $2a_{n+1}b_{n+1}a_i b_i \leq a_{n+1}^2 b_i^2 + a_i^2 b_{n+1}^2$. Ved induksjon finner vi derfor (etter å ha skrevet ut definisjonen av a_i og b_i)

$$\frac{\left|\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)\right|}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_y)^2}} \leq 1$$

og derfor at $-1 \leq \hat{\rho}_{xy} \leq 1$. At $\hat{\rho}_{xx} = 1$ følger direkte fra definisjonen. $\square$

Når to dataserier har positiv korrelasjon, så tenderer de til å følge hverandre opp og ned, mens negativ korrelerte serier tenderer til å gå i motsatt retning av hverandre. Ukorrelerte data, data der $\hat{\rho}_{xy} = 0$, vil ikke ha noe med hverandre å gjøre. Vi bruker ofte også den historiske korrelasjonen til å si oss hvordan fremtidige målinger vil korrelere.

Har man d serier med data, $x_{1,1}, x_{2,1}, \dots, x_{n,1}, x_{1,2}, x_{2,2}, \dots, x_{n,2}, \dots, x_{1,d}, x_{2,d}, \dots, x_{n,d}$, så forenkler vi notasjonen til kovariansen og korrelasjonen med å bruke notasjonen

$\hat{c}_{k\ell}$ og $\hat{\rho}_{k\ell}$ for kovariansen, henholdsvis korrelasjonen, mellom dataserie k og ℓ . Vi kan lage oss en *kovariansmatrise* C , som samler opp alle kovarianser;

$$(8) \quad C = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{c}_{12} & \cdots & \hat{c}_{1d} \\ \hat{c}_{21} & \hat{\sigma}_2^2 & \cdots & \hat{c}_{2d} \\ \cdot & \cdot & \cdots & \cdot \\ \hat{c}_{d1} & \hat{c}_{d2} & \cdots & \hat{\sigma}_d^2 \end{bmatrix}$$

Vi ser at C blir en symmetrisk $d \times d$ matrise. Tilsvarende finner vi en $d \times d$ -matrise av korrelasjoner som

$$(9) \quad \Gamma = \begin{bmatrix} 1 & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1d} \\ \hat{\rho}_{21} & 1 & \cdots & \hat{\rho}_{2d} \\ \cdot & \cdot & \cdots & \cdot \\ \hat{\rho}_{d1} & \hat{\rho}_{d2} & \cdots & 1 \end{bmatrix}$$

Hvis vi definerer en matrise av standardavvik til å være

$$(10) \quad \Sigma = \begin{bmatrix} \hat{\sigma}_1 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & 0 & \hat{\sigma}_d \end{bmatrix}$$

ser vi ved matrisemultiplikasjon at

$$(11) \quad C = \Sigma \Gamma \Sigma.$$

Noen ganger er dette en hensiktsmessig måte å representere kovariansmatrisen på som egner seg for tolkning.

2. BESTE INVESTERINGSPORTEFØLJER OG ESG-KRAV

Vi skal i denne seksjonen studere problemet med å investere penger i aksjer eller andre usikre verdipapirer, der man som investor sikter mot å høste mest mulig avkastning på sine investeringer men samtidig unngå å ta for stor risiko. Høy risiko betyr at man kan tape mange penger, mens en lavere risiko sørger for at investeringene er sikrere og sjansen for å tape stort er mye lavere. Men hva skjer med risiko og avkastning hvis vi utelater noen sektorer fra våre investeringsmuligheter, som for eksempel at vi ikke vil investere i oljeselskaper?

La $\mathbf{a} = (a_1, a_2, \dots, a_d)^\top$ være en d -dimensjonal vektor (kolonne-vektor!) som beskriver hvor mange aksjer du har investert i d ulike selskaper. Aksjekursen på et tidspunkt frem i tid til hvert selskap er gitt ved $\mathbf{S} = (S_1, S_2, \dots, S_d)^\top$. Tidspunktet kan være i morgen (lik 1 dag), eller 1 uke, eller ett år, og vi kjenner ikke hva aksjekursene da blir.¹ Vi kaller tidspunktet for *tidshorisonten* til investeringen vi gjør. Den fremtidige verdien av porteføljen din blir

$$W = \mathbf{a}^\top \mathbf{S} = \sum_{k=1}^d a_k S_k$$

Notasjonen W er hentet fra engelsk, og refererer til *wealth*, eller *formue* på norsk. I finanslitteratur noteres gjerne aksjekursen med S , for det engelske *stock*.

¹Vi er ikke orakler med evne til å se hva fremtiden vil bringe, men vi kan "spå" fremtiden basert på hva vi har observert historisk!

I finans er man mest opptatt av *avkastning*. Avkastningen til aksje k er definert som

$$R_k = \frac{S_k - S_{k,0}}{S_{k,0}}$$

der $S_{k,0}$ er dagens aksjekurs. Dagens aksjekurser kjenner vi, og man kan lese dem av på en dataskjerm koblet til en børs. På engelsk snakker man om *return*, derav notasjonen R_k . Hvis du kjøper aksjen i dag, så koster det $S_{k,0}$. Selger du den senere, vil det gi en inntekt (eller tap) $S_k - S_{k,0}$, og avkastningen er dermed hva du tjener eller taper relativt til investeringen du har gjort. Merk at hver av de d aksjene vil ha sin avkastning, henholdsvis, $R_1, R_2, \dots, R_d$.

Eksempel 2.1. *Vestas, et stort dansk vindenergiselskap, hadde en aksjekurs på DKK 167,06 den 10. november i 2023, mens en måned senere hadde kurset steget til DKK 186,76 (11. desember, 2023). Hvis du kjøpte en aksje i Vestas 10. november, som koster deg $S_0 = 167,06$, og solgte den 1 måned senere til $S = 186,76$, ville du ha tjent $S - S_0 = 19,70$ danske kroner. Avkastningen din ville ha blitt $R = 19,70/167,06 = 0,12$, eller 12%.*

La oss regne ut hva avkastningen til porteføljen din blir:

$$\begin{aligned} R &= \frac{W - W_0}{W_0} \\ &= \frac{\mathbf{a}^\top (\mathbf{S} - \mathbf{S}_0)}{W_0} \\ &= \sum_{k=1}^d \frac{a_k}{W_0} (S_k - S_{k,0}) \\ &= \sum_{k=1}^d \frac{a_k S_{k,0}}{W_0} \left(\frac{S_k - S_{k,0}}{S_{k,0}} \right) \\ &= \sum_{k=1}^d \frac{a_k S_{k,0}}{W_0} R_k \end{aligned}$$

Her er

$$W_0 = \sum_{k=1}^d a_k S_{k,0}.$$

W_0 angir hvor mange penger du har investert i dag i porteføljen din (altså din formue i dag!). Merk nå at $a_k S_{k,0}$ er summen penger du har investert i aksje k , og dermed blir $a_k S_{k,0}/W_0$ *andelen* av din formue som er investert i aksje k . La oss innføre notasjonen $\mathbf{w} = (w_1, w_2, \dots, w_d)^\top$ som vektoren av andeler, der

$$w_k = \frac{a_k S_{k,0}}{W_0}.$$

Vi ser at

$$\sum_{k=1}^d w_k = \sum_{k=1}^d \frac{a_k S_{k,0}}{W_0} = \frac{1}{W_0} \sum_{k=1}^d a_k S_{k,0} = \frac{W_0}{W_0} = 1$$

Andelene summerer seg til 1, naturlig nok. Vi konkluderer med at porteføljens avkastning blir

$$(12) \quad R = \sum_{k=1}^d w_k R_k = \mathbf{w}^\top \mathbf{R}$$

der $\mathbf{R} = (R_1, R_2, \dots, R_d)^\top$.

Eksempel 2.2. Tenk deg at du har 100 kroner investert i aksjer fra to selskaper, der 40 kroner er brukt til å kjøpe aksjer i det første selskapet (som vi kaller selskap 1), mens de resterende 60 kroner er investert i det andre selskapet (som kalles selskap 2). For enkelthets skyld, var aksjekursene $S_{1,0} = 40$ og $S_{2,0} = 60$, slik at du altså kjøpte én aksje i hvert av de to selskapene. Ett år senere så har aksjekursen til selskap 1 økt til 46 kroner, mens selskap 2 sin kurs har falt til 58 kroner. Dette betyr at din portefølje bestående av de to aksjene er verdt $46 + 58 = 104$ kroner. Investeringen din har en avkastning på $R = (104 - 100)/100 = 0,04$, eller 4%. På den annen side, avkastningen på aksje 1 er $R_1 = (46 - 40)/40 = 3/20$, eller 15%, mens aksje 2 har avkastning $R_2 = (58 - 60)/60 = -1/30$, eller omlag $-3,3\%$. Andelen din som er investert i 1 er $w_1 = 40/100 = 2/5$, og dermed er selvsagt $w_2 = 3/5$. Fra formelen i (12) finner vi da

$$w_1 R_1 + w_2 R_2 = \frac{2}{5} \times \frac{3}{20} + \frac{3}{5} \times \left(-\frac{1}{30}\right) = \frac{1}{25} = 0,04$$

Ikke overraskende er dette det samme som den avkastningen R vi regnet ut over.

Forventet avkastning og risiko til porteføljen din er de neste begrepene vi ønsker å se på. Fra forrige seksjon brukte vi historisk gjennomsnitt til å si noe om hva vi forventer vil være fremtidig gjennomsnittsverdi. Fra historiske aksjekurser kan du derfor regne ut avkastninger, som gir deg en serie med data for historiske avkastninger. Fra disse kan du regne ut gjennomsnittsverdien til avkastningene fra selskapet, og bruke det som beste gjett på hva den fremtidige forventete avkastningen til aksjen blir. Fra den samme historiske serien av avkastninger, kan du regne ut variansen til dataene.

Eksempel 2.3. I Figur 2 ser vi de daglige aksjekursene til vindenergiselskapet Vestas, notert på børsen i Danmark, observert over ett år med start 1. desember 2022 og slutt 1. desember 2023. Tilsammen har vi 252 observerte aksjekurser. Vi regner ut avkastninger over en dag, som gir oss et datasett av 251 daglige avkastninger $r_1, r_2, \dots, r_{251}$. Her regner vi ut r_i ved formelen $r_i = (s_{i+1} - s_i)/s_i$. I Figur 3 har vi plottet avkastningene, som vi ser varierer fra omlag $-0,07$ til $0,10$, eller i spennet -7% til 10% . Gjennomsnittet av disse 251 avkastningsdataene er $\hat{r} = 0,00045$,² som betyr $0,045\%$ i forventet daglig avkastning. Annualisert (altså over ett år), blir den forventete avkastningen $252 \times 0,00045 = 0,11$, eller 11% . Merk at vi multipliserer med 252, som angir antallet handledager på børsen i løpet av et år. Standardavviket til de daglige avkastningene er $\hat{\sigma} = 0,024$. Ofte oppgis dette tallet også annualisert, som i dette tilfellet blir $\sqrt{252} \times 0,024 = 0,38$, eller 38% . Avkastningen på Vestas-aksjer har altså en årlig forventning på 11% , med en volatilitet på 38% . Det betyr at avkastningen kan variere med ett standardavvik mellom $11\% - 38\% = -27\%$ og $11\% + 38\% = 49\%$.

²Fra diskusjonen i forrige seksjon burde vi ha brukt notasjonen $\hat{\mu}_r$ for denne estimerte gjennomsnittsverdien. Vi skal etter dette eksempelet samle samme en vektor av slike gjennomsnittsverdier, og for å få enkel notasjon, så skriver vi rett og slett $\hat{r}$.



FIGURE 2. Daglige kurser (såkalt "closing prices") til Vestas fra 1. desember 2022 til 1. desember 2023.

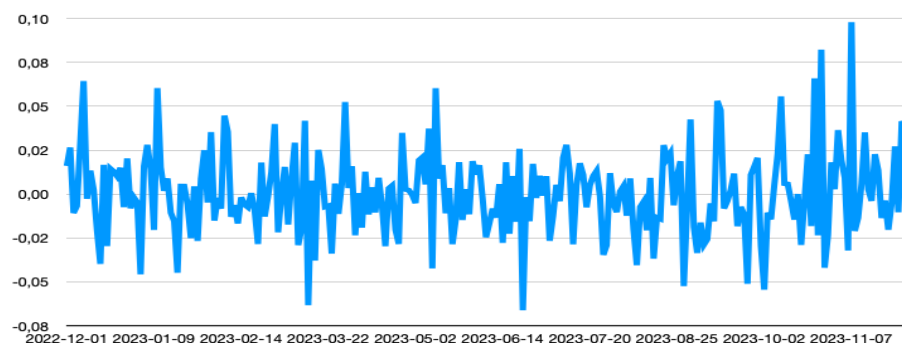


FIGURE 3. Daglige avkastninger til Vestas fra 1. desember 2022 til 1. desember 2023.

La oss anta nå at du har regnet ut forventningsverdien til alle aksjene, og samlet disse inn i en vektor $\hat{\mathbf{r}} = (\hat{r}_1, \dots, \hat{r}_d)^\top$. Videre så har du samlet kovariansene mellom avkastningene til de d aksjene i en $d \times d$ -matrise C . Vi kan da vise:

Teorem 2.4. *Forventet avkastning til porteføljen er*

$$\hat{\mu}_W = \mathbf{w}^\top \hat{\mathbf{r}}$$

og variansen er

$$\hat{\sigma}_W^2 = \mathbf{w}^\top C \mathbf{w}$$

Proof. La $r_1, r_2, \dots, r_n$ være avkastninger fra porteføljen mens $r_{1,k}, r_{2,k}, \dots, r_{n,k}$ er avkastningene fra aksje k . Fra formelen for porteføljes avkastning, finner vi at forventet avkastning er

$$\hat{\mu}_W = \frac{1}{n} \sum_{i=1}^n r_i = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^d w_k r_{i,k} = \sum_{k=1}^d w_k \frac{1}{n} \sum_{i=1}^n r_{i,k} = \sum_{k=1}^d w_k \hat{r}_k = \mathbf{w}^\top \hat{\mathbf{r}}$$

Dette viser første del av teoremet. Variansen til porteføljen er per definisjon gitt som

$$\hat{\sigma}_W^2 = \frac{1}{n} \sum_{i=1}^n (r_i - \hat{\mu}_W)^2 = \left(\frac{1}{n} \sum_{i=1}^n r_i^2 \right) - \hat{\mu}_W^2$$

Men fra formelen for porteføljens avkastning, så finner vi

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n r_i^2 &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^d w_k r_{i,k} \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k,\ell=1}^d w_k w_\ell r_{i,k} r_{i,\ell} \\ &= \sum_{k,\ell=1}^d w_k w_\ell \left(\frac{1}{n} \sum_{i=1}^n r_{i,k} r_{i,\ell} \right) \end{aligned}$$

Men,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n r_{i,k} r_{i,\ell} &= \frac{1}{n} \sum_{i=1}^n (r_{i,k} - \hat{\mu}_k + \hat{\mu}_k)(r_{i,\ell} - \hat{\mu}_\ell + \hat{\mu}_\ell) \\ &= \frac{1}{n} \sum_{i=1}^n (r_{i,k} - \hat{\mu}_k)(r_{i,\ell} - \hat{\mu}_\ell) + \frac{1}{n} \sum_{i=1}^n (r_{i,k} - \hat{\mu}_k) \hat{\mu}_\ell \\ &\quad + \frac{1}{n} \sum_{i=1}^n \hat{\mu}_k (r_{i,\ell} - \hat{\mu}_\ell) + \hat{\mu}_k \hat{\mu}_\ell \\ &= \hat{c}_{k\ell} + \hat{\mu}_k \hat{\mu}_\ell \end{aligned}$$

Setter vi disse uttrykkene sammen, finner vi

$$\begin{aligned} \hat{\sigma}_W^2 &= \sum_{k,\ell=1}^d w_k w_\ell (\hat{c}_{k\ell} + \hat{\mu}_k \hat{\mu}_\ell) - \hat{\mu}_W^2 \\ &= \sum_{k,\ell} w_k w_\ell \hat{c}_{k\ell} + \left(\sum_{k=1}^d w_k \hat{\mu}_k \right)^2 - \hat{\mu}_W^2 \\ &= \mathbf{w}^\top C \mathbf{w}. \end{aligned}$$

Teoremet er bevist. □

Det neste vi ser på er den beste porteføljen du kan sette sammen, som minimerer risikoen til en gitt forventet avkastning.

2.1. Porteføljeteorien til Markowitz. Markowitz³ betraktet følgende problem. Anta at du vil finne den porteføljen som har minst risiko gitt en forventet avkastning r . Matematisk, så søker vi å løse

$$(13) \quad \min_{\mathbf{w} \in \mathbb{R}^d} \frac{1}{2} \mathbf{w}^\top C \mathbf{w}, \quad \text{gitt} \quad \mathbf{w}^\top \hat{\mathbf{r}} = r \quad \text{og} \quad \mathbf{w}^\top \mathbf{1} = 1$$

³Markowitz fikk Nobels minnepris i økonomi i 1990 for denne teorien. For mer informasjon om prisen og hans virke, se <https://www.nobelprize.org/prizes/economic-sciences/1990/markowitz/facts/>

Her er $\mathbf{1} = (1, 1, \dots, 1)^\top$ en d -dimensjonal vektor, og $\mathbf{w}^\top \mathbf{1} = 1$ er kravet om at andelen summerer seg til 1. Koeffisienten $1/2$ foran funksjonen som skal minimeres er der kun fordi vi får littegranne enklere uttrykk senere.

For å løse dette kvadratiske minimeringsproblemet med to sidebetingelser, gjør vi følgende: først innfører vi en ny kvadratisk funksjon som inneholder sidebetingelsene og to parametere λ og γ :⁴

$$L(\mathbf{w}) = \frac{1}{2} \mathbf{w}^\top C \mathbf{w} + \lambda(r - \mathbf{w}^\top \hat{\mathbf{r}}) + \gamma(1 - \mathbf{w}^\top \mathbf{1})$$

Deretter finner vi minimum av denne,

$$\min_{\mathbf{w} \in \mathbb{R}^d} L(\mathbf{w})$$

Den $\mathbf{w}$ som minimerer L vil avhenge av parameterne λ og γ , og noteres derfor som $\bar{\mathbf{w}}(\lambda, \gamma)$. Det neste vi gjør er å finne γ og λ som løser likningene

$$(14) \quad \bar{\mathbf{w}}(\lambda, \gamma)^\top \hat{\mathbf{r}} = r$$

og

$$(15) \quad \bar{\mathbf{w}}(\lambda, \gamma)^\top \mathbf{1} = 1.$$

Løsningene, $\bar{\lambda}$ og $\bar{\gamma}$, vil være slik at sidebetingelsene er oppfylt for $\bar{\mathbf{w}}(\bar{\lambda}, \bar{\gamma})$. Vi setter $\mathbf{w}^* = \bar{\mathbf{w}}(\bar{\lambda}, \bar{\gamma})$, som blir løsningen til (13). La oss følge denne oppskriften:

Først, vi finner optimum av $L(\mathbf{w})$ ved å sette de partiellderiverte med hensyn på $w_1, w_2, \dots, w_d$ av L lik null. Etter litt regning (vis dette selv, og observer også at $1/2$ er gunstig her...), ser vi at

$$\nabla L(\mathbf{w}) = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \\ \vdots \\ \frac{\partial L}{\partial w_d} \end{bmatrix} = C\mathbf{w} - \lambda\hat{\mathbf{r}} - \gamma\mathbf{1} = 0$$

gir løsningen⁵

$$\bar{\mathbf{w}}(\lambda, \gamma) = \lambda C^{-1} \hat{\mathbf{r}} + \gamma C^{-1} \mathbf{1}.$$

Det neste på planen er å løse likningene (14) og (15), det vil si

$$(\lambda C^{-1} \hat{\mathbf{r}} + \gamma C^{-1} \mathbf{1})^\top \hat{\mathbf{r}} = r$$

og

$$(\lambda C^{-1} \hat{\mathbf{r}} + \gamma C^{-1} \mathbf{1})^\top \mathbf{1} = 1.$$

Siden C er symmetrisk, og dermed også den inverse, så gir dette oss to likninger med to ukjente

$$\lambda(\hat{\mathbf{r}}^\top C^{-1} \hat{\mathbf{r}}) + \gamma(\mathbf{1}^\top C^{-1} \hat{\mathbf{r}}) = r$$

$$\lambda(\hat{\mathbf{r}}^\top C^{-1} \mathbf{1}) + \gamma(\mathbf{1}^\top C^{-1} \mathbf{1}) = 1.$$

⁴ λ og γ kalles ofte for Lagrangemultiplikatorer.

⁵Dette er et globalt minimum, siden Hessematriksen til L er lik C , som må være positiv definit. Dette er for de mer matematisk skolerte...

Vi kan løse dette likningssettet, men før vi gjør det, la oss introdusere litt notasjon for å gjøre beregningene litt enklere. Denne notasjonen vil vi holde på videre også. Definer

$$(16) \quad a = \hat{\mathbf{r}}^\top C^{-1} \mathbf{1} = \mathbf{1}^\top C^{-1} \hat{\mathbf{r}}$$

$$(17) \quad b = \hat{\mathbf{r}}^\top C^{-1} \hat{\mathbf{r}}$$

$$(18) \quad c = \mathbf{1}^\top C^{-1} \mathbf{1}$$

Bemerkning 2.5. Normalt er $a, c > 0$ og $bc > a^2$, selv om dette ikke alltid er tilfelle når man analyserer data og regner ut de tre konstantene. I våre analyser videre tenker vi oss at dette er tilfelle.

En løsning av likningene er

$$\bar{\lambda} = \frac{cr - a}{bc - a^2} \quad \text{og} \quad \bar{\gamma} = \frac{b - ar}{bc - a^2}$$

Vi konkluderer med at den optimale porteføljen er gitt som

$$(19) \quad \mathbf{w}^* = \left(\frac{cr - a}{bc - a^2} \right) C^{-1} \hat{\mathbf{r}} + \left(\frac{b - ar}{bc - a^2} \right) C^{-1} \mathbf{1}$$

Vanligvis omskriver man dette med hensyn på r , og får dette resultatet:

Teorem 2.6. Den optimale porteføljen er

$$\mathbf{w}^* = C^{-1} (\mathbf{w}_0 + \mathbf{w}_r \cdot r)$$

der

$$\mathbf{w}_0 = \frac{b\mathbf{1} - a\hat{\mathbf{r}}}{bc - a^2}$$

og

$$\mathbf{w}_r = \frac{c\hat{\mathbf{r}} - a\mathbf{1}}{bc - a^2}$$

Bevis. Vi samler leddene som har med r i uttrykket til $\mathbf{w}^*$ i (19):

$$\begin{aligned} \mathbf{w}^* &= \frac{cr}{bc - a^2} C^{-1} \hat{\mathbf{r}} - \frac{a}{bc - a^2} C^{-1} \hat{\mathbf{r}} + \frac{b}{bc - a^2} C^{-1} \mathbf{1} - \frac{ar}{bc - a^2} C^{-1} \mathbf{1} \\ &= C^{-1} \left(\frac{b\mathbf{1} - a\hat{\mathbf{r}}}{bc - a^2} + \frac{c\hat{\mathbf{r}} - a\mathbf{1}}{bc - a^2} \cdot r \right) \end{aligned}$$

Dette viser uttrykket for $\mathbf{w}^*$ i teoremet.

For å vise at $\mathbf{w}^*$ er den optimale allokeringen til vårt optimeringsproblem, gjør vi et såkalt motsigelsesbevis. Vi antar det motsatte, og viser at det fører til en motsigelse. Anta at det finnes en $\tilde{\mathbf{w}}$ som er slik at $\tilde{\mathbf{w}}^\top \hat{\mathbf{r}} = r$ og $\tilde{\mathbf{w}}^\top \mathbf{1} = 1$, og, $\tilde{\mathbf{w}}^\top C \tilde{\mathbf{w}} < \mathbf{w}^{*\top} C \mathbf{w}^*$. Med andre ord, så gjør $\tilde{\mathbf{w}}$ variansen mindre enn $\mathbf{w}^*$. Vi husker at også $\mathbf{w}^*$ tilfredsstiller sidebetingelsene. Dette gir oss,

$$\begin{aligned} \tilde{\mathbf{w}}^\top C \tilde{\mathbf{w}} &= \tilde{\mathbf{w}}^\top C \tilde{\mathbf{w}} + \bar{\lambda}(r - \tilde{\mathbf{w}}^\top \hat{\mathbf{r}}) + \bar{\gamma}(1 - \tilde{\mathbf{w}}^\top \mathbf{1}) \\ &< \mathbf{w}^{*\top} C \mathbf{w}^* + \bar{\lambda}(r - \tilde{\mathbf{w}}^\top \hat{\mathbf{r}}) + \bar{\gamma}(1 - \tilde{\mathbf{w}}^\top \mathbf{1}) \\ &= \mathbf{w}^{*\top} C \mathbf{w}^* + \bar{\lambda}(r - \mathbf{w}^{*\top} \hat{\mathbf{r}}) + \bar{\gamma}(1 - \mathbf{w}^{*\top} \mathbf{1}) \end{aligned}$$

Men da ser vi at $L(\tilde{\mathbf{w}}) < L(\mathbf{w}^*)$, som er en motsigelse siden $\mathbf{w}^*$ nettopp er minimum av $L(\mathbf{w})$! Dermed har vi vist at $\mathbf{w}^*$ er løsningen på vårt minimeringsproblem. $\square$

2.2. Effisient porteføljefront. La oss regne ut risikoen, det vil si variansen til den optimale markowitzporteføljen: Fra Teoremene 2.4 og 2.6 følger

$$\begin{aligned}\hat{\sigma}_W^2 &= \mathbf{w}^{*\top} C \mathbf{w}^* \\ &= (\mathbf{w}_0^\top + \mathbf{w}_r^\top \cdot r) C^{-1} C C (\mathbf{w}_0 + \mathbf{w}_r \cdot r) \\ &= \mathbf{w}_0^\top C^{-1} \mathbf{w}_0 + 2\mathbf{w}_0^\top C^{-1} \mathbf{w}_r \cdot r + \mathbf{w}_r^\top C^{-1} \mathbf{w}_r \cdot r^2.\end{aligned}$$

Legg merke til at siste uttrykk er et andregradspolynom i r . Etter litt (mye) algebra, der du setter inn definisjonen til $\mathbf{w}_0$ og $\mathbf{w}_r$(gjør selv!).....vil du finne at

$$\hat{\sigma}_W^2 = c \cdot \frac{\left(r - \frac{a}{c}\right)^2}{bc - a^2} + \frac{1}{c}.$$

For å forenkle notasjonen skriver vi σ i stedet for $\hat{\sigma}_W$. Vi ender da opp med en likning som beskriver relasjonen mellom *risikoen* σ og den *forventete avkastningen* r :

$$(20) \quad \sigma^2 = c \cdot \frac{\left(r - \frac{a}{c}\right)^2}{bc - a^2} + \frac{1}{c}$$

Dette er den matematiske formelen for den såkalte *effisiente porteføljefronten*.

Bemerkning 2.7. Den effisiente porteføljefronten kan brukes på to måter: Hvis en investor har bestemt seg for en forventet avkastning r , så kan investoren regne ut den minimale risikoen σ som er mulig å oppnå i markedet ved å investere med porteføljen $\mathbf{w}^*$. Men, man kan også bruke formelen ved å si at man ikke ønsker mer risiko enn σ . Da kan man regne ut den beste forventete avkastningen som er mulig å oppnå. Etter å ha regnet ut denne r 'en, så kan man finne $\mathbf{w}^*$ som beskriver nettopp den investeringen du må gjøre i markedet.

Den effisiente porteføljefronten er best beskrevet grafisk, som en hyperbel i $\sigma - r$ -planet.⁶ La oss dele begge sider i (20) med $1/c$, og flytte det første leddet på høyresiden over til venstre:

$$(21) \quad \frac{\sigma^2}{c^{-1}} - \frac{\left(r - \frac{a}{c}\right)^2}{c^{-2}(bc - a^2)} = 1$$

I Figur 4 har vi illustrert grafisk likningen i (21), der x -aksen løper over verdier av risikoen σ fra om lag 18% til 50%. Den tilhørende forventete avkastningen er vist på y -aksen med verdier som går fra omlag 5,7% til omtrent 7,8%. På figuren har vi en blå og en rød gren av hyperbelen. Til en gitt risiko, for eksempel $\sigma = 30\%$, kan du lese av verdiene til r til å være enten $r = 6,3\%$ eller $r = 7,3\%$ (sånn cirka litt avrundet...). Du kan velge mellom disse to avkastningene, og tar derfor selvsagt den største. Den effisiente porteføljefronten er dermed den *øverste* av de to grenene, som vist på Figur 5. Den effisiente porteføljefronten gir oss hva vi *minimalt* kan ha av risiko til en gitt forventet avkastning, eller motsatt, hva den *maksimale* forventete avkastningen er til en gitt risiko. Vi ser at til 40% risiko kan vi maksimalt klare 7,6% i forventet avkastning. De røde kryssene markerer en risiko og forventet avkastning til de tre verdipapirene som er brukt i analysen i dette eksempelet. Ett av verdipapirene har en forventet avkastning på 7,5% til 40% risiko, som er lavere enn hva vi kan oppnå med markowitzporteføljen til samme risiko. Markowitzporteføljen har, som sagt, 7,6% forventet avkastning til

⁶ σ på x -aksen, og r på y -aksen.

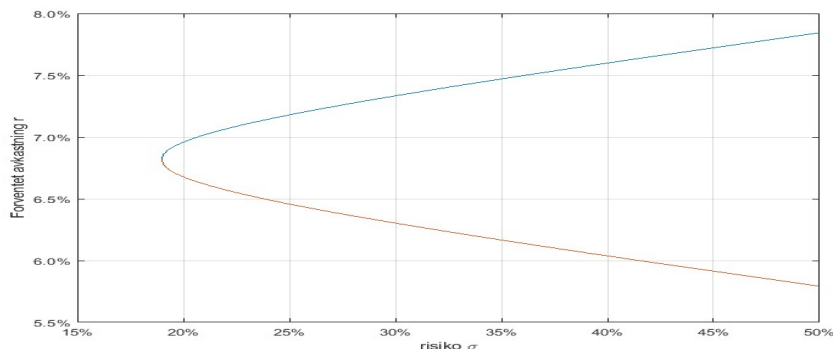


FIGURE 4. Begge ”grener” av den effisiente porteføljefronten med 3 ulike verdipapirer. Med hodet på skakke mot venstre, ser du en hyperbolsk kurve.

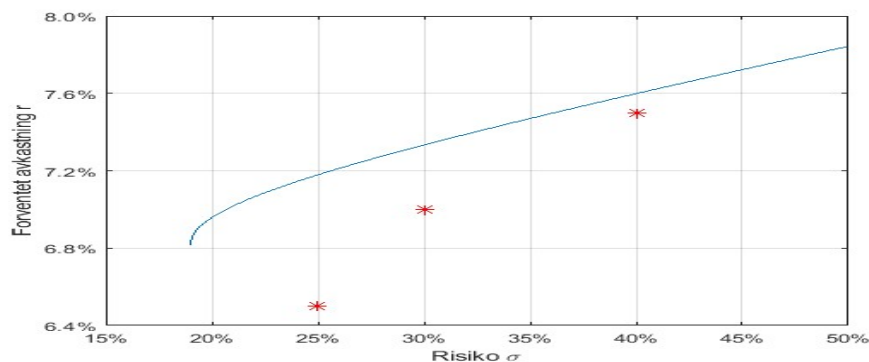


FIGURE 5. Den effisiente porteføljefronten med tre ulike verdipapirer.

denne risikoen. Nå høres kanskje ikke 0,1% veldig mye ut, men for Oljefondet så var 0,1% lik cirka 160 milliarder kroner i 2023.⁷ Vi legger merke til at alle de tre verdipapirene vil gi en dårligere forventet avkastning enn den effisiente porteføljen.

2.3. ESG, eller begrensning av investeringsmuligheter. La oss tenke oss at vi forvalter et fond som investerer i 4 ulike verdipapirer (du kan tenke på 4 ulike sektorer, som for eksempel farmasi, transport, data og elektronikk, og olje). Den effisiente porteføljefronten kan fremstilles grafisk som den svarte kurven i Figur 6. Så vil fondet rette seg etter ESG-krav, og ta bort investeringene i oljesektoren. Fondet investerer sine midler i 3 ulike verdipapirer, og den effisiente fronten blir den grønne grafen. Her er de tre røde stjernene risiko og forventet avkastning til verdipapirene i farmasi, transport og data og elektronikk, mens den svarte stjernen er verdipapiret i olje. Historisk har oljeaksjer vært en meget god investering, og gitt en høy avkastning til en grei risiko.

⁷Oljefondet hadde i desember 2023 en verdi i overkant av 16.000 milliarder kroner. I Statsbudsjettet for 2024 la til sammenligning Regjeringen opp til en bruk av 410 milliarder kroner fra Oljefondet.

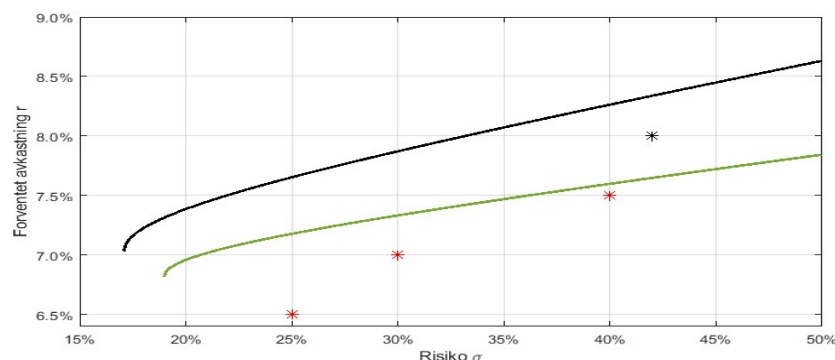


FIGURE 6. Den effisiente porteføljefronten med tre (grønn kurve) og fire (svart kurve) ulike verdipapirer.

Vi legger merke til at den svarte fronten ligger over den grønne. Dette er naturlig siden den optimale investeringen i de 3 verdipapirene gir en mulig investering i de fire papirene når du velger null allokering i olje. Du har større fleksibilitet, og kan dermed oppnå mindre risiko til samme forventete avkastning. Det interessante er at til samme risiko, for eksempel $\sigma = 30\%$, så vil fondet oppleve en nedgang i den forventete avkastningen fra omlag 7,8% til 7,3% når det reduserer investeringsmulighetene fra 4 til 3 verdipapirer. Man taper 0,5%, som for Oljefondet sin del ville bety 800 milliarder kroner. Vi kan se på denne summen som den forventete kostnaden ved å gjøre fondet ESG-vennlig. Det er verdt å merke seg at i denne grafikken ligger olje (markert ved den svarte stjernen) over den grønne grafen, noe som betyr at hvis du setter alle pengene i olje så får du bedre forventet avkastning enn hvis du optimalt investerer i de tre andre til samme risiko. For å oppnå samme forventete avkastning med de tre verdipapirene må du ta en risiko mellom 55-60% (vises ikke i denne grafen), som er betydelig høyere enn hva risikoen til olje er satt til i dette eksempelet.

Bemerkning 2.8. Tallene bak Figur 6 er satt tilfeldig, og er ikke basert på noen analyse av hva vi kan forvente i markedet. Men at olje historisk har vært en god investering fra et finansståsted er det ingen tvil om. Likevel, er nettopp oljebransjen en sektor der historisk avkastning ikke nødvendigvis gir noe godt bilde på hva den fremtidige avkastningen blir. Kan du tenke deg grunner til denne påstanden? Sjekk ut hva man mener med "stranded assets".

3. OPTIMAL PLASSERING AV VIND- OG SOLPARKER

Det foregår en stor omstilling av strømproduksjonen i Europa (og ellers i verden), der man beveger seg bort fra kull og gass og erstatter dette i stor grad med vind og sol. Også i Norge settes det opp store vindmølleparker, både på land og til havs. Dette fører til en variabel strømproduksjon, siden vindkraft kun genereres når det blåser og solkraft når det er solskinn. Problemet man står ovenfor er at man ikke lenger har en jevn generering av strøm (slik gass og kull kan sørge for), men en varierende produksjon av kraft som avhenger av været, og som kan være veldig i utakt med behovet for strøm.

Værforhold som vind og skydekke kan være veldig forskjellig fra sted til sted. For eksempel blåser det ofte sterkere langs kysten enn på land, og vær-situasjonen nord i Norge kan være helt forskjellig fra sør. Spørsmålet man kan stille seg, og som vi skal stille her, er:

Kan man finne steder for vind og solparker som jevner ut produksjonen mest mulig?

En forenklet versjon av denne problemstillingen er: er det bedre å bygge en stor vindpark på ett sted, eller burde jeg bygge to mindre på to helt forskjellige steder? Går vi tilbake til analysen rundt porteføljer i den forrige seksjonen, vet vi at det er mindre risiko å fordele formuen i to aksjer enn å sette alle pengene i én aksje. Vi skal argumentere her for at dette er en innsikt vi også kan benytte oss av når vi skal installere vind- og solkraft.

Anta at $\mathbf{R}$ er vektoren av *kapasitetsfaktorer* for vindkraft i d forskjellige geografiske lokasjoner. Disse lokasjonene er for eksempel steder i Norge der man har mulighet til å bygge en vindpark. Hvis vi nå lar $\mathbf{w}$ være vektoren som beskriver hvor mye kapasitet vi installerer i hver lokasjon, så har vi at produksjonen vil bli

$$R = \mathbf{w}^\top \mathbf{R}$$

Nå kan vi forestille oss at R står for *renewable*, og ikke *return* som i aksjeanalysen. Vi vet ikke hva kapasitetsfaktorene i de ulike lokasjonene ender opp med å bli, men vi kan anslå hva deres forventete verdier er og risikoen rundt disse.

Hva ønsker vi oss fra energisystemet? Vi ønsker oss en jevnt stabil produksjon, altså en produksjon som i snitt holder et visst nivå og ikke avviker for mye fra det. Fra foregående seksjon vet vi at avvik fra snittet kan måles med varians. Siden produksjonen er varierende, kan vi ikke kreve mer enn at den i gjennomsnitt er på et visst nivå. Vi ønsker derfor å løse følgende problem: Finn den optimale installeringen av kapasitet i de d lokasjonene som løser

$$(22) \quad \min_{\mathbf{w} \in \mathbb{R}^d} \frac{1}{2} \mathbf{w}^\top C \mathbf{w}, \quad \text{gitt} \quad \mathbf{w}^\top \hat{\mathbf{r}} = r \quad \text{og} \quad \mathbf{w} \geq 0 \quad \text{og} \quad \mathbf{w}^\top \mathbf{1} = \kappa$$

Observer at dette problemet likner veldig på Markowitz optimeringen fra forrige seksjon, formulert i (13). Men, det er noen forskjeller, som vi nå diskuterer.

I vår formulering av Markowitz i likning (13), så er det ingen begrensinger på *fortegnet* til andelene i $\mathbf{w}$. Selv om disse andelene skal summere seg til 1, så kan vi ha negative andeler (som da samtidig åpner for andeler større enn 1!). For eksempel kunne vi i en tenkt situasjon med tre aksjer ha $w_1 = -2$, $w_2 = 2$ og $w_3 = 1$. Dette ville bety at vi med 100 kroner til rådighet, bruker 200 kroner til aksje 2, og 100 kroner til aksje 3. Vi bruker altså 300 kroner, men har jo bare 100 til rådighet? Andelen $w_1 = -2$ sier oss at vi skal "*shorte*" aksje 1, noe som betyr at vi selger den tilsvarende 200 kroner. Selger vi den, får vi inn de manglende 200 kronene. Å *shorte* en aksje er det samme som å ta opp et banklån, der det vi skal betale tilbake tilsvarende aksjekursen. Når det gjelder bygging av vindparker, så kan vi naturlig nok ikke bygge en negativ park..... Å sette opp "-2" vindmøller er meningsløst. Det er derfor et krav i (22) at $\mathbf{w} \geq 0$.

Bemerkning 3.1. Mange investorer går ikke *short*, så positivitetskravet til $\mathbf{w}$ gir mening også i problem (13).

Videre, siden $\mathbf{w}$ ikke betegner andeler, men faktisk installert kapasitet, er det ikke lenger nødvendig at summen av alle w_k 'ene skal bli 1. Isteden har vi et krav

om at en viss mengde kapasiteter skal installeres, beskrevet ved κ . Dette tallet kan være et måltall fra myndighetene om hvor mye kapasitet som skal installeres totalt.

Med ulikhetskravet $\mathbf{w} \geq 0$ kan vi ikke lenger løse problemet med bruk av Lagrange-multiplikatorer som i seksjonen over. Vi må benytte oss av numeriske metoder for såkalte kvadratiske optimeringsproblemer med ulikhetsbetingelser, eller *kvadratisk programmering*. I Matlab finner du rutinen som kalles **quadprog** som kan løse et problem som (22) numerisk. Du finner tilsvarende rutiner i andre verktøy og pakker.

Nå kan det hende at man har en *indre løsning*, som betyr at minimum av $\mathbf{w}^\top C \mathbf{w}$ gitt $\mathbf{w}^\top \hat{\mathbf{r}} = r$ og $\mathbf{w}^\top \mathbf{1} = \kappa$ finnes for en $\mathbf{w}^* > 0$. Da er positivitetskravet automatisk oppfylt. Men hvis optimum *uten* positivitetskravet er gitt for en $\mathbf{w}^*$ der en eller flere koordinater er negative, så har vi en utfordring. Da er ikke dette en optimal løsning for det problemet vi er interesserte i.

3.1. Beste allokering av kapasiteter som møter etterspørsel. Et alternativt allokeringsproblem er å finne beste plassering og installering av fornybar energi som følger etterspørselen etter strøm best mulig. Vi er med andre ord ute etter å minimere avstanden mellom det som genereres av kraft og det vi trenger av kraft. La E være etterspørsel etter kraft, som vi tenker oss er usikker. Forventet etterspørsel sier vi er $\hat{\mu}_E$, mens variansen er $\hat{\sigma}_E^2$. Hvis vi beholder notasjonen fra foregående analyser, så lar vi produksjonen av kraft være notert som $R = \mathbf{w}^\top \mathbf{R}$, der vektoren $\mathbf{w}$ sier hvor mye kapasitet vi har installert i hver lokasjon, mens $\mathbf{R}$ er vektoren av kapasitetsfaktorer. Hvis det på tidspunkt i produseres r_i , mens etterspørselen er e_i , ønsker vi at det totale avviket skal være minst mulig over tidspunktene $i = 1, \dots, n$, det vil si, vi vil finne en $\mathbf{w}$ slik at $\frac{1}{n} \sum_{i=1}^n (r_i - e_i)^2$ blir minst mulig.

Lemma 3.2. *Anta at $\text{Cov}(R_k, E) = 0$ for $k = 1, \dots, d$. Da vil*

$$\frac{1}{n} \sum_{i=1}^n (r_i - e_i)^2 = \mathbf{w}^\top C \mathbf{w} + (\mathbf{w}^\top \hat{\mathbf{r}} - \hat{\mu}_E)^2 + \text{Var}(E)$$

Proof. I beviset bruker vi notasjonen $\hat{\mu}_P := \mathbf{w}^\top \hat{\mathbf{r}}$. Fra kvadratsetningene ser vi at

$$\begin{aligned} (r_i - e_i)^2 &= (r_i - \hat{\mu}_R)^2 + 2(r_i - \hat{\mu}_R)(\hat{\mu}_R - \hat{\mu}_E) - 2(r_i - \hat{\mu}_R)(e_i - \hat{\mu}_E) \\ &\quad + (\hat{\mu}_R - \hat{\mu}_E)^2 - 2(\hat{\mu}_R - \hat{\mu}_E)(e_i - \hat{\mu}_E) + (e_i - \hat{\mu}_E)^2 \end{aligned}$$

Summer vi opp over alle i og deler på n , finner vi

$$\frac{1}{n} \sum_{i=1}^n (r_i - e_i)^2 = \text{Var}(R) - 2\text{Cov}(R, E) + (\hat{\mu}_R - \hat{\mu}_E)^2 + \text{Var}(E)$$

Fra forrige seksjon så vet vi at $\text{Var}(R) = \mathbf{w}^\top C \mathbf{w}$. Videre, finner vi at

$$\begin{aligned} \text{Cov}(R, E) &= \text{Cov}(\mathbf{w}^\top \mathbf{R}, E) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^d w_k (r_{k,i} - \hat{r}_k) \right) (e_i - \hat{\mu}_E) \\ &= \sum_{k=1}^d w_k \frac{1}{n} \sum_{i=1}^n (r_{k,i} - \hat{r}_k) (e_i - \hat{\mu}_E) \\ &= \sum_{k=1}^d w_k \text{Cov}(R_k, E) = 0 \end{aligned}$$

Resultatet følger. □

Dette gir oss det kvadratiske minimeringsproblemet

$$(23) \quad \min_{\mathbf{w} \in \mathbb{R}^d} \mathbf{w}^\top C \mathbf{w} + (\mathbf{w}^\top \hat{\mathbf{r}} - \hat{\mu}_E)^2 + \text{Var}(E), \quad \text{gitt} \quad \mathbf{w} \geq 0$$

Igen ser vi at vi krever positive verdier på hvor stor kapasitet vi installerer.

La oss se på problemet der vi *ikke* krever at $\mathbf{w} \geq 0$. Definerer vi funksjonen

$$L(\mathbf{w}) = \mathbf{w}^\top C \mathbf{w} + (\mathbf{w}^\top \hat{\mathbf{r}} - \hat{\mu}_E)^2 + \text{Var}(E)$$

ser vi at vektoren av de partiell-deriverte (gradienten) blir

$$\nabla L(\mathbf{w}) = 2C\mathbf{w} + 2(\mathbf{w}^\top \hat{\mathbf{r}} - \hat{\mu}_E)\hat{\mathbf{r}}$$

Etter litt vektorregning vil

$$(\mathbf{w}^\top \hat{\mathbf{r}})\hat{\mathbf{r}} = (\hat{\mathbf{r}}\hat{\mathbf{r}}^\top)\mathbf{w}$$

og dermed finner vi en $\bar{\mathbf{w}}$ som er slik at $L(\bar{\mathbf{w}}) = 0$ gitt som

$$(24) \quad \bar{\mathbf{w}} = \hat{\mu}_E (C + \hat{\mathbf{r}}\hat{\mathbf{r}}^\top)^{-1} \hat{\mathbf{r}}$$

Vi merker oss at hvis dette optimumet tilfredsstiller positivitetsbetingelsen, har vi en løsning av (23)!

Eksempel 3.3. *I et studium⁸ av solinnstråling i Tyskland, fant man at kapasitetsfaktorene i byene Hamburg, Berlin og München i gjennomsnitt var 0,382, 0,393 og 0,456, med tilhørende standardavvik 0,244, 0,247 og 0,265. Vi merker oss at sør i Tyskland er det en betydelig høyere kapasitetsfaktor enn i nord (ikke uventet, kanskje). Med en solcellepark på 1MW, vil man i München kunne produsere 0,456MWh i snitt over en time mens i Hamburg 0,382MWh. Korrelasjonen mellom München og Hamburg er bare 0,05, mens mellom München og Berlin er den 0,16. Korrelasjonen mellom Berlin og Hamburg er relativt stor, 0,38, noe som sier oss at solinnstrålingen i disse to byene samsvarer relativt ofte, mens München og Hamburg ser ut til å være nesten uavhengige av hverandre. Eller, med andre ord, det er ofte skyer i både Hamburg og Berlin, mens skydekket i München ikke har noe samsvar med skydekke i Hamburg, men delvis med Berlin.*

⁸Se FE Benth: PPA investments of minimal variability. In *Quantitative Energy Finance – Recent Trends and Developments*, Springer Verlag, 2024.

La oss tenke oss at vi vil installere solcellepaneler i de tre byene som tilsammen skal møte en forventet etterspørsel etter strøm på $\hat{\mu}_E = 100\text{MWh}$ med minst mulig variasjon. Vi finner i dette tilfellet at

$$\bar{\mathbf{w}} = \begin{bmatrix} 68 \\ 54 \\ 84 \end{bmatrix}$$

Vi oppnår minst variasjonen ved å bygge en solcellepark i Hamburg på 68MW, i Berlin på 54MW mens i München på 84MW. Alle disse kapasitetene er positive, så dette er en løsning av (23).

Det er interessant å merke seg at München har et litt større standardavvik i sin kapasitetsfaktor enn de to andre byene, men i prosent er denne mindre. Det er altså gunstigere forhold i München for solpaneler enn i de to andre byene. Men likevel er det ikke optimalt å installere $100/0,456=219\text{MW}$ i bare München, men bedre å spre dette rundt på de tre byene. Med 219MW i München, vil vi i snitt produsere eksakt det som er forventet etterspørsel. Med den optimale $\bar{\mathbf{w}}$ blir snittproduksjonen betydelig lavere, faktisk bare 86MWh. Men, med $\text{Var}(E) = 100$, blir $L(\bar{\mathbf{w}}) = 1559$ mens $L([0, 0, 219]) = 3220$, altså mer enn dobbelt så stor. En mer rettferdig sammenligning ville være å installere totalen 206MW (som er summen av kapasitetene i de tre byene) i München alene. Dette ville gi en feil 3051, som fortsatt er betydelig større enn den optimale. Til slutt, vi kunne også tenkt oss at vi installerte 189MW i München, som er hva vi trenger for å produsere i snitt 86MWh. Avviket fra etterspørselen ville nå bli 2941, som fortsatt er veldig mye høyere enn den optimale.