

Examensarbete

**Modellerande av förhållande mellan P/E-tal och nedgångar  
på OMXS30.**

Jon Hedström, Johan Vidlund

LITH - MAT - EX - - 2014 / 08 - - SE



# Modellerande av förhållande mellan P/E-tal och nedgångar på OMXS30.

MAI, Linköpings Universitet

**Jon Hedström, Johan Vidlund**

LITH - MAT - EX - - 2014 / 08 - - SE

Examensarbete: **16 hp**

Level: **G2**

Handledare: **M. Singull**,  
MAI, Linköpings Universitet

Examinator: **M. Singull**,  
MAI, Linköpings Universitet

Linköping: **oktober 2014**



# Sammanfattning

Den rapport du just ska till att läsa är ett kandidatarbete i matematisk statistik skrivet vid matematiska institutionen, Linköpings Universitet. Det område som undersöks är att om man med hjälp av P/E-tal kan förutsäga kraftiga börsnedgångar (börskrascher) på OMXS30. För att definiera en börskrasch har vi använt måttet *Value at Risk* (VaR). Detta mått är vedertaget hos finansiella institutioner som ett riskmått men i denna rapport används det som sagt för att definiera nivån för en börskrasch. VaR har beräknats med diverse olika metoder som presenteras i rapporten.

Efter att en börskrasch definierats har vi använt *logistisk regression* med P/E-tal som förklaringsvariabel för att undersöka om dessa nedgångar har ett samband med höga P/E-tal. Denna undersökning har lett fram till ett starkt resultat som säger att om en börsnedgång definieras med ett *VaR* mått som bygger på normalfördelningsantagande där volatiliteten är simulerad med *GARCH(1,1)* så kan vi konstatera att det finns ett säkerställt samband mellan höga P/E-tal och börskrascher.

Slutsatserna som dragit från undersökningen är att man genom att inkorporera en logistisk regression mot P/E-talet kan förstärka sitt VaR mått givet de antaganden som presenterats. Författarna uppmuntrar vidare forskning på området för att se om resultatet kan generaliseras till olika börsindex och även till specifika bolag.

**Nyckelord:** P/E-tal, nedgång, riskmått, logistisk regression

**URL för elektronisk version:**

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-77777>



# Förord

Under detta arbete har vi fått en utökad förståelse för statistisk modellering samt börsbeteenden. Vi har utvecklat våra förmågor i datainsamling samt rapportskrivning.

Vi vill passa på att tacka vår handledare M. Singull för den hjälp vi fått under arbetets gång.

Vi hoppas att detta arbete kan uppbringa intresse och berika med lärdom.

Mycket nöje!





# Innehåll



# Kapitel 1

## Inledning

### 1.1 Bakgrund

#### 1.1.1 P/E-tal

Handeln på börsen är till stor del baseras på förväntningar på framtiden.  $P/E$ -tal är ett mått som indikerar marknadens tro på framtiden, vilket således bör ha en stor inverkan på börspriserna. Höga  $P/E$ -tal betyder alltså att marknaden har en stor förväntning på företagens framtida intjäningsförmåga, trots att man i dagsläget inte kan se några realiserade vinster. Detta fick oss att fundera på huruvida det finns en koppling mellan stora börsnedgångar och höga  $P/E$ -tal. Koppling skulle, enligt vårt sett att se det, bero på att investerare reagerar kraftigare på negativa nyheter då de vet att de, i alla fall historiskt, betalat ett överpris på aktien. Denna grundläggande tanke ledde till att vi ville undersöka hur detta mått påverkar de nedgångar som identifieras på marknaden. Vi ville alltså undersöka huruvida höga  $P/E$ -tal kan *flagga* för börskrascher.

Under ett samtal med vår handledare, M. Singull, framkom att *logistisk regression* var ett modelleringsätt som skulle kunna ge oss vad vi sökte. Detta då sambandet är, om existerande, olinjärt.

#### 1.1.2 OMXS30

OMXS30 är ett av NASDAQ's index på stockholmsbörsen. Det är en portfölj av viktade innehav av Sveriges 30 mest omsatta aktier. Detta index ger en uppfattning om marknaden och är i denna rapport tänkt att symbolisera den svenska aktiemarknaden.

### 1.2 Syfte

Att undersöka och modellera sambandet mellan  $P/E$ -tal och stora nedgångar av OMXS30 index via *logistisk regression*.

### 1.3 Frågor att besvara

För att kunna besvara syftet i rapporten har detta brutits ned till ett antal frågor som måste besvaras.

1. Hur skall vi definiera en börskrasch?
2. Hur valideras de börskrascher vi definerat?
3. Hur skall detta praktiskt implementeras?
4. Hur skall  $P/E$ -talen grupperas på ett logiskt sätt?

### 1.4 Upplägg

Denna rapport består av 4 kapitel och 1 bilaga med *MatLab*-kod. Dispositionen beskrivs nedan

**Kapitel 1** Bakgrund och beskrivning.

**Kapitel 2:** Teoretisk bakgrund till samtliga behandlade matematiska ämnen delges. Detta för att ge en teoretisk grund till de praktiska utförandet som genomförs.

**Kapitel 3:** Datainsamling och testerna utförs. Detta kapitel kommer leda läsaren genom utföranden så att experimentet kan replikeras.

**Kapitel 4** Resultat och slutsatser presenteras.

## Kapitel 2

# Teoretisk referensram

Den här delen i rapporten behandlar den teori som använts för att både förstå problemet och för att genomföra undersökningen. I de fall det varit möjligt har information tagits från originalkällor och i de fall originalkällor inte hittats poängteras noga att informationen är tagen från sekundärkälla.

### 2.1 P/E-tal

I företagsvärdering använder man sig ofta av  $P/E$ -tal och i dagsläget är  $P/E$ -tal den vanligaste använda värderingsmultipeln [?]. Det talar om hur dyr aktien är i förhållande till företagets vinst. Man kan se det som ett mått på marknadens förhoppningar på företaget.  $P/E$  talet beräknas som

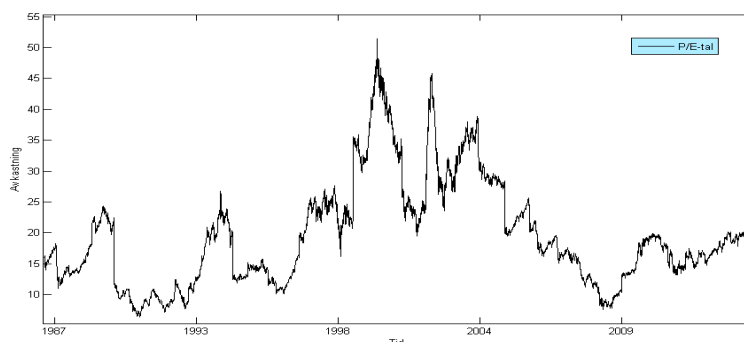
$$P/E = \text{Aktiepris} / \text{Vinst per aktie}. \quad (2.1)$$

$P/E$ -talet kan alltså tolkas på olika sätt. Ett vanligt och intuitivt sätt är att se det som hur många år det kommer ta att få tillbaka sin investering, givet intjäningsförmågan. Oavsett hur man väljer att tolka  $P/E$ -tal så är det praktiskt att använda då man vill undersöka om en aktie är dyr eller inte. Detta eftersom det ofta finns långa tidsserier så det finns möjlighet att jämföra ett bolag över tid. Dessutom kan man jämföra ett företag mot dess konkurrenter för att se hur priset i förhållande till intjäningsförmågan förhåller sig till snarlika företag.

Ett av de största problemen med  $P/E$ -talen är att det finns olika definitioner på intjäningsförmågan. De vanligaste sätten är att titta på historisk intjäning (*trailing*), nuvarande intjäning (*current*) och framtida intjäning (*forward*) [?]. För att detta skall vara konsistent har vi använt en tidsserie från *Datastream* som är en stor aktör då det gäller att leverera finansiell data. Många banker och kreditinstitut använder sig av *Datastream* och vi antar därför att den data vi hämtat är korrekt.

#### 2.1.1 Historisk utveckling av P/E-tal

Genom att bara ockulärt besiktiga den tidsserie av  $P/E$ -tal som *Datastream* erbjuder kan vi konstatera att  $P/E$ -talen sedan 1986 pendlat kraftigt. Se figur (??) nedan.



Figur 2.1: Tidsserie över  $P/E$ -tal mellan 1986-2014

Från denna tidsserie ges också en indikation om vad som tidigare sagts, nämligen att väldigt höga  $P/E$ -tal ofta följs av stora nedgångar. Det kanske tydligaste exemplet är värderingen av företag i början av 2000-talet, precis innan IT-bubblan sprack men det kan tydligt ses samma indikationer med en peak i  $P/E$ -talet runt 1990 precis innan fastighetskrisen i Sverige drog igång. Under de senaste åren har  $P/E$ -talet sakta stigit. Historiskt kan dock konstateras att ökningen inte är speciellt dramatisk, varken hastighetsmässigt eller storleksmässigt. Det ökande  $P/E$ -talet väcker dock frågor kring huruvida värderingen av bolag i dagsläget kan leda till en framtida kris.

## 2.2 Value at Risk

Value at Risk ( $VaR$ ) innebär precis som namnet antyder att ett värde är utsatt för risk.  $VaR$  definieras på ett antal olika, dock väldigt snarlika, sätt. I fortsättningen kommer definitionen som denna rapport utgår ifrån lyda *maximala förlusten som kan uppkomma under ett given tidsperiod,  $T$ , som inte kommer överskridas med en given konfidenznivå,  $c$*  [?]. Detta kan vidare formuleras matematiskt. Man pratar då ofta om två typer av  $VaR$ , nämligen relativt  $VaR$  och absolut  $VaR$ . Definitionerna kan utläsas i ekvation (??) och (??). Låt

$$\begin{aligned} V_p &= \text{Portföljens värde idag,} \\ r_p &= \text{Portföljavkastning,} \\ r_c &= \text{Lägsta portföljavkastningen vid konfidenznivå } c. \end{aligned}$$

$$VaR^{rel} = -r_c V_p + \bar{r}_p V_p, \quad (2.2)$$

$$VaR^{abs} = -r_c V_p. \quad (2.3)$$

Oavsett om man är intresserad av att mäta ett relativt eller absolut  $VaR$  finns en definition som alltid håller. Genom att låta  $f(r_p)$  beteckna täthetsfunktionen för  $r_p$  definierar vi  $r_c$  som

$$P(r_p \leq r_c) = \int_{-\infty}^{r_c} f(r_p) dr_p = 1 - c \quad (2.4)$$

$VaR$  är det i dagsläget populäraste riskmåttet för finansiella institutioner [?],[?]. Fördelen med  $VaR$  är att man komprimerar en stor mängd information från en

fördelning till ett tal som är lätt att tolka [?]. I denna rapport avses emellertid inte *VaR* användas i riskhanteringssyfte utan istället för att hjälpa oss definiera en tillräckligt stor nedgång för att vi skall kunna kalla den börskrasch.

### 2.2.1 Hur hittar vi det bästa *VaR*-mättet

Förutom det vedertagna användandet av *VaR* finns en annan anledning till varför vi väljer att använda *VaR*. Att definiera en börskrasch i termer av *VaR* gör att måttet för börskrasch anpassar sig till rådande marknadsförhållanden. Detta motiveras av *stylized facts of financial data* [?], som diskuteras mer på sida ??, samt det faktum att den logistiska regressionen kräver oberoende data. Tanken är att *VaR*-mått skall hjälpa till med skapa oberoende överskridanden, då det uppdateras till rådande marknadsförhållanden.

Eftersom finansiella tidsserier inte är stationära [?] ansätts att de metoder med vilka vår börskrasch skall beräknas utifrån inte heller är stationära. Tidsvarierande modeller bör baseras på frekvent samplade datapunkter [?]. Alexander [?] menar vidare att om man använder tidsvarierande tidsserier inom finans bör avkastningar samplas minst varje dag, eftersom att lägre sampling inte tar hänsyn till klustring som är något som syns i de absolut flesta finansiella tidsserier.

För att skapa tillfredsställande *VaR* mått har två metoder för att beräkna volatilitet valts ut. De volatilitetsmått som valts är *EWMA* och *GARCH*(1,1). Dessutom har två metoder valts för att bestämma *VaR*. Den ena metoden antar att avkastningarna är normalfördelade medan den andra bygger på historisk simulering med volatilitetsuppdateringar.

### EWMA - bakgrund och matematisk beskrivning

EWMA (*Exponentially-Weighted Moving Average*) är en modell som bygger på att skatta volatilitet med geometriskt avtagande vikter[?]. Detta innebär alltså att dagens volatilitet får större vikt än gårdagens när morgondagens volatilitet skall skattas. Fördelen med detta är att nyare och mer aktuell information ges större vikt vid skattningen av volatiliteten[?]. EWMA används flitigt på marknaden, bland annat rekommenderas den av JP Morgans Risk Metrics. Matematisk uppbyggnad av modellen följer nedan. Låt

$$\begin{aligned} S_i &= \text{Värdet på portföljen vid tid } i = 1, \dots, T \\ r_i &= \frac{S_i - S_{i-1}}{S_{i-1}} \end{aligned}$$

En vanlig variansskattning ges av

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^m r_{n-i}^2, \quad (2.5)$$

men genom att byta ut  $\frac{1}{m}$  mot  $\alpha_i$ , där  $i = 1..m$  och  $\sum_1^m \alpha_i = 1$  kan vi skriva

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i r_{n-i}^2. \quad (2.6)$$

Genom att sätta  $\lambda_i = \frac{\alpha_{i+1}}{\alpha_i}$  så har vi åstadkommit en exponentiell viktning. Insättning i (2.6) ger

$$\sigma_n^2 = \alpha_1 \sum_{i=1}^m \lambda^{i-1} r_{n-i}^2 \quad (2.7)$$

Vi ser att  $\lambda^{i-1}$  är en geometrisk serie och det följer då att  $\sum_1^m \lambda^{i-1} = \frac{1}{1-\lambda}$  då  $m \rightarrow \infty$ . Utnyttjar vi faktumet att  $\sum_1^m \alpha_i = 1$  får vi att  $\alpha_1 \frac{1}{1-\lambda} = 1$  så gäller att

$$\sigma_n^2 = (1 - \lambda) \sum_1^m \lambda^{i-1} r_{n-i}^2. \quad (2.8)$$

Det är sedan relativt enkelt att visa att man från detta uttryck kan ta sig till

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) r_{n-1}^2, \quad (2.9)$$

då  $m \rightarrow \infty$ . Man vet att  $\lambda = 0,94$  passar bra till en stor bredd av finansiella tidsserier [?] och därför kommer även denna rapport utgå från detta.  $\lambda$  kan dock skattas med Maximum Likelihood (ML). Detta kommer dock inte göras eftersom resultatet i någon mån skall vara universellt och en ML skattning gör att  $\lambda$  är väldigt anpassad till den specifika datan.

### GARCH(1,1)

*GARCH*( $p, q$ ) beskriver en generaliserad autoregressivt betingad heteroskedastitet. Modellen är från början utvecklad av R. Engle som endast en autoregressivt betingad heteroskedastitet (ARCH) modell men generaliserades 1986 av Tim Bollerslev [?]. För finansiella applikationer används ofta *GARCH*(1,1) och det är denna vi kommer använda i denna rapport.

Anledningen till att modellen är populär inom finans är att man genom empiriska studier av logaritmerade avkastningar kunnat notera ett antal fakta om finansiell data. Slutsatserna kallas ofta för "stylized facts of financial data" och innebär att

1. Serieberoende syns i data
2. Volatilitet ändras över tid
3. Avkastningsfördelningen har tjocka svansar,  
är asymmetrisk och är således inte normalfördelad

På grund av dessa fakta lämpar sig inte en vanlig slumpvandring väldigt väl och det är också därför som ARCH och slutligen GARCH modellen utvecklades [?]. Det kommer i denna rapport inte redogöras för beviset bakom varken ARCH eller GARCH, däremot är originalrapporterna med som källor och bevisen finns att läsa där i sin helhet. Däremot kommer en konceptuell överblick ges av hur GARCH modellen fungerar. Först definieras *GARCH*(1,1) processen.

Låt  $(Z_t)$  vara en sekvens av likformigt, oberoende fördelade slumpstal så att  $Z_t \sim N(0, 1)$ .  $(r_t)$  kallas då en *GARCH*(1,1) - process om

$$r_t = \sigma_t Z_t, t \in Z, \quad (2.10)$$

där  $\sigma_t$  är en ickenegativ process sådan att

$$\sigma_t^2 = \gamma V_l + \alpha r_t^2 + \beta \sigma_{t-1}^2, \quad (2.11)$$



där  $\alpha$ ,  $\gamma$  och  $\beta$  är parametrar och  $V_t$  kan tolkas som variansens långsiktiga medelvärde, kallat *long run average*.

### 2.2.2 Normalfördelningsantagandet

I många fall används ett antagande om normalfördelning för att simulera en aktie eller en portfölj av aktier. Detta kan enkelt ses genom sortera avkastningar och sedan skapa histogram av dessa. Problemen som uppstår med dessa antaganden går framför allt att referera till de tidigare nämnda *stylized facts of financial data*. Det är också av den anledningen som de tidigare diskuterade volatilitetsskattningarna används. Med dessa skattningar går det att visa, till exempel med hjälp av *QQ*-plottar, att normalfördelningsantagandet är relevant, även om det inte är helt korrekt. Således kommer normalfördelning att antas när vi beräknar *VaR* och både *EWMA* och *GARCH*(1,1) användas till att skatta volatiliteter.

### 2.2.3 Historisk simulering med volatilitetsuppdateringar

I stället för att anta att en datamängd följer en viss sannolikhetsfördelning kan man istället använda sig av historisk simulering. Med detta menas att man undersöker scenarier för en viss tidsperiod. Dessa scenarier sorteras sedan i stigande ordning och man kan då för en given konfidenznivå enkelt välja ut vilket scenario som korresponderar mot den konfidenznivån. Se [?] för mer detaljer kring metoden.

Problemet med historisk simulering är, precis som tidigare diskuterats, att finansiella tidsserier inte är stationära. Av den anledningen kommer denna rapport behandla historisk simulering genom att uppdatera volatiliteten som föreslagits av [?]. Modellen bygger på att beräkna den potentiella avkastningen genom att skala dagens avkastning med volatiliteten vid tiden  $t$  för tillgång  $j$  med kvoten mellan avkastningen för samma tillgång vid tiden  $i$ , d.v.s.

$$r_{i,j}^* = r_{i,j} \frac{\sigma_{t,j}}{\sigma_{i,j}}. \quad (2.12)$$

[?] argumenterar med hjälp av tidsserier från bland annat valutakurser och aktieindex att denna metod är överlägsen "vanlig" historisk simulering, eftersom den tar hänsyn till mer aktuell information. En annan fördel med modellen är att den är intuitivt lätt att förstå [?].

Även om historisk simulering inte förutsätter någon sannolikhetsfördelning måste volatiliteten beräknas. Vi följer här rekommendationen från [?] och undersöker modellen med både *EWMA* och *GARCH*(1,1) volatilitet.

### 2.2.4 Kontrollera oberoende med Christoffersen test

Eftersom vår logistiska regressionsmodell bygger på att observationerna skall vara oberoende införs en teststorhet,  $\Lambda$ , med syfte att mäta oberoende mellan överskridanden av *VaR*-mättet. I denna rapport kommer vi använda en modell som introducerades av [?]. Vi ställer upp teststorheten som

$$\Lambda(x) = \frac{\max[L(\theta; x); \theta \in \Theta_0]}{\max[L(\theta; x); \theta \in \Theta]}. \quad (2.13)$$

Utifrån detta skapas en log likelihood kvot som är approximativt  $\chi^2$ -fördelad med *en* frihetsgrad, där frihetsgraderna svara mot skillnaden i antalet begränsningar mellan nollmodellen, täljare i (??), och den fullständiga modellen, nämnare i (??), [?]. Med avstamp i detta landar vår teststorhet till slut i att

$$\begin{aligned} -2\ln(\Lambda(x)) = & -2\ln[(1-\pi)^{(n_{00}+n_{10})}\pi^{(n_{01}+n_{11})}] \\ & +2\ln[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}] \end{aligned} \quad (2.14)$$

där

$$\pi = \frac{n_{01} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \quad (2.15)$$

$$\pi_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad (2.16)$$

$$\pi_{11} = \frac{n_{11}}{n_{10} + n_{11}} \quad (2.17)$$

och  $n_{ij}$  är antalet observerade överskridanden då mellan tillstånd,  $(i, j) \in (0, 1)$ , där 0 motsvarar inget överskridande och 1 motsvarar ett överskridande

## 2.3 Logistisk regression

Logistisk regression är en välkänd statistisk klassifikationsmodell [?]. Modellen använder sig av en binär responsvariabel,  $Y$ , samt en förklaringsvariabel,  $X$ , som kan anta samtliga positiva värden [?].  $Y$  antar binära värdena,  $Y = 1$  för en viss händelse och  $Y = 0$  om händelsen ej inträffar. Detta leder till att  $E[Y] \in [0, 1]$ . Således kan inte det linjära förhållandet  $E[Y] = \alpha + \beta X$  upprättas [?].  $E[Y]$  modelleras istället logistiskt med

$$\text{logit}(E[Y]) = \alpha + \beta X, \quad (2.18)$$

där *logit* står för den logistiska funktionen  $\text{logit}(p) = \log(\frac{p}{1-p})$  där  $p \in [0, 1]$ . Den logistiska regressionen utförs i tre steg

1. Sätt upp den logistiska modellen
2. Sätt upp likelihood-funktionen
3. Skatta parametrarna

### 2.3.1 Modell

Från ekvation (??) ges att

$$E[Y_i] = \frac{1}{1 + e^{-(\alpha + \beta X_i)}}. \quad (2.19)$$

Då  $Y_i$  är binär följer att variablerna är *Bernoullifördelade*. Utfallen skall även vara oberoende,  $Y_i \sim \text{ober.Bernoulli}(E[Y_i])$ .

Ekvation (??) beskriver en familj av sigmoidala kurvor. Kurvan av  $E[Y_i]$  kommer därmed att bli S-formad om,  $\beta \neq 0$ , och stigande respektive sjunkande beroende

på om  $\beta > 0$  respektive  $\beta < 0$ .

I en linjär regression,  $E[Y_i] = \alpha + \beta X_i$ , beskriver  $\beta$  lutningen av kurvan, vilket enkelt kan ses av förhållandet. Alltså beskriver  $\beta$  förändringen av  $Y_i$  vid en ökning av en enhet av  $X_i$  i alla typer av regressioner. Vi använder beteckningen

$$E[Y_i] = \pi(X_i), \quad (2.20)$$

eftersom  $E[Y_i]$  är en sannolikhet mellan  $[0, 1]$ . Lutningen av den logistiska kurvan kan alltså beskrivas som

$$\beta = \text{logit}(\pi(X_{i+1})) - \text{logit}(\pi(X_i)). \quad (2.21)$$

Med hjälp av benämningen *odds*, där  $\text{odds} = \frac{\pi}{1-\pi}$ , kan man få ett intryck av vad  $\beta$  innebär

$$\begin{aligned} \beta &= \log(\text{odds}(X_{i+1})) - \log(\text{odds}(X_i)) \\ &= \log\left(\frac{\text{odds}(X_{i+1})}{\text{odds}(X_i)}\right). \end{aligned}$$

Således kan  $\beta$  tolkas som en logaritmerad oddsrelation mellan  $\pi(X_{i+1})$  och  $\pi(X_i)$  [?].

### 2.3.2 Likelihood-funktionen för den logistiska regressionen

Eftersom  $P(Y_i = 1) = E[Y_i] = \pi(X_i)$  och  $P(Y_i = 0) = 1 - \pi(X_i)$  ger detta täthetsfunktionen

$$f(Y_i, \pi(X_i)) = \pi(X_i)^{Y_i} (1 - \pi(X_i))^{1-Y_i}. \quad (2.22)$$

Likelihoodfunktionen blir därmed

$$L = \prod_{i=1}^N f(Y_i, \pi(X_i)) = \prod_{i=1}^N \pi(X_i)^{Y_i} (1 - \pi(X_i))^{1-Y_i} = \prod_{i=1}^N \left(\frac{\pi(X_i)}{1 - \pi(X_i)}\right)^{Y_i} (1 - \pi(X_i)). \quad (2.23)$$

Från ekvation (??) och ekvation (??) följer att

$$\frac{\pi(X_i)}{1 - \pi(X_i)} = e^{\alpha + \beta X_i}$$

och

$$1 - \pi(X_i) = (1 + e^{\alpha + \beta X_i})^{-1}.$$

Detta i ekvation (??) ger likelihood-funktionen

$$L = \prod_{i=1}^N e^{Y_i(\alpha + \beta X_i)} (1 + e^{\alpha + \beta X_i})^{-1}.$$

Således blir log likelihood-funktionen

$$l = \log(L) = \sum_{i=1}^N Y_i(\alpha + \beta X_i) - \log(1 + e^{\alpha + \beta X_i}). \quad (2.24)$$

### 2.3.3 Maximera likelihood-funktionen

Vi vill få ut de parametrar som maximerar likelihood-funktionen,  $L$ , i ekvation (??). Då detta är svårgenomfört kan log likelihood-funktionen,  $l$ , istället maximeras. Då  $L$  är monoton kommer denna maximering ge samma parametrar som att maximera likelihood-funktionen. Parametrarna vi vill skatta är alltså  $\alpha$  och  $\beta$ . Skattningarna  $(\hat{\alpha}, \hat{\beta})$  är de parametrar som maximerar (??), dvs. som gör att

$$\begin{cases} \frac{dl}{d\alpha} = 0, \\ \frac{dl}{d\beta} = 0. \end{cases} \quad (2.25)$$

Vi deriverar  $l$  med avseende på  $\alpha$  och  $\beta$  vilket ger

$$\begin{cases} \frac{dl}{d\alpha} = \sum_{i=1}^N (Y_i - \pi(X_i)) = 0, \\ \frac{dl}{d\beta} = \sum_{i=1}^N X_i (Y_i - \pi(X_i)) = 0. \end{cases} \quad (2.26)$$

Ekvation (??) ger att vi får ett olinjärt ekvationssystem som, i de allra flesta fall, måste lösas numeriskt. Detta görs via att ta fram hessianen till log likelihood funktionen och identifiera att variansen av  $\hat{\alpha}$  och  $\hat{\beta}$  är  $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$ , där

$$\mathbf{V} = \text{var}(\mathbf{Y}) = \text{diag}[\pi(X_i)(1 - \pi(X_i))]$$

och

$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_N \end{pmatrix}.$$

Detta resultat bygger upp för smidiga algoritmer för att bestämma  $\hat{\alpha}$  och  $\hat{\beta}$  [?]. Exempelvis så kan Newton-Raphsons algoritm användas, då den garanteras konvergera mot maximum likelihood.

I denna rapport används MatLabfunktionen *glmfit* för att utföra regressionen. Mer om denna funktion går att finna på Math Works.

### 2.3.4 Teststorheter för att evaluera modellen

För att se om vår modell, statistiskt sett, ger några indikationer om ett samband mellan stora nedgångar och  $P/E$ -talet har två teststorheter valts. Den första av dessa är ett klassiskt t-test. Testet går ut på att man sätter upp en nollhypotes,  $H_0$ , där man antar att parametern man skattar är noll. Detta testar man mot hypotesen  $H_1$ , att parametern är nollskild.

Det andra testet är devians som utgår från en likelihoodkvot. Man tar då en kvot mellan den modell som tagits fram och som skattas av en parameter och en modell som förklarar datan med alla tillgängliga parametrar. Det är uppenbart att modellen som tar hänsyn till alla parametrar alltid kommer förklara datan bättre. Det man undersöker är om modellen som skapats är tillräckligt bra för att förklara datan.

### Hypotesprövning

En hypotesprövning testar om koefficienten i regressionsmodellen är nollskild eller ej. Intuitivt kan vi se att om  $\beta$  är nollskild så kommer  $E[Y_i]$  att bero av  $X_i$ , dvs. vår modell verkar ha hittat ett samband. Om  $\beta$  däremot inte kan sägas vara nollskild så finns inget fastställt samband och modellen är således inte relevant. För att undersöka huruvida koefficienten är nollskild eller inte ställs följande teststorheten upp

$$t = \frac{\hat{\beta} - 0}{\sqrt{V\hat{A}R(\hat{\beta})}}. \quad (2.27)$$

Det att visa att denna storhet, under  $H_0$  är approximativt normalfördelad och alltså gäller att

$$t = \frac{\hat{\beta} - 0}{\sqrt{V\hat{A}R(\hat{\beta})}} \approx N(0, 1). \quad (2.28)$$

Vi vill nu undersöka om  $\beta$  är nollskild eller inte. Hypotesen som är av intresse i denna rapport är huruvida  $\beta \leq 0$  eller om  $\beta > 0$  eftersom tesen är att  $E[Y]$  är strängt växande m.a.p.  $X$ . Det test som ställs upp för att motsvara dessa krav formuleras då genom

$$t = \frac{\hat{\beta} - 0}{\sqrt{V\hat{A}R(\hat{\beta})}} > z_c, \quad (2.29)$$

där  $z_c$  motsvaras av värdet från den inverterade kummulativa normalfördelningen för en given konfidensgrad  $\alpha$ . Vi säger alltså att om värdet på vår teststorhet,  $t$ , är större än detta värde, för en given konfidensgrad,  $c$ , kan vi utesluta att  $\beta \leq 0$  och verkar således påverka modellen positivt.

### Devians

Som beskrivet ställer man, när man evaluerar deviansen, upp två hypoteser,  $H_0$  och  $H_1$ . Dessa kan defineras som

$H_0$ : Den mindre modellen med  $p$  parametrar är lika bra som den maximala modellen med  $m$  parametrar

$H_1$ : Den maximala modellen är bättre

Deviansen defineras då som

$$D = 2(l(\hat{\beta} : y) - l(\hat{\alpha} : y)) \quad (2.30)$$

och man kan visa att denna, under  $H_0$ , är approximativt  $\chi^2$ -fördelad med  $m - p$  frihetsgrader så att

$$D \approx \chi^2(m - p). \quad (2.31)$$

Deviansen kommer alltså beskriva om vår modellen är tillräckligt bra, jämfört med den maximala modellen. På samma sätt som ovan kan detta testas med

olika konfidenznivåer. D.v.s. det kommer kunna utläsas om man kan förkasta att vår modell beskriver sambandet lika bra som den maximala modellen.

## Kapitel 3

# Genomförande

### 3.1 Insamling av data

Till utförandet av denna undersökning hämtades relevant finansiell historik från *Thomson Reuters Eikon* (*Reuters*) och *Datastream*. Först hämtas tidsserier från stängningskursen,  $c_t$ , av *OMXS30(RIC:0#OMXS30:)* mellan 1987-05-04 och 2014-05-11. Det långa spannet valdes för att få en så stor, heltäckande datamängd som möjligt.

De absoluta samt relativa avkastningarna,  $r_t^{abs}$  respektive  $r_t^{rel}$ , beräknades via

$$\begin{aligned} r_t^{abs} &= c_t - c_{t-1}, \\ r_t^{rel} &= \frac{c_t - c_{t-1}}{c_{t-1}}. \end{aligned}$$

Detta för att kunna beräkna både  $VaR^{abs}$  och  $VaR^{rel}$ .

Historik på  $P/E$ -talen hämtades sedan från *Datastream*. Denna data hämtades mellan samma tidsintervall som stängningskursen. Däremot var somliga datum skiljda mellan dataserierna. Då  $P/E$ -talen hämtats från *Datastream* och stängningspriset från *Reuters* föll det sig att  $P/E$ -talen även är definierade för röda kalenderdagar vilket ej stängningspriserna var. Detta löstes genom att i *Matlab* korrigera för datum som ej matchade och plocka bort dessa mätpunkterna. Scriptet för detta finns i Appendix ??.

Datamängderna delas in i två delar, där första delen är tänkt att uppskatta modellen på och den andra delen att testa modellen på. Första delen är mellan 1989-09-19 till 2005-09-19, medan valideringsperioden sträcker sig mellan 2005-09-20 och 2014-05-09.

### 3.2 Beräkning av $VaR$

Vid beräkning av  $VaR$  har framför allt *Matlab* använts. Koden som ligger till grund för beräkningarna finns i Appendix ?? och dessutom så är de teoretiskt förankrade i den teoretiska referensramen. Av den anledningen kommer beräkningarna utelämnas i denna del. Däremot kommer metoderna för att få fram resultatet presenteras.

### 3.2.1 Volatilitetsberäkningar

Volatilitetsberäkningarna som använts är de som presenterades i den teoretiska referensramen. För att beräkna dessa har historiska priser importerats från *Reuters* till *Excel*. Volatilitetsberäkningarna har sedan gjorts i *Excel* och de färdiga resultatet har sedan importerats till *MatLab*. Metoderna som använts för respektive volatilitet beskrivs närmare nedan.

#### GARCH(1,1)

För att ta fram parametrarna till  $GARCH(1,1)$  volatiliteten maximerades log likelihood funktionen. För att göra detta på ett effektivt sätt användes problemlösaren i *Excel*.

Eftersom det optimeringsproblemet som uppstår då MLE-funktionen ska maximeras är långt ifrån linjärt så uppstår även frågeställningen huruvida den lösning som använts är ett lokalt eller globalt optima. Av den anledningen användes både *icke-linjär GNG* och *Evolutionary* metoder. Dessutom används olika startlösningar för varje test. Det visade sig dock att datat som undersöktes var relativt enkelt att arbeta med och efter att provat ett antal olika startlösningar kunde konstateras att ett globalt maximum hittats. För att vara konsistent med teorin användes logaritmerade avkastningar istället för de tidigare presenterade relativa avkastningarna. Avkastningarna beräknades alltså genom

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right). \quad (3.1)$$

#### EWMA

Även volatiliteten skattad med EWMA togs fram i *Excel*. Detta kräver inte lika mycket numeriskt arbete då vi antagit [?] råd om att välja  $\lambda = 0,94$ .

#### Jämföring av volatiliteterna

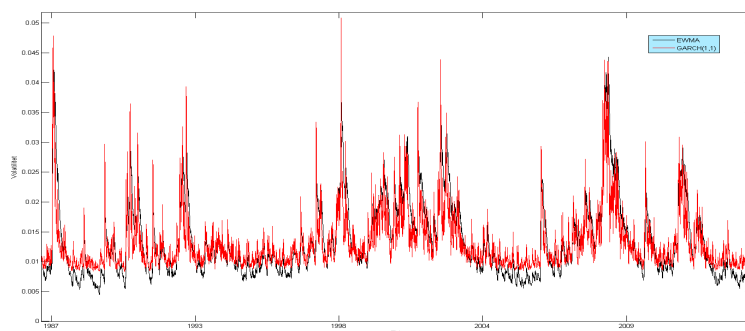
Efter att beräknat både volatiliteten kunde konstateras att  $GARCH(1,1)$  generellt var aningen högre än vad EWMA-metoden förutspådde. Detta illustreras grafiskt i figur (??). Dessutom kan vi se vissa tidsperioder med väldigt hög volatilitet. Vi kan dock se att de absolut flesta av dessa sammanfaller med perioder då börsen rört sig kraftigt.

### 3.2.2 Historisk simulering med volatilitetsuppdatering

För att genomföra den historiska simuleringen skapas *fönster* på 600 dagar som sedan simuleras till scenarion. Anledningen till detta var att då kunde det 15e sämsta värdet plockas ut, som motsvarar vår konfidsensgrad på 2,5%. Konfidsensgraden är vald tillräckligt låg för att indikera en extrem nedgång, men tillräckligt hög för att generera en tillräckligt stor datamängd. På grund av detta fönster tappades 600 datapunkter. Detta var dock kalkylerat från början, varför tidsserien som använts är längre än vad som från början planerats.

En funktion skapades i *matlab* som kan ses i Appendix ?? . Detta utfördes för både  $GARCH(1,1)$  och  $EWMA$  volatilitet och således skapades två nya *VaR*-mått.





Figur 3.1: Jämförelse mellan EWMA och GARCH(1,1) volatilitet

## 3.3 Logistisk regression

### 3.3.1 Modellering

För att på bästa sätt genomföra en logistisk regression opererades lite med datamängderna. Den logistiska regressionsmodellen arbetar efter att titta på andelen utfall under specifika värden av förklaringsvariabeln,  $X$ . Därför måste dessa,  $P/E$ -talen i detta fall, ordnas i intervall så värdena representeras av en mängd. *MatLab*-funktionen för detta går att se i appendix ??.

Med de framberäknade avkastningarna och  $VaR$ -mått utformades en binär process  $Y$  där  $Y_i = 1$  representerar att avkastningen för tidpunkt  $i$  har överskridit  $VaR$ -måttet för samma tidpunkt och  $Y_i = 0$  innebär således att  $VaR$ -måttet ej överskridits. Sedan ordnas  $P/E$ -intervallen i stigande ordning och  $Y$  sorteras för att matcha tidpunkterna i  $X$ .  $Y$  förskjuts nämligen ett tidssteg då en krasch vill jämföras mot  $P/E$ -talet dagen innan. Detta eftersom både  $P/E$ -talet och indexpriset tas från stängningsdatat. Eftersom  $P/E$ -talet linjärt beror av priset, se ekvation (??), leder detta till att  $P/E$ -talets stängningvärde blir lika relativt förändrat som indexpriset. Eftersom  $P/E$ -talet föregående dag ungefär svarar mot det ingående värdet den efterkommande dagen, känns detta som ett rimligt resonemang.  $Y$  summeras sedan ihop på varje intervall för att representera antalet överskridanden på varje  $P/E$ -intervall.

De vektorer som plockas in i *MatLabs* *glmfit*-funktion är

$X$  = De storlekssorterade intervallen  
 $Y$  = Antalet överskridanden per intervall  
 $n$  = Antalet datapunkter per intervall

Ur denna funktion ges en del data där resultat av hur bra datat passar i modellen kan utläsas. Filerna som bildas är

$b$  : Skattningarna  $\hat{\alpha}$  och  $\hat{\beta}$   
 $DEV$  : Deviansen  
 $STAT$  : En struktur innehållande testparametrar, bl.a.:  
 $se$  — Standardfelet av  $\hat{\alpha}$  och  $\hat{\beta}$   
 $t$  — Resultatet av normalfördelningstest  
 $p$  — P-värde för normalfördelningstest

### 3.3.2 Test

När parametrarna,  $\hat{\alpha}$  och  $\hat{\beta}$ , var beräknade användes valideringsdatat som tidigare uppdelats till att testa modellen.  $E[Y]$  modellerades som beskrivet i ekvation (??) på sida (??). Då  $P(Y_i = 1) = E[Y_i]$  ger resultatet en sannolikhet för att  $VaR$  överskrids. Dessa kan matchas mot de verkliga överskridandena och därmed undersöka om modellen *flaggar* för ett överskridande.

Vidare beräknas medelvärdet av det modellberäknade väntevärdet,  $\mu_{i \in Y_i=1}(E[Y_i])$ , på de datapunkter där  $Y = 1$ . Detta jämförs med det totala medelvärdet av  $E[Y]$ ,  $\mu(E[Y_i])$ , för att undersöka om det generellt är ett högre väntevärde då överskridanden har observerats. Detta kan utläsas av kvoten  $\Gamma$ , där

$$\Gamma = \frac{\mu_{i \in Y_i=1}(E[Y_i])}{\mu(E[Y_i])}. \quad (3.2)$$

## 3.4 Utförande

Samtliga steg i den logistiska regressionen utfördes ett flertal gånger med olika intervallsuppdateringar av  $P/E$ -talen. Det bestämdes dock ett förhållande för intervalluppdelningarna som uppdaterades

$$\left\{ \begin{array}{ll} P/E = \text{round}[\frac{2P/E}{k}]^{\frac{k}{2}}, & P/E < 10, \quad k = 1, \dots, 20, \\ P/E = 2 * \text{round}[\frac{P/E}{2}], & 10 \leq P/E < 25, \\ P/E = \text{round}[\frac{2P/E}{k}]^{\frac{k}{2}}, & 25 \leq P/E < 40, \quad k = 1, \dots, 20, \\ P/E = k * \text{round}[\frac{P/E}{k}], & 40 \leq P/E, \quad k = 1, \dots, 20. \end{array} \right. \quad (3.3)$$

Detta uppdelningssätt har valts med hjälp av det statistiska utseendet av  $P/E$ -talen, som kan ses i figur (??). Utgångspunkten var att  $P/E$ -talen inträffar oftare i de mellersta intervallen, och mer sällan högt och lågt. Resonemanget som indelningen bygger på är att då  $P/E$ -talet är i den övre delen av sin antagna mängd, anses det redan vara så pass högt att en förändring inte är av lika stor vikt som vid andra nivåer, då talet fortfarande bedöms som högt. Samma resonemang används för låga  $P/E$ -tal, medan de mellanliggande samlas i snävrare intervall då en förändring i  $P/E$ -talet är av större vikt.

Då samtliga tester körts, för  $k = 1, \dots, 20$ , samlades data för  $\alpha$  och  $\beta$  samt resultat av  $t$ -test och  $\Gamma$ . Alla anpassningskurvor plottades ut för varje  $VaR$ -mått och varje  $k = 1, \dots, 20$ . Sedan inleddes en utvärdering av parametrarna för att hitta den bäst passande modellen för att modellera vår datamängd.

## Kapitel 4

# Resultat och slutsatser

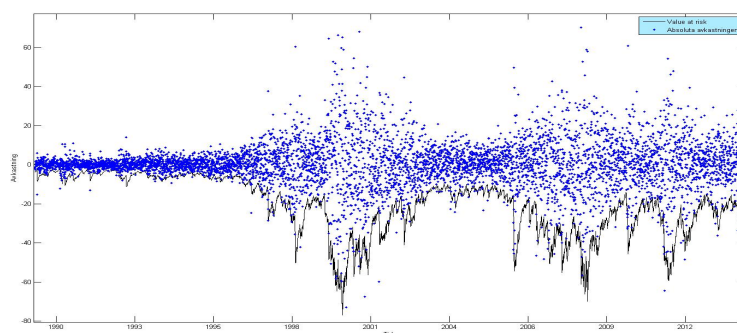
Denna del i rapporten är uppdelad i två delar. Den första delen, resultat, kommer inrikta sig på att beskriva våra resultat i form av tabeller och grafer. Den andra delen, slutsatser, kommer tolka resultaten och se om det finns några tydliga slutsatser att dra från de tester vi utfört.

### 4.1 Resultat

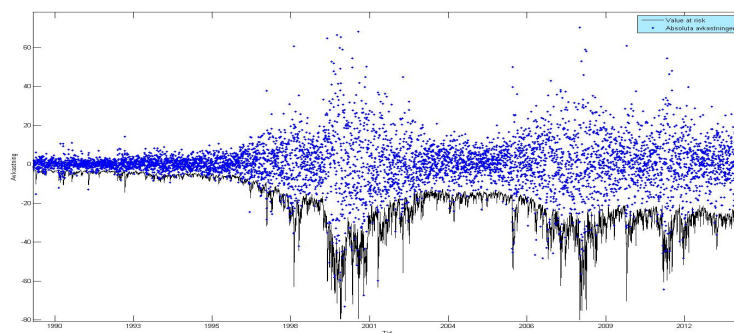
#### 4.1.1 $VaR$ – mått

De  $VaR$ -mått som beräknats illustreras nedan grafiskt i figurer ?? till ?. För den eventuellitet att någon vill se hela tidsserierna numeriskt finns all dokumentation från *MatLab* bifogad i bilaga så att det är enkelt att replikera.

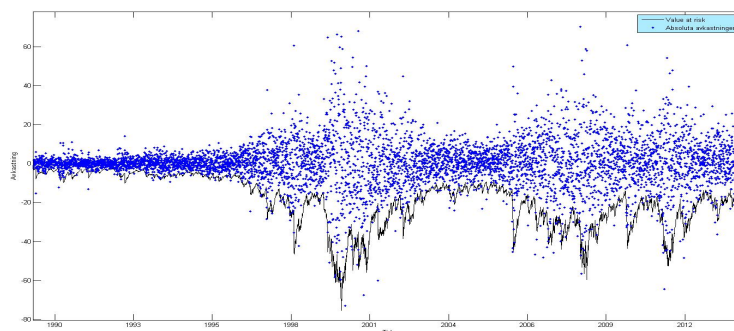
Det som syns på graferna är dels det beräknade  $VaR$ -måttet (svart linje) samt de dagliga avkastningarna (blåa punkter). De punkter som ligger nedanför, eller är mer negativa än den svarta linjen, är alltså på de ställen där  $VaR$ -måttet överskridits.



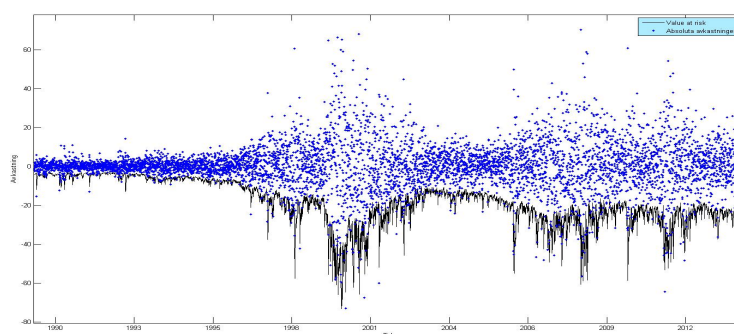
Figur 4.1: Historisk simulering med EWMA volatilitet



Figur 4.2: Historist simulering med GARCH volatilitet



Figur 4.3: Normalfördelningsantagande med EWMA volatilitet



Figur 4.4: Normalfördelningsantagande med GARCH volatilitet

Som tidigare beskrivt mäts här också endagarsberoende mellan överskridanden. Dessa presenteras i tabellen nedan.

| Volatilitetsberäkning           | Sannolikhet för oberoende |
|---------------------------------|---------------------------|
| EWMA Historisk simulering       | 0,90                      |
| GARCH(1,1) Historisk simulering | 0,95                      |
| EWMA normalfördelning           | 0,95                      |
| GARCH(1,1) normalfördelning     | 0,90                      |

Tabell 4.1: Resultat av Christofferson test på de olika  $VaR$ -mått

### 4.1.2 Logistisk Regression

Vi analyserade parameterarna vi fått ut av den logistiska regressionen för  $P/E$ -talens olika intervalluppdelning i ekvation (??) på sida (??). Från dessa presenteras parametrar nedan i tabell (??) till (??) nedan.

| k  | $\hat{\alpha}$ | $\hat{\beta}$ | t      | D       |
|----|----------------|---------------|--------|---------|
| 1  | -3,7571        | 0,0077        | 0,8070 | 55,6658 |
| 5  | -3,7344        | 0,0067        | 0,6967 | 12,6785 |
| 10 | -3,7583        | 0,0077        | 0,8050 | 9,8373  |
| 15 | -3,6920        | 0,0048        | 0,5215 | 13,7086 |
| 20 | -3,8090        | 0,0099        | 1,0110 | 5,3606  |

Tabell 4.2: VaR med Historisk simulering med  $EWMA$ -volatilitet

| k  | $\hat{\alpha}$ | $\hat{\beta}$ | t      | D       |
|----|----------------|---------------|--------|---------|
| 1  | -3,7571        | 0,0045        | 0,4933 | 77,3121 |
| 5  | -3,5746        | 0,0040        | 0,4324 | 40,3338 |
| 10 | -3,5707        | 0,0038        | 0,4102 | 36,2373 |
| 15 | -3,5280        | 0,0018        | 0,2106 | 33,5023 |
| 20 | -3,6392        | 0,0068        | 0,7259 | 18,4173 |

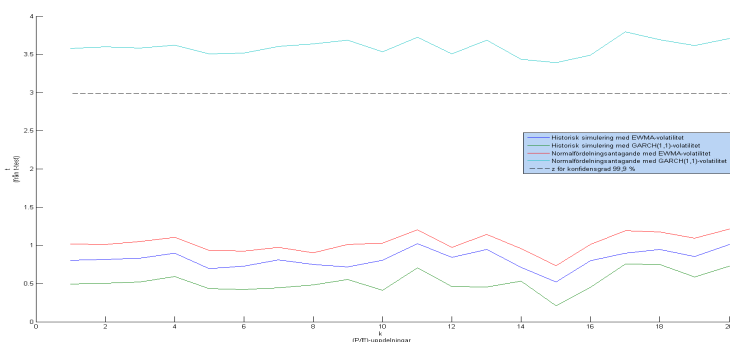
Tabell 4.3: VaR med Historisk simulering med  $GARCH(1,1)$ -volatilitet

| k  | $\hat{\alpha}$ | $\hat{\beta}$ | t      | D       |
|----|----------------|---------------|--------|---------|
| 1  | -3,6821        | 0,0092        | 1,0163 | 49,1287 |
| 5  | -3,6665        | 0,0085        | 0,9347 | 18,7396 |
| 10 | -3,6867        | 0,0093        | 1,0294 | 12,4233 |
| 15 | -3,6186        | 0,0063        | 0,7326 | 16,4852 |
| 20 | -3,7320        | 0,0113        | 1,2139 | 8,9229  |

Tabell 4.4: VaR med normalfördelningsantagande med  $EWMA$ -volatilitet

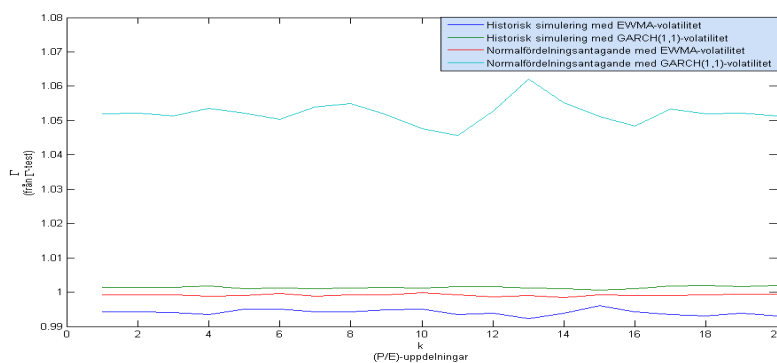
Det identifierades en tydlig överlägsen prestation av modellen då denna modellerades mot  $VaR$  måttet framtaget med normalfördelningsantagande tillsammans med  $GARCH(1,1)$ -volatiliteterna. I figur (??) nedan illustreras resultatet av t-testet för samtliga modeller.

| k  | $\hat{\alpha}$ | $\hat{\beta}$ | t      | D       |
|----|----------------|---------------|--------|---------|
| 1  | -4,1332        | 0,0307        | 3,5761 | 70,2069 |
| 5  | -4,1209        | 0,0302        | 3,5035 | 31,9994 |
| 10 | -4,1257        | 0,303         | 3,5358 | 30,1740 |
| 15 | -4,0643        | 0,0277        | 3,3909 | 28,4062 |
| 20 | -4,1932        | 0,0329        | 3,7073 | 22,7532 |

Tabell 4.5: VaR med normalfördelningsantagande med  $GARCH(1,1)$ -volatilitetFigur 4.5: Resultat av *hypotesprövning* för olika *VaR* mått.

### 4.1.3 Test på valideringsdatat

Ur våra resultat kan vi även läsa ut värdena från vårt  $\Gamma$ -test. Detta illustreras på samma sätt som  $t$ -testet i figur(?) nedan.

Figur 4.6: Resultat av  $\Gamma$  – test för olika *VaR* mått.

Det kan tydligt utläsas ur figur (?) ovan att *VaR*-mättet framtaget med  $GARCH(1,1)$ -volatilitet och normalfördelningsantagande beskrev valideringsdatat bäst. Utifrån dessa resultat valde vi att fokusera oss på modelleringen baserat på detta *VaR*-mått. Resterande delar av resultatframtagningen utgår därav ifrån dessa.

#### 4.1.4 GARCH

Då  $GARCH(1,1)$  med normalfördelningsantagandet var det som passade vår modell bäst redovisas statistiska teststorheter för denna nedan.

| k  | Hypotesprövning | D       | Christofferson | $\Gamma$ |
|----|-----------------|---------|----------------|----------|
| 1  | 3,5761          | 70,2069 | 0,90           | 1,0519   |
| 5  | 3,5035          | 31,9994 | 0,90           | 1,0521   |
| 10 | 3,5358          | 30,1740 | 0,90           | 1,0476   |
| 15 | 3,3909          | 28,4062 | 0,90           | 1,0511   |
| 20 | 3,7073          | 22,7532 | 0,90           | 1,0513   |

Tabell 4.6: Statistiska teststorheter med normalfördelningsantagande med  $GARCH(1,1)$ -volatilitet

## 4.2 Slutsatser

Den främsta slutsatsen vi kan dra är att genom att ta ut  $Var$ -mättet med  $GARCH(1,1)$  skattad volatilitet och normalfördelningsantagande passar överlägset bäst in i den logistiska regressionsmodellen. Detta styrks framför allt av de hypotesprövningar vi utfört. Resultaten från hypotesprövningen var signifikant högre för denna modell än de övriga och i och med att värdena av hypotesprövningarna är högre än 3,3 betyder det att regressionsparametern  $\hat{\beta}$  är större än noll med konfidensgrad över 99,9%. Detta betyder att  $P/E$ -talet med väldigt hög sannolikhet påverkar kraftiga börsnedgångar och kan således vara en bra förklaringsvariabel då man vill förutse dessa.

Vidare kan vi se att  $\Gamma$ -kvoten som skapats också är störst för  $Var$  skattat med  $GARCH(1,1)$  och normalfördelningsantagandet. Även detta indikerar att denna skattning fungerar bäst för att förutse börsnedgångar med hjälp av logistisk regression då  $P/E$ -talet är förklaringsvariabel.

### 4.2.1 Slutord

Precis som beskrivet i kapitel ?? så är  $Var$ -mättet det mest använda riskmättet hos finansiella institutioner. Resultatet från denna rapport är att detta mått, givet att det tagits fram med  $GARCH(1,1)$ -volatiliteter och normalfördelningsantagande, kan uppdateras och förbättras med hjälp av en logistisk regression med  $P/E$ -talet som förklaringsvariabel.

Då det test som gjorts är på OMXS30 under två olika tidsperioder finns självklart osäkerhet huruvida detta resultat kan transfereras till andra finansiella tidsserier. För att bekräfta detta resultat uppmans därför vidare forskning på ämnet.





# Litteraturförteckning

- [1] Philippe Jorion. 2007. *Value at Risk*, third edition  
ISBN-13: 978-0-07-146495-6
- [2] Hull John C. 2012. *Risk Management and Financial Institutions*, Third Edition  
ISBN 978-1-118-28638-8
- [3] Carol Alexander. 2001. *Market Models – a guide to financial data analysis*,  
ISBN 0471 89975 5
- [4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.  
(Hämtad 2014-04-25)
- [5] Posedel, P. (2005). Properties and Estimation of Garch (1, 1) model. *Metodoloski zvezki*, 2(2), 243-257.  
(Hämtad 2014-05-03)
- [6] Aswath Damodaran. 2012. *Investment Valuation*, Third Edition  
ISBN 978-1-118-20654-6
- [7] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning*.  
sid.205  
ISBN 978-0-387-31073-2
- [8] Charles E. McCulloch, Shayle R. Searle and John M. Neuhaus. 2008. *Generalized, Linear and Mixed Models, 2nd edition*, John Wiley & sons, Inc., New Jersey.
- [9] Appricon. 2010. Logistic Regression Analysis and Interpretation. <http://www.appricon.com/index.php/logistic-regression-analysis.html>  
(Hämtad 2014-04-29)
- [10] Longerstaey, J., Spencer, M. (1996). RiskMetrics™? Technical Document. Morgan Guaranty Trust Company of New York: New York.  
(Hämtad 2014-03-21)
- [11] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841-862.  
(Hämtad 2014-05-09)

- [12] Hull, J., White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1), 5-19.  
(Hämtad 2014-05-05)

# Bilaga A

## MatLab-kod

### A.1 Huvudprogram

```
clear all
close all
load 'tidsserier'
L = 20; % Antal olika uppdelningar av P/E intervall
matt = 4; % Antal olika VaR-mått
gamma = zeros(matt,L);
Beta = zeros(matt,L);
t = zeros(matt,L);
dev = zeros(matt,L);
OBER = zeros(matt,L);
for uppdel = 1:L
load 'tidsserier'
% Historiskt pris OMXS30 från 1987-05-04 --> 2014-05-11
% Absoluta avkastningarna
% Relativa avkastningar
% Volatiliteten skattat med EWMA
% Volatiliteten skattat med GARCH(1,1)
% P/E-tal
% Skapa intervall av PE talen för att få samlade utfall
PE = intervall (PE,uppdel);%funktionen rundar PE talen i bestämda intervall

% Här beräknar vi VaR med historisk simulering. Då vi använder ett fönster
% på N = 600 så tappar vi 600 mätvärden. Vi kortar därför ner samtliga
% tidsserier med 600 värden för att få rätt match.
PE = PE(1:end-600);
VaR = histuppd(vol,R_rel,pris_omxs30);

% Antag Normalfördelning och beräkna VaR med de skattade volatiliteterna
% GARCH(1,1) och EWMA
for i = 1:length(vol)-600 % -600 för att få samma längd som de andra VaR-måtten
VaR(3,i) = -norminv(0.975,0, 1)*vol(i,1)*pris_omxs30(i);
VaR(4,i) = -norminv(0.975,0, 1)*vol(i,2)*pris_omxs30(i);
end

% Kolla vilka punkter av avkastningarna som överskrider VaR och
% sätt dem till 1 i "over"-vektorn
comp = ones(4,1)*R_abs(1:length(VaR))'-VaR;
over = zeros(size(VaR));
ind = find(comp<0);
over(ind) = 1;
%Christoffersons test för att undersöka oberoenden
ober = chris(over)';
OBER(:,uppdel) = ober;
%% Nu kan vi inleda den logistiska regressionen
% Vi delar först in datat i estimerings- och valideringsdata
% Estimering: 1989-09-19 --> 2005-09-19
% Test: 2005-09-20 --> 2014-05-09
over(:,end) = [];
PE(1) = [];
```

```

PE_est = PE(2171:end);
X = PE_est; % Vi döper förklaringsvariabeln till X för att
over_est = over(:,2171:end); % underlätta sytes med rapport
Y = over_est; %Av samma anledning döper vi responsvariabeln till Y
X_val = PE(1:2170);
Y_val = over(:,1:2170);

% Sortera förklaringsvariabeln, X, i växande ordning och korrigera
% Y så idexeringen motsvarar förklaringsvariabel.
% Summera sedan de yi som är i varje P/E-intervall
[x,y,n] = sortsum(X,Y);
% Utför den logistiska regressionen
[b, DEV, STATS] = logistreg(x,y,n);
Beta(:,uppdel) = b(:,2);
%% Tillsist vill vi testa modellen på valideringsdatat och
% kolla om vi kan identifiera några av överskridandena
[Gamma] = sannolik(X_val,Y_val,b);
gamma(:,uppdel) = Gamma;
for i = 1:matt
t(i,uppdel) = STATS(i).t(2);
dfe(i,uppdel) = STATS(i).dfe;
end
dev(:,uppdel) = DEV;
end
%Vi plottar upp resultaten av t-testet från logistiska regressionen
figure()
plot(t')
plot(2.99*ones(1,L),'k--')
legend('Historisk simulering med EWMA-volatilitet',
'Historisk simulering med GARCH(1,1)-volatilitet',
'Normalfördelningsantagande med EWMA-volatilitet',
'Normalfördelningsantagande med GARCH(1,1)-volatilitet',
'z för konfidensgrad 99,9 %')
xlabel('k (P/E)-uppdelningar')
ylabel('t (från t-test)')
% Vi plottar upp resultatet av vår \Gamma-kvot
figure()
plot(gamma')
legend('Historisk simulering med EWMA-volatilitet',
'Historisk simulering med GARCH(1,1)-volatilitet',
'Normalfördelningsantagande med EWMA-volatilitet',
'Normalfördelningsantagande med GARCH(1,1)-volatilitet',
'z för konfidensgrad 99,9 %')
xlabel('k (P/E)-uppdelningar')
ylabel('\Gamma från \Gamma-test')

% Vi plottar upp hur VaR- mätten förhåller sig till avkastningarna
% Vi konstruerar en datum-axel
startdate = datenum('1989-09-19');
enddate = datenum('2014-05-11');
xData = linspace(startdate,enddate,length(VaR));
for i = 1:matt
figure()
plot(xData,fliplr(VaR(i,:)),'k')
hold on
plot(xData,fliplr(R_abs(1:end-600)),'.')
legend('Value at risk','Absoluta avkastningen')
hold off
xlabel('Tid')
ylabel('Avkastning')
datetick('x','yyyy','keepsicks')
end

```

## A.2 Övriga script

### A.2.1 Synka datum

```

clear all
load tidsserier
PE_ny = [];

```

```

skjut = 0;
for i = 1:length(datum_PE)

    if datum_PE(i) == datum_pris(i-skjut)
        PE_ny(i-skjut)=PE(i);
    else
        skjut = skjut+1;
    end

end
PE = PE_ny;
save ('PE','PE')

```

## A.3 Funktioner

### A.3.1 Uppdelning i P/E-intervall

```

function [PE]=intervall(old,gr)
PE = [];
for i = 1:length(old)
    if old(i) < 10
        PE(i) = round(old(i)/(gr/2))*(gr/2);
    elseif old(i)>= 10 && old(i) < 25
        PE(i) = round(old(i)/2)*2;
    elseif old(i) > 40
        PE(i) = round(old(i)/gr)*gr;
    else
        PE(i) = round(old(i)/(gr/2))*(gr/2);
    end
end
end

```

### A.3.2 VaR beräkning med historisk simulering

```

function [VaR_hist] = histuppd(vol,R_rel,pris_omxs30)
VaR_hist=[];
langd = length(vol);
for i=1:langd-600
    temp1=sort(R_rel(i+1:i+600)./vol(i+1:i+600,1));
    temp2=sort(R_rel(i+1:i+600)./vol(i+1:i+600,2));
    VaR_hist(1,i) = pris_omxs30(i)*vol(i,1)*temp1(15);
    VaR_hist(2,i) = pris_omxs30(i)*vol(i,2)*temp2(15);
end
end

```

### A.3.3 Backtesting med Christoffersson's metod

```

function [ober] = chris(over)
ober = [];
for i = 1:length(over(:,1))
    n00=0;
    n01=0;
    n10=0;
    n11=0;
    comp1 = [over(i,:) 0] - [0 over(i,:)];
    comp2 = [over(i,:) 0] .* [0 over(i,:)];
    comp1(1) = [];
    comp2(1) = [];
    for j = 1:length(comp1)
        if comp1(j) == -1
            n10 = n10+1;
        elseif comp1(j) == 1
            n01 = n01+1;
        elseif comp1(j) == 0 && comp2(j) == 1
            n11 = n11+1;
        elseif comp1(j) == 0 && comp2(j) == 0
            n00 = n00+1;
        end
    end
    PI = (n01+n11)/(n00+n01+n10+n11);
    pi01 = n01/(n00+n01);
end

```

```

pi11 = n11/(n10+n11);
lambda = -2*((n00+n01)*log(1-PI)+(n01+n11)*log(PI))+
2*(n00*log(1-pi01)+n01*log(pi01)+n10*log(1-pi11)+n11*log(pi11));

chitwo = 2.706; % chi2 - cdf för 10%
if (chitwo - lambda) > 0
    ober(i) = 1;
else
    ober(i) = 0;
end
end
end

```

### A.3.4 Sortera vektorerna och summera Y

```

function [x,y,n] = sortsum(X,Y)
[Xsort,ind] = sort(X);
Ysort = zeros(size(Y));
for i = 1:length(ind)
    Ysort (:,i) = Y(:,ind(i));
end

x = Xsort(1);
y = [];
n = [1];
plats = 1;
for i = 1:length(Xsort)-1

    if Xsort(i+1)==Xsort(i)
        n(plats) = n(plats)+1;
    else
        n(plats+1)=1;
        plats = plats+1;
        x = [x Xsort(i+1)];
    end
end
start = 1;
for j = 1:length(n)
    y(1,j) = sum(Ysort(1,start:start+n(j)-1));
    y(2,j) = sum(Ysort(2,start:start+n(j)-1));
    y(3,j) = sum(Ysort(3,start:start+n(j)-1));
    y(4,j) = sum(Ysort(4,start:start+n(j)-1));
    start = start + n(j);
end
end

```

### A.3.5 Logistisk regression

```

function [B,Dev,Stats] = logistreg(x,y,n)
M = length(y(1,:));
B = zeros(M,2);
Dev = zeros(M,1);
struktur = [];
figure()
for i = 1:M
    [b,DEV,STATS] = glmfit(x,[y(:,i) n], 'binomial', 'link', 'logit');
    temp = struct2cell(STATS)';
    B(i,:) = b; Dev(i) = DEV; struktur = [struktur;temp];
    yfit = glmval(B(i,:), x, 'logit', 'size', n);
    subplot(2,2,i)
    plot(x, y(:,i)./n, 'o', x, yfit./n, '-', 'LineWidth', 2)
end
f = {'beta', 'dfe', 'sfit', 's', 'estdisp', 'cvob', 'se', 'coeffcorr', 't', 'p',
    'resid', 'residp', 'residd', 'resida', 'wts'};
Stats = cell2struct(struktur,f,2);

```

### A.3.6 Beräkning av $\Gamma$ -testet

```

function [Gamma] = sannolik(X_val, Y_val, b)
expect = [];
M = length(b(:,1));
gamma = zeros(M,1);

```

---

```
for j = 1:M
    for i = 1:length(X_val)
        expect(i) = 1/(1+exp(-b(j,1)+b(j,2)*X_val(i))); % E[Y]
    end
    hitta = find(Y_val(j,:) == 1);
    gamma(j) = mean(expect(hitta))/mean(expect);
end
Gamma = gamma;
```





(20,20)(0,0)

#### Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years from the date of publication barring exceptional circumstances. The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: <http://www.ep.liu.se/>

#### Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår. Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art. Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>

© 2014, Jon Hedström, Johan Vidlund