Eilif Lande Wekre og Madeleine Heiberg

Pairs Trading

Er det mulig å oppnå positiv avkastning ved å benytte en statistisk arbitrasjestrategi på danske og tyske statsobligasjoner?

FIN3900: Masteroppgave i Finansiell Økonomi
Institutt for Samfunnsøkonomi
Fakultet for økonomi

Master i Finansiell Økonomi
Trondheim, Juni 2019
Veileder: Knut Anton Mork

NTNU Norges teknisk-naturvitenskapelige universitet

Forord

Denne masteroppgaven markerer slutten av et 2 år langt masterstudie i Finansiell Økonomi på Norges teknisk- naturvitenskaplige universitet.

I arbeidet med masteroppgaven har vi andvendt faglig teori fra emner på masterstudiet i praksis, samt tilegnet oss nye kunnskaper innen økonometrisk metode. Vi har også tatt i bruk nytt verktøy for databehandling som har vært tidkrevende i denne prosessen. Arbeidet har vært veldig lærerikt og en god erfaring vi vil ta med oss videre.

Vi ønsker å takke Knut Anton Mork for veiledning dette semesteret, med verdifulle tilbakemeldinger underveis i prosessen. Vi ønsker også å takke Mathilde Hotvedt for betydningsfull opplæring i programmeringsspråk, samt familie, venner, medstudenter og medhjelpere for nyttig tilskudd til masteroppgaven.

Sammendrag

Pairs trading er en forholdsvis ny arbitrasjestrategi hvor svingninger i ulike markeder blir utnyttet for å oppnå profitt i et nullsumspill. Vi har undersøkt utfallet av en pairs trading strategi i obligasjonsmarkedet ved bruk av den danske og tyske statsobligasjonen. Analysen har blitt utført på 3- og 10-års renter med daglige observasjoner. Vi har brukt avstandstilnærmingen og kointegrasjonstilnærmingen som metode for å vurdere om den danske og tyske renten er et optimalt obligasjonspar, og for gjennomførelsen av strategien. Tilnærmingene har blitt sammenlignet og resultatet viser en positiv avkastning i både 3- og 10-års obligasjonene. For hele datasettet gir 3-års obligasjonen en positiv kontantstrøm på 0,0837 og 0,1528 i henholdsvis avstands- og kointegrasjonstilnærmingen. 10-års obligasjonen gir høyere total kontantstrøm på 0,2952 og 0,3342 i avstands- og kointegrasjonstilnærmingen. Videre har vi sett nærmere på finanskrisen og eurokrisen og deres innvirkning på strategien. Å utelukke krisene fra analysen reduserte antall handler betydelig og resulterte i lavere total kontantstrøm i de fleste tilfeller. Vi kan dermed konkludere at strategien trigges når det er større svingninger og uro i markedet.

Abstract

Pairs trading is a relatively new arbitrage strategy which is used to take advantage of market variations to achieve profit in a zero sum game. We have investigated the outcome of the strategy in the bond market using the Danish and German government bond. The analysis has been performed on both the 3- and the 10-year bond with daily observations. Two approaches, the distance approach and the cointegration approach, have been used to valuate if the Danish and German government bond is an optimal pair, and for the implementation of the strategy. The two approaches were compared against each other and the results show a positive cashflow for both approaches in the 3- and 10-years bond. Using the whole dataset, the 3-years bond resulted in a positive cashflow of 0,0837 and 0,1528 for the distance approach and the cointegration approach, whereas the 10-years bond resulted a higher cashflow with 0,2952 and 0,3342 respectively. In addition, the financial crisis and the euro crisis and their influence on the strategy have been investigated. It was found that excluding the crises from the analysis reduced the number of trades significantly and in most cases resulted in lower total cash flow. We therefore conclude that the strategy is triggered when the market is volatile.

Innhold

1 Introduksjon

1.1 Bakgrunn

Pairs Trading er en selvfinansierende handelsstrategi som ble introdusert på Wall Street på begynnelsen av 1980-tallet. En gruppe bestående av forskere, matematikere og fysikere ble samlet av Morgan Stanley & Co for å studere arbitrasjemulighetene i aksjemarkedet. Ved å bruke avanserte statistiske modeller og utvikle et automatisk handelsprogram for utnyttelse av ubalanser i markedet, introduserte de det vi i dag kjenner som pairs trading (Gatev et al., 2006).

Pairs trading er et statistisk arbitrasje-konsept som tar utgangspunkt i to verdipapirer som historisk har fulgt hverandre tett i bestemte tidsrom. Analysen går ut på å følge utviklingen i avviket mellom verdipapirenes relative priser. Handelsmulighetene oppstår i det verdipapirene avviker fra gjennomsnittlig avstand, og en posisjon åpnes. Posisjoner formes ved å selge kort verdipapiret som overpresterer og kjøpe lang verdipapiret som underpresterer. Profitten oppnås når prisene konvergerer tilbake til sitt gjennomsnitt, og vil bli summen av den lange og korte posisjonen. Pairs trading er en markedsnøytral strategi da profitten er uavhengig av markedstrender. Strategien bidrar ikke til verdiskapning, da det alltid vil sitte noen på andre siden av veddemålet som taper det vi tjener og tjener det vi taper. Pairs trading er derfor et nullsumspill. Den bidrar imidlertid til bedre risikoallokering.

En kan ikke garantere for at verdipapirene konvergerer etter en posisjon er åpnet. Derfor kan strategien gi både suksess i perioder hvor verdipapirene konvergerer og store tap i perioder med langvarig divergens. Dette gjør at det ikke er en ren arbitrasje, men det er knyttet en risiko til strategien. Ettersom teknologien har blitt stadig mer utviklet og tilgjengelig, har strategien blitt aktuell og enklere å benytte for flere investorer de siste 10-årene.

1.2 Motivasjon for oppgaven

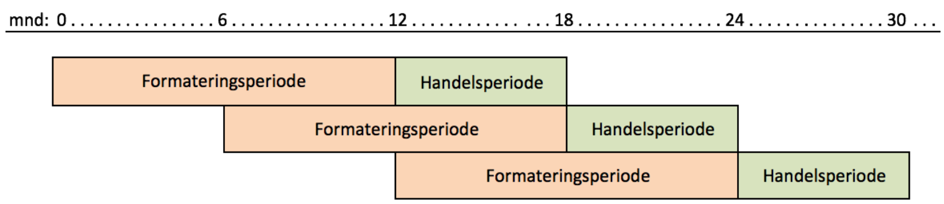
Det har bitt gjort flere studier av pairs trading strategier i aksjemarkedet de siste årene. Vi ønsker å se på muligheten for å utnytte en arbitrasje på andre verdipapirer enn aksjer, og har derfor valgt å se på to europeiske statsobligasjoner. En statsobligasjon er et rentebærende verdipapir utstedt av staten. Når man kjøper en statsobligasjon låner man ut penger til staten, og staten betaler obligasjonskjøperen årlige renter i form av kuponger, tilsvarende dividendeutbetalinger fra aksjeinvesteringer. Når obligasjonens løpetid er over tilbakebetaler staten obligasjonens pålydende verdi. Prisen på obligasjoner bestemmes av nåverdien av fremtidige kontantstrømmer. Nåverdien bestemmes av diskonteringsrenten. Siden tilhørende kontantstrømmer er satt, er det kun renten som kan variere og påvirke prisen på obligasjonen. Dersom renten stiger, vil obligasjonen falle i pris. Dersom renten synker vil obligasjonen stige i pris. Hvis en investor får kjøpt en obligasjon på rabatt vil han få høyere avkastning enn kupongrenten, men hvis han må betale en premie vil avkastningen bli lavere enn kupongrenten. Statsobligasjoner er sett på som likvide verdipapirer, og siden statsobligasjoner utstedes av staten, er disse obligasjonene forbundet med lav risiko og rente. Det er allikevel to faktorer som gjør at vi ønsker å gjennomføre en pairs trading strategi mellom danske og tyske statsobligasjoner. Den første er at finanskrisen og eurokrisen resulterte i uroligheter i flere markeder, blant annet ble statsobligasjonsmarkedet mer volatilt i disse periodene enn ellers. Større varians og usikkerhet i markedet øker sannsynligheten for at verdipapirene beveger seg ut av likevekt og åpner for potensielle fortjenestemuligheter med en pairs trading strategi. For det andre har rentene i Danmark og Tyskland sunket betraktelig de siste 20 årene, fra 6-7% til tidvis å være negative. Pairs trading strategien kan gi fortjeneste også i nedgangstider. Dette fordi vi går kort i det relativt dyre verdipapiret og dersom avviket mellom verdipapirene reduseres til tross for at begge verdipapirene faller i verdi, vil gevinsten av den korte posisjonen være større enn tapet på den lange posisjonen. Vi ser det derfor som sannsynlig at vi kan oppnå positiv kontantstrøm ved statistisk arbitrasje, selv om europeiske statsobligasjoner antas å være i et velfungerende marked og dermed i utgangspunktet skal være riktig priset.

1.3 Struktur

I denne oppgaven vil vi først ha en grundig presentasjon av de mest populære og velprøvde tilnærmingene til pairs trading. Teorien bak tilnærmingene gir en detaljert fremstilling av hvordan strategien bygges opp, og forklarer metoden for hvordan hver enkelt av tilnærmingene gjennomfører handel med strategien. Deretter vil vi presentere datagrunnlaget for vår oppgave. Her trekker vi inn grunnleggende teori fra tilnærmingene for å se om dataene har de nødvendige egenskapene for gjennomførelse av en pairs trading analyse, og begrunner hvorfor vi har valgt den danske og tyske statsobligasjonen. Vi vil gjennomgå stegvis bruk av pairs trading i praksis i analysedelen der vi har en detaljert gjennomgang av hvordan kontantstrømmene beregnes. For å analysere dataene og gjøre beregninger av metodene, har vi benyttet dataverktøyene Stata og Python. Hvilke resultater vi har fått blir deretter presentert, etterfulgt av en konklusjon.

2 Teori

Det finnes flere ulike metoder for utvelgelse av par og gjennomføring av pairs trading. Det som er felles for tilnærmingene er at de benytter seg av en formateringsperiode hvor egnede par velges ut, og rammeverket for strategien settes. Videre følger det en handelsperiode hvor kontantstrømmen av en pairs trading strategi mellom verdipapirene beregnes. Handelsperioden overlapper ikke med formateringsperioden, men starter i det formateringsperioden er over dersom et par blir valgt. I følge normen i akademisk litteratur er handelsperioden halvparten så lang som formateringsperioden (Gatev et al., 2006; Do et al., 2006). Lengden på formateringsperioden spiller en avgjørende rolle i en pairs trading strategi på grunn av sensitiviteten på avkastningene. Formateringsperioder kan vare alt fra flere år helt ned til timer og minutter på høyfrekvens pairs trading.



Figur 2.1: Formaterings- og handelsperioder

Figur 2.1 er et eksempel på formateringsperiode på tolv måneder, etterfulgt av handelsperioder på seks måneder. I illustrasjonen begynner neste formateringsperiode halvveis i forrige formateringsperiode, men dette er ingen regel. Investoren står fritt til å velge selv hvordan formateringsperiodene skal rulleres.

Vi vil videre presentere to av de mest brukte tilnærmingene, samt nevne to mindre kjente tilnærminger for pairs trading.

2.1 Avstandstilnærmingen

Avstandstilnærmingen er den mest kjente og oftest brukte metoden for empirisk arbeid med pairs trading. Tilnærmingen ble introdusert av Gatev et al. (1999, 2006), og er en metode basert på samtaler med investorer som aktivt benytter en pairs trading strategi. Gatev et al. gjennomførte en analyse av pairs trading på det amerikanske aksjemarkedet fra 1962 til 2002. I analysen finner de en signifikant meravkastning for strategien og hevder den skyldes en hittil ukjent systematisk risikofaktor (Gatev et al., 2006).

Avstandstilnærmingen velger ut potensielle par ved å beregne summen av kvadrerte avvik mellom normaliserte priser i formateringsperioden. Normalisering av priser betyr at man tar utgangspunkt i første pris i perioden og ser hvordan prisene utvikler seg i forhold til denne. Et eksempel på en matematisk formel for normalisering av priser er (Engelberg and Jagannathan, 2009; Gatev et al., 2006):

$$P_{At}^{N} = \frac{P_{At}}{P_{A0}} \tag{2.1}$$

Her er P_{At} og P_{A0} henholdsvis prisen på verdipapir A på tidspunkt t og tidspunkt 0. P_{At}^{N} er den normaliserte prisen til verdipapir A på tidspunkt t. P_{At}^{N} er dermed et forholdstall mellom renten på ulike tidspunkt og gir en tidsserie som begynner på 1 på tidspunkt 0. Tilsvarende normalisering benyttes for verdipapir B i paret. Vi følger kontinuerlig utviklingen i avvikene mellom de normaliserte prisene i formateringsperioden.

For å beregne summen av kvadrerte avvik trekker vi den ene normaliserte prisserien fra den andre, og kvadrerer avvikene slik at positive og negative avvik ikke utligner hverandre. Deretter summerer vi avvikene. Målet er å finne par med lavest mulig sum kvadrerte avvik, altså minst differanse mellom de normaliserte prisene. Denne minimumsdifferansen brukes til å rangere parene, der man velger å benytte de som kommer best ut til en pairs trading strategi. Dette vil være parene med høyest samvariasjon.

2.1.1 Handel med avstandstilnærmingen

Når et par er funnet egnet for pairs trading i formateringsperioden, beregnes det en terskelverdi basert på standardavviket til avvikene mellom de normaliserte prisene i formateringsperioden. Et vanlig og ofte brukt mål er to ganger standardavvik. Videre vil neste steg være å beregne de normaliserte prisene for handelsperioden. Normaliseringen i handelsperioden gjøres på tilsvarende måte som i formateringsperioden, ved at normaliseringen starter på nytt med første prisobservasjon i handelsperioden som P_{A0}. Dette gir oss en ny tidsserie som begynner på 1 og deretter er større eller mindre avhengig av om prisen er høyere eller lavere enn på starttidspunktet. Deretter beregnes avviket mellom de normaliserte prisene i handelsperioden. Dersom absoluttverdien av avviket overstiger absolutt terskelverdi beregnet i formateringsperioden, åpnes en posisjon denne dagen. Formelt har vi at en posisjon åpnes dersom:

$$|S_t| > |T| \quad , \quad S_t = P_{At}^N - P_{Bt}^N \tag{2.2}$$

Her er S_t avviket mellom de normaliserte prisene på tidspunkt t i handelsperioden, T er terskelverdien, mens P_{At}^N og P_{Bt}^N er de normaliserte prisene til henholdsvis verdipapir A og B på tidspunkt t. Kurser er presentert med en kjøpspris (bid) og salgspris (ask). For å enklest kunne beregne riktig avvik, er P_{At}^N og P_{Bt}^N snittet av bid og ask prisen. Et verdipapir kjøpes lang til bid-pris, og selges kort til ask-pris. En posisjon åpnes ved at vi investerer et beløp i det verdipapiret som er blitt relativt billig og går kort et tilsvarende beløp i det verdipapiret som har blitt relativt dyrt. Det er under dette punktet det er viktig at vi skiller på bid- og ask-priser. Det verdipapiret som er blitt relativt billig investerer du med bid-prisen, og følger utviklingen av verdipapirets ask-pris. Verdipapiret som er blitt relativt dyrt vil man selge til ask-pris og følge utviklingen til verdipapirets bid-pris. Posisjonen lukkes når de normaliserte prisene krysser hverandre, altså $S_t = 0$. Dersom prisene ikke konvergerer til null innen handelsperioden er over, lukkes posisjonen siste handelsdag. Siden posisjonen lukkes ved neste kryssing vil en høy terskelverdi gi høy potensiell profitt per handel, mens en lavere terskelverdi vil gi flere handler og dermed flere muligheter for profitt. Det er med andre ord vanskelig å forutsi hva som er den mest lønnsomme terskelverdien.

2.2 Kointegrasjonstilnærmingen

Kointegrasjonstilnærmingen ble introdusert av Vidyamurthy i 2004 (Vidyamurthy, 2004). Forfatteren presenterer anvendelsen av kointegrasjon som først ble introdusert av Engle og Granger i 1987 (Engle and Granger, 1987). Her blir konseptet kointegrasjon brukt til å identifisere par på en ny måte, hvor prisdifferansen mellom to verdipapirer antas å være i likevekt over tid, og unormale avvik fra denne gir handelsmuligheter. Deretter presenterer han i dybden hvordan pairs trading kan brukes i aksjemarkedet som statistisk arbitrasje. Boken er en praktisk veiledning til gjennomførelse av pairs trading med kointegrasjonstilnærmingen.

Kointegrasjonstilnærmingen er en statistisk metode som blant annet omhandler stasjonære og ikke-stasjonære prosesser. Vi har følgende forutsetninger for å kunne definere en tidsserie Y som en stasjonær prosess (Harris and Sollis, 2003):

$$E(Y_t) = \text{konstant for alle t} \tag{2.3}$$

$$Var(Y_t) = \text{konstant for alle t} \tag{2.4}$$

$$Cov(Y_t, Y_{t+n}) = \text{konstant for alle t} \tag{2.5}$$

I en pairs trading strategi vil det være relevant å se på om differansen mellom to verdipapirer har egenskapene til en stasjonær prosess. Dersom differansen er en stasjonær prosess, er det en sannsynlighet for at paret er kointegrert og kan brukes i en pairs trading strategi, som bygger på en antagelse om at differansen mellom verdipapirene i paret burde finne tilbake til likevekt.

I diskusjonen om stasjonære tidsserier, må seriene testes for eksistensen av enhetsrøtter for å unngå falsk regresjon. En enhetsrot er en tallstørrelse som angir i hvilken grad en variabel avhenger av tidligere verdier av samme variabel. Dersom en variabel har enhetsrøtter er serien ikke-stasjonær. Faren med at variabler har enhetsrøtter er at det kan bli falske resultater om en meningsfull sammenheng. Eksistensen av enhetsrøtter kan beregnes ved å regressere variabelen man vil teste på forrige periodes verdi av variabelen og deretter teste for enhetsrøtter. Dette vil se slik ut: (Harris and Sollis, 2003).

$$X_t = \rho X_{t-1} + \epsilon_t \tag{2.6}$$

ρ er koeffisienten til forrige periodes verdi av variabelen X. Dersom $\rho = 1$ vil dette tilsi at variabel X har en enhetsrot. Hvis dette stemmer vil man ikke kunne benytte sentralgrenseteoremet. Derfor må man bruke den asymptotiske distribusjonen kalt Dickey-Fuller distribusjon. Vi kan teste om dette er tilfellet ved å anvende en Dickey-Fuller test. For å bruke denne må man differensiere variabel X.

$$X_t - X_{t-1} = \rho X_{t-1} - X_{t-1} + \epsilon_t \tag{2.7}$$

$$\Delta X_t = (\rho - 1)X_{t-1} + \epsilon_t \tag{2.8}$$

$$\Delta X_t = \gamma X_{t-1} + \epsilon_t \tag{2.9}$$

Nullhypotesen i Dickey-Fuller er at en serie har enhetsrøtter, mens alternativhypotesen er at det ikke eksisterer enhetsrøtter og serien er stasjonær. Vi tester dermed nullhypotesen $\gamma = 0$, mot alternativet $\gamma < 0$ som betyr at X_t er en stasjonær prosess. Dickey-Fuller testen har følgende testobservator:

$$DF = \frac{\hat{\gamma}}{se(\hat{\gamma})} \tag{2.10}$$

$\hat{\gamma}$ er den estimerte verdien av γ med standardfeil, $se(\hat{\gamma})$. Dickey-Fuller fordelingen beregnes ved hjelp av Monte Carlo-simuleringer, og Dickey-Fuller finner at de kritiske verdiene varierer med størrelse på utvalget og formen på regresjonen. De skiller på om vi har en regresjon uten konstant og uten trend, med konstant uten trend eller med både konstant og trend. Med en regresjon uten konstant eller trend, er de kritiske verdiene til testobservatoren -1,61, -1,95 og -2,6 for henholdsvis 10%, 5% og 1% signifikansnivå med 100 observasjoner. Jo større utvalg og mer som inkluderes i regresjonen, jo mer negative blir de kritiske verdiene. Nullhypotesen, at enhetsrøtter eksisterer, forkastes når testobservatoren er mer negativ enn kritisk verdi for ønsket signifikansnivå.

Dickey-Fuller testen er ikke kun et mål på om enhetsrøtter eksisterer, men også et mål på hvilken orden en serie er integrert av. Ved å beregne hvilken orden serien er integrert av, finner man ut om variablene er ikke-stasjonære og hvor mange ganger variablene må differensieres for å få en stasjonær serie. Hvis en serie må differensieres d antall ganger før den blir stasjonær, har serien d enhetsrøtter og er integrert av orden d, I(d) (Harris and Sollis, 2003).

Vi har følgende definisjon for at en tidsserie er en I(1)-prosess (Lin et al., 2006): *Definisjon 2.2.1: A time series X_t is called an I(1) series if the first difference of the time series forms a stationary series, denoted by I(0).*

Tidsserier som er integrert av samme orden, kan bli presentert som en VAR-modell (vektor autoregressiv modell). En VAR-modell er en AR-modell på vektor form, et matrisesystem bestående av ulike avhengige og forkarende variabler. For en AR(1)-prosess kan en generell VAR-modell skrives på følgende måte:

$$X_t = A_1 X_{t-1} + \varepsilon_t. \tag{2.11}$$

X_t, X_{t-1} og ε_t er $(n \cdot 1)$-vektorer, A_1 er en $(n \cdot n)$-matrise av parameterne.

Vi vil presentere en VAR-modell med to variabler på følgende matriseform:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{2.12}$$

Som navnet antyder så vil denne metoden identifisere par med bakgrunn i at seriene har en kointegrert sammenheng. Hvis det eksisterer en lineær sammenheng (β) mellom de to tidsseriene, kan vi konkludere at seriene kan være kointegrerte og det er muligheter for at paret kan brukes i en pairs trading strategi.

Kointegrasjon er basert på arbeid gjort av Engle & Granger (1987), som finner at to eller flere tidsserier kan ha felles faktorer som driver deres utvikling. De felles faktorene sikrer at en lineær kombinasjon av de to seriene med en stasjonær likevekt kan oppdages (Lin et al., 2006). Engle-Granger har følgende definisjon på kointegrasjon (Engle and Granger, 1987):

Komponentene av vektoren x_t er kointegrert av orden d,b, definert CI(d,b), med $b > 0$ dersom:
1) alle komponentene av x_t er integrert av orden d, definert I(d) og
2) en vektor $\beta \neq 0$ eksisterer slik at βx_t er integrert av orden d-b, definert I(d-b). Vektoren β er den kointegrerte vektoren.

Hvor d er hvilken orden seriene er integrert av og b er en konstant større enn null, som uttrykker forskjellen i orden av integrasjon mellom x_t-seriene og beta-sammenhengen.

Sammenhengen mellom to serier kan presenteres med følgende regresjon:

$$X_{2,t} = \beta X_{1,t} + u_t \tag{2.13}$$

Seriene er kointegrerte hvis β har en reell verdi forskjellig fra null og forventningsverdien til u_t er null. Kointegrasjonskoeffisienten β er forventet økning i prisen av $X_{2,t}$ når prisen på $X_{1,t}$ øker med 1. Det estimerte restleddet u_t i kointegrasjonslikningen må være en stasjonær prosess, I(0), med forvente verdi lik null. u_t er også et uttrykk for differansen mellom to verdipapirer. Da får vi at (Lin et al., 2006):

$$X_{2,t} - \beta X_{1,t} = u_t \tag{2.14}$$

I likhet med avstandtilnærmingen opererer vi her med bid- og ask-priser, der $X_{1,t}$ er gjennomsnittlig kurspris mellom bid- og ask-verdien til verdipapir 1, og $X_{2,t}$ har også en gjennomsnittlig verdi av bid- og ask-prisen til verdipapir 2.

Vidamurthy (Vidyamurthy, 2004) velger i sin analyse å ta logaritmen til prisseriene og deretter kjøre regresjonen på de logaritmiske prisene. Vidamurthy gir ingen god begrunnelse for hvorfor han gjør dette, men Markus Harlacher (Harlacher, 2016) peker i sin doktorgrad på at en fordel med log-seriene er at de bedre kan fange opp ikke-lineære sammenhenger. Med logaritmiske tidsserier får regresjonen følgende form:

$$lnX_{2,t} - \beta lnX_{1,t} = u_t \tag{2.15}$$

En forutsetning for at to verdipapirer kan være kointegrerte er at de begge er integrert av samme orden og ingen av variablene er I(0). Da kan man fortsette med videre beregninger om i hvilken grad de er kointegrerte. Det finnes flere måter å teste for kointegrasjon. To av de mest kjente metodene er Engle-Granger metode (Engle and Granger, 1987) og Johansen-metoden (Johansen, 1988).

2.2.1 Engle-Granger metode

Engle og Granger foreslo en kointegrasjonstest som består av to hoveddeler. Først estimeres en regresjon med minste kvadraters metode, for å deretter teste residualene for enhetsrøtter (Engle and Granger, 1987). Testen er forholdsvis rett frem hvor meningen er å finne ut om to serier y_t og z_t er kointegrert av orden CI(1,1). Gjennomføringen av Engle-Granger 2-stegs modell for kointegrasjon gjøres på følgende måte (Enders, 2014):

Steg 1: I dette steget estimeres det langsiktige likevektsforholdet mellom seriene. Dersom vi finner en langsiktig likevekt og begge seriene er I(1), vil det si at y_t og z_t er kointegrert av orden CI(1,1).

Det langsiktige likevektsforholdet finnes ved å estimere følgende regresjon med minste kvadraters metode:

$$y_t = \beta_0 + \beta_1 z_t + e_t \tag{2.16}$$

Hvis variablene er kointegrerte vil denne regresjonen med minste kvadraters metode være en konsistent estimator for de kointegrerte parameterene β_0 og β_1. La $\hat{e}_t$ være den estimerte residualen fra denne regresjonen. Deretter utfører man en utvidet Dickey Fuller test på residualen for å sjekke hvilken orden residualen er integrert av. En utvidet Dickey-Fuller test er utviklet for å kunne teste for enhetsrøtter i ARMA(p,q) modeller av ukjent rang. For dette benytter de en estimert fordeling (Eric Zivot, 2006). Siden $\hat{e}_t$ er estimert må vi bruke kritiske verdier som tar hensyn til dette. Disse kalles MacKinnon kristiske verdier (MacKinnon, 1990) og er noe strengere enn de originale Dickey Fuller kritiske verdier, da de er mer negative. MacKinnon kristiske verdier er -3,054 , -3,350 og -3,921 for henholdsvis 10%, 5% og 1% signinfikansnivå. Hvis $\hat{e}_t$ er integrert av orden 0, I(0), det vil si at $\hat{e}_t$ er stasjonær og variablene er kointegrerte, går man videre til steg 2.

Steg 2: Her estimeres feilkorreksjonsmodellen (ECM). Siden variablene er kointegrerte kan vi bruke residualene fra likevektsregresjonen (2.15) til å estimere en feilkorreksjonsmodell. Hvis y_t og z_t seriene er CI(1,1), vil variablene ta en feilkorreksjonsform. En feilkorreksjonsmodell kjennetegnes ved at den inkluderer forrige periodes avvik fra den langsiktige likevekten, residualene fra likevektslikningen i periode t-1, til å forklare den kortsiktige endringen. Vi estimerer dermed en modell på følgende form:

$$\Delta y_t = \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1} \alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i)\Delta y_{t-i} + \epsilon_{yt} \tag{2.17}$$

$$\Delta z_t = \alpha_2 + \alpha_z \hat{e}_{t-1} + \sum_{i=1} \alpha_{21}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i)\Delta z_{t-i} + \epsilon_{zt} \tag{2.18}$$

Siden alle leddene i likning 2.17 og 2.18 er stasjonære, kan vi estimere feilkorreksjonslikningene som en VAR-modell.

Til slutt sjekker vi om modellen er tilstrekkelig. Dette gjøres ved å sjekke om residualen fra feilkorreksjonsmodellen i steg 2 er hvit støy. Dersom residualene er hvit støy, har vi en modell som forklarer sammenhengen tilstrekkelig. Det vil si at vi har ingen signifikant uforklarte restledd og vi kan konkludere med at den langsiktige likevekten kan forklares med kointegrasjonskoeffisientene α.

2.2.2 Johansen-metoden

Johansen-metoden er den mest populære metoden for å estimere feilkorreksjonsmodeller (ECM), og er en mer omfattende og presis metode enn Engle-Granger metoden. Engle-Granger er en to stegs-prosedyre der feil fra det første steget forfølges videre. Påvirkningen på tilpasningshastigheten, α, kan være misvisende fordi estimatene i Engle-Granger er basert på estimerte residualer og ikke det originale støy-restleddet. Johansen-metoden er en forbedret metode og unngår denne to-stegs prosedyren ved at den estimerer hele systemet av likninger simultant som en VAR-modell. Dette fører til at en test for kointegrasjon ikke avhenger av vilkårlige valgte forutsetninger for kointegrasjonsvektoren.

Johansen-prosedyren er en multivarierende generalisering av Dickey-Fuller testen som tidligere introdusert. En enkel generalisering for n variabler er som følger:

$$X_t = A_1 X_{t-1} + \varepsilon_t \tag{2.19}$$

$$\Delta X_t = (A_1 - I) X_{t-1} + \varepsilon_t \tag{2.20}$$

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t \tag{2.21}$$

Der X_t, X_{t-1} og ε_t er $(n \cdot 1)$-vektorer, A_1 er en $(n \cdot n)$-matrise av parameterne, I er en $(n \cdot n)$-identitetsmatrise og $\Pi = A_1 - I$.

Johansen-metoden brukes for å identifisere antall kointegrerte sammenhenger, og se om to tidsserier samvarierer over tid og vender tilbake til likevekt. Vi kan teste for om det eksisterer og hvor mange kointegrerte vektorer det er, ved å se på rangen til Π-matrisen. La oss anta at antallet lineært uavhengig kointegrerte sammenhenger, r, ligger mellom 0 og K-1, hvor K er antallet avhengig variabler i X_t. Med kointegrasjon kan vi skrive Π-matrisen som $\alpha\beta'$ og denne har rang r. En estimering av $\alpha\beta'$ er også et estimat på antall kointegrerte sammenhenger. Johansen and Juselius (1990) foreslo en traseobservator som hjelpemiddel til å finne r.

Testobservatoren er gitt ved følgende formel:

$$\lambda_{trase} = -T \sum_{i=r+1}^{K} ln(1 - \hat{\lambda}_i) \tag{2.22}$$

Hvor T er utvalgsstørrelsen og $\hat{\lambda}_i$ er den i-te egenverdien av $\alpha\beta'$. Nullhypotesen er at det ikke er flere enn r kointegrerte sammenhenger, mot alternativhypotesen at det er flere enn r kointegrerte sammenhenger. Under nullhypotesen er alle λ-ene i traseobservatoren lik 0, så traseobservatoren er også 0 under nullhypotesen. Johansen utviklet en metode for trase-testen hvor man tester for nullhypotesen at r=0 mot alternativhypotesen at r>0. Dersom nullhypotesen forkastes tester vi for r≤1 mot alternativhypotesen at r>1, og videre til vi beholder nullhypotesen. Fra traseobservatorene får vi at antall kointegrerte vektorer er større enn r-1 og mindre eller lik r.

For å med sikkerhet kunne si at antall kointegrerte vektorer er r, brukes en annen metode for å identifisere antall kointegrerte sammenhenger. Denne kalles maksimal egenverditest. Maksimal egenverditesten har følgende testobservator:

$$\lambda_{max} = -Tln(1 - \hat{\lambda}_{r+1}) \tag{2.23}$$

Som for traseobservatoren er her T utvalgsstørrelsen mens $\hat{\lambda}_{r+1}$ er egenverdi r+1. I denne testen gjennomføres det en rimelighetstest av nullhypotesen om det er nøyaktig r kointegrerte sammenhenger mot alternativhypotesen at det er nøyaktig r+1 kointegrerte sammenhenger.

Ved å bruke Johansen-metoden for å finne kointegrasjon mellom serier, bruker man rangen til en matrise til å finne kointegrerte sammenheng eller ei, men også hvor stor grad seriene er kointegrerte. I vår oppgave er vi kun interessert i å finne kointegrasjon uavhengig av i hvilken grad. Derfor aksepterer vi å bruke alle serier som er gir en Π-matrise med rang større enn null.

2.2.3 Handel med kointegrasjonstilnærmingen

Når vi handler med kointegrasjonstilnærmingen benytter vi en test for kointegrasjon for å identifisere mulige fortjenestemuligheter. Kointegrasjon bygger på asymptotisk gyldighet, en uendelig lang tidsserie. Det å kun se på en formateringsperiode på et år vil dermed gi for lite data til å konkludere om en objektiv sannhet, enten vi finner bevis for eller mot kointegrasjon i perioden. Dersom resultatene i metodene som presentert over indikerer kointegrasjon i formateringsperioden, vil paret bli valgt for videre handelsperiode. I formateringsperioden estimeres likevektslikningen i (2,14). Fra denne får vi en estimert likevektskoeffisient β og differansen u_t. Vi bruker standardavviket til u_t for å beregne en terskelverdi q, som er utgangspunkt for åpning av posisjon i handelsperioden. I handelsperioden estimerer vi en ny u_t som er basert på estimert beta fra formateringsperioden. En posisjon åpnes når absoluttverdien til u_t i handelsperioden overstiger absoluttverdien til q. Dersom $u_t < -q$, vil $X_{1,t}$ være overpriset sammenlignet med $X_{2,t}$. Det vil si at dersom differansen mellom verdipapirene faller under den negative terskelverdien, vil man kjøpe lang 1 dollar av $X_{2,t}$ og selge kort β dollar av $X_{1,t}$, hvor betaverdien er hentet fra regresjonen i kointegrasjonstesten. I motsatt tilfelle, hvis $u_t > q$ vil $X_{1,t}$ være underpriset sammenlignet med $X_{2,t}$. Differansen stiger høyere enn den positive terskelverdien, og man vil da selge kort $\frac{1}{\beta}$ dollar av $X_{2,t}$ og kjøpe 1 dollar av $X_{1,t}$ (Rada et al., 2015). Dersom vi får en samvariasjon som gir negativ betaverdi, vil vi enten gå lang eller kort i begge verdipapirene (Harlacher, 2016). Som i avstandstilnærmingen er det også viktig i handel med kointegrasjonstilnærmingen å skille mellom bid- og ask-priser. Man investerer lang i bid-pris og følger utviklingen i ask-pris, samtidig som man går kort i ask-pris og følger utviklingen i bid-pris. Posisjonene lukkes når differansen u_t går tilbake til null, som betyr at paret går tilbake til sin langsiktige likevekt, eller når handelsperioden slutter.

2.3 Andre tilnærminger

I tillegg til avstandstilnærmingen og kointegrasjonstilnærmingen som er de to mest brukte metodene innenfor pair trading, finnes det også andre tilnærminger som er noe mindre kjente. To av disse tilnærmingene er presentert under.

2.3.1 Copula-tilnærmingen

Pairs trading strategier som benytter kointegrasjons- eller avstandstilnærmingen er utbredt i finansbransjen. Det er anerkjent blant både praktikanter og teoretiker at en pairs trading strategi kan gi positiv profitt, men de tradisjonelle strategiene er ikke uten svakheter. Copula-tilnærmingen prøver å løse utfordringen de tradisjonelle strategiene har som følge av at de kun bruker korrelasjon eller kointegrasjon som mål på samvariasjon. Copula er en multivariabel sannsynlighetsfordeling hvor marginalsannsynlighetsfordelingen for hver variabel er uniform. Copulaer er vesentlig mer robuste og realistisk enn korrelasjon da den i større grad forklarer den sanne sammenhengen mellom de to variablene. Dette gjør den dermed mer egnet for å utvikle handelsregler til en pairs trading strategi. Dersom dataene er normalfordelt, vil en enkelt korrelasjonskoeffisient fullt ut forklare avhengigheten, men dette er svært skjelden tilfellet ved finansielle data (Liew and Wu, 2013).

Copulaer er brukt til å beskrive avhengigheten mellom tilfeldige variabler og brukes mye i kvantitativ finans for å modellere og minimere hale-risiko. De betingede sannsynlighetene er helt sentrale for copula strategien da disse bestemmer når det skal åpnes en posisjon. En handel gjennomføres dersom en av de betingede sannsynlighetene er nær 1 (Liew and Wu, 2013). Det er foreløpig svakheter også ved copula-tilnærmingen. Den er foreløpig en relativt ny tilnærming i finansbransjen og det er gjort lite empiriske analyser med copula.

2.3.2 Stokastisk kontrolltilnærming

Robert J. Elliott, John Van Der Hoek og William P. Malcolm (Elliott et al., 2005) var først ute med å bruke Ornsteín Uhlenbeck-prosessen, som er en stokastisk prosess med anvendelse i finansmatematikk og naturvitenskap, til å modellere en tilbakevendende oppførsel på avviket mellom to verdipapirer. Stokastisk kontrolltilnærming har i likhet med copula-tilnærmingen en ikke normalfordelt sannsynlighetsfordeling. Do Binh, Robert Faff og Kais Hamza (Do et al., 2006) generaliserte modellen i en standard verdipapirprisingteori der man kunne kontrollere for faktorvariablene. Verken Elliot et al. eller Do et al. anvendte teorien i praksis, men Baronyan et al brukte teorien på Dow Jones 30 indeks med ukentlig data. I formateringsperioden kjørte han regresjon på prognosen til en uke fremover i tid. Det som ga høyest kumulativ avkastning ble brukt som parametre i handelsperioden, og denne tilnærmingen vil i likhet med avstandstilnærmingen og kointegrasjonstilnærmingen åpne en posisjon på to ganger standardavvik. Parametrene blir tilpasset hver enkelt observasjon. Fordelen med å oppdatere parametrene etter observasjonene er at vi ikke får en konstant gjennomsnittsdifferanse mellom verdipapirene. Med bakgrunn i at gjennomsnittsdifferansen ikke har en konstant verdi over tid indikeres det at stokastisk kontrolltilnærming kan gi et tryggere bevis på optimal handel.

2.4 Valg av tilnærminger

Vi har valgt å ta i bruk avstandstilnærmingen og kointegrasjonstilnærmingen i vår pairs trading analyse. Det finnes mye litteratur rundt begge tilnærmingene, og det eksisterer dermed mange gode praktiske eksempler på bruk av pairs trading med disse tilnærmingene i andre markeder. Copula-tilnærmingen og stokastisk kontrolltilnærming er verdt å nevne i oppgaven da disse også benyttes i litteraturen, dog i et mindre omfang enn avstands- og kointegrasjonstilnærmingen. Copula- og stokastisk kontrolltilnærmingen er lite anvendt og mer kompliserte strategier å implementere. Vi ser det derfor som lite hensiktsmessig å benytte disse i denne oppgaven.

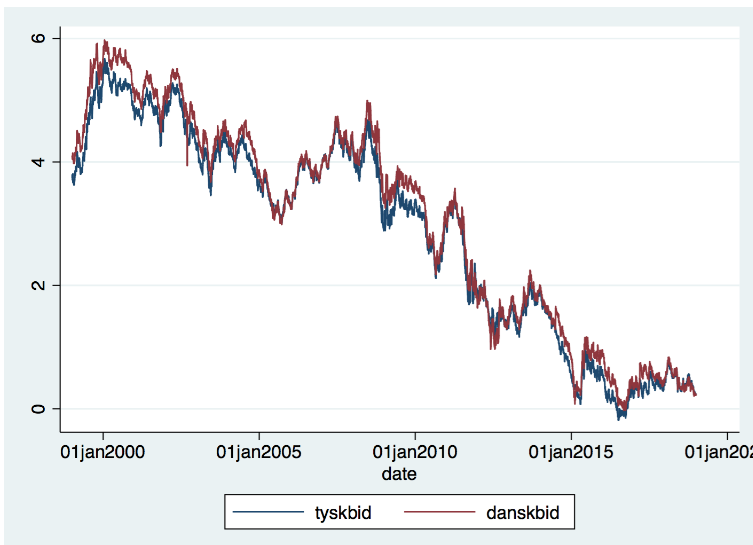
3 Data

3.1 Datautvelgelse

Statsobligasjonsmarkedet er et likvid og velfungerende marked. Det vil si at det er lett å kjøpe og selge obligasjoner når man skulle ønske, og prisen skal til en hver tid være riktig markedspris. Prisen på statsobligasjonen bestemmes blant annet av renten og landets økonomiske og politiske situasjon. I Europa har noen land relativt like forutsetninger. Landene er ansett som stabile og trygge betalere og hovedhandelspartnere er som regel andre europeiske land. Derfor er det flere statsobligasjoner i Europa som følger hverandre tett. Det kreves en historisk sammenheng mellom verdipapirer over en viss periode for å kunne gjennomføre en pairs trading analyse, og vi har antagelser om at statsobligasjoner tilfredsstiller dette kravet.

I vår oppgave ser vi i hovedsak på den danske og tyske statsobligasjonen da de følger hverandre tett over tid. Danmark har ført en fastkurspolitikk siden 1982. Den danske kronen var først bundet mot den tyske D-marken, og deretter mot Euro fra 1999. Vekselsforholdet mellom den danske kronen og euroen er dermed hele tiden det samme. Da det er svært krevende å fastholde kursen helt nøyaktig, så godtas svingninger i kursen med 2,25 % på hver side av den faste kursen. Det er svært sjeldent kursen svinger i så stor grad, vanligvis holder den seg jevnt tett opp mot euroen. Sentralbankens viktigste verktøy for å påvirke valutakursen er styringsrenten. Styringsrenten i to land med bundet valutakurs vil vanligvis være lik, eventuelt med en liten risikopremie til en av partene for å være i likevekt. Etterspørselen etter å plassere penger i de to landene vil dermed være likt. Fastkurspolitikken har sine fordeler ved at eksportnæringen kan være sikre på at de kan veksle betalingen i euro til danske kroner til en fast kurs. Mesteparten av Danmarks eksport går til euroland, og med å sikre kursen reduseres risikoen knyttet til valutasvingninger for danske eksportører og importører.

Analysene i oppgaven vil bli utført for 3- og 10-års statsobligasjoner. For 10-års rentene er analysen over en 20-års periode fra 1999 til 2018, mens 3-års rentene er over 15 år fra 2004 til 2018 grunnet manglende data i perioden før 2004. Ved å ha lange dataperioder får man et mest mulig representativt datagrunnlag. Ved å bruke en kortere dataperiode som for eksempel 5 til 10 år, utelater vi viktige hendelser som finanskrisen (2007-2009) og eurokrisen (2011) fra datasettet. Finanskrisen og eurokrisen skapte store svingninger og usikkerhet i mange markeder. Vi ser av figuren under hvordan den tyske og danske 10-års renten følger hverandre fra 1999 til 2018.



Figur 3.1: Dansk og tysk bid-rente

Figur 3.1 viser utviklingen av den 10-årige statsobligasjonen med daglige observasjoner. Den tyske renten er markert med blå linje og den danske renten med rød linje. Vi ser en svært høy korrelasjon mellom tysk og dansk bid-rente (0.9967). Vi ser også at det er en tendens til at rentene hele veien søker mot hverandre etter perioder med økende avvik i utviklingen. Dette er et godt utgangspunkt for en pairs trading strategi. Vi ser også hvordan rentene stiger markant under kriseårene i 2007-2009 og 2011.

Bakgrunnen for valg av daglige data er at det ble for lange passive perioder med månedlige data. Daglig data kan være noe påvirket av støy, men vi så på det som unaturlig at man venter i månedslange perioder før man foretar seg noe i en ren veddemålstrategi. En investor som benytter seg av en pairs trading strategi vil følge utviklingen på porteføljen hyppig, og vil ikke sitte avventende fra måned til måned dersom veddemålet slår feil midt i perioden. Vi kan se på pairs trading som en strategi for å utnytte støy. Uten denne støyen ville vi kanskje ikke fått de nødvendige svingningen for å benytte handelsstrategien. Som nevnt i teorien bruker vi både bid- og ask-rente for å kunne gi et mest mulig riktig bilde av de faktiske kjøps- og salgsprisene.

For å få et bilde om den danske og tyske statsobligasjonen er et optimalt valg av par for en pairs trading strategi med avstandstilnærmingen, gjør vi ulike sammenligninger med andre parsammensetninger for å se hvordan den danske og tyske statsobligasjonen kommer ut. Vi har innhentet data på daglige observasjoner av avkastninger på norske og franske 10-års statsobligasjoner fra Eikon-databasen. Vi valgte disse landene da det virket rimelig at de kom til å være sterkt korrelerte med vår statsobligasjoner. Alle fire landene er viktige handels- og samarbeidspartnere. Norge skiller seg ut blant disse landene siden de har sin egen valuta og ikke har fast kurs mot euro som den danske kronen. Norge er også en ekstraordinær økonomi med stor formue til tross for å være et lite land med åpen økonomi som i stor grad påvirkes av handelspartnere rundt om i verden. Vi har skaffet oss en oversikt over parsammensetningenes korrelasjon, standardavvik og sum kvadrerte avvik for 10-års rentene. Vi fikk følgende verdier presentert i tabell 3.1:

Par	**Korrelasjon BID**	**Korrelasjon ASK**	**Std.avvik avvik BID**	**Std.avvik avvik ASK**	**Sum kvadrerte avvik BID**	**Sum kvadrerte avvik ask**
Dansk - Tysk	0,992	0,993	0,079	0,087	206,306	239,955
Dansk - Fransk	0,957	0,958	0,408	0,420	485,521	507,821
Fransk - Tysk	0,975	0,976	0,179	0,001	168,489	162,125
Tysk - Norsk	0,963	-	0,116	-	307,253	-

* Daglig data fra 1.januar 2009 til 31.desember 2018

Tabell 3.1: Sammenligninger av obligasjonspar

Som Tabell 3.1 viser, har samtlige par høy korrelasjon, noe som vi også tidligere i oppgaven har antatt. Det betyr at landene er samvarierte og at det er en statistisk sammenheng mellom deres statsobligasjoner. For at sammenligningen skal bli så representativ som mulig er alle beregningene gjort på data fra 1. januar 2009 til 31. desember 2018. For den norske renten har vi kun tilgjengelig data for én pris, ikke en bid- og en ask-pris. Dette var tilfellet både i Eikon databasen og i tidsserien som er gjort tilgjengelig på Norges Banks hjemmeside. Derfor er det ikke ført opp verdier i kolonnene som gjelder beregninger med norske ask-renter.

Beregningen av sum kvadrerte avvikene er mellom normaliserte renter. Blant våre par har den franske og tyske 10 års-renten lavest sum kvadrerte avvik på 168,489 og 162,125 for henholdsvis bid- og ask-rente. Den danske og tyske 10-års renten har nest lavest sum kvadrerte avvik på 206,306 og 239,955. De resterende parene i vår analyse har betydelig høyere verdier. Om vi ser på standardavviket til de kvadrerte avvikene er dette tilnærmet null for det fransk-tyske paret. For det tysk-danske paret er standardavviket noe høyere med 0,087, men fortsatt svært lavt. På bakgrunn av vårt par kommer godt ut i sammenligningene, ser vi det som høyst sannsynlig at danske og tyske 10-års obligasjoner kan benyttes i en lønnsom pairs trading strategi med avstandstilnærmingen. Vi vil i tillegg benytte de danske og tyske 3-års statsobligasjonene i analysen. Dette for å kunne sammenligne effekten av ulik løpetid på obligasjonene. 3-års rentene viser svært høy korrelasjon (0,977), men høyere kvadrerte avvik (863,26) og standardavvik (0,219) enn 10-års renten.

Som innledet i teorien er kointegrasjonstilnærmingen en mye brukt metode for pairs trading. For å se om det er sannsynlig at vi kan bruke en slik tilnærming sjekker vi om de danske og tyske statsobligasjonene gir bevis for kointegrasjon når vi ser på hele dataperioden. Her bruker vi Johansen-metoden fordi denne metoden gir mer robuste resultater enn Engle-Granger. Før vi kan sjekke om parene har kointegrerte vektorer, er det nødvendig å finne ut om de er integrert av samme orden. Dette gjør vi ved hjelp av en Dickey-Fuller test. Dickey-Fuller testen gjøres på hver av dansk og tysk, 3- og 10-års rente.

Dickey-Fuller testen gir oss følgende resultater:

Rente	t-observator	5% kritisk verdi	p-verdi	Nullhypotese om eksistens av enhetsrøtter
Tysk 10 år bid	-0,324	-2,860	0,9221	Behold
Tysk 10 år bid 1. differensiert	-69,164	-2,860	0,0000	Forkast
Tysk 10 år ask	-1,567	-2,880	0,5001	Behold
Tysk 10 år ask 1. differensiert	-17,279	-2,880	0,0000	Forkast
Tysk 3 år bid	-0,731	-2,860	0,8385	Behold
Tysk 3 år bid 1. differensiert	-65,605	2,860	0,0000	Forkast
Tysk 3 år ask	-0,726	-2,860	0,8398	Behold
Tysk 3 år ask 1. differensiert	-64,944	-2,860	0,0000	Forkast
Dansk 10 år bid	-0,411	-2,860	0,9082	Behold
Dansk 10 år bid 1. differensiert	-76,252	-2,860	0,0000	Forkast
Dansk 10 år ask	-0,858	-2,880	0,8014	Behold
Dansk 10 år ask 1. differensiert	-15,783	-2,880	0,0000	Forkast
Dansk 3 år bid	-0,815	-2,860	0,8146	Behold
Dansk 3 år bid 1. differensiert	-69,930	-2,860	0,0000	Forkast
Dansk 3 år ask	-0,986	-2,860	0,7584	Behold
Dansk 3 år ask 1. differensiert	-71,313	-2,860	0,0000	Forkast

Tabell 3.2: Resultater fra Dickey-Fuller test

Dickey-Fuller testen på ikke differensierte variabler gir oss en testobservator som er mindre enn absoluttverdien til 5% kritisk verdi, og befinner seg derfor ikke i forkastningsområdet. Det er 95% sannsynlighet for at nullhypotesen er sann, det finnes minst en enhetsrot, og rentene er minst integrert av orden 1, I(1). Den samme konklusjonen gjelder også for ask-rentene.

En Dickey-Fuller test på de første-differensierte rentene gir oss en testobservator større enn absoluttverdien til 5% kritisk verdi, og befinner seg derfor i forkastningsområdet. Dette vil si at den differensierte tidserien til renten med 95% sannsynlighet ikke har enhetsrøtter og dermed er integrert av orden 0, I(0). Dette er også tilfellet for ask-rentene. Vi kan dermed konkludere med at seriene våre er I(1)-prosesser.

Nå som vi har beviser for at begge rentene er integrert av samme orden, kan vi nå finne ut om de har felles kointegrerte vektorer. Dette gjøres med Johansen-metoden for trase- og maksimeringstest:

```
. vecrank tyskbid danskbid, max

                       Johansen tests for cointegration
Trend: constant                                        Number of obs =    5007
Sample:  3 - 5009                                               Lags =       2
-------------------------------------------------------------------------------
                                                              5%
maximum                                         trace    critical
  rank    parms      LL        eigenvalue   statistic     value
    0       6     18653.447         .         38.3891     15.41
    1       9      18672.59     0.00762        0.1027*     3.76
    2      10     18672.642     0.00002
-------------------------------------------------------------------------------
                                                              5%
maximum                                           max    critical
  rank    parms      LL        eigenvalue   statistic     value
    0       6     18653.447         .         38.2864     14.07
    1       9      18672.59     0.00762        0.1027      3.76
    2      10     18672.642     0.00002
-------------------------------------------------------------------------------
```

Figur 3.2: Johansen-metoden for kointegrerte vektorer mellom dansk og tysk 10-års rente

I trase-testen er nullhypotesen at rangen er null, mens alternativhypotesen er at rangen er større enn null. Rangen her referer til antall lineært uavhengige vektorer i Π-matrisen. Vi beholder nullhypotesen hvis traseobservatoren er mindre enn 5% kritisk verdi. Siden traseobservatoren er større enn 5% kritisk verdi, forkaster vi nullhypotesen og rangen er større enn null. Videre tester vi nullhypotesen om at rangen er mindre eller lik 1 mot alternativhypotesen at rangen er større enn 1. Dette gir oss en traseobservator som er mindre enn 5% kritisk verdi, og vi beholder vår nullhypotese. Rangen er dermed mindre eller lik 1.

For å få bekreftet nøyaktig hvilken rang testen har, ser vi videre på maksimal egenverdi-testen. Her er nullhypotesen at rangen er null og alternativhypotesen at rangen er 1. Vi forkaster nullhypotesen for at rangen er null siden maksimeringsobservatoren er større enn 5% kritisk verdi i nullhypotesen og konkluderer derfor at rang lik 1 og rentene er integrert med én kointegrert vektor. Det samme resultatet gjelder for danske og tyske 3-års renter.

Med bakgrunn i datautvelgelsen vil hovedfokuset i denne oppgaven være på de danske og tyske statsobligasjonene. De danske og tyske statsobligasjonene har som nevnt høy korrelasjon, lav sum kvadrerte avvik og er kointegrerte i følge Johansen-metoden. 10-års renten har betydelig lavere sum kvadrerte avvik enn 3-års renten. Dersom teorien stemmer, at det best egnede paret er det med lavest sum kvadrerte avvik, bør 3-års renten gi betydelig lavere avkastning.

3.2 Databehandling

Dataene på statsobligasjonene viser avkastning til forfall, ikke kurser. Datasettene med kurser har store mangler og vi klarte ikke å finne komplette tidsserier for kursene i noen av databasene vi har tilgjengelig (Eikon (Thomson Reuters), Macrobond, Titlon). I en pairs trading strategi er det utviklingen av kurser man bruker for beregning av kontantstrømmer. Derfor ønsket vi å transformere avkastningen til kurser manuelt. Transformasjonen er gitt:

$$P = \left[1 - \frac{1}{(1+y)^n}\right] \times \frac{C}{y} + \frac{F}{(1+y)^n} \tag{3.1}$$

Der P er prisen på statsobligasjonen, C er kupongutbetaling, y er renteavkastningen ved å holde obligasjonen til forfall og F er pålydende verdi. Det første leddet uttrykker nåverdien av fremtidige kupong utbetalinger. Det siste leddet er nåverdien av å få pålydende verdi utbetalt ved forfall. Vi ser av transformasjonen at en økning i y vil gi en reduksjon i prisen og vice versa.

Transformasjonen var heller ikke mulig å gjennomføre fordi kupongrente for hver obligasjon ikke var tilgjengelig. For de obligasjonene vi fant kupongrente var det heller ikke alltid samsvar mellom den transformerte prisen og enkelte kurspriser som Eikon-databasen (Thomson Reuters) hadde gitt oss. Men siden kursen er en transformasjon av renten, kan vi beregne endringene i kursen på bakgrunn av svingningene i renten. Dette gjør at vi kan benytte renten i analysen. Det er imidlertid viktig å merke seg at vår analyse ikke fanger effekten av endringer i kupongrenten. Endring i kupongrenten kan gi et hopp i positiv eller negativ retning da avkastningen av å holde obligasjonen til forfall vil endres.

4 Analyse

Analysen vår går ut på å se om en tilbaketest av pairs trading på danske og tyske statsobligasjoner med avstands- og kointegrasjonstilnærmingen gir en positiv meravkastning. Vi ønsker også å se hvilken effekt ekstraordinære hendelser i økonomien har på avkastningen til strategien.

Pairs trading er en strategi som bruker differansen mellom to verdipapirer, i vårt tilfelle danske og tyske statsobligasjoner, for å oppnå en fortjeneste på unormale avvik. For å enklere kunne beregne avviket vil vi benytte gjennomsnittet av dansk bid- og ask-rente og tysk bid- og ask-rente, som en approksimasjon for den sanne verdien til renten. Disse beregnes på følgende måte:

$$r^{Dt} = \frac{r^{Dt}_{bid} + r^{Dt}_{ask}}{2} \tag{4.1}$$

$$r^{Tt} = \frac{r^{Tt}_{bid} + r^{Tt}_{ask}}{2} \tag{4.2}$$

Her er r^{Dt}_{bid} dansk bid-rente, r^{Dt}_{ask} dansk ask-rente og r^{Dt} snittrenten. Tilsvarende beregninger er gjort for den tyske renten.

Neste steg i analysedelen er å periodisere rentedataene. De 15 og 20 år lange dataene periodiseres i flere formateringsperioder og handelsperioder. Nicolas Huck publiserte i 2013 en artikkel hvor han ser på hvilken effekt forskjellig lengde på formateringsperioden har på meravkastningen. Han kommer frem til at meravkastningen øker ved lengre formateringsperioder og at det da er sannsynlig at resultatene er økonomisk signifikante (Huck, 2013). På bakgrunn av dette har vi valgt å sette formateringsperioden til ett år, og følgelig handelsperioden til seks måneder. Med dette oppnår vi en kontinuerlig, men ikke overlappende handelsperiode.

I datautvelgelsen har vi sett at danske og tyske statsobligasjoner er egnet for pairs trading med avstandstilnærmingen fordi dette paret er høyt korrelert og har lave sum kvadrerte avvik for de ti siste årene i dataperioden. Vi antar derfor at hver av de enkelte formateringsperiodene tilfredsstiller dette kravet, og gjennomfører dermed en handelsperiode etter samtlige formateringsperioder.

Avstandstilnærmingen involverer som nevnt normaliserte priser. Siden vi benytter oss av renter i vår analyse, vil vi derfor gjennomføre følgende normalisering av snittrentene for videre analyse i denne tilnærmingen:

$$\tilde{r}^{Dt} = r^{Dt} - r^{D0} \tag{4.3}$$

$$\tilde{r}^{Tt} = r^{Tt} - r^{T0} \tag{4.4}$$

$\tilde{r}^{Dt}$ er normalisert dansk snittrente og $\tilde{r}^{Tt}$ er normalisert tysk snittrente på tidspunkt t. Vi har valgt å ikke normalisere rentene på formen til Gatev et al. (Gatev et al., 2006) som er utledet i likning 2.1 i teoridelen, fordi vi bruker renter i analysen og ikke kurser. Det er dermed naturlig å heller bruke differansen i stedet for forholdet i rentene mellom tidspunkt t og tidspunkt 0. Normaliseringen vil da starte på 0 første dag i formateringsperioden og vi følger utviklingen i differansen fra første dag i perioden. Deretter følger samme prosess for den påfølgende handelsperioden.

Differansen mellom rentene beregnes som avstanden mellom normalisert dansk og tysk snittrente:

$$S^{DT} = \tilde{r}^{Dt} - \tilde{r}^{Tt} \tag{4.5}$$

Differansen mellom snitt-rentene, S^{DT}, danner grunnlaget for å beregne en terskelverdi i formateringsperioden. I handelsperioden følger vi S^{DT} for å bestemme når en posisjon åpnes og lukkes.

Vi har følgende tabellarisk oppsett over den normaliserte danske og tyske 10-års renten i avstandstilnærmingen:

DATO	DANSK BID	DANSK ASK	DANSK SNITT	TYSK BID	TYSK ASK	TYSK SNITT	N.DANSK SNITT	N.TYSK SNITT	AVVIK
04-Jan-1999	4,095	4,092	4,094	3,748	3,624	3,686	0,000	0,000	0,000
05-Jan-1999	4,085	4,079	4,082	3,791	3,673	3,732	0,011	-0,046	-0,058
06-Jan-1999	4,057	4,050	4,054	3,768	3,688	3,728	0,040	-0,042	-0,082
07-Jan-1999	4,048	4,036	4,042	3,711	3,725	3,718	0,051	-0,032	-0,083
08-Jan-1999	4,061	4,049	4,055	3,714	3,802	3,758	0,039	-0,072	-0,111
11-Jan-1999	4,143	4,141	4,142	3,724	3,777	3,751	-0,048	-0,065	-0,016
12-Jan-1999	4,121	4,109	4,115	3,805	3,830	3,818	-0,022	-0,132	-0,110
13-Jan-1999	4,097	4,085	4,091	3,672	3,785	3,729	0,002	-0,043	-0,045
21-Des-1999	5,559	5,547	5,553	5,216	5,274	5,245	-1,460	-1,559	-0,099
22-Des-1999	5,592	5,578	5,585	5,257	5,241	5,249	-1,492	-1,563	-0,071
23-Des-1999	5,582	5,570	5,576	5,246	5,210	5,228	-1,483	-1,542	-0,059
24-Des-1999	5,587	5,576	5,582	5,241	5,188	5,215	-1,488	-1,529	-0,040
27-Des-1999	5,556	5,545	5,551	5,271	5,148	5,210	-1,457	-1,524	-0,067
28-Des-1999	5,611	5,600	5,606	5,293	5,140	5,217	-1,512	-1,531	-0,019
29-Des-1999	5,619	5,607	5,613	5,276	5,198	5,237	-1,520	-1,551	-0,032
30-Des-1999	5,656	5,644	5,650	5,330	5,175	5,253	-1,557	-1,567	-0,010
03-Jan-2000	5,772	5,761	5,767	5,468	5,179	5,324	**0,000**	**0,000**	**0,000**
04-Jan-2000	5,781	5,769	5,775	5,505	5,195	5,350	-0,008	-0,026	-0,018
05-Jan-2000	5,821	5,810	5,816	5,487	5,200	5,344	-0,049	-0,020	0,029
06-Jan-2000	5,866	5,854	5,860	5,548	5,221	5,385	-0,093	-0,061	0,032
07-Jan-2000	5,778	5,767	5,773	5,497	5,249	5,373	-0,006	-0,049	-0,043
10-Jan-2000	5,748	5,735	5,742	5,436	5,271	5,354	0,025	-0,030	-0,055
11-Jan-2000	5,853	5,843	5,848	5,473	5,296	5,385	-0,081	-0,061	0,020
12-Jan-2000	5,886	5,875	5,881	5,563	5,301	5,432	-0,114	-0,109	0,005
21-Jun-2000	5,691	5,677	5,684	5,198	5,226	5,212	0,083	0,112	0,029
22-Jun-2000	5,712	5,698	5,705	5,233	5,229	5,231	0,062	0,093	0,031
23-Jun-2000	5,738	5,724	5,731	5,233	5,238	5,236	0,036	0,088	0,052
26-Jun-2000	5,730	5,716	5,723	5,286	5,245	5,266	0,043	0,058	0,015
27-Jun-2000	5,737	5,723	5,730	5,241	5,260	5,251	0,037	0,073	0,037
28-Jun-2000	5,731	5,718	5,725	5,242	5,241	5,242	0,042	0,082	0,040
29-Jun-2000	5,698	5,684	5,691	5,246	5,206	5,226	0,075	0,097	0,022
30-Jun-2000	5,703	5,689	5,696	5,223	5,205	5,214	0,071	0,110	0,039

Tabell 4.1: Normalisering av renter i en formaterings- og handelsperiode

I Tabell 4.1 ser vi hvordan de normaliserte rentene begynner på null 4. januar 1999, og utvikler seg frem til siste dag i formateringsperioden 30. desember 1999. Normaliseringen begynner på nytt i handelsperioden fra 3.januar 2000 og varer til 30. juni 2000.

Ved kointegrasjonstilnærmingen bruker vi lik periodisering som for avstandstilnærmingen, med ett års formateringsperiode og seks måneders handelsperiode. Forskjellen mellom metodene er at kointegrasjonstilnærmingen velger ut et statsobligasjonspar dersom de er kointegrerte i formateringsperioden, i motsetning til avstandstilnærmingen som bruker minste kvadraters metode.

Vi har i datautvelgelsen funnet at datsettet som helhet gir bevis for kointegrasjon med Johansen-metoden mellom de danske og tyske statsobligasjonene. Denne metoden skal som nevnt i teorien gi mer robuste resultater enn for eksempel Engle-Granger metoden. For å sjekke om et statsobligasjonspar er kointegrert i de spesifikke formateringsperiodene har vi igjen brukt Johansen-metoden. Ved bruk av dataverktøyet Python finner vi bevis for kointegrasjon i 17 av 37 formateringsperioder for 10-års renten og i 11 av 29 formateringsperioder for 3-års renten. Siden en handelsperiode følger etter formateringsperioden kun dersom det er funnet bevis for kointegrasjon, har vi like mange handelsperioder.

I motsetning til avstandstilnærmingen som beskriver sammenhengen mellom statsobligasjonene med differansen mellom de normaliserte rentene, beskriver kointegrasjonstilnærmingen sammenhengen mellom statsobligasjonene med en regresjonslikning. Regresjonslikningen med de danske og tyske statsobligasjonene tar følgende form:

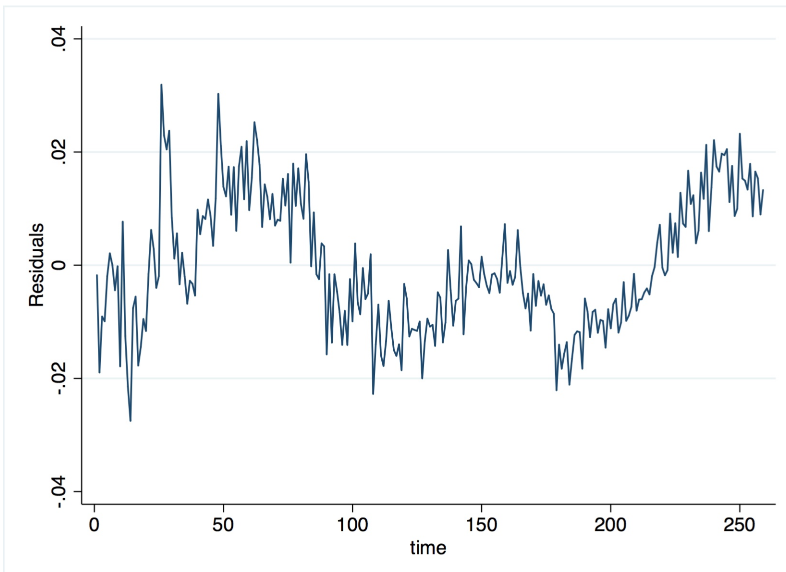
$$r_D = \alpha + \beta r_T + u_t \tag{4.6}$$

Her er r_D og r_T henholdsvis den danske og tyske snitt-renten på brutto form. Vi har valgt å ikke sette regresjonen på logaritmisk form, da vi ser på renter og ikke kurser, og $ln(1 + r) \approx r$. Kointegrasjonskoeffisienten β sier hvor mange prosent dansk rente endres, dersom tysk rente øker med 1%. Alfa er en konstant.

Den estimerte u_t er restleddet i regresjonslikningen, og er et uttrykk for differansen mellom rentene. Vi kan skrive differansen slik:

$$u_t = r_D - \alpha - \beta r_T \tag{4.7}$$

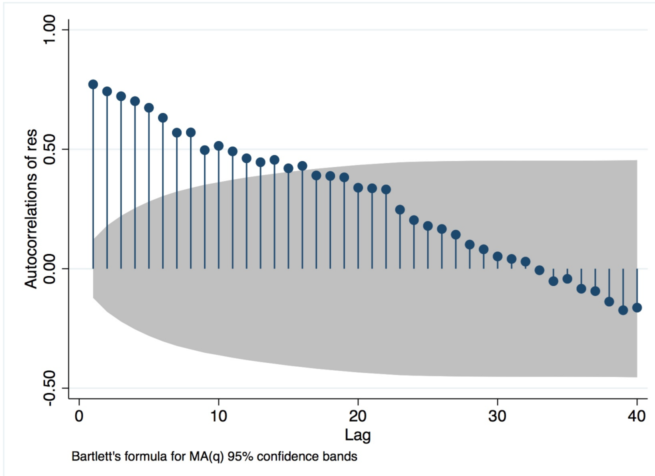
Vi estimerer likning 4.6 og får estimater for α og β. Ved å beregne u_t som vist i likning 4.7 i formateringsperioden fra 1.juli 1999 til 30.juni 2000, får vi følgende graf for u_t over tid:



Figur 4.1: Residual

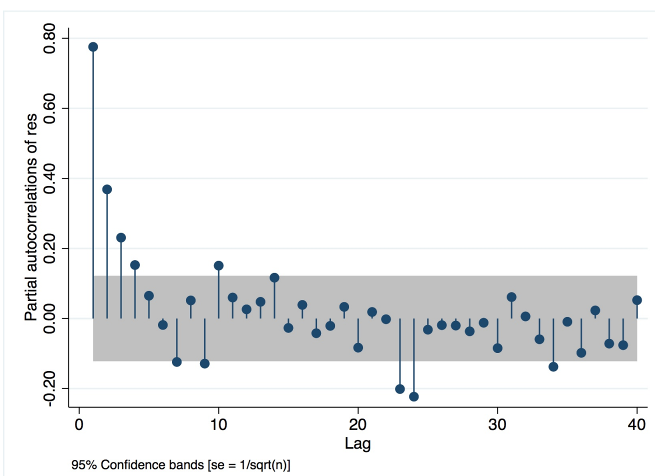
For å estimere u_t i handelsperioden må vi se hva slags prosess som beskriver u_t-tidsserien. For dette benytter vi ACF (autocorrelation function) og PACF (partial autocorrelation function). Autokorrelasjon, også kjent som seriekorrelasjon, er likhetene mellom observasjoner som en funksjon av tiden i mellom.

Vi får med dette følgende fremstilling av seriekorrelasjonen til u_t:



Figur 4.2: ACF

En jevnt avtagende ACF tilsier at prosessen er en AR(p)-prosess. PACF måler den delvise korrelasjonen mellom en stasjonær tidsserie med sine egne tidligere verdier (lags). I motsetning til autokorrelasjonsfunksjonen som ikke kontrollerer for andre tidspunkter, tar PACF høyde for at det kan finnes andre tidligere verdier med numerisk samvariasjon til variabelen. PACF-funksjonen ble introdusert som en del av Box-Jenkins tilnærming til tidsserier og er sentral for å bestemme riktig antall tidspunkter (p) i en AR(p) modell (Box et al., 2008).



Figur 4.3: PACF

Prosessen viser en hoppende PACF med ett markant signifikant hopp i første tidspunkt. På bakgrunn av ACF og PACF kan prosessen til u_t beskrives som en AR(1)-prosess. Dermed skal formen på den estimerte u_t i handelsperioden også være en AR(1)-prosess (Harlacher, 2016).

4.1 Gjennomførelse

Vi tar et utdrag fra den danske og tyske 10-års renten i handelsperiode 15, som strekker seg fra 2. juli 2007 til 28. desember 2007, for å vise hvordan en pairs trading strategi gjøres i praksis. Tabellene 4.2 og 4.3 under viser utdraget av rentenivået og hvordan rentene normaliseres i avstandstilnærmingen.

DATO	DANSK BID RENTE	DANSK ASK RENTE	TYSK BID RENTE	TYSK ASK RENTE
02-juli-2007	4,566	4,559	4,555	4,482
03-juli-2007	4,633	4,626	4,484	4,563
04-juli-2007	4,674	4,667	4,564	4,486
05-juli-2007	4,707	4,701	4,612	4,526
06-juli-2007	4,739	4,732	4,654	4,681
09-juli-2007	4,733	4,725	4,687	4,653
10-juli-2007	4,651	4,643	4,663	4,585

Tabell 4.2: Danske og tyske renter

DATO	DANSK SNITT	TYSK SNITT	N. DANSK SNITT	N. TYSK SNITT	DIFF
02-juli-2007	4,563	4,519	0	0	0
03-juli-2007	4,630	4,524	0,067	0,005	0,062
04-juli-2007	4,671	4,525	0,108	0,007	0,102
05-juli-2007	4,704	4,569	0,142	0,051	0,091
06-juli-2007	4,736	4,668	0,173	0,149	0,024
09-juli-2007	4,729	4,670	0,166	0,152	0,015
10-juli-2007	4,647	4,624	0,085	0,106	-0,021

Standardavvik: 0,03805
Terskelverdi: $\pm$ 0,0761

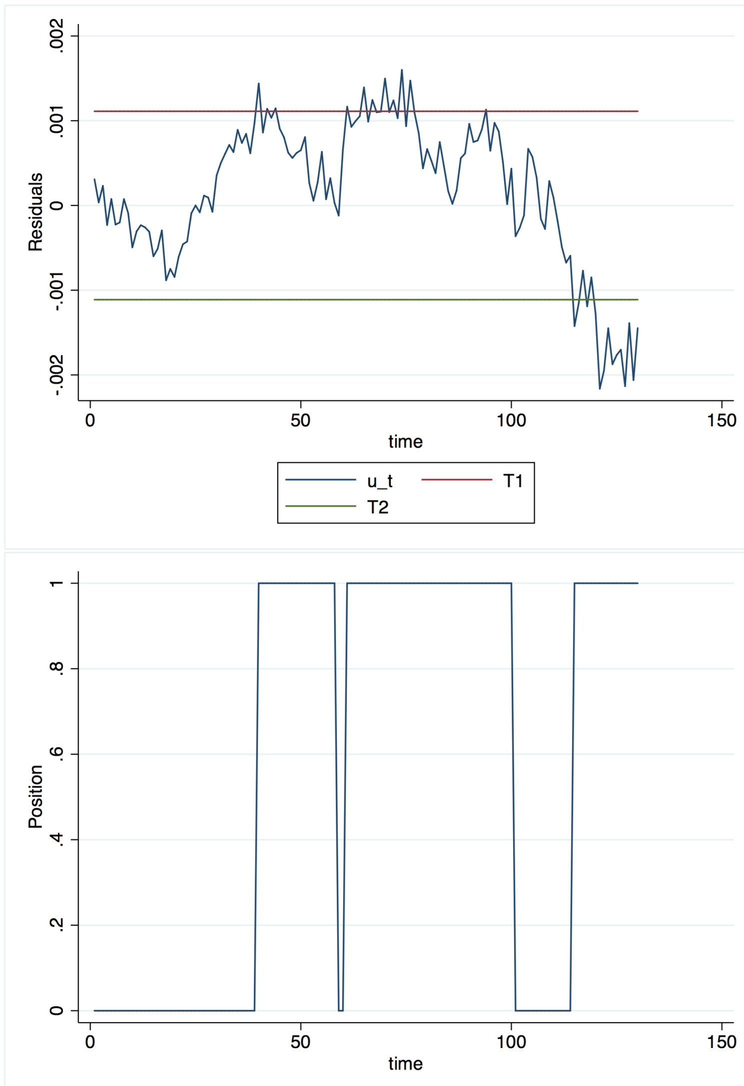
Tabell 4.3: Normaliserte renter og differansen

En posisjon åpnes når differansen mellom de normaliserte snittrentene i handelsperioden er større en absoluttverdien til 2 ganger standardavviket til differansen i formateringsperiode 15, altså at de overstiger terskelverdien på $\pm 0,0761$. Vi velger å benytte 2 ganger standardavvik som terskelverdi da det er normen i tilgjengelig litteratur om pairs trading. Dette er også et enkelt mål å forholde seg til.

I det en posisjon åpnes må det gjøres en vurdering av hvilken av obligasjonene som har blitt relativt billig eller relativt dyr. Den billige statsobligasjonen er den kursen som enten har sunket mest eller steget minst fra den ene dagen til den andre dagen hvor posisjonen åpnes. Den dyre statsobligasjonen er den kursen som har steget mest eller sunket minst i tilsvarende periode. Utviklingen i kurs og renter går motsatt vei, det betyr at en økning i rente gir en nedgang i kursen og motsatt. Derfor vil en billig statsobligasjon være den renten som har steget mest eller sunket minst, og den dyre statsobligasjonen være den renten som har sunket mest eller steget minst. Vi velger så å gå 1 krone kort i den dyre statsobligasjonen og 1 krone lang i den billige statsobligasjonen i håp om at differansen mellom de normaliserte rentene returnerer til null. Den lange posisjonen vil da selges dyrere enn den opprinnelig er kjøpt for og den korte posisjonen kjøpes billigere enn den er solgt for. I perioder kan kursene bevege seg i samme retning, men i ulik grad. Dersom den relativt dyre renten synker raskere eller den relativt billige renten stiger raskere, vil differansen allikevel kunne komme tilbake i likevekt. Vi vil da tape på den ene delen av veddemålet, men tjene tilsvarende mer på den andre, og sitte igjen med en gevinst tilsvarende det opprinnelige avviket fra normalen. Verdien til posisjonen man er lang i vil følge kursutviklingen i ask-prisen over tid da man selger til ask-pris, mens i den korte posisjonen vil man følge utviklingen i bid-prisen over tid da man kjøper tilbake til bid-pris. Når differansen mellom de normaliserte rentene treffer null eller bytter fortegn, lukkes posisjonen. Da vil avkastningen på posisjonen være positiv, og verdien av den tyske og danske statsobligasjonen har utviklet seg i tråd med veddemålet. Hvis posisjonen fortsatt er åpen når handelsperioden er slutt, lukkes også posisjonen. Her avhenger avkastningen av om avviket er større eller mindre enn når posisjonen ble åpnet. For å redusere risikoen for å måtte bære store tap dersom avviket fortsetter å øke etter at en posisjon er åpnet, velger de fleste investorer å benytte en stopp-loss. Vi benytter en stopp-loss på fire ganger standardavvik, altså når avviket har doblet seg fra det tidspunktet vi åpnet posisjonen. En stopp-loss vil si at man setter en grenseverdi for hvor stort man tør at avviket blir før man selger posisjonen. Dersom avviket overstiger denne grenseverdien, lukkes posisjonen. Siden endringen i avviket fra terskelverdien for å åpne til stopp-loss tilsvarer avviket fra utgangsposisjonen på null til åpning vil investoren realisere et tap tilsvarende potensiell fortjeneste (Lin et al., 2006).

For kointegrasjonstilnærmingen følger vi de samme prinsippene som for avstandstilnærmingen når det gjelder å åpne og lukke en posisjon, men utregningen av terskelverdiene og avviket er forskjellig. Siden kointegrasjonstilnærmingen ikke normaliserer rentene, vil vi ikke kunne regne avviket kun som den ene renten trukket fra den andre renten slik som S_t i avstandstilnærmingen. I stedet benytter vi i kointegrasjonstilnærmingen beregninger for avviket u_t som definert i likning 4.7 i teoridelen. Terskelverdien er to ganger standardavviket til u_t fra formateringsperioden. Vi bruker α- og β-verdiene estimert i formateringsperioden til å estimere en ny serie med u_t i handelsperioden. Dersom u_t i handelsperioden overstiger absoluttverdien til to ganger standardavviket beregnet i formateringsperioden åpnes en posisjon. Vi har valgt å holde fast ved en stopp-loss på to ganger terskelverdi, og lukker posisjonen når u_t overstiger stopp-loss, krysser null eller handelsperioden er slutt.

Grafen under viser utviklingen i u_t i handelsperiode 15, der vi ser hvilke tidspunkter u_t krysser terskelverdiene som er utgangspunkt for åpning av posisjon:



Figur 4.4: u_t i handelsperiode 15 med terskelverdier og åpning/lukking av posisjon

I figur 4.4 ser vi hvordan det estimerte avviket u_t utvikler seg over tid. T1 og T2 er henholdsvis den positive og negative terskelverdien. Vi ser at en posisjon åpnes dag 39 i handelsperioden fordi den krysser den positive terskelverdien og varer til u_t krysser null på dag 58. En posisjon åpnes på nytt dag 60 og lukkes på dag 100. En siste posisjon åpnes på dag 114 fordi u_t krysser den negative terskelverdien, og lukkes ved periodens slutt.

Når det gjelder investeringene i kointegrasjonstilnærmingen investerer vi nå 1 krone i den lange og går kort henholdsvis β kroner og $\frac{1}{\beta}$ kroner avhengig av hvilken obligasjon som er relativt dyr eller billig. Betaverdien er den estimerte β fra formateringsperioden. Dersom den danske obligasjonen er relativt billig, går vi 1 krone lang i dansk og β kroner kort i tysk. Dersom det er den tyske obligasjonen som er billig, går vi 1 krone lang i tysk og $\frac{1}{\beta}$ kroner kort i dansk. Når en posisjon lukkes, beregnes kontantstrøm og avkastning med lik fremgangsmåte i begge tilnærmingene. Vi vil gjennomgå denne fremgangsmåten i neste delkapittel.

I tillegg til å se på de to ulike tilnærmingen ønsker vi å se hvilken effekt ekstraordinære hendelser har for resultatet av strategiene vår. Dataperioden inkluderer flere ekstraordinære perioder. Dataperioden dekker blant annet den internasjonale finanskrisen (2007-2009) og eurokrisen (2011). Dette kan ha stor betydning for resultatene våre. Det er ulik praksis blant forskere på om ekstraordinære hendelser skal inkluderes eller ikke når man gjør beregninger basert på historiske data. Barro advarer i artikkelen "Rear disasters and asset markets in the twentieth century" mot å utelate perioder som krig og finanskriser i beregningene. Han argumenterer for at vi må ta høyde for sjeldne begivenheter og at investorer tar med dette i beslutningen om å investere eller ikke investere. De krever høyere kompensasjon for risiko enn hva volatiliteten tilsier på grunn av faren for katastrofer (Barro, 2006). Vi velger derfor å se på forskjellene i utfallene ved å inkludere og ekskludere disse periodene.

4.2 Avkastning

Generelt når det utarbeides en avkastning av en pairs trading strategi kan man beregne kontantstrømmen ved å se på endringene i kursen gjennom holdingperioden. I og med at vi har renter i vår data, må vi finne et uttrykk for endringen i kursen da det ikke blir riktig å kun se på endringen i renten gjennom holdingperioden. For å få et uttrykk for hvordan kursen endres som følge av en endring i renten, deriverer vi transformasjonen i likning (3.1) med hensyn på renten på følgende måte:

$$\frac{\partial P}{\partial y} = -[1-(1+y)^{-n}]\frac{C}{y^2} + n(1+y)^{-n-1}\frac{C}{y} - n(1+y)^{-n-1}F \tag{4.8}$$

Videre antar vi at derivasjonen er evaluert i et punkt der $C = Py$ hvor kupongrenten c er lik avkastning til forfall. Da har vi at:

$$n(1+y)^{-n}F = P - [1-(1+y)^{-n}]\frac{C}{y} = [1-1+(1+y)^{-n}]P = (1+y)^{-n}P \tag{4.9}$$

Setter inn for dette i derivasjonen og oppnår:

$$\frac{\partial P}{\partial y} = -[1-(1+y)^{-n}]\frac{P}{y} \tag{4.10}$$

Vi sitter nå igjen med en separabel differensiallikning, hvor vi kan finne utviklingen ved å integrere mellom avkastning til forfall på kjøps- og salgstidspunktet på begge sider. Vi får da følgende uttrykk:

Venstre side:

$$\int_{p_{start}}^{p_{slutt}} \frac{1}{P}\partial P = ln(p_{slutt}) - ln(p_{start}) = ln\left(\frac{p_{slutt}}{p_{start}}\right) \tag{4.11}$$

Høyre side for 3-års statsobligasjon, n = 3:

$$\int_{y_{start}}^{y_{slutt}} -[1-(1+y)^{-n}]\frac{1}{y}\partial y = \Big[\frac{2y+3}{2(y+1)^2} - log(y+1)\Big]_{y_{start}}^{y_{slutt}} = V \tag{4.12}$$

Høyre side for 10-års statsobligasjon, n = 10:

$$\int_{y_{start}}^{y_{slutt}} -[1-(1+y)^{-n}]\frac{1}{y}\partial y = \Big[\frac{1}{2(y+1)} + \frac{1}{2(y+1)^2} + \frac{1}{3(y+1)^3} + \frac{1}{4(y+1)^4} + \frac{1}{5(y+1)^5} + \frac{1}{6(y+1)^6} + \frac{1}{7(y+1)^7} + \frac{1}{8(y+1)^8} + \frac{1}{9(y+1)^9} - log(y+1)\Big]_{y_{start}}^{y_{slutt}} = V \tag{4.13}$$

Vi setter sammen venstre side og høyre side og får følgende:

$$ln\left(\frac{p_{slutt}}{p_{start}}\right) = V \tag{4.14}$$

$$\left(\frac{p_{slutt}}{p_{start}}\right) = e^V \tag{4.15}$$

Dette gir da bruttoavkastningen for holdingperioden, mens den nominelle avkastningen vil da være $e^V - 1$. Vi kan finne kontantstrømmen fra investeringen ved hjelp av følgende formel:

$$kontantstrøm = xe^V - x \tag{4.16}$$

hvor x er beløpet investert i obligasjonen. Kontantstrømmen til posisjonen finnes ved å bruke renten på lukkedagen som sluttrente og renten på startdagen som startrente i utregningen av brutto avkastning. Hvis man er kort i en posisjon, er y_{start} ask-verdien på åpningstidspunktet og y_{slutt}

bid-verdien på tidspunktet vi lukker. Vice versa for lang-posisjonen. Total kontantstrøm er kontantstrøm oppnådd i dansk posisjon pluss kontantstrøm oppnådd i tysk posisjon. Her tar vi ikke hensyn til eventuelle kupongutbetalinger i holdingperioden. Vi ser det som størst sannsynlig at en investor som benytter en pairs trading strategi spekulerer kun i kursutvikling og ikke kupongutbetalinger.

I vårt eksempel fra avstandstilnærmingen åpnes en posisjon 4. juli 2007. Som vi ser av Tabell 4.3 er differansen større enn terskelverdien denne dagen. I dette tilfellet vil vi gå lang i den danske obligasjonen da den danske renten har steget mer relativt til den tyske fra dagen før, og dermed falt mest i pris. Vi går derfor kort i den tyske obligasjonen. Videre følger vi utviklingen i disse investeringene i dansk ask-pris og tysk bid-pris frem til 10. juli 2007 der den gjennomsnittlige differansen krysser null. Metoden for kontantstrømberegning er vist under i Tabell 4.4. Verdien av posisjonen presenteres for hver dag fra en posisjon er åpen til den lukkes. I det posisjonen lukkes beregnes hvilken kontantstrøm som er oppnådd.

DATO	POSISJON	POSISJON DANSK	POSISJON TYSK	K.STRØM DANSK	K.STRØM TYSK	K.STRØM PAR
02-juli-2007	0	0,0000	0,0000	0,0000	0,0000	0,0000
03-juli-2007	0	0,0000	0,0000	0,0000	0,0000	0,0000
04-juli-2007	1	1,0000	-1,0000	-1,0000	1,0000	0,0000
05-juli-2007	1	0,9979	-0,9901	0,0000	0,0000	0,0000
06-juli-2007	1	0,9955	-0,9868	0,0000	0,0000	0,0000
09-juli-2007	1	0,9960	-0,9843	0,0000	0,0000	0,0000
10-juli-2007	0	1,0024	-0,9861	1,0024	-0,9861	0,0163

Tabell 4.4: Kontantstrøm

Fra Tabell 4.2 har vi at den danske startrenten er 4,674% og den tyske startrenten er på 4,486%. Verdien av posisjonen hver dag fra 4. juli er beregnet ved hjelp av avkastningsformelen for 10-års statsobligasjoner. Posisjonen lukkes 10. juli 2007 da differansen denne dagen krysser null, altså går fra å være positiv til negativ. Den danske og tyske sluttrenten er på dette tidspunktet henholdsvis 4,643% og 4,663%. Disse start- og sluttrentene gir oss en verdi på 1,0024 på dansk posisjon og -0,9861 på tysk posisjon. Den danske renten man har betalt 1 krone for vil på lukketidspunktet selges for 1,0024 kroner. Den tyske renten man har solgt for 1 krone kan nå kjøpes for 0,9861 kroner. Med dette har vi i denne posisjonen en kontantstrøm på 0,0163.

4.3 Risiko

Alle investeringsstrategier inneholder risiko, også pairs trading, selv om den søker å utnytte en statistisk arbitrasjemulighet. Ekte arbitrasje er per definisjon risikofritt (IG Group, 2017). For at det skal være attraktivt for investorene å investere i en pairs trading strategi må vi kunne vise til en positiv risikojustert meravkastning.

Det kan benyttes flere forskjellige metoder for å beregne risikoen ved en pairs trading strategi. Man kan blant annet benytte standardavvik, VaR, forventet nedside og forventet oppside for å få et best mulig bilde av risikoen på pairs trading strategien.

Standardavvik er ofte brukt som et mål på risiko, der det beskriver volatiliteten, forventet avstand fra gjennomsnittlig avkastning. Som et mål på risikoen til pairs trading strategien ser vi på standardavviket til den oppnådde kontantstrømmen. Den forventede kontantstrømmen $\bar{K}$ er gjennomsnittet av oppnådde kontantstrømmer over en periode (Jonathan Berk, 2014):

$$\bar{K} = \frac{1}{T}(K_1 + K_2 + ... + K_T) = \frac{1}{K}\sum_{t=1}^{T} K_t \tag{4.17}$$

Standardavviket for kontantstrømmene beregnes på følgende måte:

$$\sigma_K = \sqrt{\frac{1}{T-1}\sum_{i=t}^{T}(K_t - \bar{K})^2} \tag{4.18}$$

Med grunnlag i historiske data kan man si noe om forventet kontantstrøm av å inngå en pairs trading posisjon, $\bar{K}$, og at vi forventer svingninger innenfor standardavviket σ_K. Standardavviket måler hvor mye variasjon det er i kontantstrømmen fra periode til periode.

Et VaR-mål (Value at risk metric) har til hensikt å gi et mål på hvor mye penger en portefølje eller posisjon kan tape med en gitt sannsynlighet over en gitt tidsperiode (Holton, 2014). For å kunne beregne et VaR-mål, må vi fastsette tre ting:

1. Over hvor lang tid kan det potensielle tapet skje. Dette kalles VaR-horisonten. Dersom denne settes til én dag, vil det si at porteføljen ikke taper mer enn VaR-målet med en gitt sannsynlighet i løpet av én dag.

2. For hvilken sannsynlighet vårt VaR-mål er største mulige tap. Dette kan for eksempel være 90 eller 95%.

3. I hvilken valuta det potensielle tapet er uttrykt, kalt grunnvaluta.

Forventet nedside (Expected Shortfall (ES)) er definert som gjennomsnittet av alle tapene som er større enn VaR. Det vil si at forventet tap er et tall på gjennomsnittlig tap i de verste (1-p) % av tilfellene, hvor p er signifikansnivå for VaR. Dersom vi har benyttet et VaR-mål på 95%, vil forventet nedside vise gjennomsnittlig tap i de 5% verste tilfellene (Statpro, 2013).

Forventet oppside er det motsatte av forventet nedside. Der forventet nedside gir oss de verste tilfellene, gir forventet oppside gjennomsnittet av de beste tilfellene. Et mulig mål er gjennomsnittet ved de beste 5% av tilfellene, i disse tilfellene får man uventet god avkastning.

Som mål på risikojustert avkastning har vi valgt å benytte et forholdstall. Forholdstallet er gitt med følgende uttrykk:

$$\text{Forholdstall} = \frac{K}{\sigma_K} \tag{4.19}$$

Hvor K er kontantstrømmen i hele perioden og σ_K er standardavviket til kontantstrømmen. Forholdstallet gjør det enkelt å sammenlikne resultatet av strategien på ulike datasett og gir et mål på hvor høy kontantstrømmen er relativt til risikoen.

I et modifisert forholdstall benytter vi forventet nedside (Expected shortfall) i nevneren i stedet for standardavvik. Forventet nedside blir da beregnet på bakgrunn av kontantstrømmen i de 5% verste utfallene til et par.

5 Resultater

Vi har gjennomført en analyse i samsvar med avstands- og kointegrasjonstilnærmingen. For å kunne beregne resultatet av tilnærmingene, har vi implementert strategien i Python. Ved å ta utgangspunkt i lang-posisjoner på 1 krone og kort-posisjoner avhengig av tilnærming, har vi kommet frem til følgende resultater for de danske og tyske 3- og 10-årige statsobligasjonene.

5.1 Danske og tyske 10-års statsobligasjoner

Par	Total kontantstrøm	Forholdstall	Standardavvik	Antall handler	Forventet oppside*	Forventet nedside**
Avstandstilnærmingen						
Dansk - Tysk 10år	0,2952	19,8121	0,0149	46	0,0546	-0,0129
Dansk - Tysk 10år, uten finanskrise	0,2138	11,0206	0,0194	23	0,0617	-0,0164
Dansk - Tysk 10år, uten finanskrise og eurokrise	0,1922	9,7563	0,0197	22	0,0617	-0,0164
Kointegrasjonstilnærmingen						
Dansk - Tysk 10år	0,3342	28,3220	0,0118	55	0,0401	-0,0161
Dansk - Tysk 10år, uten finanskrise	0,3511	29,5042	0,0119	40	0,0456	-0,0177
Dansk - Tysk 10år, uten finanskrise og eurokrise	0,3511	29,5042	0,0119	40	0,0456	-0,0177

* Forventet gevinst i de 5% beste utfallene
** Forventet tap i de 5% dårligste utfallene

Tabell 5.1: Resultater av ordinær pairs trading strategi

5.1.1 Avstandstilnærmingen

Vi ser av Tabell 5.1 at 10-års renten gir 46 handler og en positiv total kontantstrøm på 0,2952 ved bruk av avstandstilnærmingen for hele datasettet. Den totale kontantstrømmen er beregnet som summen av kontantstrømmene til alle veddemålene. Forholdstallet mellom kontantstrøm og risiko er høyt for hele datasettet. Når krisene ekskluderes fra datasettet reduseres forholdstallet med 44,4%. Dette kommer av en kombinasjon av at kontantstrømmene reduseres med 27,6% fra 0,2952 til 0,2138, og høyere standardavvik. Ved å ekskludere finanskrisen fra datasettet reduseres antall handler med 50%. Dersom eurokrisen i tillegg fjernes fra datasettet reduseres antall handler til 22, og total kontantstrøm faller ytterligere til 0,1922. Dette kan tyde på at avstandstilnærmingen klarer å utnytte de økte svingningene under finanskrisen med flere lønnsomme veddemål. Det er med andre ord flere tidspunkter under krisene at obligasjonsprisene avviker unormalt mye fra hverandre.

5.1.2 Kointegrasjonstilnærmingen

Vi ser av samme Tabell 5.1 at kointegrasjonstilnærmingen gir en positiv total kontantstrøm på 0,3342 på hele datasettet, og et vesentlig høyere forholdstall mellom kontantstrøm og risiko enn ved avstandstilnærmingen. Kointegrasjonstilnærmingen forsterker resultatet fra avstandstilnærmingen om en positiv kontantstrøm. Ved ekskludering av finanskrisen fra datasettet fjernes 15 handler, mens total kontantstrøm øker med 5,05%. Dette kan tyde på at kointegrasjonstilnærmingen har færre lønnsomme veddemål ved økte svingninger, og gir større tap enn gevinst for periodene med kriser. Det er ingen forandring på resultatene i denne tilnærmingen ved å ekskludere begge krisene i forhold til å kun ekskludere finanskrisen. Vi har ikke funnet bevis for kointegrasjon i formateringsperioden fra eurokrisen og dermed ingen handler i den påfølgende perioden.

5.2 Danske og tyske 3-års statsobligasjoner

Par	Total kontantstrøm	Forholdstall	Standardavvik	Antall handler	Forventet oppside*	Forventet nedside**
Avstandstilnærmingen						
Dansk - tysk 3år	0,0837	12,4925	0,0067	36	0,0223	-0,0048
Dansk - tysk 3år, uten finanskrise	0,0537	9,5893	0,0056	12	0,0119	-0,0048
Dansk - tysk 3år, uten finanskrise og eurokrise	0,0537	9,5893	0,0056	12	0,0119	-0,0048
Kointegrasjonstilnærmingen						
Dansk - tysk 3år	0,1528	8,4420	0,0181	28	0,0939	-0,0115
Dansk - tysk 3år, uten finanskrise	0,0468	10,8837	0,0043	24	0,0136	-0,0025
Dansk - tysk 3år, uten finanskrise og eurokrise	0,0468	10,8837	0,0043	24	0,0136	-0,0025

* Forventet gevinst i de 5% beste utfallene

** Forventet tap i de 5% dårligste utfallene

Tabell 5.2: Resultater av ordinær pairs trading strategi

5.2.1 Avstandstilnærmingen

Vi ser av Tabell 5.2 at pairs trading strategien mellom danske og tyske 3-års obligasjoner gir en positiv total kontantstrøm på 0,0837 for avstandstilnærmingen. Forholdstallet mellom kontantstrøm og risiko er på 12,4925 når hele datasettet inkluderes. Forholdstallet reduseres når krisene ekskluderes fra datasettet, hovedsaklig fordi total kontantstrøm reduseres med 35,8% til 0,0537. Ved fjerning av finanskrisen reduseres handlene til en tredjedel av handlene med hele datasettet. Dette tyder igjen på at åpninger av posisjoner trigges ved uroligheter i markedet, og at veddemålet vårt har slått positivt ut når markedet svinger. Siden det ikke eksisterer handler i årene under eurokrisen, vil det heller ikke gi endringer i resultatene dersom eurokrisen ekskluderes.

5.2.2 Kointegrasjonstilnærmingen

Kointegrasjonstilnærmingen gir en total kontantstrøm på 0,1528 med hele datasettet som grunnlag. Selv om kointegrasjonstilnærmingen gir 82,5% høyere avkastning enn avstandstilnærmingen er forholdstallet mellom kontantstrøm og risiko lavere. Dette skyldes vesentlig mer svingninger i kontantstrømmene fra kointegrasjonstilnærmingen enn avstandstilnærmingen. Ved ekskludering av finanskrisen reduseres total kontantstrøm med 69,4% fra 0,1528 til 0,0468, det vil si at handler i denne perioden stod for brorparten av den positive kontantstrømmen i 3-års obligasjonene. Antall handler reduseres fra 28 til 24, og forholdstallet øker når vi ekskluderer krisene fra datasettet. Dette skyldes i hovedsak et svært lavt standardavvik. I likhet med 10-års renten påvirker ikke ekskludering av eurokrisen resultatet av kointegrasjonstilnærmingen.

5.3 Svakheter ved resultatene

Resultatene er basert på vurderinger vi har tatt underveis i arbeidet, og det er derfor rom for diskusjoner. Det er blant annet relativt få veddemål tatt i betraktning at vi har handlet i 20 år. Dette kan tyde på at markedet for tyske og danske statsobligasjoner er velfungerende, og det er vanskelig å utnytte noen arbitrasje. I 3-års renten er det definitivt færrest handler og lavere total kontantstrømmer. En del kan forklares med 3-års renten har noe kortere datasett. Det hadde derfor vært bedre å sammenlikne 3- og 10-års renten dersom vi hadde hatt datasett fra 1999 for 3-års renten for å få mer robuste resultater. Det kunne også vært spennende å se om endring i lengden av formaterings- og handelsperioder hadde gitt flere eller færre handler, og hvordan det hadde påvirket resultatet. Vi tar heller ikke hensyn til at en investor i virkeligheten må stille en garanti for deler av den korte posisjonen. Dersom tapet på den korte posisjonen er større enn gevinsten i den lange, sitter investor igjen med en gjeld.

Vi har valgt ut par på bakgrunn av at dataene i hele perioden hadde de egenskapene vi var på utkikk etter. For å få et mer realistisk bilde kunne man basert valg av par på egenskaper i data tilgjengelig før første handelsperiode. Siden vi gjør en tilbaketest av pairs trading har vi også hatt mulighet til å se på historiske egenskaper med dataene som en investor i sanntid ikke har hatt mulighet til. Selv om vi har prøvd å begrense datasnoking, har vi fortsatt valgt par etter å ha testet for om det har de prefererte egenskapene. En tilbaketest av spesifikt utvalgte data kan alltid vinkles og endres til å gi ønskede resultater.

6 Konklusjon

For å oppsummere gir pairs trading strategien en positiv kontantstrøm for 3- og 10-års statsobligasjoner for begge tilnærmingene, og dermed en positiv avkastning på statistisk arbitrasje. Siden vi finansierer lang-posisjonen med å selge en tilsvarende kort-posisjon, investerer vi ingen egne penger. Derfor kan vi ikke si noe om prosentvis avkastning på strategien, men resultatene gir oss summen av kontantstrømmene fra alle posisjonene.

3-års obligasjonen gir oss positive kontantstrømmer på 0,0837 og 0,1528 for avstands- og kointegrasjonstilnærmingen når vi inkluderer hele datasettet. 10-års obligasjonen gir generelt en høyere total kontantstrøm enn 3-års obligasjonen med kontantstrømmer på henholdsvis 0,2952 og 0,3341. En positiv faktor med alle resultatene er at forventet oppside alltid har en høyere positiv verdi enn forventet nedside har negativ verdi.

Det er flere handler og høyere total kontantstrøm i 10-års renten enn 3-års renten. Dette kan skyldes at effekten av endringen i renten på kursen er mye større for 10-års obligasjonen, jamfør likning 4.15. Vi har også kortere dataperiode for 3-års obligasjonene. Vi ser at antagelsen i datautvelgelsen stemmer, at paret med lavest sum kvadrerte avvik er best egnet til pairs trading. Generelt for 10-års obligasjonene avhenger effekten av ekstraordinære svingninger i markedet av hvilken tilnærming man har brukt i pairs trading strategien.

Vår analyse viser et positivt resultat ved å benytte en pairs trading strategi mellom danske og tyske statsobligasjoner. Vi ser viktigheten ved å ta med det fulle datasettet som inkluderer finanskrisen og eurokrisen, da det har stor innvirkning på resultatene. I både avstands- og kointegrasjonstilnærmingen er det betydelig flere handler når krisene inkluderes, noe som gir et tegn på at det er under urolige tider at markedet svinger og en åpning av posisjon trigges. Som videre arbeid vil det være spennende å se hvilken effekt det har å endre lengden på formaterings- og handelsperiodene. Det kunne blant annet vært aktuelt å benytte lenger formateringsperioder enn ett år for 10-års obligasjonene som er lange verdipapirer. Det vil også være interessant å gjøre pairs trading strategier på andre par med statsobligasjoner for å ha flere resultater å sammenligne med.

Bibliografi

Barro, R. J. (2006). Rear disasters and asset marrkets in the twentieth century. *The Quarter Journal of Economics 121*, 20(3):823–866.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis*. Wiley.

Do, B., Faff, R., and Hamza, K. (2006). A new approach to modeling and estimation forpairs trading.

Elliott, R. J., *, J. V. D. H., and Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5(3):271–276.

Enders, W. (2014). *Applied Econometric Time Series*. John Wiley Sons Inc.

Engelberg and Jagannathan, G. . (2009). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, pages 2–4.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Economatrica*, 55(2):251–276.

Eric Zivot, J. W. (2006). *Modelling Financial Time Series with S-PLUS*. ....

Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3):798–805.

Harlacher, M. (2016). Cointegration based algorithmic pairs trading. pages 77–105.

Harris, R. and Sollis, R. (2003). *Applied Time Series Modelling and Forecasting*. John Wiley & Sons Ltd.

Holton, G. A. (2014). *Value-at-Risk: Theory andPractice*. Published by the author at `www.value-at-risk.net`.

Huck, N. (2013). The high sensitivity of pairs trading returns. *Applied Economics Letters, 2013*, 20(14):1301–1304.

IG Group (2017). Arbitrasje - definisjon. https://www.ig.com/no/trading-ordliste/arbitrasje-definisjon#information-banner-dismiss. Accessed: 2019-04-05.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231 – 254.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration — with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52:231 – 254.

Jonathan Berk, P. D. (2014). *Corporate Finance*. Pearson Education Limited.

Liew, R. Q. and Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds*, 19(1):12–30.

Lin, Y.-X., McCrae, M., and Gulati, C. (2006). Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences*, 2006:1–14.

MacKinnon, J. G. (1990). Critical values for cointegration tests. *Long-Run Economic Relationships*, 2.

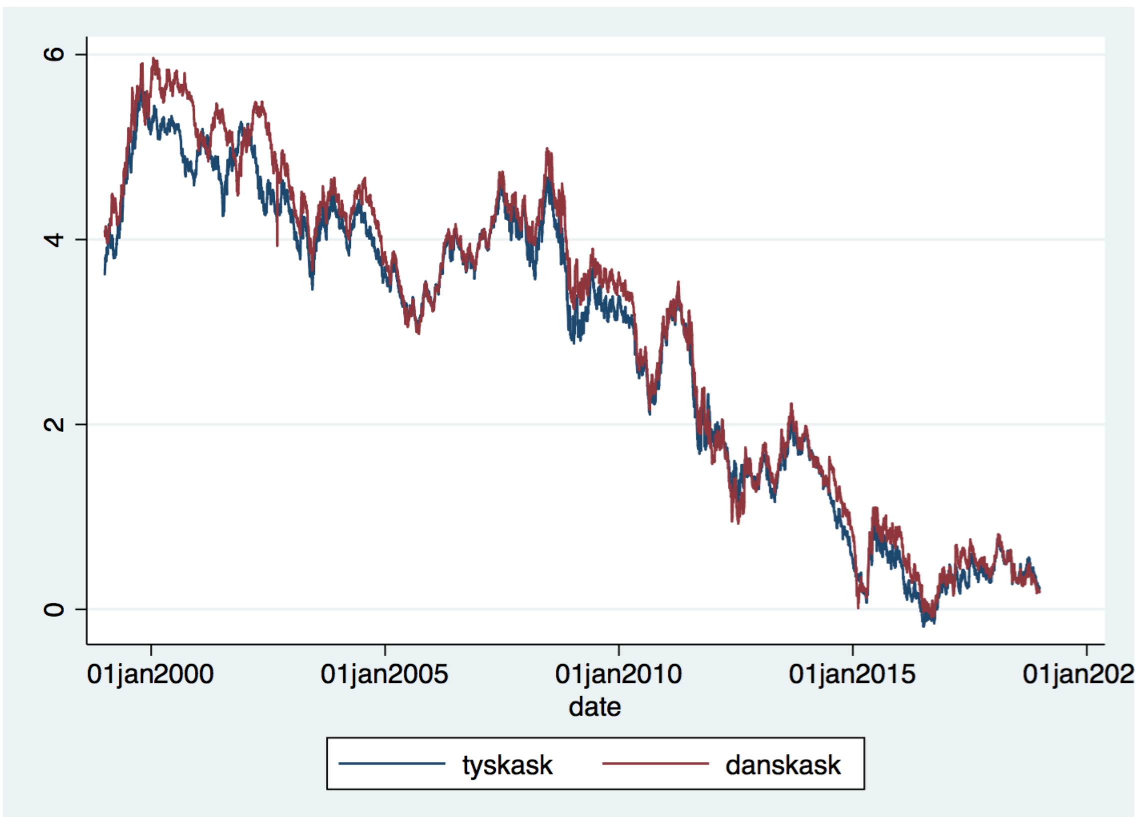
Rada, H., Lowa, R. K. Y., and Faff, R. (2015). The profitability of pairs trading strategies: distance, cointegration, and copula methods. *Quantitative Finance*, pages 1–35.

Statpro (2013). Expected shortfall (or conditional var or cvar. https://www.statpro.com/glossary/expected-shortfall-or-conditional-var-or-c-var/. Accessed: 2019-04-08.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons.

Appendiks

A - Dansk og tysk ask-rente



Figur 6.1: Dansk og tysk ask-rente

B - Kode for å sjekke for kointegrasjon

```python
# Test for cointegration between the mid-prices

Coint_G = []
Coint_D = []
for r in range(len(formation_periods)):
    for_coint_G = []
    for_coint_D = []
# Collect the data you need in two lists
    for i in range(len(GObsv_Formation_periods[r])):
        #appending the mid-prices from the danish and german observations
        for_coint_G.append((GObsv_Formation_periods[r][i]["BY"] + GObsv_Formation_periods[r][i]["AY"])/2)
        for_coint_D.append((DObsv_Formation_periods[r][i]["BY"] + DObsv_Formation_periods[r][i]["AY"])/2)
    Coint_D.append(for_coint_D)
    Coint_G.append(for_coint_G)
    print(len(for_coint_D))
for i in range (len(Coint_D)):
# create a panda object where x and y are the two dataseries you want to take cointegration on
    df = pd.DataFrame({'x': Coint_G[i], 'y':Coint_D[i]})
    #print(df)
    print ('\n formation period:',i)
# this will print to your console
    coint_johansen(df, 0, 1)
```

C - Kode for bruttoavkastning

```
def gross_return(year, yslutt, ystart):
    if year == 3:
        return math.exp((((2*(yslutt - 1)+ 3) / (2*(yslutt**2))) - math.log(yslutt)) - (
                    (((2 * (ystart - 1) + 3) / (2*(ystart**2))) - math.log(ystart))))

    if year == 10:
        return math.exp(((1/yslutt)+(1/(2*(yslutt**2)))+(1/(3*(yslutt**3)))+(1/(4*(yslutt**4)))+(1/(5*(yslutt**5)))+(1/(6*(yslutt**6)))+(1/(7*(yslutt**7)))+(1/(8*(yslutt**8)))+(1/(9*(yslutt**9)))- math.log(yslutt)) -((1/ystart)+(1/(2*(ystart**2)))+(1/(3*(ystart**3)))+(1/(4*(ystart**4)))+(1/(5*(ystart**5)))+(1/(6*(ystart**6)))+(1/(7*(ystart**7)))+(1/(8*(ystart**8)))+(1/(9*(ystart**9))) - math.log(ystart)))
```

D - Kode for handel med avstandstilnærmingen

```python

print("\nStart trading logic\n")
cashflows = []  # save in this every time you get a cashflow

for i in range(len(ind_mean_spread_Tperiods)):
    # Loop for each period
    spread_period = ind_mean_spread_Tperiods[i]
    # start in "state" close pos
    openpos_t = False
    openpos_y = False
    Gopen_yield = 0
    Dopen_yield = 0
    Gclose_yield = 0
    Dclose_yield = 0
    #invested = False  # When moving into open the first day, invest in G or D
    investmentD = 0  # 1 or -1 depending on criteria
    investmentG = 0
    short = ""

    for r in range(len(spread_period)):
        # loop through each observation in period

        # This will open position if you are in close and the |spread| > |T|
        if (not openpos_t) and (abs(threshold_values[i]*2) > abs(spread_period[r]) > abs(threshold_values[i])):
            print("\nopening position in trade period", i, "at day", r)
            openpos_t = True

        if openpos_t and not openpos_y:
            # first day of open, (today is open, yesterday is close) decide whether to short or long
            # Place investments according to criteria
```

```
            if (ind_mean_spread_Tperiods[i][r] > threshold_values[i]):
                investmentD = 1
                investmentG = -1
                short = "G"
                Gopen_yield = GObsv_Trade_periods[i][r]["AY"]
                Dopen_yield = DObsv_Trade_periods[i][r]["BY"]
            elif (ind_mean_spread_Tperiods[i][r] < -(threshold_values[i])):
                investmentG = 1
                investmentD = -1
                short = "D"
                Gopen_yield = GObsv_Trade_periods[i][r]["BY"]
                Dopen_yield = DObsv_Trade_periods[i][r]["AY"]
            print('german open yield',Gopen_yield)
            # Logic to check if moving to close when you are in open
            # Only for when you have been in open for several days and not on the first day
            # Therefore inside this if
        if openpos_t and openpos_y:
            spread_y = spread_period[r-1]
            spread_t = spread_period[r]

            # Check if cross has occurred
            cross = False
            if spread_y > 0 and spread_t < 0 or spread_y < 0 and spread_t > 0:
                cross = True

            if abs(spread_t) > abs(2*threshold_values[i]) or cross:
                # criterias for moving into close
                print("Close position in trade period", i, "at day", r)
                if short == "G":
                    Gclose_yield = GObsv_Trade_periods[i][r]["BY"]
                    Dclose_yield = DObsv_Trade_periods[i][r]["AY"]
                elif short == "D":
                    Gclose_yield = GObsv_Trade_periods[i][r]["AY"]
```

```
                Dclose_yield = DObsv_Trade_periods[i][r]["BY"]
            print("German close yield", Gclose_yield)
            G_g_r = gross_return(10, Gclose_yield, Gopen_yield )
            D_g_r = gross_return(10, Dclose_yield, Dopen_yield )

            investment_G = investmentG * G_g_r
            investment_D = investmentD * D_g_r
            cashflows.append((investment_G-investmentG)+(investment_D-investmentD))

            # Reinitialise so that you can open several times during one period
            openpos_t = False
            investmentD = 0
            investmentG = 0
            short = ""
            Gopen_yield = 0
            Dopen_yield = 0
            Gclose_yield = 0
            Dclose_yield = 0

        # loop into next day
        openpos_y = openpos_t

    # Trade period has ended, calculate cashflow if you're in openpos
    if openpos_t:
        print("Close position at end of trade period", i, "at day", r)
        if short == "G":
            Gclose_yield = GObsv_Trade_periods[i][r]["BY"]
            Dclose_yield = DObsv_Trade_periods[i][r]["AY"]
        elif short == "D":
            Gclose_yield = GObsv_Trade_periods[i][r]["AY"]
            Dclose_yield = DObsv_Trade_periods[i][r]["BY"]
        print("German close yield", Gclose_yield)
```

```python
        G_g_r = gross_return(10, Gclose_yield, Gopen_yield)
        D_g_r = gross_return(10, Dclose_yield, Dopen_yield)

        investment_G = investmentG * G_g_r
        investment_D = investmentD * D_g_r
        cashflows.append((investment_G - investmentG) + (investment_D - investmentD))

# Looping into next period

print("\nCashflows:\n", cashflows)
print("\nNumber of cashflows:", len(cashflows))
print("\nTotal cashflow:", sum(cashflows))
print("\navg. cashflow:", sum(cashflows)/ len(cashflows))
```