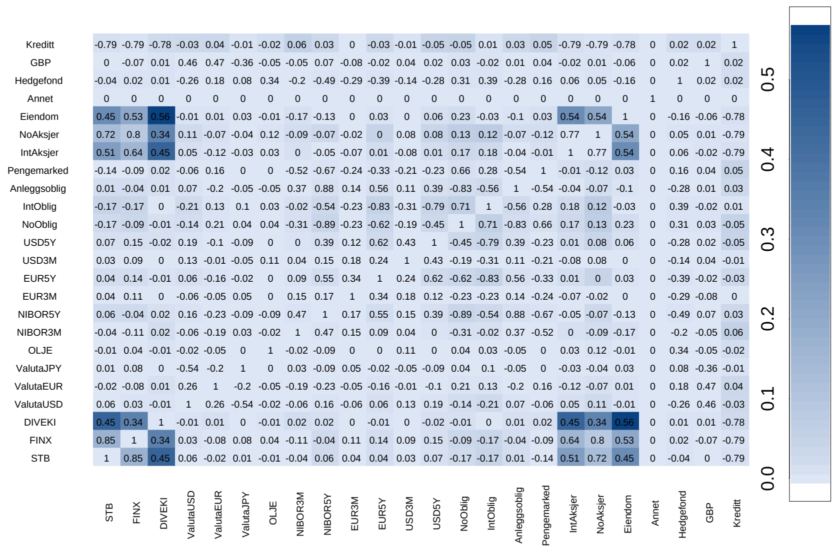
Notat

Justering til gyldig korrelasjonsmatrise

Absolute difference between original and adjusted matrix (major, with weights)



Notatnr SAMBA/11/08

Forfatter Anders Løland

Dato 12. mars 2008

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Informatikkbygningen ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistikk. Det jobbes med svært mange forskjellige problemstillinger slik som estimering av torskebestanden, finansiell risiko, beskrivelse av geologien i petroleumsreservoarer og overvåking av klimaendringer. NR er også ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Problemstillinger kan være å overvåke inntrengning i datasystemer, e-læring i skole og næringsliv, bruk av datateknologi i markedsanalyser samt anvendelser av multimedia på forskjellige plattformer. NRs visjon er forskningsresultater som brukes og synes.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Tittel	**Justering til gyldig korrelasjonsmatrise**
Forfatter	**Anders Løland** <anders.loland@nr.no>
Dato	12. mars 2008
Publikasjonsnummer	SAMBA/11/08

Sammendrag

Vi ser på problemet med å finne den "beste" gyldige korrelasjonsmatrisen fra en ikke-gyldig korrelasjonsmatrise. Vi sammenligner to metoder på justering av en korrelasjonsmatrise med 24 aktiva. Resultatene viser at det er mulig å legge mer vekt på enkelte korrelasjoner enn andre, men legges det for mye vekt på noen, må en betale ved at andre korrelasjoner må justeres tildels mye for at sluttresultatet skal bli en gyldig korrelasjonsmatrise.

Emneord	Korrelasjonsmatrise, vekter, norm
Målgruppe	
Tilgjengelighet	Åpen
Prosjekt	SFI02-Risk
Prosjektnummer	220302
Satsningsområde	Finans, forsikring og råvarer
Antall sider	17

Innhold

1 Innledning

Å finne en gyldig korrelasjonsmatrise er et svært vanlig problem i finans (Higham, 2002). En gyldig korrelasjonsmatrise er

- symmetrisk,
- positiv semidefinitt og med
- enere på diagonalen.

Årsaken til at en korrelasjonsmatrise ikke er gyldig kan være at en har estimert en (gyldig) korrelasjonsmatrise for log-avkastninger til aksjeindekser fra historiske data og i ettertid manuelt endret noen av korrelasjonene. Alternativt kan det mangle data for noen aktiva, slik at en må velge korrelasjoner basert på ekspertkunnskap. Dessuten kan det mangle noen observasjoner på ulike tidspunkter for ulike aktiva, slik at bare parvise korrelasjoner kan estimeres. I disse tre tilfellene kan en risikere at korrelasjonsmatrisen ikke er gyldig. Det er vanskelig å manuelt justere korrelasjonene slik at matrisen blir gyldig, men det finnes heldigvis ulike automatiske metoder for en slik justering.

I kapittel 2 vil vi se på en ofte brukt algoritme av Rebonato og Jäckel (1999), og sammenligne denne med en nyere og bedre algoritme (Pietersz og Groenen, 2004). Vi bruker en korrelasjonsmatrise med 24 aktiva som eksempel i kapittel 3.

2 Metoder

I dette kapittelet trenger vi først noen definisjoner, deretter beskriver vi kort algoritmene til Rebonato og Jäckel (1999) og Pietersz og Groenen (2004).

2.1 Definisjoner

Definisjon 1 (Korrelasjonsmatrise) *La* $\boldsymbol{C}$, $\dim(\boldsymbol{C}) = n \times n$, *være den reelle, symmetriske, men potensielt ikke-gyldige korrelasjonsmatrisen. La* $\widetilde{\boldsymbol{C}}$ *være en gyldig korrelasjonsmatrise som er avledet fra* $\boldsymbol{C}$.

Definisjon 2 (Frobeniusnormen) *Frobeniusnormen til en matrise,* $\boldsymbol{X}$, $\dim(\boldsymbol{X}) = n \times m$, *er*

$$||\boldsymbol{X}||_F^2 = tr(\boldsymbol{X}\boldsymbol{X}^T) = \sum_{i=1}^{n}\sum_{j=1}^{m}|x_{ij}|^2.$$

Definisjon 3 (Den vektede Frobeniusnormen) *La* $\boldsymbol{W}$ *være en positiv, symmetrisk vektmatrise, og anta nå at* $\dim(\boldsymbol{X}) = n \times n$. *Da er den vektede Frobeniusnormen*

$$||\boldsymbol{X}||_{F,\boldsymbol{W}}^2 = tr(\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\boldsymbol{W}) = \sum_{i=1}^{n}\sum_{j=1}^{n}|w_{ij}x_{ij}|^2.$$

2.2 Spektraldekomponering

Rebonato og Jäckel (1999) introduserer problemet "å finne den beste korrelasjonsmatrisen", og presenterer to teknikker for å finne en gyldig korrelasjonsmatrise: hypersfæredekomposisjon og spektraldekomposisjon (som også er kalt prinsipalkomponentanalyse (PCA)). Vi vil konsentrere oss om den sistnevnte teknikken. Den er den enkleste å implementere og derfor ofte brukt i praksis. En gyldig korrelasjonsmatrise $\widetilde{\boldsymbol{C}}$ kan da finnes med algoritme 1 på neste side. Algoritmen er svært rask.

Schöttle og Werner (2004) kritiserer denne metoden av to grunner:

1. Den gir bare en tilnærmet løsning på problemet

$$\min||\boldsymbol{C} - \widetilde{\boldsymbol{C}}||_{F,\boldsymbol{W}} \text{ slik at } \widetilde{\boldsymbol{C}} \text{ er en gyldig korrelasjonsmatrise.} \qquad (2.1)$$

2. Det er vanskelig å ta hensyn til ytterligere beskrankninger.

1: Regn ut egenverdiene λ_i og egenvektorene $\boldsymbol{s}_i$, $i = 1, 2, \ldots, n$. Da er $\boldsymbol{S}$ egensystemet til $\boldsymbol{C}$;

$$\boldsymbol{C} \cdot \boldsymbol{S} = \boldsymbol{\Lambda} \cdot \boldsymbol{S},$$

hvor $\boldsymbol{\Lambda} = \text{diag}(\lambda_i)$.

2: Sett alle negative egenverdier λ_i lik 0. De korrigerte egenverdiene kalles $\widetilde{\lambda}_i$.

3: La $\widetilde{\lambda}_i \boldsymbol{s}_i$ utgjøre kolonnene i $\widetilde{\boldsymbol{B}}$.

4: B finnes ved å normalisere *radvektorene* til $\widetilde{\boldsymbol{B}}$ til lengde 1.

Per konstruksjon er da

$$\widetilde{\boldsymbol{C}} = \widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{B}}^T$$

en gyldig korrelasjonsmatrise; positiv semidefinitt og med enere på diagonalen.

Algoritme 1. Algoritme for å finne en korrelasjonsmatrise $\widetilde{\boldsymbol{C}}$ nær en gitt matrise $\boldsymbol{C}$.

2.3 Alternativ metode

Det finnes en hel rekke metoder som løser (2.1). Higham (2002) løser problemet ved å minimere en vektet Frobeniusnorm ved hjelp av konveks analyse. Qi og Sun (2006) løser problemet ved hjelp av en Newton-metode. Forfatterne oppnår kvadratisk konvergens istedenfor lineær konvergens. Grubišić og Pietersz (2007) gir en oversikt over disse to metodene og sammenligner dem med geometrisk optimering, som viser seg å være den raskeste algoritmen. Vi har imidlertid valgt å prøve algoritmen til Pietersz og Groenen (2004), fordi den, som forfatterne påpeker, er den eneste som er

1. effektiv,
2. enkel å implementere,
3. i stand til å håndtere generelle vekter[1] og
4. er garantert å konvergere til et lokalt minimum

samtidig. Merk at det finnes metoder som klarer en eller flere av disse punktene, men ikke alle samtidig. Grubišić og Pietersz (2007) klarer i tillegg å identifisere om løsningen er et lokalt eller globalt minimum, men deres algoritme er betraktelig vanskeligere å implementere.

1. I for eksempel rentemodellering kan en ha en parametrisk form på korrelasjonen mellom to punkter på rentekurven. Vektene kalles da generelle, siden de ikke er konstante.

2.3.1 Pieterz og Groenens metode: majorisering

Pietersz og Groenen (2004) minimerer Frobeniusnormen (definisjon 2 og 3), slik at løsningen $\widetilde{\boldsymbol{C}}$ er en gyldig korrelasjonsmatrise, ved hjelp av såkalt majorisering, som er en effektiv teknikk for minimering av funksjoner. Vi vil ikke gå inn på alle detaljene her, men angir en oppskrift i algoritme 2. Merk at det er mulig å oppgi ønsket rang $d \leq n$ til $\widetilde{\boldsymbol{C}}$, og at algoritmen itererer seg fram til en løsning med en gitt nøyaktighet.

Finn startmatrise $\boldsymbol{X}$ ved hjelp av modifisert PCA-metode (algoritme 1).
$k = 0$
while normen til gradienten til f i $\boldsymbol{X}^{(k)} = \boldsymbol{X} < \epsilon_{||\nabla f||}$ og forbedringen i funksjonsverdien $f_{k-1}/f_k - 1 < \epsilon_f$ **do**
    $k = k + 1$
    for $i = 1$ to n **do**
        $\boldsymbol{B} = \sum_{i \neq j} w_{ij} \boldsymbol{x}_j \boldsymbol{x}_j^T$
        La λ være største egenverdi til $d \times d$-matrisen $\boldsymbol{B}$
        $\boldsymbol{z} = \lambda \boldsymbol{x}_i - \boldsymbol{B}\boldsymbol{x}_i + \sum_{i \neq j} w_{ij} c_{ij} \boldsymbol{x}_j$
        if $\boldsymbol{z} \neq \boldsymbol{0}$ **then**
            $\boldsymbol{x}_i = \boldsymbol{z}/||\boldsymbol{z}||_2$
        end if
    end for
end while
Resultat: $n \times n$-matrisen $\widetilde{\boldsymbol{C}} = \boldsymbol{X}\boldsymbol{X}^T$ er rang-d-tilnærmingen til $\boldsymbol{C}$.

Algoritme 2. Algoritme for å finne en korrelasjonsmatrise $\widetilde{\boldsymbol{C}}$ med lav rang lokalt nærmest en gitt matrise $\boldsymbol{C}$. $f = ||\boldsymbol{C} - \boldsymbol{X}\boldsymbol{X}^T||_{F,\boldsymbol{W}}$, c_{ij} er et element i $\boldsymbol{C}$, d er ønsket rang, $\boldsymbol{W}$ er vektmatrisen, $\epsilon_{||\nabla f||}$ er konvergenskriteriet til normen til gradienten til f og ϵ_f er konvergenskriteriet til forbedringen i funksjonsverdi. Konvergenskriteriene er konstanter som angir ønsket numerisk presisjon.

3 Resultater

Vi har tatt utgangspunkt i en korrelasjonsmatrise (figur 3.1) som er estimert på bakgrunn av historiske data, men hvor korrelasjonen mellom norske og internasjonale aksjer er endret til 0,77 og det er lagt til tre ekstra aktiva/dimensjoner:

Annet: er subjektivt satt til å være ukorrelert med alle andre aktiva.

DIVEKI: er subjektivt satt til å være ukorrelert med alle andre aktiva, bortsett fra kreditt.

Kreditt: er subjektivt satt til å være negativt korrelert (-0,8) med STB, FINX, DIVEKI, internasjonale aksjer, norske aksjer og eiendom, og ukorrelert med alle andre aktiva.

Både det at korrelasjonen mellom norske og internasjonale aksjer er endret til 0,77 og at kreditt er lagt til, fører til at matrisen ikke er en gyldig korrelasjonsmatrise: Korrelasjonsmatrisen har to negative egenverdier.

Resultatene av å benytte henholdsvis Rebonatos metode (algoritme 1) og majoriseringsmetoden (algoritme 2) uten vekter ($\boldsymbol{W} = \mathbf{1}$) på korrelasjonsmatrisen i figur 3.1 er vist i figur 3.2 og 3.3. Det er kanskje ikke så lett å se, men avstanden mellom original og justert korrelasjonsmatrise er høyest for Rebonatos metode. For å sammenligne de ulike justerte korrelasjonsmatrisene bedre, oppgir vi avstanden mellom justert og original matrise i tabell 3.1.

Vekt	Majorisering	Rebonato
Ingen vekt	94,6%	100%
Høye vekter	226%	–
Medium vekter	147%	–
Lavere vekter	179%	–

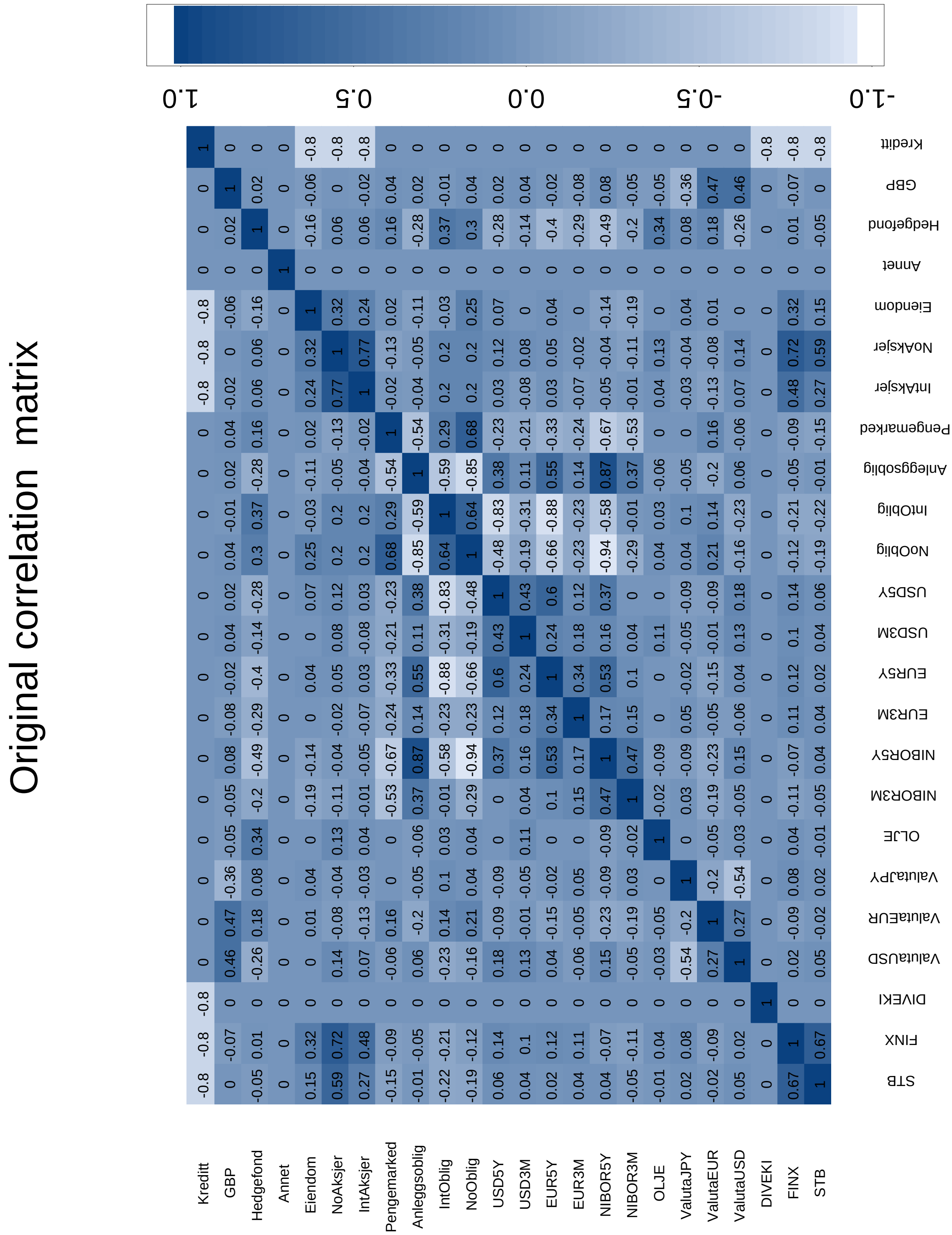
Tabell 3.1. $||\boldsymbol{C} - \widetilde{\boldsymbol{C}}||_F$; avstanden mellom original ($\boldsymbol{C}$) og justert ($\widetilde{\boldsymbol{C}}$) korrelasjonsmatrise, i forhold til Rebonatos spektraldekomposisjonsmetode (100%).

Siden det for eksempel er spesielt viktig at korrelasjonen mellom norske og internasjonale aksjer og de negative korrelasjonene mellom kreditt og resten beholdes i stor grad, har vi i figur 3.4 satt elementene i vektmatrisen til 100 for disse korrelasjonene og 1 ellers. Vi ser at vi i større grad beholder disse korrelasjonene, men må betale prisen ved at andre korrelasjoner endres mer enn kanskje ønskelig, som korrelasjonen mellom eiendom og DIVEKI som endres fra 0 til 0,56.

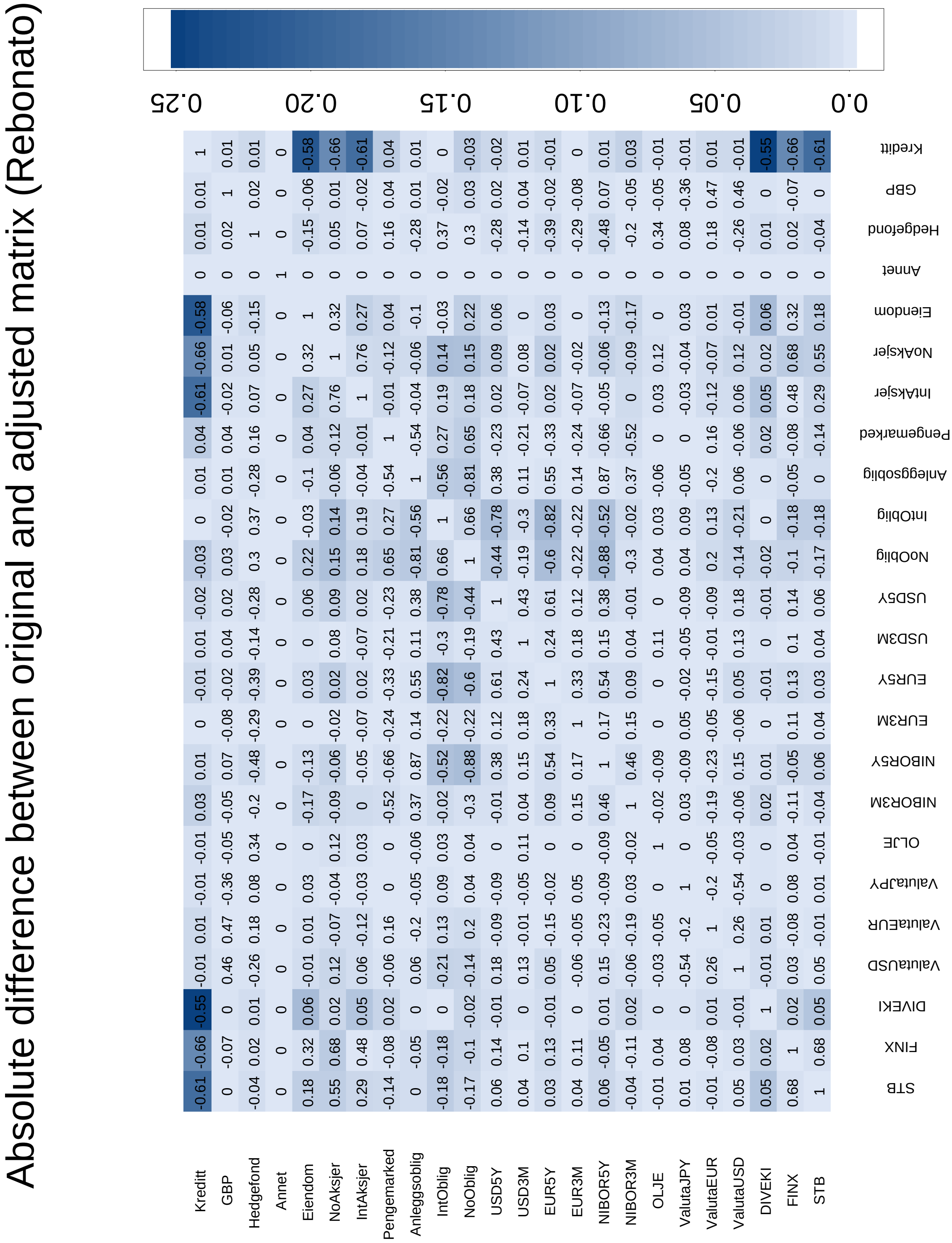
Figur 3.5 viser det samme som figur 3.4, men med vekter på 10 istedenfor 100. Vi ser at utslagene ikke blir like ekstreme som i figur 3.4. Også figur 3.6 viser det samme som figur 3.4, men med vekter på 20 istedenfor 100, samt at korrelasjonene som er 0 mellom kreditt og andre aktiva vektes med kun 0,01. Vi har med andre ord tillatt større justering av de sistnevnte korrelasjonene, fordi disse ikke er så viktige å beholde. Vi ser at utslagene ikke blir like ekstreme, men at forskjellen ikke er så stor fra figur 3.5. Av tabell 3.1 ser vi at (den uvektede) avstanden fra den opprinnelige korrelasjonsmatrisen øker når vi innfører vekter.

3.1 Oppsummering

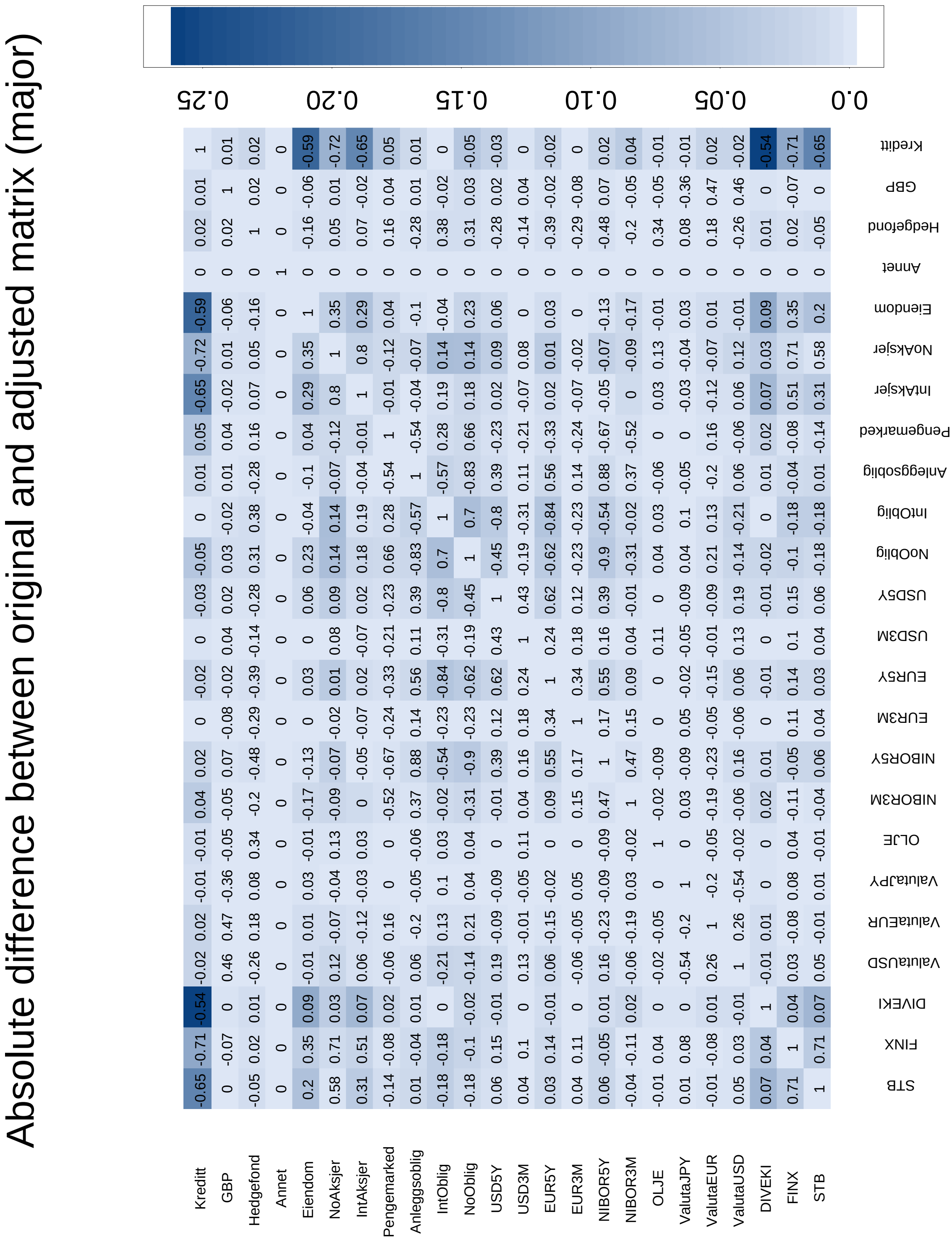
Metoden til Pietersz og Groenen er bedre enn Rebonato og Jäckels. Det er mulig å legge mer vekt på enkelte korrelasjoner enn andre, men legges det for mye vekt på noen, er kostnaden at andre korrelasjoner må justeres tildels mye for at sluttresultatet skal bli en gyldig korrelasjonsmatrise.



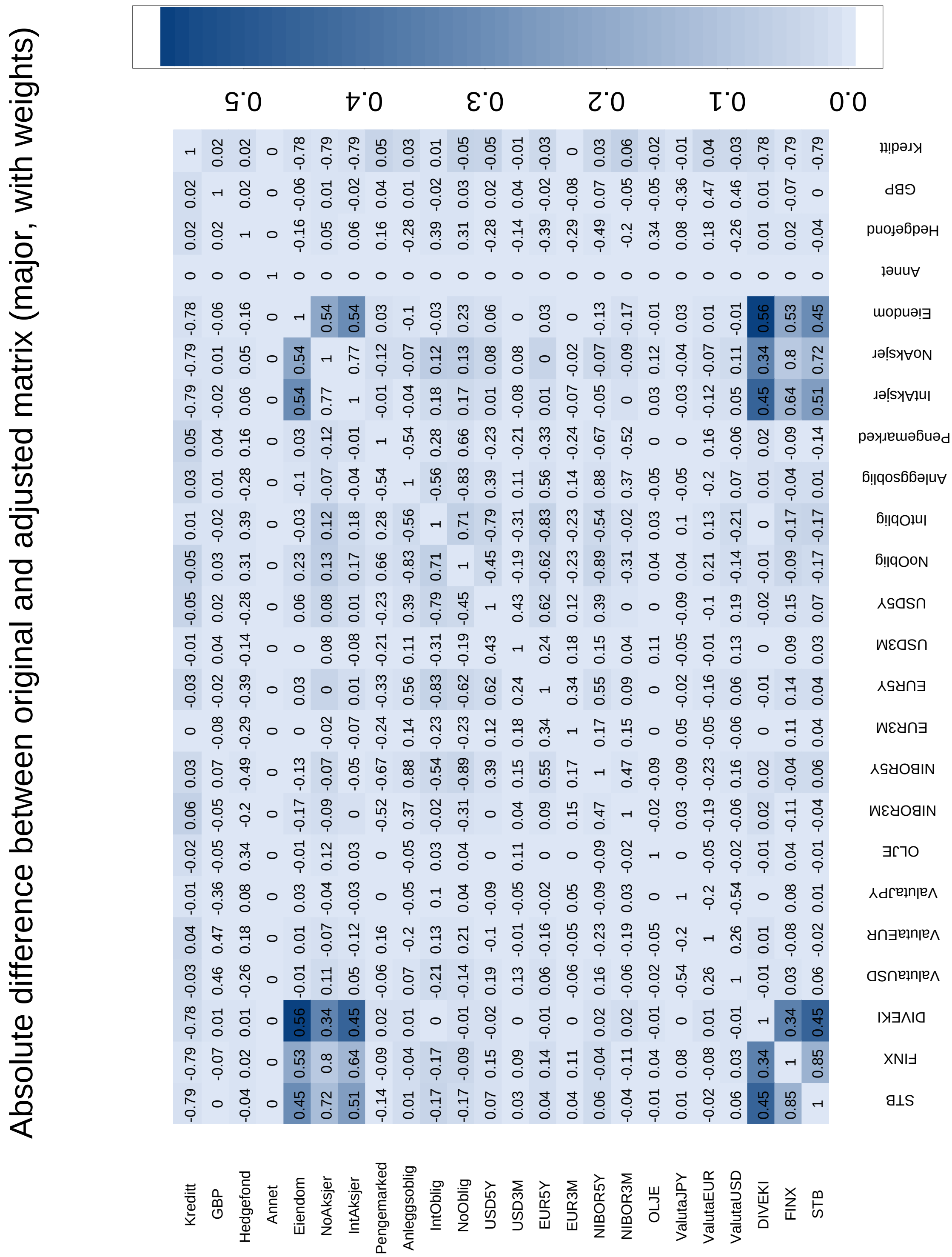
Figur 3.1. Korrelasjonsmatrise C før justering.



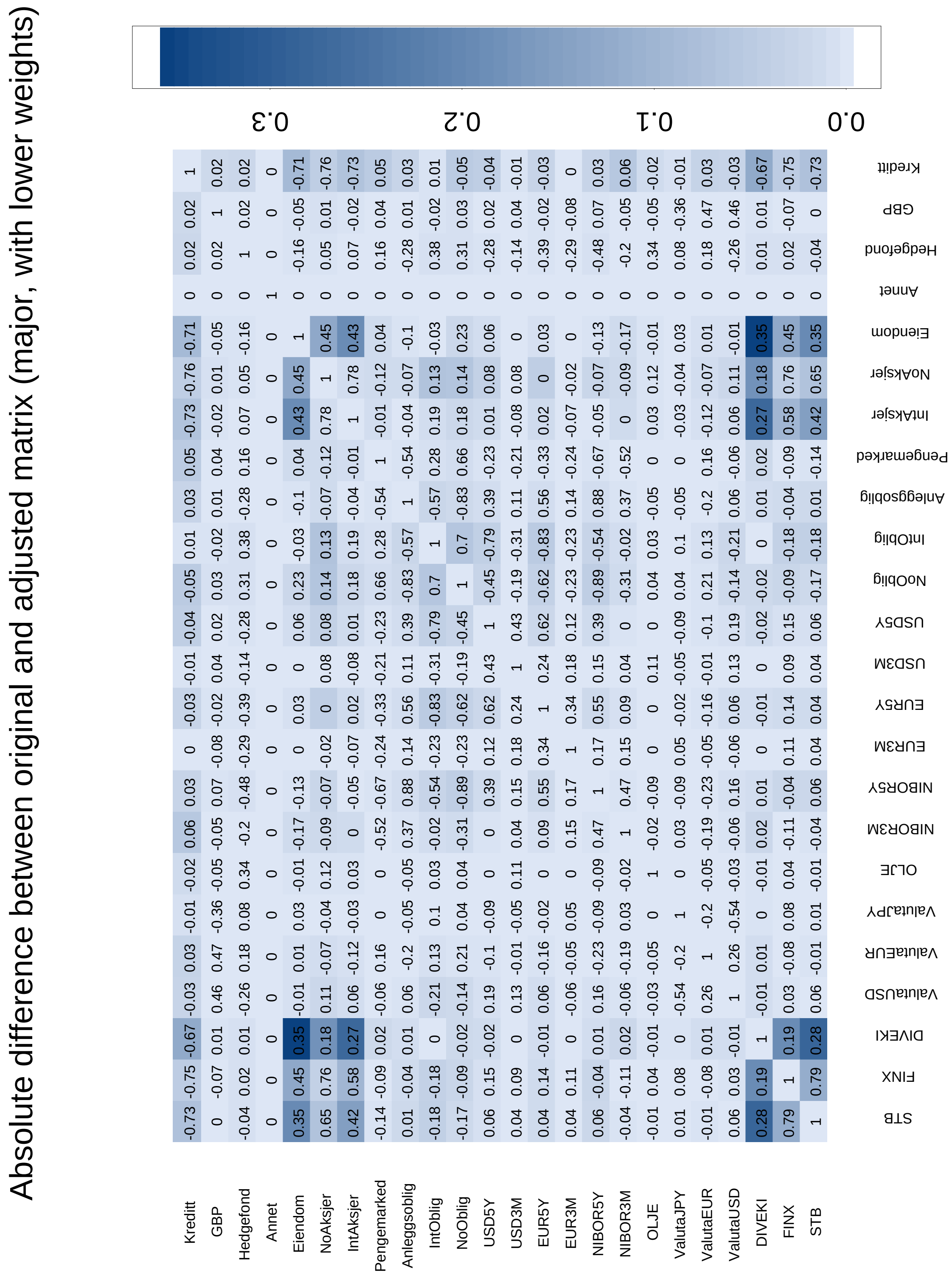
Figur 3.2. Korrelasjonsmatrise $\widetilde{C}$ etter Rebonatojustering. Tallene angir korrelasjonene etter justering. Fargeskalaen angir absoluttavvik mellom de justerte og de opprinnelige korrelasjonene; $|c_{ij} - \widetilde{c}_{ij}|$.



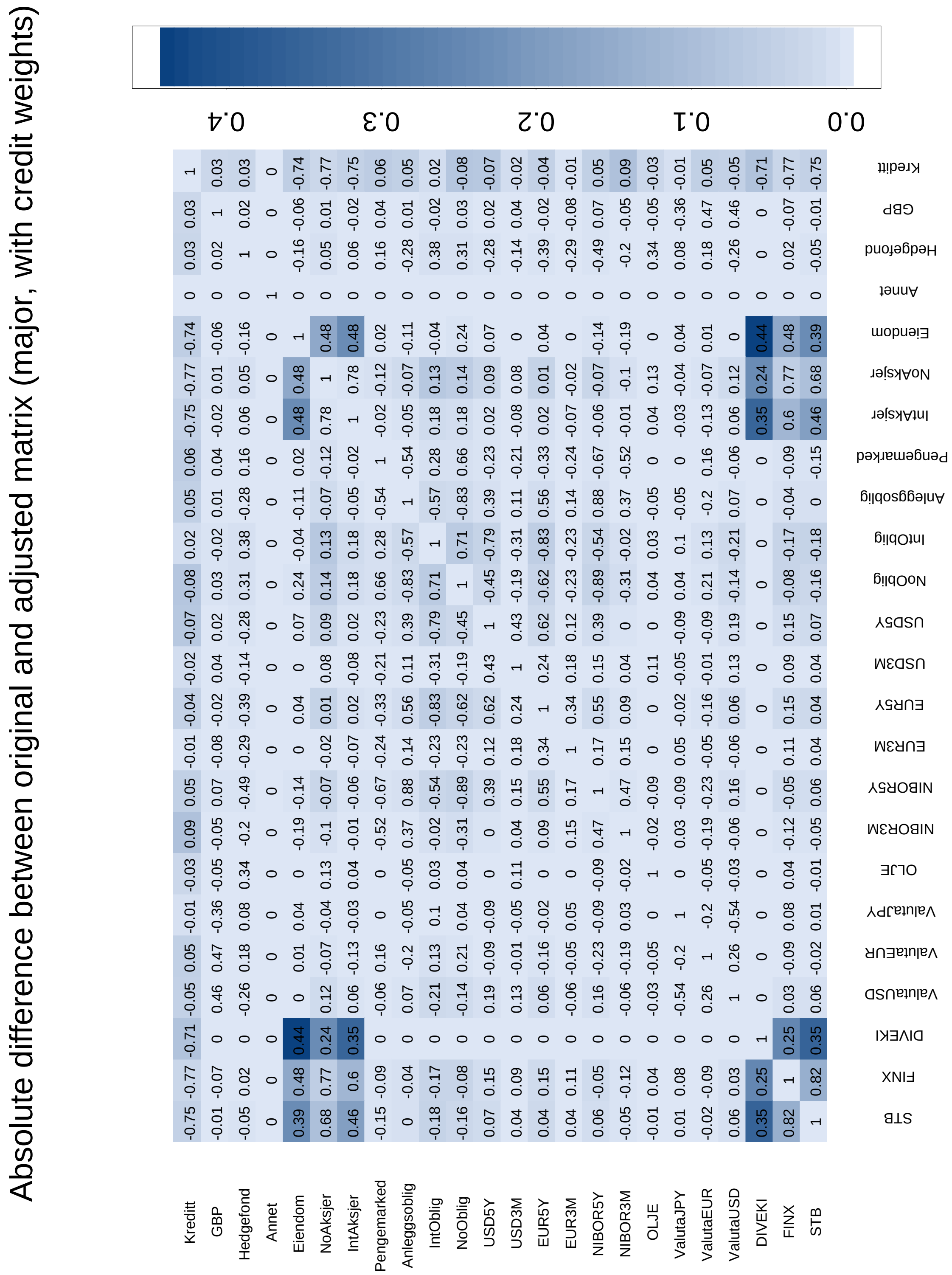
Figur 3.3. Korrelasjonsmatrise $\widetilde{C}$ etter majoriseringsjustering. Tallene angir korrelasjonene etter justering. Fargeskalaen angir absoluttavvik mellom de justerte og de opprinnelige korrelasjonene; $|c_{ij} - \widetilde{c}_{ij}|$.



Figur 3.4. Korrelasjonsmatrise $\widetilde{C}$ etter majoriseringsjustering, med vektmatrise med høye vekter. Tallene angir korrelasjonene etter justering. Fargeskalaen angir absoluttavvik mellom de justerte og de opprinnelige korrelasjonene; $|c_{ij} - \widetilde{c}_{ij}|$.



Figur 3.5. Korrelasjonsmatrise $\widetilde{C}$ etter majoriseringsjustering, med vektmatrise med medium vekter. Tallene angir korrelasjonene etter justering. Fargeskalaen angir absoluttavvik mellom de justerte og de opprinnelige korrelasjonene; $|c_{ij} - \widetilde{c}_{ij}|$.



Figur 3.6. Korrelasjonsmatrise $\widetilde{C}$ etter majoriseringsjustering, med vektmatrise med lave vekter. Tallene angir korrelasjonene etter justering. Fargeskalaen angir absoluttavvik mellom de justerte og de opprinnelige korrelasjonene; $|c_{ij} - \widetilde{c}_{ij}|$.

Referanser

Grubišić, I. og Pietersz, R. (2007). Efficient rank reduction of correlation matrices. *Linear Algebra and its Applications*, 422(2-3):629–653.

Higham, N. J. (2002). Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis*, 22:329–343.

Pietersz, R. og Groenen, P. J. F. (2004). Rank reduction of correlation matrices by majorization. *Quantitative Finance*, 4:649–662.

Qi, H. og Sun, D. (2006). A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications*, 28(2).

Rebonato, R. og Jäckel, P. (1999). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. Technical report, Quantitative Research Centre of the NatWest Group.

Schöttle, K. og Werner, R. (2004). Improving the most general methodology to create a valid correlation matrix". In Brebbia, C. A., editor, *Risk Analysis IV*, volume 9, side 701–712. WIT Press.