



Stockholms
universitet

En jämförelse av prediktionsförmågan hos GARCH-modeller med olika antal parametrar

Anna Francesconi

Kandidatuppsats 2016:22
Matematisk statistik
Juni 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

En jämförelse av prediktionsförmågan hos GARCH-modeller med olika antal parametrar

Anna Francesconi*

Juni 2016

Sammanfattning

I denna studie jämfördes två GARCH-modeller med olika antal parametrar, med avseende på hur väl de båda modellerna predikterar volatiliteten hos en vald datamängd sedd som en tidsserie. Modellerna i fråga var GARCH(1,1) och GARCH(2,3), med 3 respektive 6 parametrar som skattades med Maximum-likelihood-metoden. Jämförelsen utfördes genom att göra två backtest (ett för varje modell) där Value at Risk skattades över samma datamängd och sedan jämfördes i hur många fall förlustfunktionen överskred denna kvantil. Det visade sig att båda modellerna fungerade lika bra för att prediktera volatiliteten och därför kan vi sluta oss till att GARCH(1,1) är den bättre modellen pga. enkelheten i att ha färre parametrar.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: anna.francesconi@hotmail.com. Handledare: Filip Lindskog, Joanna Tyrcha och Mathias Lindholm.

Abstract

In this thesis two GARCH-models, with different number of parameters, were compared regarding how well they predict the volatility for a chosen dataset seen as a time series. The models were GARCH(1,1) and GARCH(2,3), with 3 and 6 parameters respectively, which were estimated with the method of Maximum-likelihood. The comparison was done by performing two backtests (one for each model) where Value at Risk were estimated over the same dataset and then checked for how many cases the loss-function violated this quantile. As it turned out, both models seemed to work just as well to predict the volatility and therefore we can draw the conclusion that GARCH(1,1) is the better model of the two because of the simplicity in having less parameters.

Förord

Jag skulle vilja rikta ett tack till mina handledare Mathias Lindholm, Filip Lindskog och Joanna Tyrcha för värdefull feedback och rådgivning under arbetets gång.

Sedan vill jag också tacka Erik, för att du är min klippa.

Innehållsförteckning

1	Inledning	4
2	Teori	6
2.1	Log-avkastningar	6
2.2	Vad menas med en tidsserie?	6
2.3	Volatilitet	7
2.4	ARCH/GARCH-processer	8
2.4.1	ARCH(p)	8
2.4.2	GARCH(p,q)	8
2.5	Maximum-likelihood-metoden	9
2.6	Value at risk (VaR)	9
2.7	Statistiska test	10
2.7.1	Autokorrelations-test	10
2.7.2	Ljung-Box-test	10
2.7.3	T-test för ett stickprov	11
2.7.4	Backtest	11
2.7.5	Tvåsidigt Binomialtest	12
3	Data	13
3.1	Apple Inc.	13
3.2	Stickprovstorlek och tidsintervall	13
3.3	Dataanalys	14
4	Metod och resultat	16
4.1	Deskriptiv analys	16
4.2	Ordning	18
4.3	Parameterskattning och validering	19
4.4	Backtest	22
5	Diskussion och slutsats	24
6	Appendix	26
6.1	Sannolikhetsfördelningar	26
6.1.1	Bernoullifördelningen	26
6.1.2	Binomialfördelningen	26
6.1.3	Normalfördelningen	26
6.1.4	t-fördelningen	26
6.2	Skevhets och toppighet	27
6.3	Akaiikes informationskriterium	27

6.4	Beräkning av Maximum-likelihood-skattningar	27
7	Källhänvisningar	29

1. Inledning

Alla ekonomiska storheter varierar med tiden. När man mäter dessa vid diskreta tidpunkter, till exempel slutkursen varje börsdag för en viss aktie under ett år, brukar denna följd av mätvärden kallas för en *tidsserie*. Eftersom mätvärdena är upprepade observationer av en slumpvariabel (motsvarande den ekonomiska storheten) så kommer dessa observationer att vara beroende av varandra och detta beroende kan man t.ex. utnyttja för att försöka prediktera framtida värden.

På finansmarknaden är det viktigt att kunna mäta ”risken” att göra en förlust och man pratar ofta om volatilitet som ett mått på risknivån. Volatiliteten brukar definieras som den betingade standardavvikelsen av tidsserien (som utgörs av den underliggande tillgångens avkastning).

Volatilitet är inte konstant över tid och därför existerar det så kallade volatilitet-skluster (dvs. volatiliteten kan vara stor för vissa tidsperioder och liten i andra). Därför har modeller som t.ex. ARCH och GARCH (som är förkortningar för den engelska betydelsen av ”autoregressiv betingad heteroskedastisk” respektive ”generaliserad ARCH”) utvecklats för att ta hänsyn till föränderlig volatilitet. I vanliga fall brukar man låta tidsserier ha en konstant varians, medan ARCH-processen, introducerad av Engle (1982) (som för övrigt tilldelades nobelpriset i ekonomi för denna modell), tillåter den betingade variansen att förändras över tid som en funktion av tidigare feltermen. Även om ARCH-modellen i sig är enkel krävs det ofta många parametrar för att adekvat beskriva volatilitetsprocessen av en tillgångs avkastning. Utvidgningen (eller generaliseringen) av ARCH-processen kallas GARCH-process, introducerad av Bollerslev (1986), och tillåter även den betingade volatiliteten att förändras över tid som en funktion av tidigare värden på volatiliteten.

Syftet med detta kandidatarbete var att jämföra GARCH-modeller med olika antal parametrar, med avseende på förmågan att prediktera volatiliteten. Eftersom det finns oändligt många kombinationer av parameterantal, måste vi begränsa oss. För att det inte skulle bli en alltför omfattande uppgift i förhållande till den korta tid som avsatts till arbetet begränsar vi oss till att jämföra endast två GARCH-modeller, vilka har olika antal parametrar, och deras förmåga att prediktera den framtida volatiliteten för Apple Inc.:s aktieutveckling. GARCH-modellernas parameterantal beror på vilken ordning dom har, och för att bestämma oss för vilka två modeller vi ska jämföra, så väljer vi dels GARCH(1,1) som den ena pga. dess utbredda användning och dels GARCH(2,3) eftersom denna modell uppvisar det lägsta AIC-värdet bland alla modeller som fanns med i en urvalsgrupp av GARCH-modeller av högre ordningar (i vilken GARCH(1,1) ej

ingick).

I avsnitt 2 går vi igenom teorin för olika begrepp och metoder som vi senare använder och refererar till i vår analys. I avsnitt 3 har vi en kortfattad presentation av Apple Inc. och av hur våra data ser ut och har för egenskaper. I avsnitt 4 går vi igenom metod och resultat av deskriptiv analys, valet av GARCH(2,3), skattningen av parametrarna i de båda GARCH-modellerna och prediktionsanalysen. Till sist i Avsnitt 5 spekulerar vi om felkällor, förbättringar och sedan för vi en diskussion om resultaten och drar en slutsats.

2. Teori

Vi kommer att använda oss av den justerade aktiekursen, vid hämtning av data, eftersom denna har justerats för eventuella utdelningar och splittar (dvs. då antalet andelar delas upp i flera för att sänka priset per aktie).

2.1 Log-avkastningar

Innan vi börjar med själva analysen gör vi om de justerade aktiepriserna till avkastningar och logaritmerar dessa. En fördel med att göra om dem till avkastningar är att sådana är skal-fria, vilket betyder att de inte beror på någon enhet (som t.ex. SEK, dollar). Avkastningen från en aktie (eller annan tillgång) beror på förändringen av aktie-priset P_t .

För en given tidshorisont (i vårt fall en dag), så ges förlusten (eller vinsten) beroende på om vi har en positiv eller negativ differens) av

$$L = P_t - P_{t-1}.$$

Medan L antas vara observerbar vid tiden t , så är den typiskt slumpmässig från och med tiden $t - 1$. För att få den aritmetiska avkastningen dividerar vi L med priset från föregående dag P_{t-1} .

Då avkastningar beror av tid är det dock enklare att använda log-avkastningar, eftersom dessa kan summeras över tid till skillnad från den aritmetiska avkastningen. Log-avkastningen r_t är den naturliga logaritmen av priset P_t dividerat med det ursprungliga priset P_{t-1} , dvs.

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

där $t = 0, 1, 2, \dots$ står för tiden (Ruppert 2004, s. 75-76).

2.2 Vad menas med en tidsserie?

Här kommer vi att följa kapitel 2 i Tsays bok (2005). Om vi behandlar vår log-avkastning r_t som en samling slumpvariabler över tid, så kan vi se dessa som en tidsserie $\{r_t\}$. Vi har alltså en enkel tidsserie-modell

$$r_t = \mu_t + a_t$$

där $\mu_t = E(r_t|F_{t-1})$ (och F_{t-1} består av alla tidigare och nuvarande värden på r_t) och a_t är feltermen (eller residualtermen). Väntevärdet μ_t kan följa t.ex. en ARMA-process eller så kan det vara en konstant.

Givet att tidsseriens första och andra moment existerar, definierar vi väntevärdessfunktionen μ_t (första momentet) och autokovariansfunktionen $\gamma(t, s)$ (andra momentet) med

$$\begin{aligned}\mu_t &= E(r_t) \\ \gamma(t, s) &= E((r_t - \mu_t)(r_s - \mu_s))\end{aligned}$$

där t och s är olika tidpunkter. Det följer att autokovariansfunktionen satisfierar $\gamma(t, s) = \gamma(s, t)$ för alla t, s och $\gamma(t, t) = Var(r_t)$.

En hel del trevliga resultat som håller för oberoende slumpvariabler håller även för tidsserier som är *stationära*. Tidsserien r_t sägs vara strikt stationär om den sammansatta fördelningen för avkastningarna inte förändras över tid. Detta villkor är dock svårt att verifiera empiriskt, därför antar vi istället att vår tidsserie är svagt stationär, dvs. att både väntevärdet av r_t och kovariansen mellan r_t och r_{t-l} är konstanta och förändras ej över tid. Eller mer specifikt, $E(r_t) = \mu$, vilket är en konstant, och $Cov(r_t, r_{t-l}) = \gamma_l$, som alltså bara beror på den tidsmässiga separationen l mellan r_t och r_{t-l} (vilket kallas *lag*, vilket är den engelska termen för tidsförskjutningen).

När man är intresserad av hur starkt det lineära beroendet (korrelationen) mellan r_t och dess tidigare värden r_{t-i} är, generaliseras konceptet till autokorrelation, som brukar betecknas ρ_l , vilken under vårt antagande om svag stationaritet är en funktion av enbart l och definieras som

$$\rho_l = \gamma_l / \gamma_0$$

där γ_0 alltså är detsamma som variansen av r_t .

Stationära tidsserier utan autokorrelation kallas för *vitt brus* och är en sekvens av oberoende och identiskt fördelade slumpvariabler med ändligt väntevärde och varians. För en vitt brus serie är alla ρ_l ekvivalenta med noll, men i praktiken räcker det att ρ_l är nära noll för att man ska kunna anta att det är en vitt brus serie.

2.3 Volatilitet

I inledningen nämnde vi volatilitet som ett mått på risken att göra en förlust. Volatilitet är ett riskmått i den mening att det mäter spridningen hos avkastningarna för aktien (eller annan typ av finansiell tillgång). En högre volatilitet innebär att värdet som priset på aktien antar ligger i ett större intervall av värden. Detta betyder alltså att priset på aktien kan ändras dramatiskt i någon riktning på kort sikt, dvs. vi har en större "risk". En lägre volatilitet innebär därmed mindre svängningar och att värdet på aktien istället förändras mer långsamt.

Volatilitet modelleras oftast som den betingade standardavvikelsen av aktiens avkastning. En speciell egenskap hos just aktie-volatilitet är att den inte är direkt observerbar eftersom det bara finns en observation per dag. Men även om volatiliteten inte är direkt observerbar, så har den en egenskap som kan ses grafiskt, nämligen volatilitetskluster som vi tidigare har nämnt. Svängningarna i volatiliteten sker inte nödvändigtvis vid specifika tidpunkter, istället kan det vara så att tidpunkterna, vid vilka svängningarna inträffar, själva är stokastiska (och det är då vi har användning av exempelvis en GARCH-modell).

2.4 ARCH/GARCH-processer

Volatilitetskluster-beteendet i data är något som vanliga lineära tidsseriemodeller brukar ha svårt att fånga upp. Om det i analysen kommer fram att denna tendens verkar finnas hos data, så är det lämpligt att använda icke-lineära tidsseriemodeller istället (såsom t.ex. ARCH/GARCH).

2.4.1 ARCH(p)

Den grundläggande idén med ARCH-modeller är att feltermen a_t för en akties avkastning är seriellt okorrelerad, men ändå beroende, och att beroendet av a_t kan beskrivas som en enkel kvadratisk funktion av dess egna tidsförskjutna värden.

Låt $a_t = r_t - \mu_t$ vara feltermen vid tiden t för vår tidsserie. Då följer a_t en ARCH(p)-modell om

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2,$$

där $\{\epsilon_t\}$ är vitt brus med väntevärde 0 och variansen 1, $\alpha_0 > 0$, och $\alpha_i \geq 0$ för $i = 1, \dots, p$ (där p är ett positivt heltal). I praktiken antas ϵ_t ofta följa standardnormalfördelningen eller t-fördelningen (Tsay 2005, s. 103).

Från modellens struktur kan man se att om de föregående kvadrerade feltermerna $\{a_{t-i}^2\}_{i=1}^p$ är stora, så implicerar det en stor betingad varians σ_t^2 för a_t . Det vill säga, sannolikheten att även a_t blir stort är större och modellen fångar alltså upp eventuellt volatilitetskluster-beteende hos data.

2.4.2 GARCH(p,q)

GARCH-processer är ”generaliserade” ARCH-processer i den meningen att den kvadrerade volatiliteten σ_t^2 tillåts att bero på tidigare kvadrerade volatiliteter och tidigare kvadrerade värden på processen. Alltså, istället för att specificera vad variansen kommer att vara vid varje specifik tidpunkt, kan vi istället modellera variansen.

Låt på samma sätt som för ARCH ovan $a_t = r_t - \mu_t$ vara feltermen vid tiden t för vår tidsserie. Då följer a_t en GARCH(p,q)-modell om

$$a_t = \sigma_t \epsilon_t$$

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

där återigen $\{\epsilon_t\}$ är vitt brus med väntevärde 0 och variansen 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ och $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. Och precis som tidigare, så antas ϵ_t ofta följa standard-normalfördelningen eller t-fördelningen (Tsay 2005, s. 114).

Ett stort värde på a_{t-1}^2 eller σ_{t-1}^2 ger upphov till ett stort σ_t^2 . Det betyder att ett stort a_{t-1}^2 tenderar att följas av ett stort a_t och fångar därmed upp volatilitetskluster-beteendet i data.

Om $q = 0$ reduceras modellen till en ARCH(p)-modell. Parametrarna α_i och β_j refereras ofta till som ARCH- respektive GARCH-parametrar.

2.5 Maximum-likelihood-metoden

När vi vet ordningen (p,q) på en GARCH-modell behöver vi göra en skattning av de olika α - och β -parametrarna som den innehåller. Det vanligaste är att använda Maximum-likelihood-metoden, som går ut på att hitta de bästa (effektivaste) skattningarna av okända parametrar i en sannolikhetsfördelning som beskriver en datamängd (ett stickprov).

I vår definition av GARCH-modellen tidigare så nämnde vi att ϵ_t ofta antas följa normal- eller t-fördelningen. Låter vi den följa t.ex. normalfördelningen så har vi att $\epsilon_t \sim N(\mu_t, \sigma_t^2)$. Variansen σ_t^2 är ju en okänd parameter i och med att våra α - och β -parametrar är okända (de ingår ju i ekvationen för σ_t^2 enligt GARCH-modellens definition). Därför kan vi utgå från täthetsfunktionen för ϵ_t när vi beräknar våra maximum-likelihood-skattningar och vi låter denna täthetsfunktion vara likelihoodfunktionen. Likelihoodfunktionen brukar sedan logaritmeras för att göra det enklare för oss. Partialderiverar vi sedan log-likelihoodfunktionen och sätter dessa derivator till noll, kan vi lösa ut de okända parametrarna som vi är intresserade av. Dessa är sedan våra Maximum-likelihood-skattningar.

För den intresserade läsaren finns det en mer utförlig förklaring av hur dessa uträkningar går till mer praktiskt i Appendix avsnitt 6.4.

2.6 Value at risk (VaR)

Volatilitet mäter, som tidigare nämnt, spridningen hos våra avkastningar, dvs. både upp- och nedgångar. Risk, å andra sidan, brukar handla om något ogynnsamt (vi har ju inget emot att priset på aktien kraftigt svänger uppåt) och därför brukar man använda sig av ett annat riskmått som beror av volatiliteten, nämligen Value-at-risk (eller som vi i fortsättningen förkortar det, VaR). Simpelt uttryckt så står VaR för den maximala förlusten som inte överskrids med en given hög sannolikhet (konfidensnivån α).

Givet någon konfidensnivå $\alpha \in (0, 1)$, så ges VaR av vår aktie (eller annan finansiell tillgång) av det minsta talet l sådant att sannolikheten att förlusten L överskrider l inte är större än $1 - \alpha$. Eller mer formellt,

$$\begin{aligned} VaR_\alpha &= \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : 1 - P(L > l) \geq \alpha\} = \\ &= \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\} \end{aligned}$$

där $\inf\{\}$ står för infimum och F_L betecknar fördelningsfunktionen för förlusten L . VaR_α står för Value at Risk vid konfidensnivån α .

Med andra ord mäter alltså VaR den procentuella kvantilen för avkastningarnas fördelning över en viss tidshorisont (i vårt fall en dag).

Antag att förlustfördelningen F_L är normal med väntevärde μ och varians σ^2 . Då ges α -kvantilen av $\mu + \sigma\Phi^{-1}(\alpha)$, och eftersom vi bara använder VaR som en procentsats (dvs. ej penningbelopp) så kan vi definiera det som:

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha)$$

där Φ betecknar standard normal fördelningsfunktionen och $\Phi^{-1}(\alpha)$ är α -kvantilen av Φ (McNeil, Frey & Embrechts 2005, s. 38-39).

2.7 Statistiska test

2.7.1 Autokorrelations-test

Vi erinrar oss att korrelationskoefficienten mellan r_t och r_{t-l} kallas för lag- l autokorrelationen av r_t och betecknas vanligen ρ_l och att definitionen för denna är (Tsay 2005, s. 31):

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}$$

För ett givet positivt heltal l , så kan man använda detta resultat för att testa nollhypotesen $H_0 : \rho_l = 0$ mot alternativhypotesen $H_1 : \rho_l \neq 0$, dvs. om vi har autokorrelation mellan r_t och r_{t-l} eller inte.

Detta test kan utföras visuellt med hjälp av ett korrelogram (se figur 5 för exempel), vilket är en plott, $\{(l, \hat{\rho}(l)) : l = 0, 1, 2, \dots\}$, som är designad för att underlätta tolkningen av stickprovs-autokorrelationsfunktionen (ACF). 95% av de skattade korrelationerna bör ligga inom intervallet $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$ (McNeil, Frey & Embrechts 2005, s. 134), och det är därför som korrelogram ritas med konfidensintervall vid dessa värden (streckade linjer i figur 5). Om mer än 5% av de skattade korrelationerna ligger utanför dessa gränser, så kan det betraktas som bevis mot nollhypotesen.

2.7.2 Ljung-Box-test

Om vi istället vill testa om det finns autokorrelation över flera lags samtidigt (dvs. nollhypotesen $H_0 : \rho_1 = \dots = \rho_m = 0$) kan man använda Ljung-Box-testet, vars statistika är (Tsay 2005, s. 32):

$$Q(m) = n(n+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{n-l}$$

där m står för antal autokorrelationer som testas samtidigt och n för stickprovets storlek. Vi förkastar nollhypotesen om $Q(m) > \chi_\alpha^2(m)$.

2.7.3 T-test för ett stickprov

Väntevärdet i en tidsserie kan som tidigare nämnt följa en ARMA-process eller vara en konstant. För att ta reda på om vi kan anta att väntevärdesprocessen är noll kan vi utföra ett vanligt t-test för ett stickprov.

Låt $x = (x_1, \dots, x_n)$ vara ett generellt stickprov från $X \sim N(\mu, \sigma^2)$. För att testa nollhypotesen $H_0 : \mu = \mu_0$ mot alternativhypotesen $H_1 : \mu \neq \mu_0$ används test-statistikan

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$$

där s är den skattade standardavvikelsen och n stickprovets (seriens) storlek. Nollhypotesen kan förkastas om $|T| > t_{\alpha/2}(n-1)$, där $t_{\alpha/2}(n-1)$ är $\alpha/2$ -kvantilen av t-fördelningen (Alm & Britton 2008, s. 334).

2.7.4 Backtest

Det är ofta svårt att tillämpa GARCH för prediktion framåt och därför kan man istället göra en simulering av GARCH på historiska data för att kunna mäta dess effektivitet. Detta angreppssätt kallas för ett "backtest" och innebär att man delar upp sina data i en del med historiska data (in-sample) som man gör själva analysen på och sedan en del som vi kan använda som ett "facit" med verkliga utfall att jämföra de predikterade värdena med (out-of-sample).

Vi följer avsnitt 4.4.3 i McNeil, Frey & Embrechts bok (2005) och skapar en förlustfunktion L_t som mäter skillnaden mellan den predikterade volatiliteten och volatiliteten i out-of-sample-delen av data. Sedan skapar vi en indikatorvariabel

$$I_t = \begin{cases} 1, & \text{om } L_t \geq VaR_\alpha \\ 0, & \text{om } L_t < VaR_\alpha \end{cases}$$

som räknar antalet gånger som förlusten är större än eller lika med VaR och t står för vilken dag i ordningen det är.

Om vi är framgångsrika när vi skattar VaR i varje tidssteg, kunde vi förvänta oss att indikatorvariablerna skulle uppföra sig som realiseringar av oberoende och likafördelade Bernoulli-försök. Indikatorvariablerna antar alltså värdet 1 med sannolikheten p .

Om vi beräknar VaR -skattningar för $t = 1, \dots, n$ förväntar vi oss att

$$\sum_{t=1}^n \hat{I}_t \sim Bin(n, p)$$

och detta kan man använda ett standard tvåsidigt binomialtest för att kontrollera (se nästa avsnitt 2.7.5). Nollhypotesen blir alltså $H_0 : p = \alpha$, där α är signifikansnivån (vi har valt $\alpha = 0.01$). Eftersom det är med sannolikheten p som prediktionsfelet blir för stort så blir p samma sak som signifikansnivån i det här fallet. Eftersom det är ett tvåsidigt test blir den alternativa hypotesen $H_1 : p \neq \alpha$. Om vi kan förkasta nollhypotesen tyder det på antingen systematisk underskattning eller överskattning av VaR .

2.7.5 Tvåsidigt Binomialtest

Låt x vara en observation från $X \sim Bin(n, p)$, där p är okänd. Enligt Alm & Britton (2008) kan p skattas med $p^* = x/n$. Om n är tillräckligt stort kan X approximeras med en normalfördelning: $X \approx N(np, np(1-p))$ och då gäller $p^*(X) = X/n \approx N(p, \frac{p(1-p)}{n})$, så att vi får ett konfidensintervall med konfidensgrad approximativt $1 - \alpha$ som

$$I_p = (p^* \pm \lambda_{\alpha/2} \cdot \sqrt{\frac{\alpha(1-\alpha)}{n}})$$

För att testa $H_0 : p = p_0$ används testvariabeln

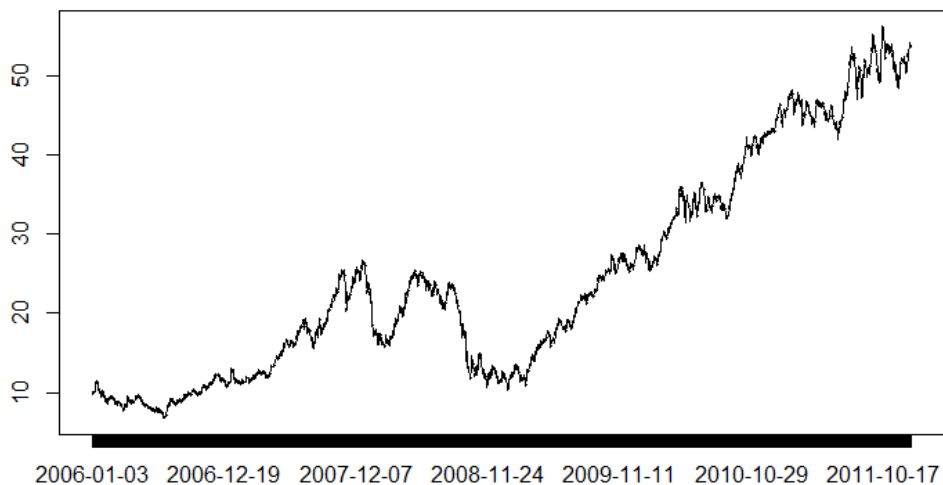
$$Z = \frac{p^* - \alpha}{\sqrt{\alpha(1-\alpha)/n}},$$

som under H_0 är en observation från $Z(X) \approx N(0, 1)$ (Alm & Britton 2008, s. 348).

3. Data

3.1 Apple Inc.

Apple Incorporated (eller i vardagstal bara Apple) är ett amerikanskt dator- och hemelektronikföretag grundat år 1976 av Steve Jobs, Steve Wozniak och Ronald Wayne. Företaget har nått stora framgångar och omsatte år 2015 nästan 234 miljarder amerikanska dollar. En av Apples mest framgångsrika produkter är mobiltelefonen Iphone (första modellen släpptes år 2007), som idag står för huvuddelen, ca 65%, av företagets intäkter (enligt Wikipedia).



Figur 1. Apple-aktiens prisutveckling över vårt valda tidsintervall.

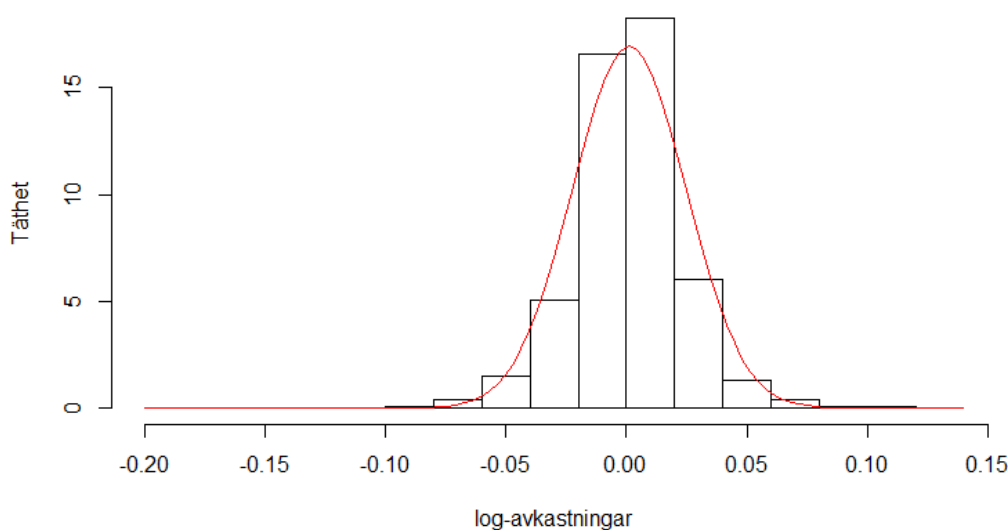
3.2 Stickprovsstorlek och tidsintervall

Vilken tidsperiod data är hämtad ifrån är av stor vikt. Vi väljer estimeringsperioden 2006-01-03 till 2011-12-31, vilket blir totalt 1512 observationer (1511 avkastningsobservationer). Vi har valt ett ganska stort stickprov för att få bättre parameterskattningar, men ett större stickprov än så är onödigt eftersom GARCH-modellen efter ett tag ”tappar minnet”.

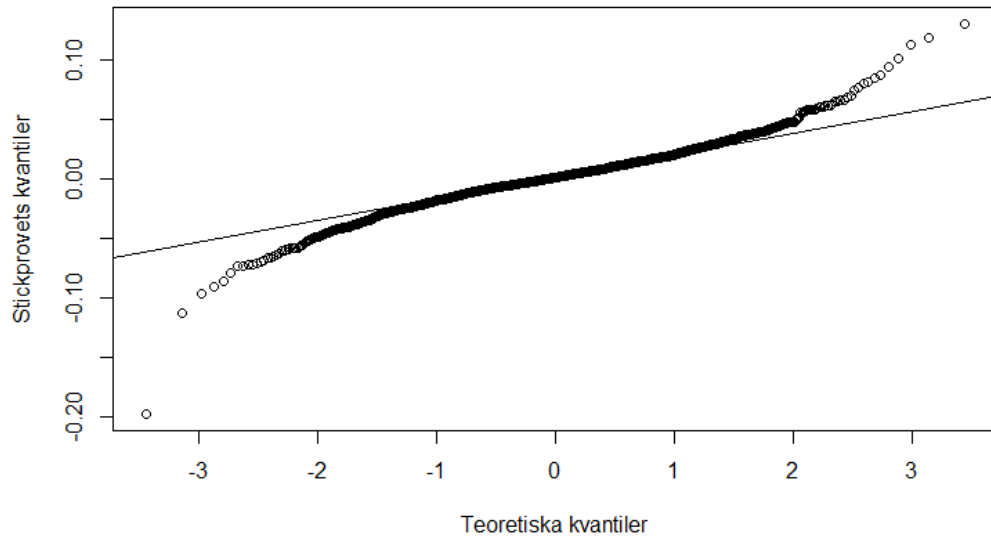
3.3 Dataanalys

Vi vill analysera data för att se om vi kan göra antagandet att data kommer från normalfördelningen. Vi undersöker detta genom att använda oss av histogram, qq-plott och mått på skevhet/toppighet (skewness/kurtosis på engelska). Skevhet och toppighet mäter tillsammans "formen" på en sannolikhetsfördelning. Skevhet mäter graden av asymmetri, där ett värde på noll betyder symmetri, positivt värde betyder tung (lång) högersvans jämfört med den vänstra och vice versa för negativt värde på skevhet. Med svansarna för en sannolikhetsfördelning menas regionerna långt från mitten. Toppighet indikerar i vilken utsträckning sannolikheten är koncentrerad i mitten och i svansarna, lågt värde betyder "låg" topp med få outlier-punkter och ett högt värde innebär det motsatta (för definitioner se Appendix 6.2).

Så kallad *leptokurtosis* - dvs. tendensen att avkastningen har en fördelning med tjocka svansar och stor toppighet vid väntevärdet - är vanligt hos aktie-avkastnings-data och måtten vi får på skevhet: -0.2245716 och toppighet: 5.181397 tyder på att detta gäller även våra data, som inte riktigt är normalfördelade, eftersom en ren normalfördelning bör ha värdet 0 för skevhet och värdet 3 för toppighet. I figur 2 nedan ser det ut som om data skulle kunna vara normalfördelade, även om de verkar ha en tyngre (längre) vänster-svans. Man har länge noterat att aktie-avkastningar har "tung-svansade" eller "outlier-benäigna" sannolikhetsfördelningar. I figur 3 nedan kan vi se tydligare att data inte riktigt är normalfördelat, utan har just så kallade "tunga svansar". Detta betyder att data innehåller mer extrema outlier-punkter än vad man kan vänta sig från en normalfördelning. En anledning till detta kan vara volatilitetskluster, dvs. den betingade variansen inte är konstant och att outlier-punkterna förekommer då variansen är stor. Faktum är att GARCH-processer uppvisar tunga svansar av denna anledning. Därför kan vi använda GARCH-modeller för att modellera både den betingade heteroskedasticiteten och tung-svansade fördelningar.



Figur 2. Histogram över log-avkastningarna med en normalfördelningskurva som har anpassats till log-avkastningarnas väntevärde och standardavvikelse.



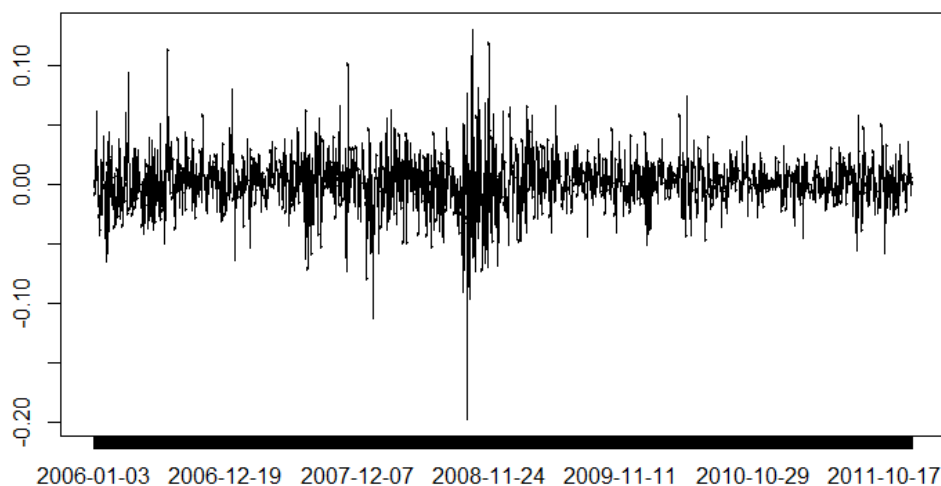
Figur 3. QQ-plott mellan de observerade kvantilerna och normal-kvantilerna.

4. Metod och resultat

All analys har utförts med statistikprogrammet R och en del av den med hjälp av det inbyggda R-paketet "rugarch".

4.1 Deskriptiv analys

I figur 4 nedan ser vi våra plottade log-avkastningar. Eftersom dessa ser ut att hålla sig på en horisontell linje (dvs. data verkar inte ha någon synlig trend eller periodicitet) antar vi att vi har en stationär tidsserie.



Figur 4. De logaritmerade avkastningarna.

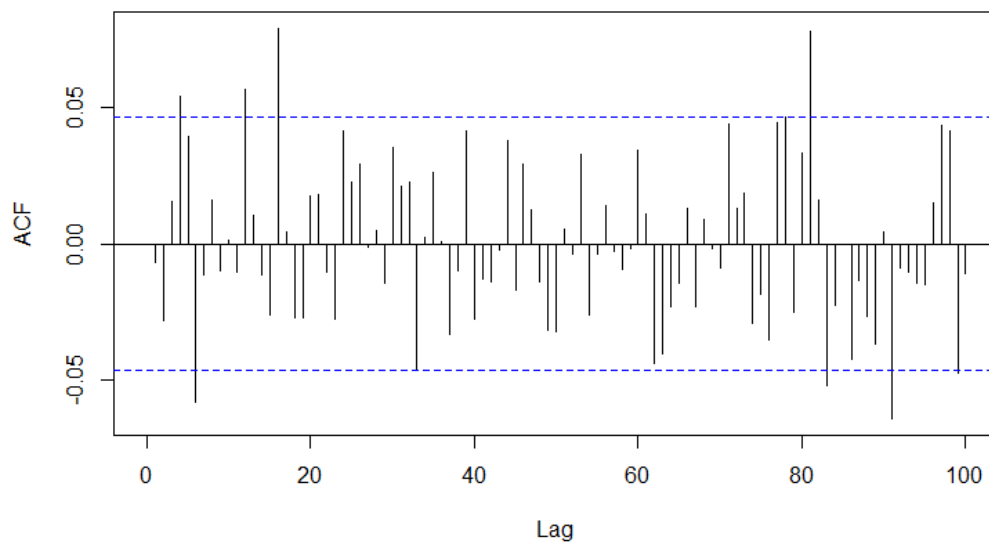
I Figur 4 ovan kan vi se exempel på volatilitetskluster, dvs. fenomenet att perioder med ett stort värde på volatiliteten ofta följs av vidare perioder med stor volatilitet och vice versa för perioder med mindre volatilitet. Det är lämpligt att använda exempelvis GARCH för att modellera volatilitetskluster.

För att utföra GARCH-modellering vill vi att väntevärdet i $r_t = \mu_t + a_t$ ska vara noll. Stickprovsmedelvärdet brukar användas som skattning av väntevärdet, och i vårt fall blir det så pass litet, 0.00133068, att vi kan anta att väntevärdet är noll. Vi utför även ett vanligt t-test för ett stickprov för att ge tyngd till detta antagande, dvs. nollhypotesen är $H_0 : \mu_t = 0$.

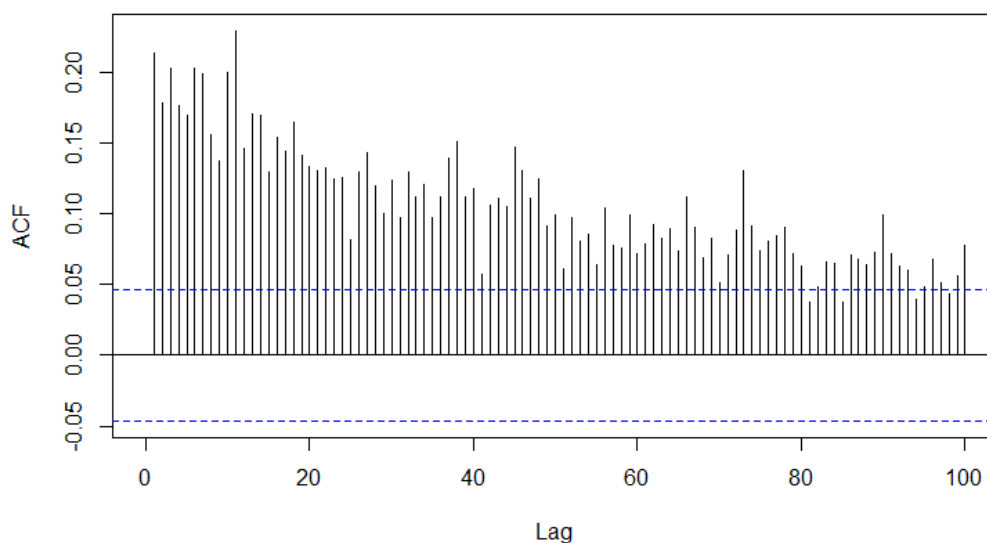
Vi får t-statistikan $t = 1.7917$ och p-värdet 0.07338. Vilket alltså betyder att vi

INTE kan förkasta nollhypotesen om att väntevärdet är noll på signifikansnivån 5%. Vi accepterar därför nollhypotesen och antar att väntevärdet är noll och reducerar vår modell.

Tittar vi på korrelogrammet för log-avkastningarna (Figur 5 nedan) så ser det ut som om det är vitt brus. Vi vill styrka detta genom att även göra ett Ljung-Box-test av de första 10 tidsstegen (lags). På signifikansnivån 5% får vi statistikan $Q(10) = 17.529$ och p-värdet 0.06344, som innebär att vi inte kan förkasta nollhypotesen och att de tio första tidsstegen ej är autokorrelerade. Om vi tittar på korrelogrammet för avkastningstidsseriens absolutbelopp (Figur 6 nedan) så ser vi ett seriellt beroende som verkar klinga av med tiden. Grundtanken bakom volatilitets-studier är att $\{r_t\}$ inte är autokorrelerad eller väldigt liten/låg ordning på autokorrelationerna, men att det ändå är en seriellt beroende serie. Så verkar vara fallet med vår avkastningsserie.



Figur 5. Korrelogram över log-avkastningar.



Figur 6. Korrelogram över absolutbeloppet av log-avkastningar.

Vi ska även testa efter ARCH-effekter (eller heteroskedasticitet i feltermerna) för att se om GARCH är en lämplig modell att tillämpa, vilket innebär att vi kollar om de kvadrerade feltermerna a_t^2 är autokorrelerade. Vi använder Ljung-Box-testet även för detta och får statistikan $Q(10) = 432.75$ och p-värdet $< 2.2e-16$, vilket betyder att vi kan förkasta nollhypotesen om att det inte finns autokorrelation. Alltså har vi ARCH-effekter (heteroskedasticitet i residualerna) och det är lämpligt att använda GARCH.

4.2 Ordning

Eftersom jämförelsen är mellan två GARCH-modeller med olika antal parametrar, så behöver vi bestämma oss för vilket antal parametrar var och en av modellerna ska ha, dvs. vi behöver bestämma ordningen.

Det var redan i förväg bestämt att den ena skulle vara GARCH(1,1) eftersom den är mest känd och använd. Istället för att bara välja en annan ordning rätt ut ur tomma luften jämför vi AIC-värde (se Appendix 6.2 för definition), där modellen som har det minsta värdet enligt kriteriet är den bäst lämpade för prediktion, mellan GARCH(p,q)-modeller med ordningar som väljs ur en grupp med 8 kombinationer på ordningen (p,q):
 ((1,2),(2,1),(2,2),(2,3),(3,2),(1,3),(3,1),(3,3)).

	AIC
GARCH(1, 2)	-4.8238
GARCH(2, 1)	-4.8238
GARCH(2, 2)	-4.8225
GARCH(3, 2)	-4.8214
GARCH(2, 3)	-4.8239
GARCH(1, 3)	-4.8227
GARCH(3, 1)	-4.8227
GARCH(3, 3)	-4.8236

Figur 7. Resultat av AIC för olika ordningar på GARCH.

Vi väljer GARCH(2,3) som vår andra modell, eftersom denna hade det lägsta AIC-värdet i denna grupp (se Figur 7).

Räknar vi ut AIC-värdet även för GARCH(1,1) så får vi -4.8249 vilket är ännu lite lägre än AIC för GARCH(2,3), och detta ger redan nu en antydning om att GARCH(1,1) eventuellt kommer att prestera bättre än GARCH(2,3).

4.3 Parameterskattning och validering

I fallet med GARCH(1,1) har vi definitionen

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

där de okända parametrarna som skall skattas alltså är $\theta = (\alpha_0, \alpha_1, \beta_1)$, och i fallet med GARCH(2,3) har vi definitionen

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \beta_3 \sigma_{t-3}^2 \end{aligned}$$

där de okända parametrarna som skall skattas alltså är $\theta = (\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)$.

Skattningen av parametrarna i de båda GARCH-modellerna gör vi med hjälp av Maximum-likelihood-metoden (Se teori-avsnitt 2.5 och Appendix avsnitt 6.4). Som vi tidigare såg, så var log-avkastningarna inte riktigt normalfördelade, utan har tunga svansar. En anledning till att log-avkastningarna inte är exakt normalfördelade kan bero på volatilitetskluster-egenskapen. För enkelhetens skull antar vi att de tunga svansarna beror på detta och kan modelleras med GARCH, och att ϵ_t följer normalfördelningen.

När vi har fått fram skattningarna ser våra modeller alltså ut som följer: GARCH(1,1):

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.000010 + 0.082038 a_{t-1}^2 + 0.900547 \sigma_{t-1}^2 \end{aligned}$$

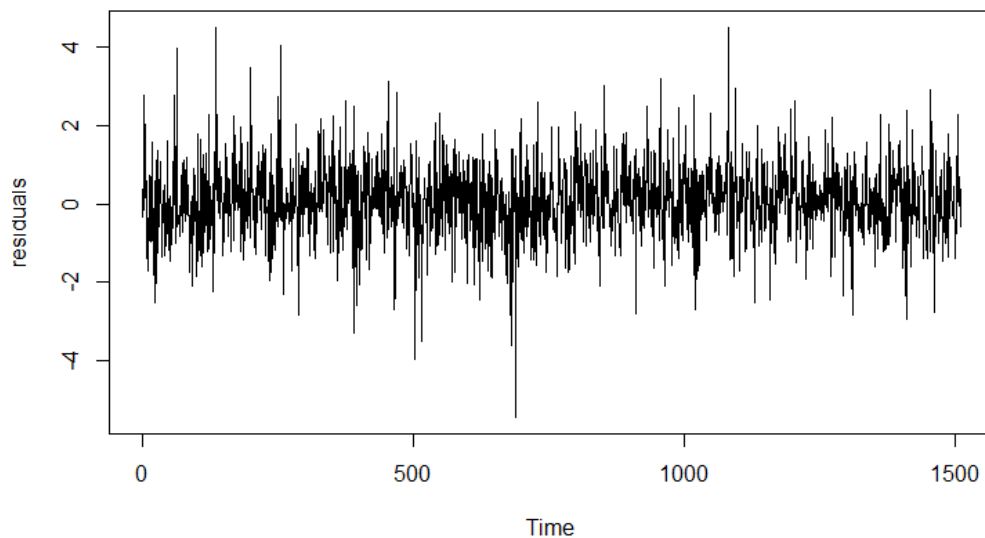
GARCH(2,3):

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.000011 + 0.094452 a_{t-1}^2 + 0.004743 a_{t-2}^2 + 0.754034 \sigma_{t-1}^2 + \\ &\quad 0.000015 \sigma_{t-2}^2 + 0.126505 \sigma_{t-3}^2 \end{aligned}$$

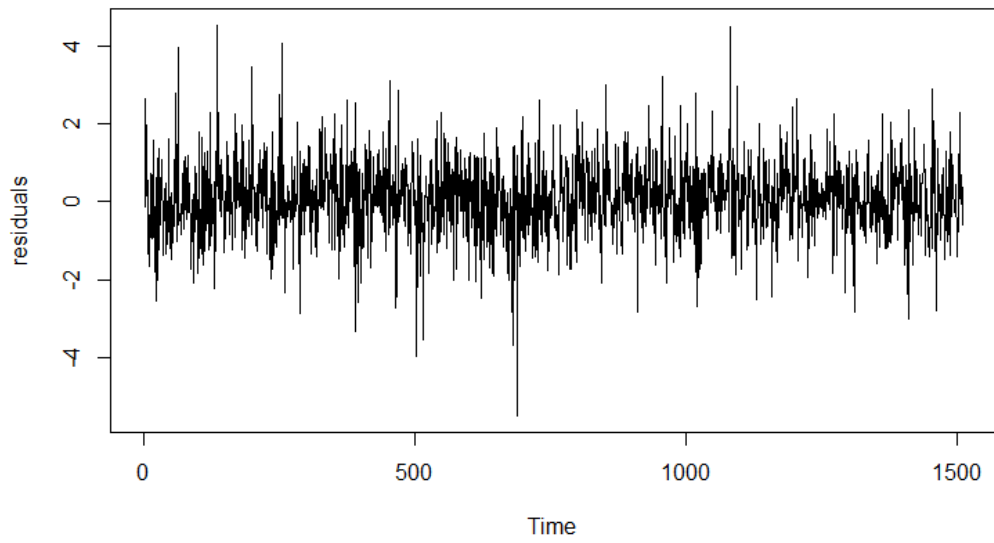
I definitionen för GARCH tidigare så hade vi parametervillkoren $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ och $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. Våra modeller uppfyller dessa villkor eftersom alla koefficienter är större än 0 och summerar sig (för respektive modell) till ett tal mindre än 1.

Det är vanligt att kontrollera anpassade GARCH-modeller genom att använda residualer. Man brukar skilja mellan ostandardiserade och standardiserade residualer (se figur 8 & 9). De förstnämnda är residualerna a_1, \dots, a_n för r_t -delen av GARCH-modellen och under den hypotetiserade modellen borde dessa bete sig som en ren GARCH-process. De senare är rekonstruerade realiseringar av det vita brus som antas driva GARCH-delen av modellen, och de beräknas från de förra genom $\epsilon_t = a_t/\sigma_t$.

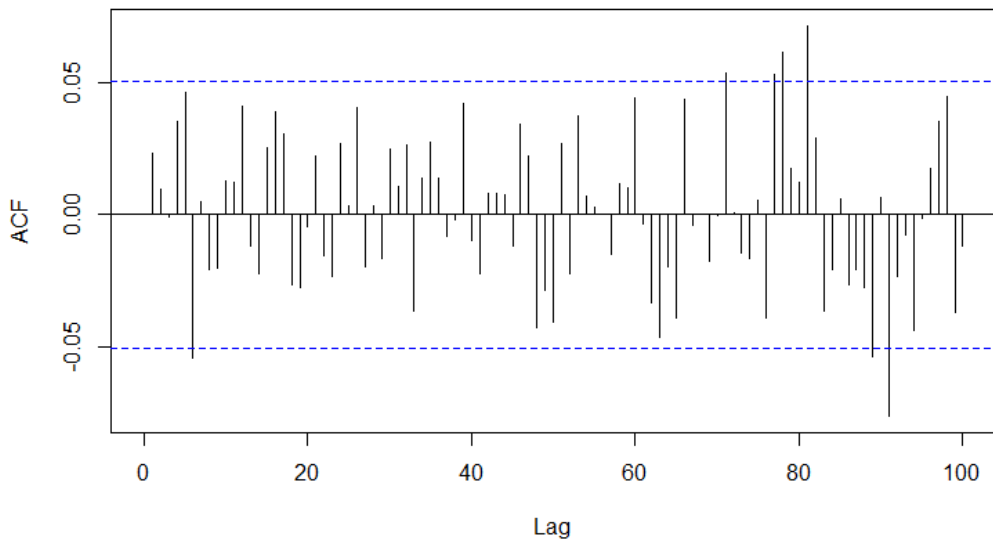
De standardiserade residualerna borde alltså uppföra sig som vitt brus om vi har en adekvat modell och detta kan undersökas genom att konstruera korrelogram (se figur 10 & 11) och tillämpa Ljung-Box-testet på dessa residualer. Vi får på signifikansnivån 5% statistikan $Q(10) = 12.075$ och p-värdet 0.2801, vilket innebär att vi inte kan förkasta nollhypotesen om att vi har vitt brus. Eftersom det dessutom för båda modellerna ser ut som att serien $\{\epsilon_t\}$ är vitt brus kan vi dra slutsatsen att båda modellerna lyckas fånga upp volatilitetsklusterbeteendet hos data.



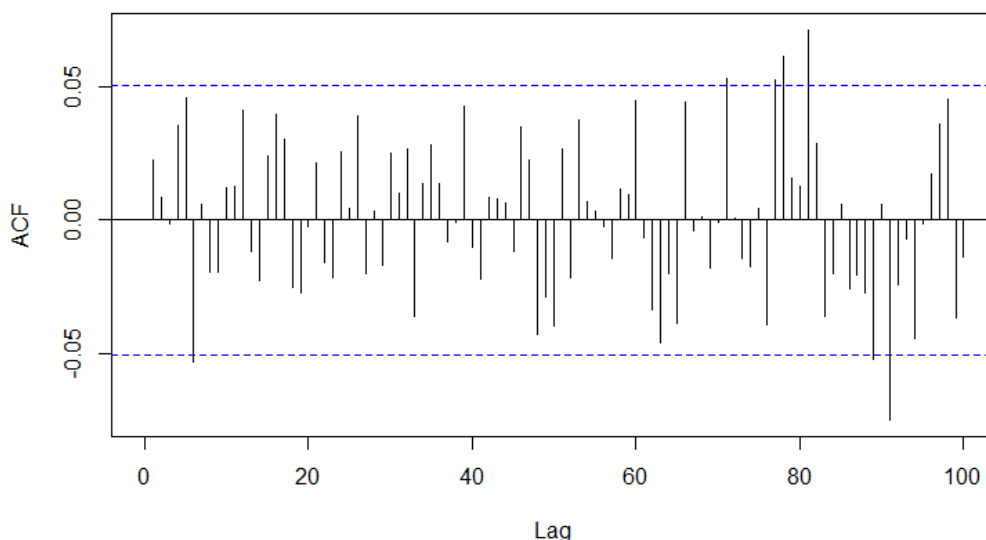
Figur 8. De standardiserade residualerna till den anpassade GARCH(1,1)-modellen.



Figur 9. De standardiserade residualerna till den anpassade $GARCH(2,3)$ -modellen.



Figur 10. Korrelogram av de standardiserade residualerna av den anpassade $GARCH(1,1)$ -modellen.



Figur 11. Korrelogram av de standardiserade residualerna av den anpassade $GARCH(2,3)$ -modellen.

4.4 Backtest

Vi kommer att göra två backtest, ett per modell, och sedan jämföra vilken av modellerna som verkar klara sig bäst i förhållande till varandra.

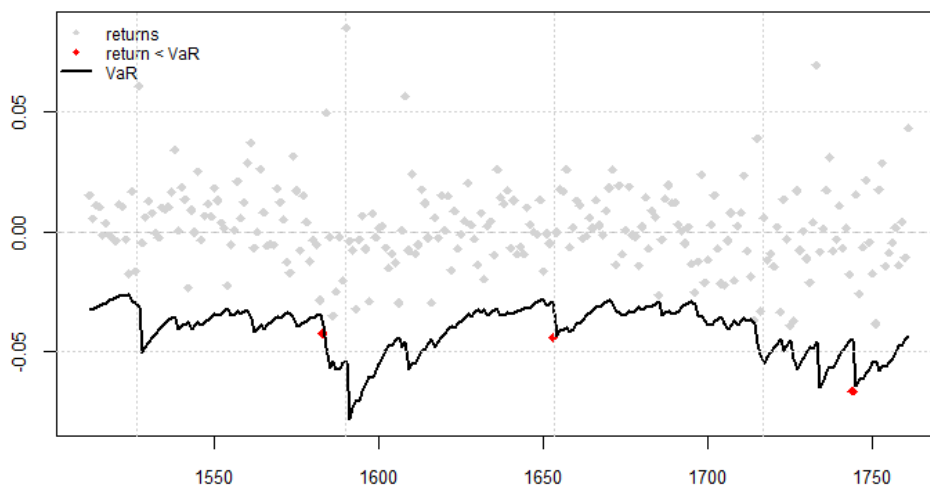
Under hela processen kommer vi att prediktera för en dag framåt (ju längre in i framtiden man vill prediktera, desto mer perfekt måste modellen vara). Det första stickprovet (in-sample) som vi hittills har gjort all analys på innehåller 1511 observationer (dagar), en observation per dag från 2006-01-04 t.om. 2011-12-30.

Detta stickprov använder vi först för att prediktera volatiliteten för dag 1512, sedan om vi vill prediktera volatiliteten för dag 1513 använder vi ett "nytt" stickprov som består av lika många observationer (1511 stycken) men där vi har släppt den "äldsta" som fanns med i det förra stickprovet och lagt till den första observationen från vårt referensmaterial (out-of-sample). En prediktion utifrån detta nya stickprov kommer att ge oss volatiliteten för dag 1513. Sedan upprepar vi bara denna procedur 250 gånger totalt (längden på vår out-of-sample-period), där vi i varje tidssteg skattar modell-parametrarna på nytt.

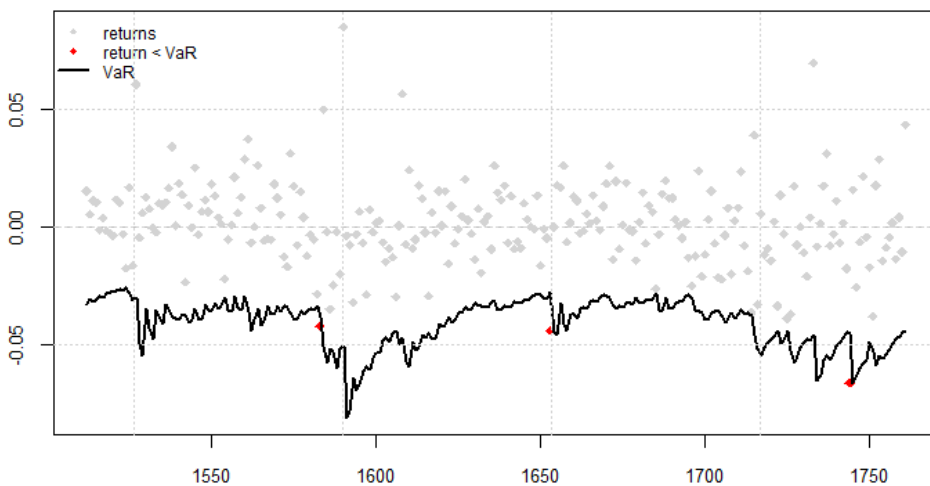
I Figur 12 resp. 13 nedan kan vi se resultatet av backtesten med signifikansnivån $\alpha = 0.01$ grafiskt. Alla log-avkastningar plottas som grå och röda punkter, där de röda är de observationer som är lägre än den skattade VaR-kvantilen (den heldragna svarta linjen i varje graf, som innehåller VaR som skattas på nytt i varje tidssteg), dessa observationer kan vi kalla för "överskridningar".

Även om de skattade VaR-kvantilerna inte är detsamma för varje modell så visar det sig att utfallet blir likadant för bägge modellerna:

Backtestlängd: 250
Signifikansnivå: 0.01
Förväntade överskridningar: 2.5 (1% av 250)
Faktiska överskridningar: 3 (1.2%)



Figur 12. Resultat av backtest för $GARCH(1,1)$ -modellen, där x -axeln visar vilken observation i ordningen som predikteras av de 250.



Figur 13. Resultat av backtest för $GARCH(2,3)$ -modellen.

Enligt det tvåsida binomialtestet med p -värde 0.7426 kan vi alltså anta att båda summorna i båda fallen kommer från en $Bin(250, 0.01)$ -fördelning. Det vill säga den predikterade volatiliteten ligger tillräckligt nära den realiserade volatiliteten i våra referensdata.

5. Diskussion och slutsats

Syftet med detta kandidatarbete var att jämföra förmågan att prediktera volatilitet mellan GARCH-modeller med olika antal parametrar. Eftersom vi får samma resultat från båda backtesten så kan vi rimligtvis anta att båda modellerna fungerar lika bra när det gäller att prediktera den framtida volatiliteten för Apple Inc.s aktie-avkastning.

Det kan finnas fler anledningar till att vi får samma resultat. Till exempel har vi kanske valt en ”snäll” datamängd som gör att alla GARCH-modeller fungerar nästan lika bra för prediktion? Log-avkastnings-serien kunde antas vara stationär och behövde inte differentieras och det tyder ju på en snäll datamängd, eftersom det i vanliga fall brukar behövas.

Frånsett att statistiska modeller aldrig exakt kan skatta eller prediktera ett problem i verkligheten, så finns det några förenklingar vi har gjort som givetvis kan ha orsakat fel i vår analys. Dessutom har vi valt en tidsperiod att hämta data ifrån under vilken Finanskrisen 2008 skedde.

För att förenkla analysen har vi till exempel ”accepterat” nollhypotesen i alla test där den inte kan förkastas, i och med detta riskerar vi att göra Typ II-fel, dvs. att man accepterar nollhypotesen trots att denna är falsk. Bland annat har vi antagit att väntevärdet är noll hos log-avkastnings-serien på detta sätt.

När man utför Ljung-Box-test så kan antalet lags, m , man väljer att testa påverka test-resultatet. Det verkar dock inte finnas mycket skrivet om hur man lämpligast väljer m , men enligt Tsay (2005) så antyder simuleringsstudier att valet av $m \approx \ln(n)$ ger ett bättre resultat och Athanasopoulos & Hyndman (2013) föreslår att man väljer $m = 10$ om man har icke-periodiska data. I vår analys valde vi det senare och eftersom det inte verkar vara någon exakt vetenskap så kan det förstås vara en felkälla.

Vi valde att anta att residualerna i GARCH-modellerna är standard-normalfördelade när vi skulle skatta parametrarna, vilket inte är självklart. Det hade antagligen blivit en bättre anpassning med en t-fördelning (som har tunga svansar), men eftersom det är prediktionsförmågan vi är intresserade av här, håller vi oss till normalfördelningen för enkelhetens skull. Kanske hade vi även fått ett mer exakt resultat på parameterskattningarna om vi hade använt kvasi-maximum-likelihood-skattning (även om R ger oss samma parametrar som vi redan har med denna metod).

Själva backtest-metoden kan ifrågasättas eftersom det är avgörande vilka data vi har valt att ha i in-sample-delen respektive out-of-sample-delen. Längden

på fönstret kan också ha spelat roll för vårt resultat. Dessutom kan det vara tveksamt att bara förlita sig på en enda statistika (som en förlustfunktion).

Vi kunde ha valt en ännu högre konfidensnivå än 99% på VaR för att få ett ännu känsligare riskmått, men då har vi problemet att normalfördelningsantagandet kanske blir alltför dåligt istället (den verkliga fördelningen har ju tunga svansar).

I vårt fall var 250 prediktioner kanske för få för att det skall bli någon skillnad mellan de båda modellerna, men efter att ha provat att göra 1000 prediktioner utifrån samma in-sample-del, blev det ändå samma slutresultat.

Efter att även ha provat utföra samma backtest fast med en GARCH(7,8)-modell så blev skillnaderna fortfarande mycket små och därför verkar det mer rimligt att GARCH(1,1) och GARCH(2,3) (som är låga ordningar i jämförelse med GARCH(7,8)) kan ge samma resultat.

Men om vi ändå ska välja en modell som är bättre än den andra så blir det GARCH(1,1) på grund av enkelheten det ger att ha färre parametrar i sin modell.

Man skulle kunna göra ett mer omfattande test för att se om vi skulle ha kommit fram till samma sak även för olika typer av data och se om vi hittar ett mönster. Är anledningen till GARCH(1,1)-modellens popularitet att den alltid är bäst?

6. Appendix

6.1 Sannolikhetsfördelningar

Vi refererar flera gånger till olika sannolikhetsfördelningar och ifall man behöver fräscha upp minnet så listar vi dom här.

6.1.1 Bernoullifördelningen

En diskret slumpvariabel X är Bernoulli-fördelad om $P(X = 0) = 1 - p$ och $P(X = 1) = p$, för något $0 \leq p \leq 1$.

6.1.2 Binomialfördelningen

Om man gör n oberoende upprepningar av ett försök där en händelse inträffar med sannolikheten p i ett enskilt försök och låter X vara antalet gånger händelsen inträffar i de n försöken så gäller det att $X \sim Bin(n, p)$.

En diskret slumpvariabel X sägs vara binomialfördelad med parametrarna n och p , där $0 \leq p \leq 1$, om sannolikhetsfunktionen ges av

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

där $k = 0, \dots, n$.

6.1.3 Normalfördelningen

En kontinuerlig slumpvariabel X sägs vara normalfördelad med parametrar μ och σ^2 , där $\sigma^2 > 0$ om täthetsfunktionen ges av

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

där $-\infty < x < \infty$.

6.1.4 t-fördelningen

En slumpvariabel X med täthetsfunktion

$$f(x) = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\Gamma(\frac{n-1}{2})} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}$$

sägs vara t-fördelad med $n - 1$ frihetsgrader då $-\infty < x < \infty$.

6.2 Skevhet och toppighet

Inom statistik används ofta skevhet och toppighet (skewness resp. kurtosis på engelska), vilket är det tredje respektive fjärde momentet av en slumpvariabel X , för att sammanfatta graden av asymmetri och svanstyngd. Dessa definieras

$$S(x) = E\left(\frac{(X-\mu_x)^3}{\sigma_x^3}\right)$$
$$K(x) = E\left(\frac{(X-\mu_x)^4}{\sigma_x^4}\right)$$

där $S(x)$ står för skevhet och $K(x)$ för toppighet.

6.3 Akaikes informationskriterium

Akaikes informationskriterium (AIC) definieras som

$$AIC = -\frac{2}{n} \cdot \ln(L(\hat{\theta})) + \frac{2}{n} \cdot (\text{antal parametrar})$$

där man har stoppat in maximum-likelihood skattningarna i likelihoodfunktionen och n är stickprovets storlek. Den första termen i AIC mäter hur god anpassningsförmåga modellen i fråga har, medan den andra termen kallas för ”straff”-funktionen eftersom den straffar en kandidering modell med antalet parametrar den har. Ju lägre värde på AIC en modell har, desto bättre lämpad för prediktion är modellen.

6.4 Beräkning av Maximum-likelihood-skattningar

För att göra det enkelt för oss utgår vi ifrån GARCH(1,1)-modellen (log-likelihoodfunktion enligt Tsay 2005, s.17):

För att räkna ut maximum-likelihood-skattningarna av våra okända parametrar $\theta = (\alpha_0, \alpha_1, \beta_1)$ så låter vi som tidigare nämnt likelihoodfunktionen vara täthetsfunktionen för ϵ_t .

Vi betecknar likelihoodfunktionen med $L(\theta)$, log-likelihoodfunktionen med $l(\theta)$ och antalet observationer med n .

$$l(\theta) = \ln(L(\theta)) = \ln\left(\prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{a_t^2}{2\sigma_t^2}\right\}\right) = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma_t^2}}\right)^n \exp\left\{-\frac{1}{2} \sum_{t=1}^n \frac{a_t^2}{\sigma_t^2}\right\}\right) =$$
$$= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^n \frac{a_t^2}{\sigma_t^2}$$

Sedan deriverar vi $l(\theta)$ och sätter derivatan till 0.

$$\frac{\partial l(\theta)}{\partial \sigma_t^2} = -\frac{1}{2} \sum_{t=1}^n \frac{1}{\sigma_t^2} \cdot \frac{\partial \sigma_t^2}{\partial \theta} + \frac{1}{2} \sum_{t=1}^n \frac{a_t^2}{(\sigma_t^2)^2} \cdot \frac{\partial \sigma_t^2}{\partial \theta} = 0$$

Eftersom σ_t^2 beror på $\theta = (\alpha_0, \alpha_1, \beta_1)$ så deriveras $\frac{\partial l(\theta)}{\partial \sigma_t^2}$ med avseende på α_0, α_1 och β_1 . Dessa derivator blir

$$\begin{aligned}\frac{\partial \sigma_t^2}{\partial \alpha_0} &= 1 - \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \alpha_0} \\ \frac{\partial \sigma_t^2}{\partial \alpha_1} &= a_{t-1}^2 - \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \alpha_1} \\ \frac{\partial \sigma_t^2}{\partial \beta_1} &= \sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1}\end{aligned}$$

som sedan stoppas in i $\frac{\partial l(\theta)}{\partial \sigma_t^2}$ för att få ML-skattningarna $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$.

7. Källhänvisningar

Alm, Sven Erick, & Britton, Tom. *Stokastik, Sannolikhets teori och statistik teori med tillämpningar* 2008. Stockholm: Liber.

Athanasopoulos, George, & Hyndman, Rob J. *Forecasting: principles and practice* 2013. OTexts: Melbourne, Australia. Hämtat den 14 maj 2016.
<https://www.otexts.org/fpp/2/6>

Bollerslev, Tim. 1986. *Generalized Autoregressive Conditional Heteroskedasticity*. *Journal of Econometrics* 31, 307-327.

Brooks, Chris. *Introductory econometrics for finance*. 2008. Cambridge university press.

Cryer, Jonathan D. & Chan, Kung-Sik. *Time Series Analysis. With applications in R*. 2008. Springer.

Engle, Robert F. 1982. *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*. *Econometrica*, 50(4), 987-1008.

Forskning & Framsteg. Nr 8/2003 *Ekonomi: Att tolka tids serier*.
<http://fof.se/tidning/2003/8/ekonomi-att-tolka-tidsserier>

Ghalanos, Alexios. *Introduction to the rugarch package.(Version 1.3-1)*. CRAN. 2015-02-20.
http://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf

Held, Leonhard & Bové, Daniel Sabanés. *Applied Statistical Inference. Likelihood and Bayes*. 2014. Springer.

Investopedia. Hämtat 4 april 2016.
<http://www.investopedia.com/terms/v/volatility.asp>
<http://www.investopedia.com/terms/b/backtesting.asp>

Köksal, Bülent. 2009. *A Comparison of Conditional Volatility Estimators for the ISE National 100 Index Returns*. *Journal of Economic and Social Research* 11(2), 1-28.

Mathworks. Hämtat 4 april 2016. <http://se.mathworks.com/help/econ/maximim-likelihood-estimation-for-conditional-variance-models.html>

McNeil, Alexander J., Frey, Rüdiger & Embrechts, Paul. *Quantitative Risk Management. Concepts, Techniques and Tools*. 2005. Princeton University Press.

Ruppert, David. *Statistics and Finance. An introduction*. 2004. Springer.

Tsay, Ruey S. *Analysis of Financial Time Series*. 2005. John Wiley & Sons.

Wikipedia. Hämtat 6 april 2016.

https://sv.wikipedia.org/wiki/Apple_Inc.

Wolfram Mathworld. Hämtat 3 maj 2016.

<http://mathworld.wolfram.com/Studentst-Distribution.html>