



NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Hvilken GARCH-modell predikerer volatilitet best?

MASTEROPPGAVE I FINANSIELL ØKONOMI

FIN3900

Jørgen Thue Lind
Magnus Robinson Johansen

Veileder
Professor Gunnar Bårdsen

Forord

Denne oppgaven ble utført ved Norges Tekniske Naturvitenskaplige Universitet (NTNU), institutt for samfunnsøkonomi, i perioden januar til juni 2016.

Arbeidet med oppgaven har vært lærerikt og interessant, og vi håper å ha tilegnet oss kunnskap som vi får nytte av senere.

Til slutt vil vi rette en stor takk til oppgaveveileder Gunnar Bårdsen for gode råd og tilbakemeldinger underveis i prosessen.

NTNU Dragvoll, 31. mai 2016

Jørgen Thue Lind

Magnus Robinson Johansen

Sammendrag

Formålet med denne oppgaven er å sammenligne tre ulike volatilitetsmodeller ved og se på hvordan de egner seg til å predikere fremtidig volatilitet. Hvor godt de ulike modellene er i stand til å predikere volatilitet vil bli målt gjennom å sammenligne fire ulike finansielle tidsserier. Nærmere bestemt vil vi se på fire forskjellige aksjeindekser; Oslo Børs Benchmark Index (OSEBX), Deutscher Aktienindex (DAX), Standard & Poors 500 (S&P 500), Shanghai Composite Index (SHCOMP).

Ved å sammenligne predikert volatilitet med den "virkelige" volatiliteten vil man få et mål på hvor god modellen er til å predikere. Gjennom bestemte feilmål finner vi den modellen hvor predikert volatilitet avviker minst fra "virkelige" volatilitet. På bakgrunn av dette er vi i stand til å rangere de ulike modellene etter predikeringsprestasjonen. Resultatene gir ingen modell som presterer best i alle tilfellene. Vi kan imidlertid konkludere med en modell som skiller seg ut som bedre enn resten og det er EGARCH. Denne modellen presterer best for OSEBX og DAX, og nest best for S&P 500.

Innhold

Forord	i
Sammendrag	ii
1 Introduksjon	2
2 Metode	4
2.1 Data	4
2.2 Predikeringsfeil og volatilitetsproxy	4
2.3 Predikering og evaluering	6
3 Teori	8
3.1 Stokastiske volatilitetsmodeller	9
3.1.1 Autoregressive conditional heteroskedasticity - (ARCH)	9
3.1.2 Generalized ARCH (GARCH)	10
3.1.3 Exponential GARCH (EGARCH)	10
3.1.4 GJR	11
3.2 Betinget forventning til avkastningene	11
3.3 Stasjonaritet	12
3.4 Autokorrelasjon	13
3.5 ARCH-effekter	13
3.6 Predikering av volatilitet	14
3.7 Statistiske fordelinger	15
3.8 Estimering	16
3.9 Evaluering	17
4 Indeksene	18
4.1 Stasjonaritet	18
4.2 DAX	19
4.3 OSEBX	24
4.4 S&P 500	29
4.5 SHCOMP	34
4.6 Sannsynlighetsfordelinger	39
4.7 Parkinsonestimatorene	41
5 Oppsummering	44
5.1 Konklusjon	44
5.2 Forslag til videre arbeid	45
Bibliografi	46

A	Appendiks	48
A.1	Deskriptiv statistikk	48

Kapittel 1

Introduksjon

Verdens aksjemarked har den siste tiden vært preget av høy usikkerhet og store svingninger. Noen av årsakene ligger i stor variasjon i oljeprisen, lav kinesisk eksport og en BNP vekst som har vært lavere enn på mange år. Et slikt marked kjennetegnes ofte som volatil.

I finans er volatilitet en viktig parameter som inngår i mange modeller, og som sier noe om risikoen og usikkerheten knyttet til et aktivum. Volatiliteten blir her målt som standardavviket til avkastningen av et aktivum.

Volatiliteten er en viktig parameter i forhold til for eksempel prising av derivater og risikomåling når det kommer til porteføljevaltning. Hvis vi ser noen tiår tilbake, finner vi flere teoretiske modeller som antar at volatiliteten til forskjellige aktivum er konstant, se Black and Scholes (1973). Andersen and Bollerslev (1997) har derimot vist at volatiliteten er tidsvarierende. Etersom at volatiliteten er en såpass viktig parameter innenfor finans har det vært mye forskning for å finne gode måter å predikere fremtidig volatilitet.

Ved modellering av finansielle data ser man ofte en tendens til at dataene ikke passer med egenskapene til en regresjon på lineær form. Det er utviklet flere ikke-lineære modeller, men det er derimot kun et fåtall av disse som har vist seg å være gode til å modellere finansielle data. Brooks (2014) mener at de mest populære modellene for finansielle data er såkalte ARCH/ GARCH modeller. Dette er modeller som tar utgangspunkt i at variansen til restleddet ikke er konstant, og som på denne måten er i stand til å fange opp bestemte

egenskaper som man ofte finner i slik data. Disse egenskapene vil bli forklart senere i oppgaven. I denne oppgaven skal vi se på hvilken GARCH-modell som best beskriver og predikerer volatiliteten til fire forskjellige børsindekser. Indeksene vi har valgt er OSEBX, S&P 500, DAX og SHCOMP, og modellene vi har brukt til å estimere og predikere er GARCH, EGARCH og GJR.

Kapittel 2

Metode

2.1 Data

Denne oppgaven går ut på å sammenligne tre forskjellige volatilitetsmodeller ved å se på hvordan de egner seg til å predikere volatilitet. Evnen til å predikere blir målt gjennom å sammenligne ulike finansielle tidsserier. Vi har valgt fire aksjeindekser som inneholder data for en tiårsperiode, fra januar 2006 til januar 2016. For hvert av datasettene har vi fått tak i daglige sluttnoteringer, som er hentet fra Macrobond og Yahoo finance.

Når man undersøker finansielle tidsserier studerer man som regel prisendringen og ikke selve prisen. Grunnen til dette er at avkastningen til et aktiva gir et komplett bilde av investeringsmulighetene, samtidig som slik data er mye enklere å håndtere på grunn av de statistiske egenskapene dataene innehar. Det finnes flere definisjoner av avkastning, men i denne oppgaven vil vi kun se på log-avkastningen. Variabelen vi er interessert i er log-avkastningen (r_t), definert som log-endringen av dagens pris (p_t) og gårsdagens pris (p_{t-1}). Daglig avkastning er derfor definert som:

$$r_t = 100(\log(p_t) - \log(p_{t-1}))$$

2.2 Predikeringsfeil og volatilitetsproxy

Når man benytter GARCH-modeller til å predikere fremtiden, vurderer man hvor gode de er ut ifra hvor nøyaktige resultater man får. Man måler ofte denne evnen ved å se på modellens predikeringsfeil, som bedømmes ved hjelp av ulike feilmål. Felles for disse feilmålene er at vi ønsker å minimere verdien på de, hvor modellen med lavest

predikeringsfeil vil være den beste modellen.

Ved å sammenligne predikert volatilitet med den ”virkelige” volatiliteten til et aktivum vil man få et mål på hvor god modellen er til å predikere. Et problem med volatiliteten er at den ikke er observerbar. En løsning på dette problemet er å bruke en proxy. En proxy er ikke i seg selv relevant, men fungerer som et instrument for en uobserverbar variabel. En volatilitetsproxy som er mye brukt innenfor finansielle tidsserie-studier er kvadrert daglig avkastning. I ligningen under er σ_t^2 proxyen for variansen, og r_t^2 den daglige kvadrerte avkastningen i tidspunkt t :

$$\sigma_t^2 = r_t^2$$

Flere empiriske studier har undersøkt hvor nøyaktig ulike GARCH-modeller er i stand til å predikere når man benytter kvadrert daglig avkastning som volatilitetsproxy. Hovedkonklusjonen er at GARCH-modeller ikke er i stand til å predikere særlig bra når man benytter denne proxyen. Andersen and Bollerslev (1998) har gitt en forklaring på hvorfor dette gir et slikt unøyaktig resultat. Kort forklart har proxyen som benyttes for mye støy og for høy varians.

Nyere studier har vist til påfallende gode resultater ved bruk av andre mer egnede volatilitets-proxyer. Ved bruk av høyfrekvensdata hvor man beregner intradaglig volatilitet, vil man få en proxy med forholdsvis lav støy. Dette fører til at man får en lavere predikeringsfeil noe som vil gi et bilde av at modellen er mer presis.

Slike data er derimot lite tilgjengelige og ofte svært kostbare. Et annet alternativ til intradaglig data, er å bruke den såkalte Parkinsonestimatoren som volatilitetsproxy. I stedet for å benytte daglige sluttnoteringer, benyttes høy- og lavnoteringer. For eksempel kan prisen variere mye i løpet av en og samme dag, men ender sluttkursen på det samme som åpningskursen vil volatiliteten ved bruk av kvadrert daglig avkastning bli null. Parkinsonestimatoren vil gi et nokså godt bilde på daglig volatilitet siden den beskriver noe av hvor store svingninger det er i løpet av en dag. Vi ønsker derfor å se på predikeringsprestasjonen ved bruk av Parkinsonestimatoren som volatilitetsproxy sammenlignet med kvadrert daglig avkastning. Denne sammenligningen vil vi kun foreta på DAX-indeksen. Parkinsonestimatoren er gitt ved, se Poon (2005):

$$\sigma_t^2 = \frac{(\ln H_t - \ln L_t)^2}{4 \ln 2}$$

Hvor σ_t^2 er proxyen for volatiliteten, og H_t , L_t betegner henholdsvis daglig høy- og lavnotering.

2.3 Predikering og evaluering

For å se på hvilken modell som presterer best deler vi tidsseriene inn i to ulike dataperioder. En estimeringsperiode som brukes for å estimere modellens parametre, og en predikeringsperiode som brukes for å evaluere modellens predikeringssevne.

Lengden på predikeringsperioden ($T-n$) har vi bestemt ut i fra formelen gitt av Tsay (2008). Hvor $n = 2T/3$. Der T er totalt antall observasjoner. Det gir en predikeringsperiode på tre år (750 dager) for hver av tidsseriene.

For hver av aksjeindeksene har vi estimert tre ulike GARCH-modeller. Notasjonen brukt i oppgaven for å skille mellom modellene er $p,q(1/0)$. p står for antall GARCH-ledd og q står for antall ARCH-ledd. Vi skriver 1 eller 0 i parentesene ettersom om modellen er estimert med eller uten et AR(1)-komponent (autoregressiv). For eksempel. Så har vi en GARCH 1.2 (1), som er en GARCH-modell med $p = 1$, $q = 2$ og AR(1)-komponent i den betingede forventningen.

GARCH-modellene vi har valgt å estimerte er GARCH, EGARCH og GJR (se avsnitt 3.1). For hver av de ulike GARCH-modellene har vi videre sett på ulike kombinasjoner av p og q . Fire modeller med ulike verdier av p og q er estimert: (1.1), (1.2), (2.1) og (2.2). Dette betyr at vi i totalt estimerer 24 modeller. (3 ulike GARCH-modeller med 4 ulike verdier av p og q , og med og uten AR(1)-komponent i den betingede forventningen).

Når vi skal vurdere prestasjonen til de ulike modellene har vi i hovedfokuset vært å se på predikeringsprestasjonen, altså hvor godt modellen er i stand til å predikere. Vi har tatt i bruk to ulike feilmål å evaluere predikeringssevnen (se avsnitt 3.9). Feilmålene vi har brukt er; "Mean Absolute Error" (MAE) og Logarithmic Loss Function (LL). Vi har også sett på hvilke modell som er best i estimeringsperioden. Dette har vi gjort for å se om det er

noen sammenheng mellom den modellen som er best i predikeringsperioden og den som er best i estimeringsperioden. For å evaluere modellene i estimeringsperioden har vi brukt informasjonskriteriene AIC og SBIC (se avsnitt 3.10).

Kapittel 3

Teori

I denne delen av oppgaven vil teorien som har blitt brukt bli presentert.

I finansielle data finnes det ofte en del egenskaper som en lineær modell ikke kan fange opp. Ved bruk av GARCH modeller er man i bedre stand til å kunne modellere disse egenskapene. Disse egenskapene kan være

- ”Leptokurtose”: Tendens til at fordelingen for finansielle avkastninger har store haler og flere observasjoner sentrert rundt forventningen. Det at fordelingen har store haler betyr at det er større sannsynlighet for å observere høye eller lave ekstremverdier. I den tradisjonelle finansielle teorien tenkte man seg at avkastningen ville være normalfordelt. Dette har dog vist seg og ikke alltid være tilfelle, noe som har blitt bekreftet av flere blant annet Andersen et al. (2001)
- Volatilitetsklumping: Tendens til at volatilitet i finansielle aktivum kommer i perioder. En periode med lav volatilitet er ofte etterfulgt av en periode med lav volatilitet, og motsatt for perioder med høy volatilitet. ARCH/GARCH modeller vil være i stand til beskrive volatilitetsklumping bedre enn andre ikke-lineære modeller.
- ”Leverage effects”: Tendens til at volatiliteten blir høyere i en periode med et høyt prisfall, enn i en periode med høy prisstigning. Dette fenomenet er beskrevet av flere akademikere, blant annet Black (1976)
- ”Lang hukommelse”: Tendens til at effekten av et volatilitetssjokk forsvinner sakte, og som fører til at man kan få problemer med autokorrelasjon. På grunn av dette fenomenet med ”lang hukommelse” kan man være i stand til å predikere fremtidige

avkastninger. Dette bryter med ”markedeffisiens hypotesen”. I følge Fama (1991) skal prisene reflektere all tilgjengelig informasjon.

3.1 Stokastiske volatilitetsmodeller

For å være i stand til å fange de egenskapene som er beskrevet over, kan man bruke modeller der man antar at variansen til feilledet ikke er konstant. Vi vil først starte med å forklare ARCH-modellen da GARCH, EGARCH og GRJ er modeller som bygger videre på ARCH modellen.

3.1.1 Autoregressive conditional heteroskedasticity - (ARCH)

ARCH modellen ble først presentert av Robert Engle (1982), og har siden den tid fått mye oppmerksomhet. I de klassiske antakelsene om lineær regresjon er variansen til feilledet homoskedastisk. Det vil si at variansen til feilledet er konstant og ikke vil variere over tid. Hvis variansen ikke er konstant, er dette kjent som heteroskedastisitet. I finansielle avkastningsdata er det ofte slik at restleddet ikke har konstant varians. For å kunne estimere slike egenskaper kan man bruke en ARCH modell. Den betingede variansen, som er forventningen til neste periodes varians gitt den informasjonen man har i nåværende periode kan skrives slik.

$$\sigma_t^2 = \text{var}(u_t | u_{t-1}, u_{t-2}\dots) = E[(u_t - E(u_t | u_{t-1}, u_{t-2}\dots))^2 | u_{t-1}, u_{t-2}\dots]$$

Antar videre at $E(u_t | u_{t-1}, u_{t-2}\dots) = 0$, slik at

$$\sigma_t^2 = \text{var}(u_t | u_{t-1}, u_{t-2}\dots) = E[u_t^2 | u_{t-1}, u_{t-2}\dots]$$

Vi ser da fra ligningen over at den betingede variansen til u_t er lik den betingede forventningen til verdien av u_t^2 . I ARCH modellen er σ_t^2 basert på tidligere kvadrerte ledd av residualene. Modellen under er en ARCH(1), da den betingede variansen bare er avhengig av forrige periodes kvadrerte feilledd. ARCH modellen kan videre bli utviklet til en ARCH (q).

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

ARCH modellen i sin helhet er beskrevet under. r_t er her regresjonen for avkastningen og σ_t^2 for variansen i periode t

$$\begin{aligned}
r_t &= \beta_0 + \beta_1 r_{t-1} + \dots + \beta_p r_{t-p} + u_t \\
u_t &= v_t \sigma_t \quad v_t \sim N(0, 1) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2
\end{aligned}$$

ARCH modellen har skapt et grunnlag for analysering av volatiliteten til en tidsserie, men modellen har ikke blitt så mye brukt de siste tiår på grunn av noen begrensninger. Dette er begrensninger som går ut på hvordan man skal bestemme hvor mange tidligere feilledd man skal inkludere, og at det ofte kreves mange feilledd for å godt fange den betingede variansen. Videre har vi begrensningen om at leddene i modellen må være positive. En negativ varians gir ingen mening. På grunn av dette er GARCH-modellen en mer brukt modell.

3.1.2 Generalized ARCH (GARCH)

Generalized ARCH er en videreføring av ARCH modellen, og ble introdusert av Tim Bollerslev (1986). GARCH modellen lar den betingede variansen avhenge av sine tidligere ledd, og modellen i sin enkleste form er:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Hvis vi setter inn for tidligere ledd av den betingede variansen $(\sigma_{t-1}^2, \sigma_{t-2}^2, \dots, \sigma_{t-q}^2)$ kan vi se at GARCH(1,1) modellen er en ARCH(∞). På grunn av dette har vi fått en enklere og mer kompakt modell som bedre fanger den betingede variansen. GARCH-modellen kan utvides til en GARCH(p,q).

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

Hvor $\alpha_0 > 0, \alpha_i > 0, \beta_i > 0$ og $\sum \alpha_i + \sum \beta_i < 1$ for at ikke negativitetsbetingelsen skal bli brutt. Brooks (2014) mener at det ofte har vist seg at en GARCH (1,1) modell vil være i god stand til å fange opp volatilitetsklumping og andre egenskaper tidserien måtte ha.

3.1.3 Exponential GARCH (EGARCH)

Eksponensiell GARCH ble utviklet av Nelson (1991). ARCH og GARCH modellen går ut ifra at positive og negative sjokk har samme effekt på volatiliteten, mens EGARCH-modellen tar hensyn til såkalte "leverage effects". En annen fordel med

EGARCH er at siden det er $\ln(\sigma_t^2)$ som blir estimert, så kan høyresideestimatorene være negative da σ_t^2 uansett vil være positiv.

$$\ln(\sigma_t^2) = w + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

3.1.4 GJR

GJR (Glosten, Jagannathan og Runkel) er en enkel utvidelse av GARCH modellen, som har fått inkludert ett ekstra ledd ($\gamma u_{t-1}^2 I_{t-1}$) for å fange opp asymmetri i variansen, se Glosten et al. (1993).

$$\sigma_t^2 = \alpha_0 + \alpha_i u_{t-1}^2 + \beta_j \sigma_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1}$$

Hvor $\gamma > 0$ hvis vi har "leverage effects". For at ikke negativitetsbetingelsen skal bli brutt må $\alpha_0 > 0, \alpha_i > 0, \beta_i > 0$ og $\alpha_i + \gamma > 0$. Så hvis feilledet i forrige periode var negativt vil dette ha en større effekt på variansen enn om det var positivt.

3.2 Betinget forventning til avkastningene

I denne oppgaven er det den betingede variansen som er hovedfokuset, men for å få gode parametere for den betingede variansen, må også den betingede forventningen estimeres. Vi skiller mellom ubetinget og betinget forventning. Den ubetingede forventningen er gjennomsnittet av tidligere avkastninger. Grunnen til at det kalles ubetinget forventning er fordi den ikke er betinget på informasjon som er tilgjengelig i dag. Den betingede forventningen er derimot forventningen vår til neste periodes avkastning betinget på den informasjonen vi har i nåværende periode. Vi har i denne oppgaven valgt to modeller for den betingede forventningen: Hvor den første er at vi forventer at avkastningen (r_t) er konstant.

$$E(r_t | I_{t-1}) = \beta_0$$

I_{t-1} representerer informasjonen som er tilgjengelig på tidspunkt $t-1$. β_0 er den forventede avkastningen betinget på I_{t-1} : $\beta_0 = E(r_t | I_{t-1})$. Den andre modellen vi har valgt er en autoregressiv modell. En autoregressiv modell vil si en modell der den avhengige

variabelen estimeres på bakgrunn av tidligere ledd av seg selv. Vi har i denne oppgaven valgt en AR(1) modell. En modell der den avhengige variabelen bare er avhengig av forrige periodes variabel.

$$E(r_t | I_{t-1}) = \beta_0 + \beta_1 r_{t-1} + u_t$$

3.3 Stasjonaritet

I tidsserieøkonometri er man opptatt av at tidsseriene skal være stasjonære. Hvis de ikke er det er det ikke mulig å gjennomføre gyldig inferens. Vi skiller mellom sterk og svak stasjonaritet. En stokastisk variabel sies å være sterkt stasjonær hvis:

$$Fr_{t1}, r_{t2}, \dots, r_{tT}(r_1, \dots, r_T) = Fr_{t1+k}, r_{t2+k}, \dots, r_{tT+k}(r_1, \dots, r_T)$$

En sterkt stasjonær variabel har den samme fordelingen på tidspunkt t som på tidspunkt $t + k$. Ved svak stasjonaritet er følgende egenskaper oppfylt. Konstant forventning, konstant varians og konstant autokovariansstruktur:

1. $E(r_t) = \beta$
2. $E(r_t - \beta)(r_t - \beta) = \sigma^2 < \infty$
3. $E(r_{t_1} - \beta)(r_{t_2} - \beta) = \gamma_{t_2-t_1} \quad \forall t_1, t_2$

Test for stasjonaritet

For å teste om dataene er stasjonære bruker vi en utvidet Dickey-Fuller test. Fremgangsmåten er beskrevet under:

Dickey-Fuller(DF) test:

1. $r_t = \beta r_{t-1} + u_t$
Differensierer regresjonen, slik at man ender med
 $\Delta r_t = \psi r_{t-1} + u_t$

2. $H_0 : \psi = 0$ vs. H_1 : Serien er stasjonær

3. Testobservatoren er gitt ved

$$Test = \frac{\hat{\psi}}{SE(\hat{\psi})} \sim DF \text{ distribution}$$

Dickey-Fuller testen er bare gyldig dersom u_t er hvit støy. Hvis den ikke er det, må man utføre en utvidet Dickey-Fuller (ADF) test.

1. $\Delta r_t = \psi r_{t-1} + \sum_{i=1}^p \Delta r_{t-i} + u_t$
2. Vi inkluderer her differensierte ledd av r for å fjerne dynamisk struktur som fører til at feilleddet er autokorrelert. For å finne antall differensierte ledd som skal tas med, bruker vi informasjonskriteriene. (se avsnitt 3.10)

3. $H_0 : \psi = 0$ vs. H_1 : Serien er stasjonær

Testobservatoren er gitt ved

$$Test = \frac{\hat{\psi}}{SE(\hat{\psi})} \sim DF \text{ distribution}$$

3.4 Autokorrelasjon

Autokorrelasjonen sier noe om samvariasjonen mellom restleddene, og er definert som:

$$Corr(u_t, u_{t-i}) = \frac{Cov(u_t, u_{t-i})}{\sqrt{Var(u_t)Var(u_{t-i})}}$$

Autokorrelasjon oppstår dersom det er en systematisk sammenheng mellom feilleddene, u_t , slik at feilleddet i tidspunkt t inneholder informasjon om restleddet i tidspunkt $t+1$. Dersom man har autokorrelasjon i restleddene, må man spesifisere en modell som fjerner denne korrelasjonen. Har man en GARCH modell som fanger fordelingen godt, skal korrelasjonen bli fjernet, se Hull (2006). Vi tester dette på de standardiserte residualene som er gitt ved u_t/σ_t^2 . Når man skal teste for autokorrelasjon kan man bruke en Ljung-Box test, Ljung and Box (1978). Ljung-Box testen er gitt ved følgende formel:

$$Q_k = T(T+2) \sum_{i=1}^k \frac{\tau_k^2}{T-i} \sim \chi_k^2$$

Hvor T er antall i utvalget, k er antall lag og τ_k er autokorrelasjonen i tidspunkt k . $Q_k \sim \chi^2$ fordelt med k frihetsgrader. Nullhypotesen ved denne testen, er at det ikke er autokorrelasjon i restleddene.

3.5 ARCH-effekter

For å sjekke om variansen til residualene inneholder ARCH-effekter kan vi utføre Engle's ARCH-test, se Engle (1982). Det at residualene inneholder ARCH-effekter vil si at

variansen til residualene ikke er konstant, altså at vi har hetroskedastisitet. Hvis man har ARCH-effekter, trenger man en ikke-lineær modell som for eksempel ARCH/GARCH modeller til å estimere parameterne.

Test for ARCH-effekter

1. Kjører først en lineær regresjonen som vist under, hvor r_t er avkastningen i tidspunkt t , og hvor r_t er basert på sine tidligere avkastninger.

$$r_t = \beta_0 + \beta_1 r_{t-1} + \beta_2 r_{t-2} + u_t$$

Lagre så residualene \hat{u}_t fra ligningen.

2. Kvadrer residualene, og kjør en regresjon på tidligere kvadrerte feilledd.

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \dots + \gamma_q \hat{u}_{t-q}^2 + v_t$$

Finn R^2 for denne regresjonen.

3. Testobservatoren er gitt ved $TR^2 \sim \chi^2(q)$

4. Null og alternativhypotesen er: $H_0 : \gamma_1 = 0, \gamma_2 = 0, \dots, \gamma_q = 0$

$$H_0 : \gamma_1 \neq 0, \gamma_2 \neq 0, \dots, \gamma_q \neq 0$$

3.6 Predikering av volatilitet

Mens den betingede variansen varierer, er den ubetingede variansen konstant. For å vise fremgangsmåten for hvordan vi bruker GARCH-modeller til å predikere volatiliteten i predikeringsperioden, tar vi utgangspunkt i en GARCH(1.1)-modell. Hovedpoenget i selve predikeringen vil være det samme for EGARCH og GJR modellen.

Vi kan tenke oss at den betingede variansen for tre tidspunkter frem i tid er gitt ved:

$$\begin{aligned}\sigma_{T+1}^2 &= \alpha_0 + \alpha_1 u_T^2 + \beta_1 \sigma_T^2 \\ \sigma_{T+2}^2 &= \alpha_0 + \alpha_1 u_{T+1}^2 + \beta_1 \sigma_{T+1}^2 \\ \sigma_{T+3}^2 &= \alpha_0 + \alpha_1 u_{T+2}^2 + \beta_1 \sigma_{T+2}^2\end{aligned}$$

La $\sigma_{1,T}^{f^2}$ være predikasjonen for en tidsperiode frem i tid for σ gjort i tidspunkt T. Denne kan vi nå lett regne ut da alle høyreside variablene er kjent.

$$\sigma_{1,T}^{f^2} = \alpha_0 + \alpha_1 u_T^2 + \beta \sigma_T^2$$

Predikering av variansen for to tidsperioder frem i tid $\sigma_{2,T}^{f^2}$ er gitt ved å ta den betingede forventningen av σ_{t+2}^2 .

$$\sigma_{2,T}^{f^2} = \alpha_0 + \alpha_1 E(u_{T+1}^2 | I_T) + \beta \sigma_T^2$$

Hvor $E(u_{T+1}^2 | I_T)$ er forventningen gjort i tidspunkt T, av u_{T+1}^2 . I_T er all tilgjengelig informasjon gitt på tidspunkt T. Vi kan skrive $E(u_{T+1}^2 | I_T) = \sigma_{T+1}^2$, men da σ_{T+1}^2 ikke er kjent på tidspunkt T, erstatter vi den med den predikerte verdien $\sigma_{1,T}^{f^2}$. Det vil si at:

$$\begin{aligned}\sigma_{2,T}^{f^2} &= \alpha_0 + \alpha_1 \sigma_{1,T}^{f^2} + \beta \sigma_{1,T}^{f^2} \\ \sigma_{2,T}^{f^2} &= \alpha_0 + (\alpha_1 + \beta) \sigma_{1,T}^{f^2}\end{aligned}$$

Predikering av verdi for den tredje tidsperioden blir da:

$$\begin{aligned}\sigma_{3,T}^{f^2} &= \alpha_0 + (\alpha_1 + \beta) \sigma_{2,T}^{f^2} \\ \sigma_{3,T}^{f^2} &= \alpha_0 + (\alpha_1 + \beta) [\alpha_0 + (\alpha_1 + \beta) \sigma_{1,T}^{f^2}] \\ \sigma_{3,T}^{f^2} &= \alpha_0 + \alpha_0 (\alpha_1 + \beta) + (\alpha_1 + \beta)^2 \sigma_{1,T}^{f^2}\end{aligned}$$

For s-perioder frem i tid blir predikert verdi:

$$\sigma_{s,T}^{f^2} = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta)^{s-1} \sigma_{1,T}^{f^2}$$

Hvis $\alpha_1 + \beta_1 < 1$ vil, $(\alpha_1 + \beta)^{s-1} \sigma_{1,T}^{f^2}$ etterhvert gå mot 0, og den predikerte verdien av den betingede variansen vil gå mot den ubetingede variansen. Den ubetingede variansen for GARCH(1.1) er gitt ved:

$$\sigma_u^2 = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

3.7 Statistiske fordelinger

I denne oppgaven har vi valgt å se på tre forskjellige sannsynlighetsfordelinger. Fordelingene er normalfordelingen, student-t og skjev student-t. Vi vil teste de forskjellige fordelingene på den modellen som har best predikeringsevne for hver av indeksene. Ettersom finansielle avkastningsdata har en tendens til å ha en fordeling med store haler og flere observasjoner sentrert rundt forventningen, er det sannsynlig at det er andre fordelinger enn normalfordelingen som fanger fordelingen best, se Mandelbrot (1997)

3.8 Estimering

Hvis vi har restledd som ikke er homoskedastiske, må vi bruke maximum likelihood for å estimere modellen. Det er flere grunner til at vi ikke kan bruke vanlig minste kvadraters metode (MKM). MKM går blant annet ut på å minimere den kvadrerte summen av residualene i den betingede forventningen, og ikke av den betingede variansen.

Maximum likelihood fungerer ved å finne verdiene til de parametrene som er mest sannsynlige, gitt de aktuelle dataene. Sagt på en annen måte så har vi en log-likelihood-funksjon som skal maksimeres, og hvor verdiene til parameterene som maksimerer bli søkt.

For å vise hvordan maximum likelihood fungerer kan vi kan for enkelhets skyld tenker oss at vi har en lineær modell hvor restleddet ikke er hetroskedastisk. Regresjonen er gitt ved:

$$r_t = \beta_0 + \beta_1 r_{t-1} + u_t$$

Antar at $u_t \sim N(0, \sigma^2)$, så $r_t \sim N(\beta_0 + \beta_1 r_{t-1}, \sigma^2)$. (Hvis vi hadde hatt hetroskedastisitet i restleddet ville vi antatt at $u_t \sim N(0, \sigma_t^2)$). Sannsynlighetsfordelingen for en normalfordelt variabel med forventning og varians er gitt ved:

$$f(r_t | \beta_0 + \beta_1 r_{t-1}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{(r_t - \beta_0 + \beta_1 r_{t-1})^2}{\sigma^2} \right\}$$

Vanligvis er r_t og r_{t-1} gitt, og det er $\beta_0, \beta_1, \sigma^2$ skal estimeres. Hvis dette er tilfelle er $f(\bullet)$ kjent som likelihood funksjonen, og gitt ved:

$$LF(\beta_0, \beta_1, \sigma^2) = \frac{1}{\sigma^T(\sqrt{2\pi})^T} \exp\left\{ -\frac{1}{2} \sum_{t=1}^T \frac{(r_t - \beta_0 + \beta_1 r_{t-1})^2}{\sigma^2} \right\}$$

Maximum likelihood estimering går ut på å velge de parameterverdiene $(\beta_0, \beta_1, \sigma^2)$ som maksimerer funksjonen. Vi tar videre logaritmen av funksjonen for å få den på formen:

$$LLF = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\sigma^2) - \frac{1}{2} \sum_{t=1}^T \frac{(r_t - \beta_0 + \beta_1 r_{t-1})^2}{\sigma^2}$$

For å maksimere funksjonen over bruker man vanligvis et statistikkprogram som for eksempel OxMetrics eller Eviews. Programmet vil generere de parameterne som maksimerer LLF funksjonen, å beregne standardfeilene.

3.9 Evaluering

For å evaluere modellene i estimeringsperioden bruker vi informasjonskriterier. Akaike's(1974)- Schwarz's(1978) informasjonskriterie:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}$$
$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T} \ln T$$

Hvor $\hat{\sigma}^2$ er feilleddets varians, $k = p + q + 1$ er antall av estimerte parametere og T er antall i utvalget. Den beste modellen er den som minimerer informasjonskriteriene. Forskjellen mellom AIC og SBIC er at SBIC blir påvirket i større grad dersom det inkluderes en ekstra parameter i modellen.

For å evaluere modellen i predikeringsperioden bruker vi valgt å bruke Mean Absolute Error (MAE) og Logarithmic Loss function (LL) som feilmål, se Souza et al. (2002):

$$MAE = \frac{1}{T} \sum_{t=1}^T | \sigma_t^2 - \sigma_t^{f^2} |$$
$$LL = \sqrt{\frac{1}{T} \sum_{t=1}^T (\log(\sigma_t^2) - \log(\sigma_t^{f^2}))^2}$$

Hvor T er lengden på predikeringsperioden, som i vårt tilfelle er 750. σ_t^2 er den "virkelige" volatiliteten, mens $\sigma_t^{f^2}$ er den predikerte variansen. Med MAE ser vi på absoluttverdien til forskjellen mellom "virkelig" og predikert verdi, mens for LL ser vi på kvadratroten av til det kvadrerte avviket mellom "virkelig" og predikert verdi.

Kapittel 4

Indeksene

4.1 Stasjonaritet

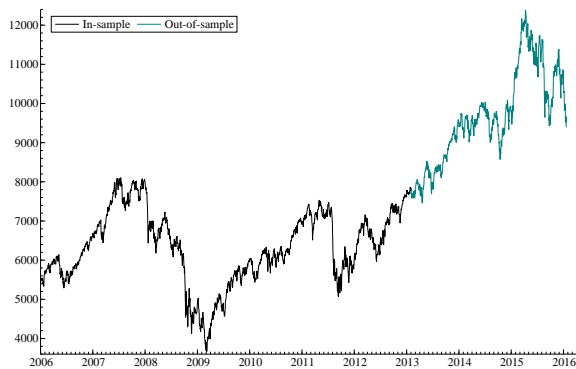
For at vi skal være i stand til å utføre gyldig inferens, er dataen nødt til å være stasjonær. Under har vi utført en utvidet Dickey-Fuller test (se avsnitt 3.1). Vi ser at dataen vi har brukt for alle indeksene er stasjonære på alle signifikantnivåene i tabellen under.

Tabell 4.1: Utvidet Dickey-Fuller test

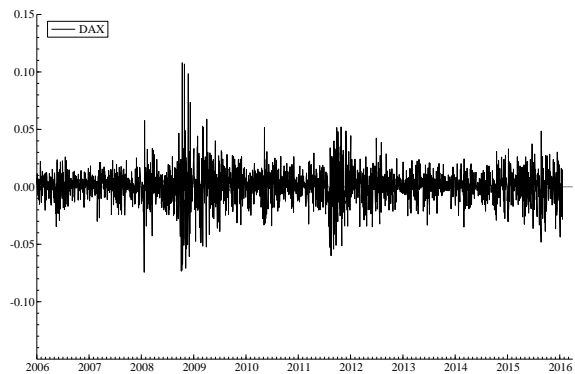
	OSEBX	DAX	S&P 500	SHCOMP
ADF Statistics	-25.55	-30.70	-29.92	-27.85
	1%	5%	10%	
Critical values	-2.56572	-1.94093	-1.61663	

4.2 DAX

Deutscher Aktienindex (DAX) er en aksjeindeks som består av 30 store tyske selskaper som er notert på Frankfurt-børsen. Datasettet består av 2563 observasjoner fra 2 januar 2006 til 20. Januar 2016. Estimeringsperioden går fram til 1. Februar 2013. Predikeringsperioden går fra 4. Februar 2013 til 20. Januar 2016.



Kursutvikling DAX



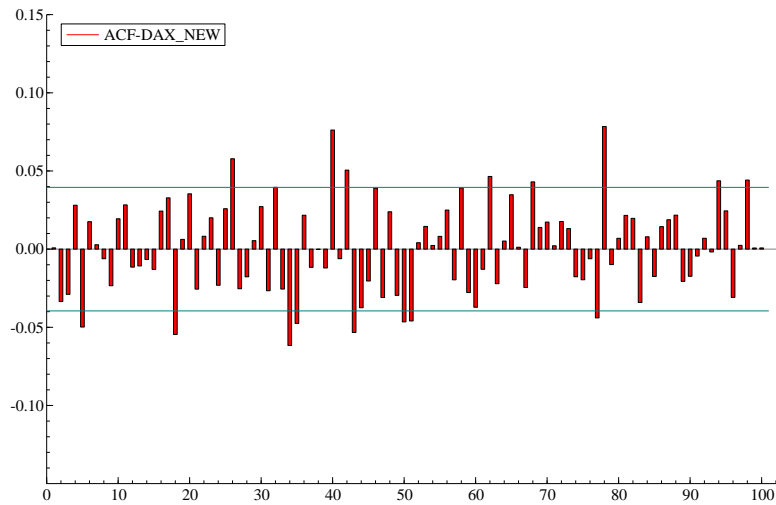
Avkastning DAX

Deskriptiv statistikk

Gjennomsnitt	0,00021533
Standardavvik	0,014474
Skjevhet	0,014448
Kurtose	5,7574
Maksimum	-0,074335
Minimum	0,10797

I tabellen over kan vi se den deskriptive statistikken for avkastningene til DAX. Vi ser her at kurtose og skjevhet er henholdsvis 5,7574 og 0,014448. Dette innebærer at fordelingen har større haler enn normalfordelingen (kurtose=3), og at fordelingen er høyreskjev. Den deskriptive statistikken er her basert på alle observasjonene, både for estimerings- og predikeringsperioden.

Autokorrelasjon



Figur 4.2: Autokorrelasjon av avkastningene

Ser vi på figuren over ser vi at vi kan ha et problem med autokorrelasjon i dataen. Figuren viser autokorrelasjonen fra lag 1-100, og et konfidensintervall på 95%.

Tabell 4.2: Ljung-Box test av avkastningene

$Q(5) =$	12.9769	$[0.0235960]^*$
$Q(10) =$	17.4322	$[0.0653315]$
$Q(20) =$	36.7484	$[0.0125418]^*$
$Q(50) =$	138.135	$[0.0000000]^*$

Tabellen over viser en Ljung-Box test for autokorrelasjon. Vi ser her som forventet at vi har et problem med autokorrelasjon i ledd 5, 20 og 30. Vi forkaster her nullhypotesen om ingen autokorrelasjon i dataen.

Engle's ARCH-test

Tabell 4.3: ARCH-test av avkastningene

ARCH 1-2 test:	$F(2,2558) = 104.99$	$[0.0000]**$
ARCH 1-5 test:	$F(5,2552) = 83.004$	$[0.0000]**$
ARCH 1-10 test:	$F(10,2542) = 52.689$	$[0.0000]**$

For å se om restleddet er hetroskedastisk, anvender vi en ARCH test. Nullhypotesen er at ARCH-effekter ikke er tilstedeværende. Vi ser i tabellen over at nullhypotesen forkastes for alle testene. Vi har her ARCH-effekter i dataen og en ikke-lineær modell må anvendes for å fange opp disse effektene.

Estimering og evaluering

Som vi ser av figuren øverst på neste side er hver av GARCH-modellene estimert med ulike verdier av p og q . Tallene oppgitt er verdiene for de ulike feilmålene, hvor de to første radene viser verdiene for informasjonskriteriene brukt for å evaluere estimeringsperioden, mens de to nederste radene inneholder verdiene for predikeringsfeilene. For hver rad angir grønn farge beste modell, mens rød farge angir dårligste modell.

Den standard GARCH-modellen som gir best resultat for predikeringsperioden er GARCH 1.1(1) med en (MAE) 2,05 og en (LL) på 10,38.

Den GARCH-modellen som gir best resultat totalt er EGARCH 2.2(1) med en (MAE) på 1,679 og en (LL) på 9,024. Modellen er her best i følge begge predikeringsfeilene.

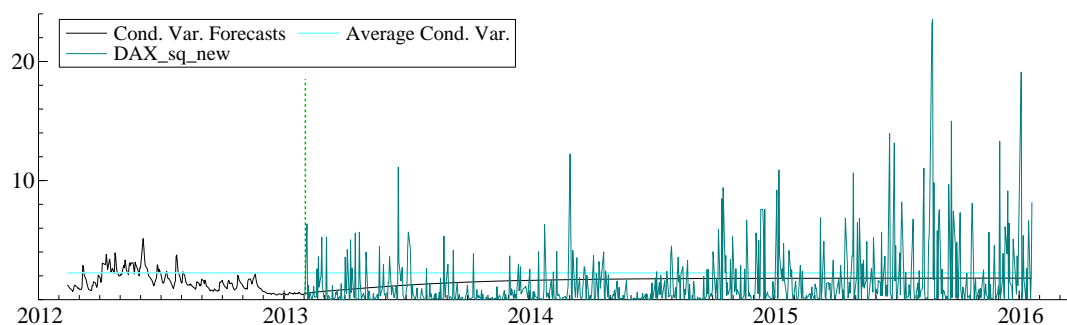
DAX

GARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,347468	3,348258	3,343365	3,344327	3,347264	3,348127	3,344725	3,34566
Schwarz (BIC)	3,359609	3,363434	3,358541	3,362538	3,36244	3,366338	3,362936	3,366906
Mean Absolute Error(MAE)	2,052	2,05	2,101	2,1	2,062	2,061	2,093	2,092
Logarithmic Loss Function(LL)	10,38	10,38	10,54	10,54	10,42	10,42	10,52	10,52

EGARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,291826	3,292839	3,286682	3,287711	3,293212	3,294104	3,282321	3,283365
Schwarz (BIC)	3,310037	3,314085	3,307928	3,311992	3,314458	3,318385	3,306602	3,310682
Mean Absolute Error(MAE)	1,732	1,731	1,695	1,693	1,702	1,699	1,681	1,679
Logarithmic Loss Function(LL)	9,265	9,261	9,099	9,093	9,138	9,128	9,033	9,024

GJR(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,297284	3,298381	3,286387	3,287488	3,298451	3,299546	3,278514	3,289178
Schwarz (BIC)	3,31246	3,316592	3,307633	3,31177	3,316662	3,320792	3,302795	3,316494
Mean Absolute Error(MAE)	1,739	1,741	1,722	1,723	1,733	1,736	1,681	#N/A
Logarithmic Loss Function(LL)	9,299	9,309	9,231	9,235	9,277	9,289	9,04	#N/A

Figur 4.3: Evaluering av DAX



Figur 4.4: Predikering av volatiliteten til DAX med EGARCH 2.2(1)

Tabell 4.4: Koeffisientene, standardfeilene og t-verdiene til DAX EGARCH 2.2(1)

	Koeffisient	Std.feil	t-verdi	t-sanns.
Cst(M)	0,021519	0,021999	0,9782	0,3281
Cst(V)	0,596982	0,21776	2,741	0,0062
ARCH(Alpha1)	-0,71739	0,3697	-1,94	0,0525
ARCH(Alpha2)	-0,194125	0,30101	-0,6449	0,5191
GARCH(Beta1)	1,807995	0,1251	14,45	0,0000
GARCH(Beta2)	-0,809816	0,12332	-6,567	0,0000
EGARCH(Theta1)	0,18193	0,047999	-3,79	0,0002
EGARCH(Theta2)	0,12965	0,027625	4,693	0,0000

Tabell 4.5: Ljung-Box test av de standardiserte residualene

Q(5) =	3.41840	[0.6357693]
Q(10) =	5.81875	[0.8302546]
Q(20) =	14.5895	[0.7993968]
Q(50) =	60.8336	[0.1402203]

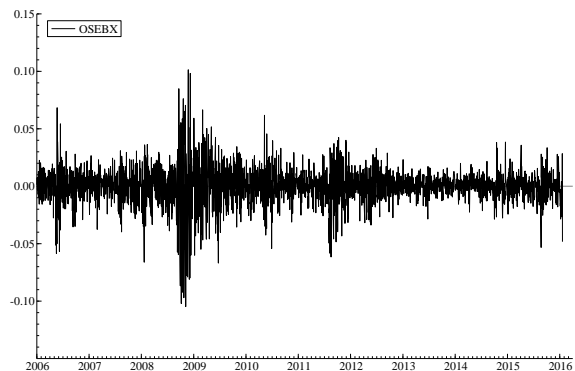
Ser vi nå på autokorrelasjonen til de standardiserte residualene til EGARCH 2.2(1) som er gitt ved u_t/σ_t^2 , ser vi at den er fjernet og at modellen godt fanger opp autokorrelasjonen som var i restleddet. I figuren over ser vi at vi beholder nullhypotesen for alle de fire testene.

4.3 OSEBX

OSEBX er en aksjeindeks som inneholder et representativt utvalg av alle aksjer på Oslo Børs. Per 1 januar 2016 består denne indeksen av 58 ulike aksjer. Datasettet inneholder daglige data fra 2. Januar 2006 til 20. Januar 2016 og består av til sammen 2554 observasjoner. Estimeringsperioden strekker seg fra 1 januar 2006 til 21. Januar 2013, mens predikeringsperioden består av 750 datapunkt og dekker de gjenstående tre årene fram til 20. Januar 2016.



Kursutvikling OSEBX



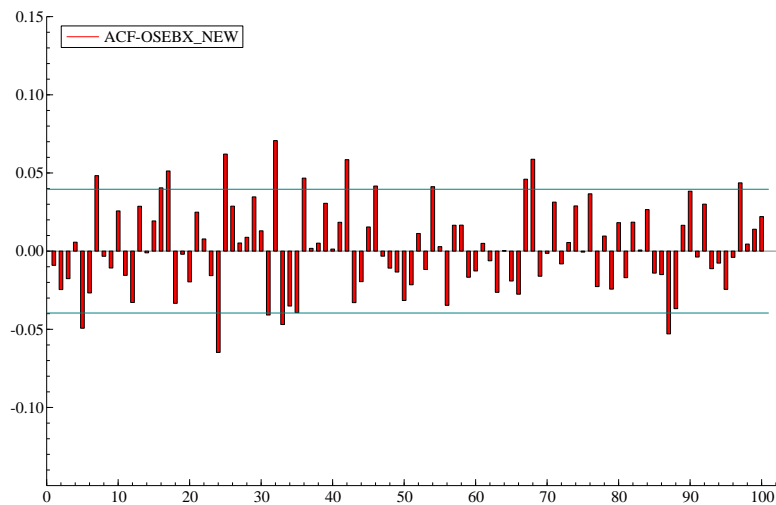
Avkastning OSEBX

Deskriptiv statistikk

Gjennomsnitt	0,00017361
Standardavvik	0,016517
Skjevhet	-0,56914
Kurtose	6,1447
Maksimum	-0,10478
Minimum	0,10136

I tabellen over kan vi se den deskriptive statistikken for avkastningene til OSEBX. Vi ser her at kurtose og skjevhet er henholdsvis 6,1447 og -0,56914. Dette innebærer at fordelingen har større haler enn normalfordelingen (kurtose=3), og at fordelingen er venstreskjev. Den deskriptive statistikken er her basert på alle observasjonene, både for estimerings- og predikeringsperioden.

Autokorrelasjon



Figur 4.6: Autokorrelasjon av avkastningene

Ser vi på figuren for autokorrelasjonen over ser vi at vi kan ha et problem med autokorrelasjon i dataen. Figuren viser autokorrelasjonen fra lag 1-100, og et konfidensintervall på 95%.

Tabell 4.6: Ljung-Box test av avkastningene

$Q(5) =$	6.85987	[0.2312720]
$Q(10) =$	15.4168	[0.1175884]
$Q(20) =$	34.8119	[0.0211293]*
$Q(50) =$	115.695	[0.0000004]*

Tabellen over viser en Ljung-Box test for autokorrelasjon. Vi ser her som forventet at vi har et problem med autokorrelasjon i ledd 20 og 30. Vi forkaster her nullhypotesen om ingen autokorrelasjon i dataen.

Engle's ARCH-test

Tabell 4.7: ARCH-test av avkastningene

ARCH 1-2 test:	$F(2,2549) = 189.48$	$[0.0000]**$
ARCH 1-5 test:	$F(5,2543) = 202.41$	$[0.0000]**$
ARCH 1-10 test:	$F(10,2533) = 114.93$	$[0.0000]**$

For å se om restleddet er hetroskedastisk, anvender vi en ARCH test. Nullhypotesen her er at ARCH-effekter ikke er tilstedeværende. Vi ser i tabellen over at nullhypotesen forkastes for alle testene. Vi har her ARCH-effekter i dataen og en ikke-lineær modell må anvendes for å fange opp disse effektene.

Estimering og evaluering

Ved bruk av standard GRACH er GARCH (1.1)(1) med en (MAE) på 2,441 og en (LL) på 13,39 den modellen som gir best resultat for predikeringsperioden. Det er liten variasjon mellom modellene i hvor godt de presterer når det kommer til ulike verdier av p og q ved bruk av standard GARCH.

Den GARCH-modellen med som har lavest predikeringsfeil og dermed presterer best totalt er EGARCH 2.1(0). Med en (MAE) 1,204 og en (LL) på 9,397. Alle variantene av EGARCH presterer for øvrig bedre en den beste GARCH-modellen. EGARCH 2.1(0) er også den modellen med best resultat i estimeringsperioden. Modellen som gir dårligst resultat er GARCH 2.2(0) med en (MAE) på 2,456 og en (LL) på 13,44.

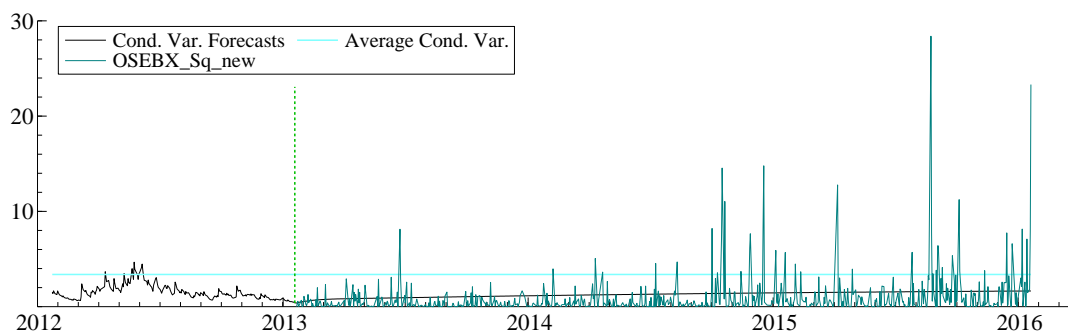
OSEBX

GARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,669046	3,669417	3,666707	3,667278	3,66874	3,669238	3,668104	3,668647
Schwarz (BIC)	3,681237	3,684654	3,681944	3,685564	3,683978	3,687523	3,686389	3,68998
Mean Absolute Error(MAE)	2,446	2,441	2,45	2,441	2,449	2,441	2,456	2,448
Logarithmic Loss Function(LL)	13,4	13,39	13,42	13,4	13,42	13,4	13,44	13,42

EGARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,646271	3,647057	3,645435	3,646053	3,633015	3,633716	3,633033	3,633918
Schwarz (BIC)	3,664556	3,66839	3,666768	3,670434	3,654348	3,658097	3,657414	3,661346
Mean Absolute Error(MAE)	1,888	1,909	1,873	1,893	1,204	1,21	1,211	1,21
Logarithmic Loss Function(LL)	11,95	12,01	11,89	11,94	9,397	9,432	9,434	9,428

GJR(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,641041	3,641974	3,638052	3,638937	3,642486	3,643427	3,628035	3,628506
Schwarz (BIC)	3,656278	3,660259	3,659385	3,663317	3,660771	3,66476	3,652416	3,655934
Mean Absolute Error(MAE)	1,918	1,907	1,944	1,925	1,919	1,908	1,83	1,805
Logarithmic Loss Function(LL)	12,1	12,07	12,17	12,11	12,1	12,07	11,8	11,73

Figur 4.7: Evaluering av OSEBX



Figur 4.8: Predikering av volatiliteten til OSEBX EGARCH 2.1(0)

Tabell 4.8: Koeffisientene, standardfeilene og t-verdiene til OSEBX med EGARCH 2.1(0)

	Koeffisient	Std.feil	t-verdi	t-sanns.
Cst(M)	0,049142	0,042589	1,758	0,0789
Cst(V)	1,120491	0,24512	4,571	0,0000
ARCH(Alpha1)	-0,103194	1,8716	64,40	0,0000
GARCH(Beta1)	1,259494	0,88526	1,423	0,1551
GARCH(Beta2)	-0,276384	0,86268	103,5	0,0000
EGARCH(Theta1)	-0,073496	0,06525	-6,152	0,0000
EGARCH(Theta2)	0,154452	0,12223	6,379	0,0000

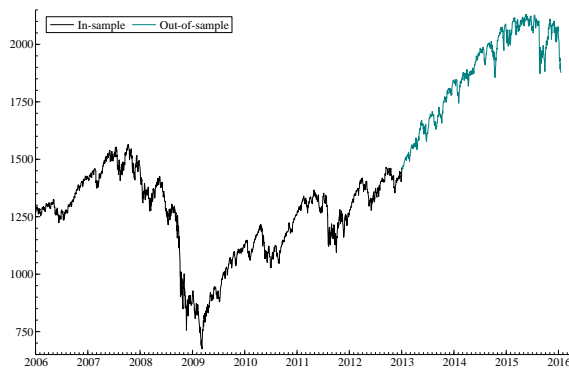
Tabell 4.9: Ljung-Box test av de standardiserte residualene

Q(5) =	2.88025	[0.7184413]
Q(10) =	5.46315	[0.8581736]
Q(20) =	10.4460	[0.9593455]
Q(50) =	54.7949	[0.2976453]

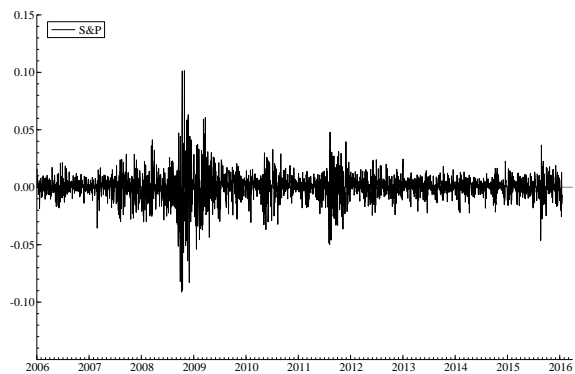
Ser vi nå på autokorrelasjonen til de standardiserte residualene til EGARCH 2.1(0) som er gitt ved u_t/σ_t^2 , ser vi at den er fjernet og at modellen godt fanger opp autokorrelasjonen som var i restleddet. I figuren under ser vi at vi beholder nullhypotesen for alle de fire testene.

4.4 S&P 500

Standard & Poors 500 (S&P 500) er en aksjeindeks som består av over 500 store selskaper som er notert på New York Stock Exchange. Datasettet består av 2528 observasjoner fra 3. januar 2006 til 19. januar 2016. Predikeringsperioden består av de siste 750 observasjonene og starter 28. januar 2013.



Kursutvikling S&P 500



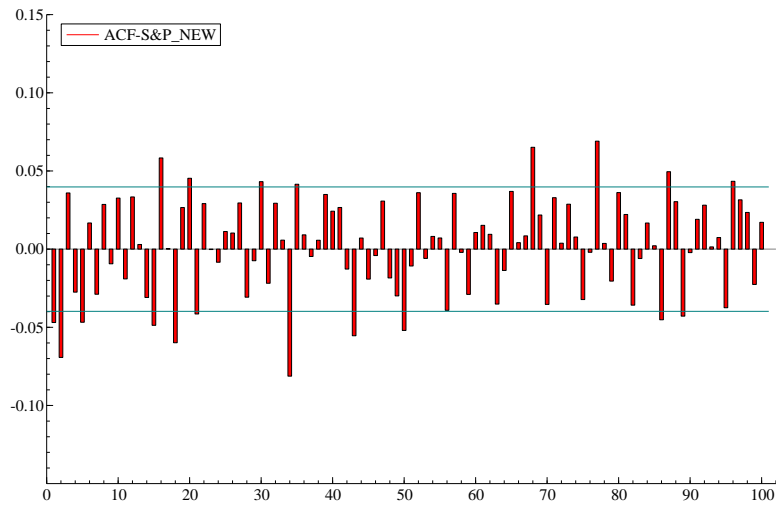
Avkastning S&P 500

Deskriptiv statistikk

Gjennomsnitt	0,00016118
Standardavvik	0,012366
Skjevhet	-0,3621
Kurtose	10,424
Maksimum	-0,09115
Minimum	0,10139

I tabellen over kan vi se den deskriptive statistikken for avkastningene til S&P 500. Vi ser her at kurtose og skjevhet er henholdsvis 10,424 og -0,3621. Dette innebærer at fordelingen har større haler enn normalfordelingen (kurtose=3), og at fordelingen er venstreskjev. Den deskriptive statistikken er her basert på alle observasjonene, både for estimerings- og predikeringsperioden.

Autokorrelasjon



Figur 4.10: Autokorrelasjon av avkastningene

Ser vi på figuren for autokorrelasjonen over ser vi at vi kan ha et problem med autokorrelasjon i dataen. Figuren viser autokorrelasjonen fra lag 1-100, og et konfidensintervall på 95%.

Tabell 4.10: Ljung-Box test av avkastningene

$Q(5) =$	25.7226	[0.0001010]*
$Q(10) =$	32.1826	[0.0003733]*
$Q(20) =$	61.0864	[0.0000048]*
$Q(50) =$	125.721	[0.0000000]*

Tabellen over viser en Ljung-Box test for autokorrelasjon. Vi ser her som forventet at vi har et problem med autokorrelasjon i ledd 5, 10, 20 og 30. Vi forkaster her nullhypotesen om ingen autokorrelasjon i dataen.

Engle's ARCH-test

Tabell 4.11: ARCH-test av avkastningene

ARCH 1-2 test:	$F(2,2523) = 329.10$	$[0.0000]**$
ARCH 1-5 test:	$F(5,2517) = 190.84$	$[0.0000]**$
ARCH 1-10 test:	$F(10,2507) = 108.61$	$[0.0000]**$

For å se om restleddet er hetroskedastisk, anvender vi en ARCH test. Nullhypotesen her er at ARCH-effekter ikke er tilstedeværende. Vi ser i tabellen over at nullhypotesen forkastes for alle testene. Vi har her ARCH-effekter i dataen og en ikke-lineær modell må anvendes for å fange opp disse effektene.

Estimering og evaluering

Ved standard GARCH er GARCH 1.2(1) den modellen som presterer best, med en (MAE) på 1,115 og en (LL) på 11,5.

Når det kommer til den modellen som gir best resultat når man inkluderer alle modellene er GJR 1.2(1) best. Modellen er her best i begge feilmålene. Med en (MAE) på 0,8673 og en (LL) på 10,1 så slår den så vidt nest beste modell som er EGARCH 2.2(1) med en (MAE) på 0,8719 og en (LL) på 10,11. Når man ser på forskjell i prestasjon ved ulike verdier av p og q er det totalt sett liten forskjell mellom GJR og EGARCH.

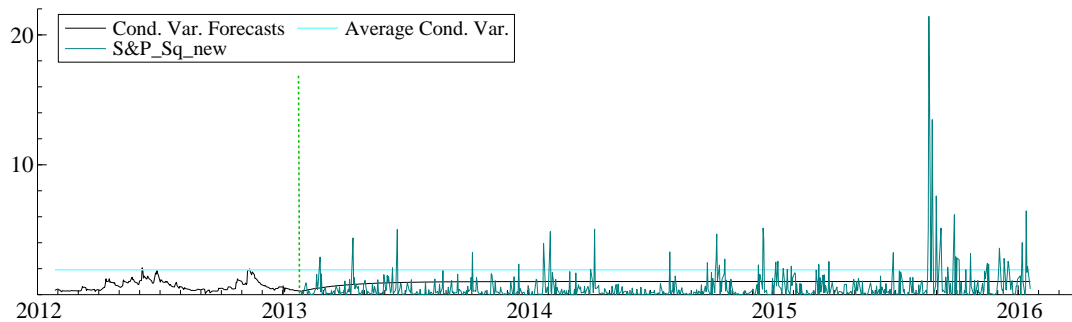
S&P 500

GARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	2,92955	2,930496	2,914753	2,91499	2,919872	2,920637	2,916327	2,916567
Schwarz (BIC)	2,941886	2,945915	2,930173	2,933493	2,935292	2,939141	2,93483	2,938155
Mean Absolute Error(MAE)	1,145	1,145	1,118	1,115	1,191	1,188	1,119	1,117
Logarithmic Loss Function(LL)	11,64	11,64	11,51	11,5	11,84	11,83	11,51	11,5

EGARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	2,874584	2,875693	2,875273	2,876378	2,871115	2,872196	2,872792	2,873869
Schwarz (BIC)	2,893088	2,897281	2,896861	2,90105	2,892703	2,896868	2,897463	2,901624
Mean Absolute Error(MAE)	0,9477	0,9498	0,9412	0,9433	0,8776	0,876	0,8731	0,8719
Logarithmic Loss Function(LL)	10,59	10,6	10,55	10,57	10,15	10,14	10,12	10,11

GJR(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	2,871007	2,872117	2,861592	2,862392	2,872025	2,873132	2,863039	2,863841
Schwarz (BIC)	2,886427	2,890621	2,883179	2,887063	2,890529	2,89472	2,887711	2,891597
Mean Absolute Error(MAE)	0,9143	0,9107	0,8801	0,8673	0,9059	0,9027	0,8837	0,8707
Logarithmic Loss Function(LL)	10,4	10,38	10,18	10,1	10,35	10,33	10,21	10,12

Figur 4.11: Evaluering av S&P 500



Figur 4.12: Predikering av volatiliteten til S&P 500 med GJR 1.2(1)

Tabell 4.12: Koeffisientene, standardfeilene og t-verdiene til S&P 500 GJR 1.2(1)

	Koeffisient	Std.feil	t-verdi	t-sanns.
Cst(M)	0,024845	0,020665	1,202	0,2294
AR(1)	-0,016261	0,025027	-0,6497	0,5159
Cst(V)	0,025243	0,0057626	4,38	0,0000
ARCH(Alpha1)	-0,075991	0,0071228	-10,67	0,0000
ARCH(Alpha2)	0,050381	0,014397	3,499	0,0005
GARCH(Beta1)	0,891988	0,018255	48,86	0,0000
GJR(Gamma1)	0,093669	0,030635	3,058	0,0023
GJR(Gamma2)	0,123832	0,041052	3,016	0,0026

Tabell 4.13: Ljung-Box test av de standardiserte residualene

Q(5) =	5.49371	[0.3586360]
Q(10) =	16.7964	[0.0789918]
Q(20) =	25.3278	[0.1891758]
Q(50) =	59.2308	[0.1742600]

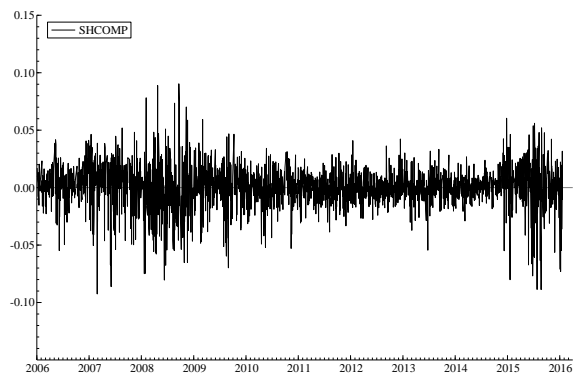
Ser vi nå på autokorrelasjonen til de standardiserte residualene til GJR 1.2(1) som er gitt ved u_t/σ_t^2 , ser vi at den er fjernet og at modellen godt fanger opp autokorrelasjonen som var i restleddet. I figuren under ser vi at vi beholder nullhypotesen for alle de fire testene.

4.5 SHCOMP

Shanghai Composite Index (SHCOMP) er en aksjeindeks som består av alle aksjene som er notert på Kinas største børs, Shanghai-børsen. Datasettet består av 2447 observasjoner fra 4. januar 2006 til 20. januar 2016. Predikeringsperioden går fra 25. desember 2012 til 20. januar 2016.



Kursutvikling SHCOMP



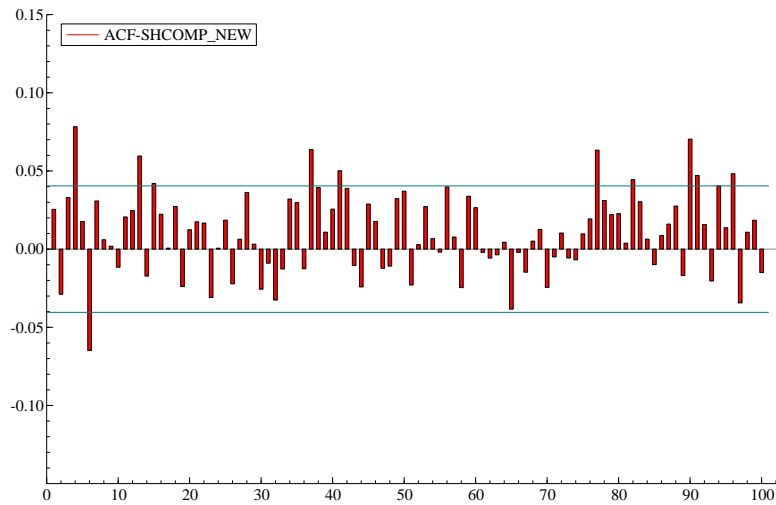
Avkastning SHCOMP

Deskriptiv statistikk

Gjennomsnitt	0,00038475
Standardavvik	0,018108
Skjevhet	-0,58343
Kurtose	3,5851
Maksimum	-0,092561
Minimum	0,090345

I tabellen over kan vi se den deskriptive statistikken for avkastningene til SHCOMP. Vi ser her at kurtose og skjevhet er henholdsvis 3,5851 og -0,58343. Dette innebærer at fordelingen har større haler enn normalfordelingen (kurtose=3), og at dataen er venstreskjev. Den deskriptive statistikken er her basert på alle observasjonene, både for estimerings- og predikeringsperioden.

Autokorrelasjon



Figur 4.14: Autokorrelasjon av avkastningene

Ser vi på figuren for autokorrelasjonen over ser vi at vi kan ha et problem med autokorrelasjon i dataen. Figuren viser autokorrelasjonen fra lag 1-100, og et konfidensintervall på 95%.

Tabell 4.14: Ljung-Box test av avkastningene

Q(5) =	9.77662	[0.0818179]
Q(10) =	18.1223	[0.0529352]
Q(20) =	42.4889	[0.0023870]*
Q(50) =	99.5615	[0.0000389]*

Tabellen over viser en Ljung-Box test for autokorrelasjon. Vi ser her som forventet at vi har et problem med autokorrelasjon i ledd 20 og 30. Vi forkaster her nullhypotesen om ingen autokorrelasjon i dataen.

Engle's ARCH-test

Tabell 4.15: ARCH-test av avkastningene

ARCH 1-2 test:	$F(2,2442) = 55.062$	$[0.0000]**$
ARCH 1-5 test:	$F(5,2436) = 42.093$	$[0.0000]**$
ARCH 1-10 test:	$F(10,2426) = 25.679$	$[0.0000]**$

For å se om restleddet er hetroskedastisk, anvender vi en ARCH-test. Nullhypotesen her er at ARCH-effekter ikke er tilstedeværende. Vi ser i tabellen over at nullhypotesen forkastes for alle testene. Vi har her ARCH-effekter i dataen og en ikke-lineær modell må anvendes for å fange opp disse effektene.

Estimering og evaluering

Den standard GARCH modellen med best resultat er GARCH 2.1(0) med en (MAE) på 3,941 og en (LL) på 11,34. Det er også denne modellen som gir best resultat totalt.

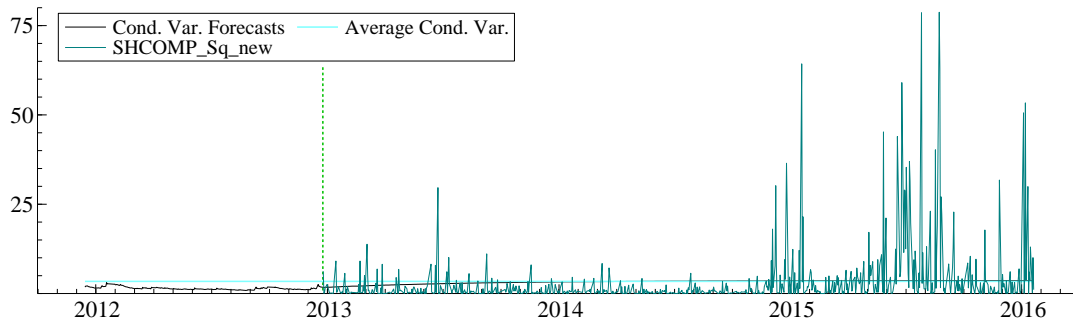
SHCOMP

GARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,842406	3,843572	3,843285	3,844449	3,843533	3,844686	3,843992	3,84512
Schwarz (BIC)	3,855221	3,85959	3,859303	3,863671	3,859551	3,863908	3,863214	3,867545
Mean Absolute Error(MAE)	3,945	3,945	3,951	3,951	3,941	3,942	3,963	3,963
Logarithmic Loss Function(LL)	11,35	11,35	11,36	11,36	11,34	11,34	11,39	11,39

EGARCH(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,840473	3,841652	3,839024	3,840177	3,84165	3,842828	3,840078	3,841232
Schwarz (BIC)	3,859695	3,864077	3,86145	3,865806	3,864076	3,868458	3,865707	3,870065
Mean Absolute Error(MAE)	4,958	4,956	5,024	5,028	4,957	4,957	5,024	5,029
Logarithmic Loss Function(LL)	13,06	13,06	13,15	13,15	13,06	13,06	13,15	13,16

GJR(p,q)	1.1 (0)	1.1 (1)	1.2 (0)	1.2 (1)	2.1 (0)	2.1 (1)	2.2 (0)	2.2 (1)
Akaike (AIC)	3,843584	3,84475	3,844597	3,845692	3,844707	3,845861	3,840291	3,84138
Schwarz (BIC)	3,859602	3,863972	3,867023	3,871321	3,863929	3,868287	3,86592	3,870213
Mean Absolute Error(MAE)	3,949	3,95	3,986	3,986	3,951	3,951	3,99	3,988
Logarithmic Loss Function(LL)	11,36	11,36	11,44	11,44	11,36	11,36	11,45	11,44

Figur 4.15: Evaluering av SHCOMP



Figur 4.16: Predikering av volatiliteten til SHCOMP med GARCH 2.1(0)

Tabell 4.16: Koeffisientene, standardfeilene og t-verdiene til SHCOMP GARCH 2.1(0)

	Koeffisient	Std.feil	t-verdi	t-sanns.
Cst(M)	0,027789	0,039891	0,6966	0,4861
Cst(V)	0,031979	0,022275	1,436	0,1513
ARCH(Alpha 1)	0,071313	0,037688	1,892	0,0586
GARCH(Beta 1)	0,595371	0,69928	0,8514	0,3947
GARCH(Beta 2)	0,324529	0,66166	0,4905	0,6239

Tabell 4.17: Ljung-Box test av de standardiserte residualene

Q(5) =	11.8240	[0.0372802]*
Q(10) =	23.4860	[0.0090882]*
Q(20) =	38.5530	[0.0075740]*
Q(50) =	74.6927	[0.0133848]*

Ser vi nå på autokorrelasjonen til de standardiserte residualene til GARCH 2.1(0) som er gitt ved u_t/σ_t^2 , ser vi at vi ikke har klart å fjerne autokorrelasjonen i dataen gitt den beste modellen innenfor vårt rammeverk. Selv hvis vi øker koeffisientene til modeller som GARCH 5.5(1) og GARCH 10.10(1), ser vi at vi fortsatt ikke klarer å fjerne autokorrelasjonen i de standardiserte residualene. Mye av grunnen for dette er den ekstremt volatile perioden fra 2007 til 2010.

4.6 Sannsynlighetsfordelinger

Ved estimering av de ulike GARCH-modellene i denne oppgaven har vi brukt standard normalfordeling. Vi ønsker nå å undersøke om bruk av andre sannsynlighetsfordelinger vil gi bedre predikeringsresultat i form av lavere verdi på feilmålene. Vi er altså interessert i å se om andre sannsynlighetsfordelinger er i stand til å beskrive dataene bedre og på denne måten gi en predikering med større presisjon.

Sannsynlighetsfordelingene vi har valgt å se på i tillegg til standard normalfordeling er; student-t- og skjev student-t fordeling. For hver av indeksene har vi valgt å estimere den beste modellen og sammenligne effekten av de forskjellige fordelingene. I figuren under angir grønn farge beste modell, mens rød farge angir dårligste modell

For DAX indeksen hvor EGARCH 2.2(1) er beste modell ser vi i figuren under at skjev student-t fordeling er å foretrekke ut ifra begge predikeringsmålene. Noe som betyr at ved bruk av student-t fordeling istedenfor standard normalfordeling vil vi få bedre predikeringsresultater.

For OSEBX indeksen hvor EGARCH 2.1(0) er beste modell, ser vi at vi får samme resultat som for DAX. Skjev student-t fordeling er foretrukket og gir best resultat ifølge alle feilmålene.

For S&P 500 indeksen er det GJR 1.2(1) som er den beste modellen. Her gir student-t fordelingen best resultat ifølge begge predikeringsmålene.

For SHCOMP er GARCH 2.1(0) den beste modellen. Dette er den eneste modellen hvor standard normalfordeling gir best resultat i predikeringsperioden, og det til tross for at skjev student-t er best i estimeringsperioden.

DAX	Normal	Student-t	Skewed student-t
EGARCH(p,q)	2.2 (1)	2.2 (1)	2.2 (1)
Akaike (AIC)	3,283365	3,266817	3,258711
Schwarz (BIC)	3,310682	3,297169	3,292097
Mean Absolute Error(MAE)	1,679	1,477	1,35
Logarithmic Loss Function(LL)	9,024	7,943	6,655

OSEBX	Normal	Student-t	Skewed student-t
EGARCH(p,q)	2.1 (0)	2.1 (0)	2.1 (0)
Akaike (AIC)	3,633015	3,643987	3,6278
Schwarz (BIC)	3,654348	3,668368	3,655227
Mean Absolute Error(MAE)	1,204	1,602	0,8628
Logarithmic Loss Function(LL)	9,397	11,05	6,063

S&P 500	Normal	Student-t	Skewed student-t
GJR(p,q)	1.2 (1)	1.2 (1)	1.2 (1)
Akaike (AIC)	2,862392	2,832112	2,820095
Schwarz (BIC)	2,887063	2,859867	2,850934
Mean Absolute Error(MAE)	0,8673	0,8353	4,983
Logarithmic Loss Function(LL)	10,1	9,865	19,47

SHCOMP	Normal	Student-t	Skewed student-t
GARCH(p,q)	2.1 (0)	2.1 (0)	2.1 (0)
Akaike (AIC)	3,843533	3,790993	3,784618
Schwarz (BIC)	3,859551	3,810215	3,807043
Mean Absolute Error(MAE)	3,941	4,996	4,755
Logarithmic Loss Function(LL)	11,34	13,1	12,77

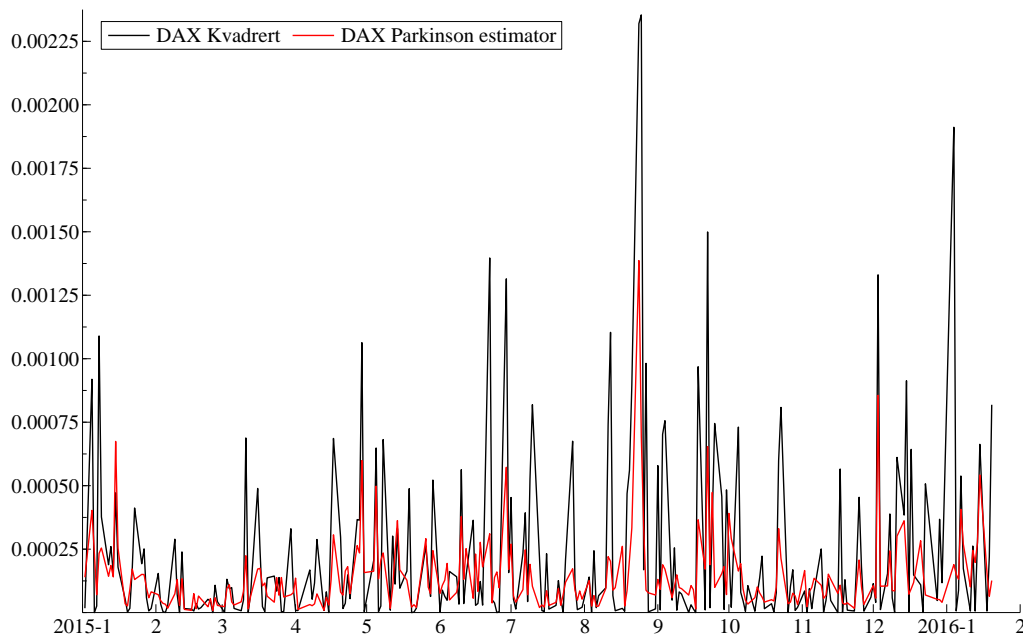
Figur 4.17: Evaluering av sannsynlighetsfordelingene

Konklusjonen er at for alle indeksene gir skjev student-t fordeling best resultat for estimeringsperioden. Dette kan vi tolke som at denne fordelingen er i stand til å fange opp variasjonene i avkastningsdataene bedre enn noen av de andre fordelingene. Med hensyn til indeksenes deskriptive statistikk gir disse resultatene samsvar. Alle indeksene har en kurtose høyere enn tre og en skjevhet forskjellig fra null. Dette innebærer at andre fordelinger enn normalfordelingen er i stand til å beskrive dataene bedre, noe som vi ser gir utslag i mer presise predikeringer.

4.7 Parkinsonestimatoren

Ved estimering av GARCH-modellene i denne oppgaven har vi benyttet oss av kvadrert daglig avkastning som volatilitetsproxy. Denne proxyen har derimot en del støy som fører til tilsynelatende upresise predikeringer. Ved å benytte en volatilitetsproxy med mindre støy vil man i teorien få resultater med lavere predikeringsfeil, noe som vil gi et bilde av en mer presis predikering. Vi ønsker derfor å se på predikeringsprestasjonen til modellene ved bruk av Parkinsonestimatoren som volatilitetsproxy sammenlignet med kvadrert daglig avkastning.

Som vi ser av figuren under, innehar Parkinsonestimatoren ikke like mye støy som kvadrert daglig avkastning. Den svinger ikke like mye og har dermed en lavere varians. Ved bruk av Parkinsonestimatoren får vi en proxy som er mer lik den volatiliteten vi ville fått dersom vi hadde brukt intradaglig data, se Ebens et al. (1999)



For å begrense datamengden har vi valgt å sammenligne de to proxyene ved å kun se på DAX-indeksen. Under estimeringene gjort tidligere i oppgaven har vi multiplisert avkastningen med 100 for å ”oppskalere” resultatet, slik at vi har fått penere resultater i form av høyere tallverdi på feilmålene. Dette er ikke gjort her siden det da ikke ville vært mulig å sammenligne de to proxyene. Avkastningen er nå gitt ved:

$$r_t = (\log(p_t) - \log(p_{t-1}))$$

DAX - GARCH

		1,1 (0)	1,1 (1)	1,2 (0)	1,2 (1)	2,1 (0)	2,1 (1)	2,2 (0)	2,2 (1)
Parkinson	GARCH(p,q)								
	Mean Absolute Error(MAE)	0,00015669	0,00015685	0,00016314	0,00016324	0,00015804	0,00015791	0,00016217	0,00016216
	Logarithmic Loss Function(LL)	2,971	2,973	3,077	3,079	2,997	2,995	3,062	3,062
Kvadrert	GARCH(p,q)								
	Mean Absolute Error(MAE)	0,00020488	0,00020499	0,00020985	0,00020993	0,00020598	0,00020588	0,00020911	0,0002091
	Logarithmic Loss Function(LL)	10,283	10,286	10,441	10,443	10,321	10,318	10,418	10,418

DAX - EGARCH

		1,1 (0)	1,1 (1)	1,2 (0)	1,2 (1)	2,1 (0)	2,1 (1)	2,2 (0)	2,2 (1)
Parkinson	EGARCH(p,q)								
	Mean Absolute Error(MAE)	0,00012097	0,00012053	0,0001206	0,0001204	0,00010601	0,00010526	0,00010828	0,00010323
	Logarithmic Loss Function(LL)	2,429	2,422	2,423	2,419	2,152	2,138	2,194	2,099
Kvadrert	EGARCH(p,q)								
	Mean Absolute Error(MAE)	0,00017973	0,00017941	0,00017947	0,00017932	0,00016869	0,00016814	0,00017037	0,00016669
	Logarithmic Loss Function(LL)	9,447	9,435	9,437	9,432	9,001	8,977	9,072	8,915

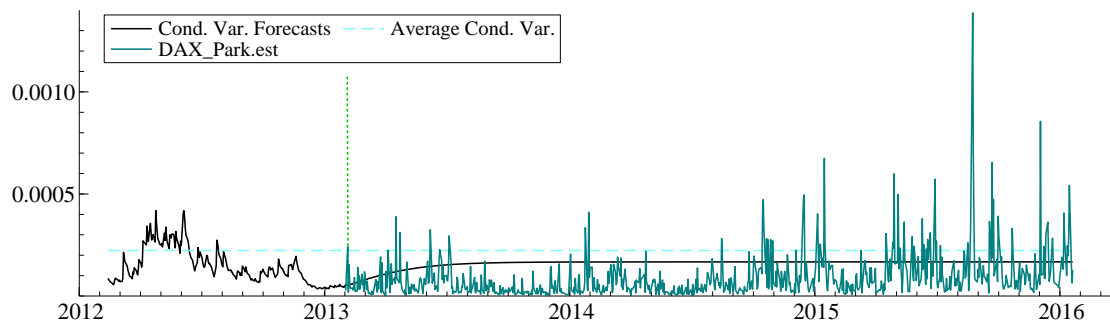
DAX - GJR

		1,1 (0)	1,1 (1)	1,2 (0)	1,2 (1)	2,1 (0)	2,1 (1)	2,2 (0)	2,2 (1)
Parkinson	GJR(p,q)								
	Mean Absolute Error(MAE)	0,00011303	0,0001135	0,00011063	0,00011098	0,0001123	0,00011284	0,00010557	0,00010436
	Logarithmic Loss Function(LL)	2,281	2,290	2,240	2,246	2,268	2,278	2,127	2,105
Kvadrert	GJR(p,q)								
	Mean Absolute Error(MAE)	0,00017377	0,00017411	0,00017207	0,00017232	0,00017324	0,00017363	0,00016795	0,00016708
	Logarithmic Loss Function(LL)	9,213	9,226	9,145	9,155	9,191	9,207	8,963	8,926

Figur 4.18: Parkinsonestimatorens sammenlignet med kvadrerte daglig avkastning

Vi ser av sammenligningsresultatet at bruk av Parkinsonestimatorens som proxy ikke gir noen endring i hvilken modell som er foretrukket. Vi ser derimot at denne proxyen gir lavere verdi på alle feilmålene, noe som gir et bilde av en mer presis predikering. En proxy med lavere støy resulterer derfor i en mer nøyaktig predikering.

Figuren under viser den GARCH-modellen med best predikeringsresultat for DAX-indeksen. EGARCH 2.2(1).



Figur 4.19: Predikering av volatiliteten til DAX med EGARCH 2.2(1)

Kapittel 5

Oppsummering

5.1 Konklusjon

Formålet med denne oppgaven var å sammenligne ulike volatilitetsmodeller ved og se på hvordan de egner seg til å predikere volatilitet. Sammenligner man alle aksjeindeksene etter hvor godt de predikerer, så er det ingen modell som presterer best i alle tilfellene. Det er derimot en modell som skiller seg ut som bedre enn resten, og det er EGARCH. Denne modellen presterer best med tanke på predikeringsprestasjonen for indeksene OSEBX og DAX, og nest best for S&P 500. Noe ødeleggende for konklusjonen er at resultatene til SHCOMP avviker betydelig fra resultatet gitt de tre aksjeindeksene OSEBX, DAX og S&P500. EGARCH presterer nest dårligst av GARCH-modellene når det kommer til SHCOMP. En av grunnene til dette er at SHCOMP-indeksen er den klart mest volatile indeksen, noe som fører til at modell-parameterne ikke er signifikante.

Ved estimering av GARCH-modellene i oppgaven har vi benyttet standard normalfordeling som sannsynlighetsfordeling og kvadrert daglig avkastning som volatilitetsproxy. Vi ønsket så å se om bruk av andre sannsynlighetsfordelinger eller en mer egnet volatilitetsproxy gav utslag i mer presise predikeringsresultater. Konklusjonen er at andre fordelinger, og da i hovedsak skjev student-t fordeling, er i stand til å beskrive dataene bedre og på denne måten gi mer presise predikeringer. Videre så vi at Parkinsonestimatoren viste seg å være en bedre egnet volatilitetsproxy, i det den gav lavere verdi på feilmålene og dermed et mer nøyaktig resultat.

5.2 Forslag til videre arbeid

Denne oppgaven har kun tatt for seg en bestemt predikeringshorisont på $h=750$ dager. Andre måter å angripe en slik oppgave på er å se på ulike predikeringshorisonter, og da fortrinnsvis en kortere tidshorisont som vil være mer hensiktsmessig med tanke på større presisjon.

Bibliografi

- Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.
- Torben G Andersen and Tim Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2):115–158, 1997.
- Chris Brooks. *Introductory econometrics for finance*. Cambridge university press, 2014.
- Torben G Andersen and Tim Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905, 1998.
- Ser-Huang Poon. *A practical guide to forecasting financial market volatility*. John Wiley & Sons, 2005.
- Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Heiko Ebens. The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76, 2001.
- F Black. Studies of stock market changes, proceedings of the american statistical association. In *Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section (American Statistical Association, Alexandria, VA)*, 1976.
- Eugene F Fama. Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617, 1991.
- Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.
- Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.
- Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.
- John C Hull. *Options, futures, and other derivatives*. Pearson Education India, 2006.
- Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

Benoit B Mandelbrot. *The variation of certain speculative prices*. Springer, 1997.

Leonardo Souza, Alvaro Veiga, and Marcelo C Medeiros. *Evaluating the Forecasting Performance of GARCH Models Using White's Reality Check*. Pontifícia Universidade Católica de Rio de Janeiro, Departamento de Economía, 2002.

Heiko Ebens et al. Realized stock volatility. *Department of Economics, Johns Hopkins University*, 1999.

Tillegg A

Appendiks

A.1 Deskriptiv statistikk

Her vil det bli gitt en kort introduksjon av den deskriptive statistikken som er bukt.

Forventningsverdi

$$E(r_t) = \frac{1}{T} \sum_{t=1}^T r_t$$

Varians

$$\hat{\sigma}_r = \frac{1}{T-1} \sum_{t=1}^T (r_t - E(r_t))^2$$

Standardavvik

$$\hat{\sigma}_r = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - E(r_t))^2}$$

Kovarians

$$\hat{\sigma}_{r,x} = \frac{1}{T-1} \sum_{t=1}^T (r_t - E(r_t))(x_t - E(x_t))$$

Skjevhet

$$\hat{S}_r = \frac{1}{T-1} \sum_{t=1}^T \frac{(r_t - E(r_t))^3}{\hat{\sigma}^3}$$

Verdien på skjevheten kan ta en verdi mellom -1 til 1. Hvis man har en positiv skjevhet, dvs. $\hat{S}_r > 0$ så vil man ha flere verdier som er høyere enn gjennomsnittet enn de som er under. Normalfordelingen har ikke skjevhet, altså $\hat{S}_r = 0$.

Kurtose

$$\hat{K}_r = \frac{1}{T-1} \sum_{t=1}^T \frac{(r_t - E(r_t))^4}{\hat{\sigma}^4}$$

Verdien på kurtosen sier noe om fordelingen har større haler, såkalt leptokurtose. Normalfordelingen har en kurtose på 3, så hvis dataen har en kurtose som er høyere enn dette, vil det si at fordelingen til dataen har fetere haler enn normalfordelingen.