LØSNINGSFORSLAG MET4 V24

Til sensor: Hver deloppgave teller likt og gir maksimalt 10 poeng.

Oppgave 1

(a) Avkastningen i normale måneder har en relativt symmetrisk og klokkeformet fordeling sentrert rundt 5%. Fordelingen til avkastningen i måneder med høy risiko er asymmetrisk med en tung venstre hale og vi ser et balansepunkt rundt 0%. Dette betyr at man både kan forvente seg lavere avkastning og at det er høyere risiko for store tap på SP500 i måneder med høy risiko sammenlignet med normale måneder.

(b) Vi skal her teste nullhypotesen om lik varians:

$$H_0 : \sigma^2_{high} = \sigma^2_{normal} \quad \text{mot} \quad H_1 : \sigma^2_{high} \neq \sigma^2_{normal}$$

der σ^2_{high} og σ^2_{normal} er populasjonsvariansen til månedlig avkastning i måneder med hhv. høy og normal finansiell risiko.

Her er det lurte å sette den største estimerte variansen i telleren av testobservatoren slik at en en kun trenger å sammenligne testobservatoren mot kritisk verdi i høyre hale av F-fordeling (0.975 percentilen).

Testobservatoren tar verdien

$$F = \frac{S^2_{high}}{S^2_{normal}} = 8.60^2/4.83^2 = 3.17$$

Kritisk verdi er gitt ved 0.975 percentilen i en F-fordeling med $162 - 1$ og $231 - 1$ frihetsgrader. Her kan vi bruke en tabell eller følgende R-utskrift i Vedlegg 1:

```r
qf(0.975, df1 = 161, df2 = 230)
```

```
[1] 1.325851
```

Siden $F = 3.17 > k^{161,230}_{0.975} = 1.33$, kan vi forkaste nullhypotesen om lik varians.

(c) Det fremgår ikke av oppgaven at vi ønsker å teste i en spesifikk retning, så det er ikke noen spesiell grunn til å gjøre ensidige tester. Vi utfører derfor følgende tosidige hypotesetest:

$$\mu_{high} = \mu_{normal} \quad \text{mot} \quad \mu_{high} \neq \mu_{normal}$$

der μ_{high} og μ_{normal} er forventet månedlig avkastning i måneder med henholdsvis høy og normal finansiell risiko. Fra oppgave b) vet vi at det er rimelig å anta ulik varians. Vi bruker derfor følgende testobservator:

$$T = \frac{\bar{X}_{high} - \bar{X}_{normal}}{\sqrt{S^2_{high}/n_{high} + S^2_{normal}/n_{normal}}} = \frac{0.68 - 5.52}{\sqrt{8.60^2/162 + 4.83^2/231}} = -6.48$$

Eksakt antall frihetsgrader er gitt ved Welch's formel, men i dette tilfellet er antall observasjoner så stort at vi uansett får kritisk verdi svært nær 1.96. Siden $|T| > 1.96$ kan vi forkaste H_0. Det ser ut til at det forskjell på avkastning i de to periodene. Ut fra den målte effekten $0.68 - 5.52 = -4.84$, tyder dette på at det er en mye større avkastning i normale perioder.

Oppgave 2

(a)

- **Smitte-effekt:** I modellen ser vi at antall konkurser forrige måned "smitter" over og har en effekt på antall konkurser neste måned med leddet ϕY_{t-1}. Smitte-effekten er derfor representert ved koeffisienten ϕ.
- **Systematisk risiko:** Den systematiske risikoen er reprentert ved å inkludere den makroøkonomiske variabelen `GDP` og den finansielle variabelen `SP500vol` i modellen, og effekten av disse er derfor β_1 og β_2.
- **Signifikante?:** Fra utskriften i Vedlegg 2 ser vi at estimatene $\hat{\phi}$, $\hat{\beta}_1$ og $\hat{\beta}_2$ er alle signifikant forskjellige fra null. Basert på denne modellen kan dette tyde på at både smitte-effekten og den systematiske risikoen er tilstede.

  Merk: Her kan vi ikke tolke koeffisientene på vanlig måte pga av det autoregressive leddet. Se løsningforslag av oppgave f).

(b)

- **Residualplott:** Antydning til både heteroskedastisitet og feil ved lineærantagelsen ved at det både er variasjon i spredningen rundt 0 og symmetrien rundt 0. Det mest iøyenfallende ved residualplottet er de systematiske stripene noe som også tyder på modellfeil.

  (Stripene har følgende forklaring, som ikke kreves for full pott: Vi har her log-transformert en variabel som bare kan ta heltall (antall konkurser per måned + 1) og i datasettet er det svært mange 0ere, 1ere, 2ere og 3ere, som da svarer til de tre nederste "stripene" i residualplottet. Vi har ikke lært om dette i MET4, men vi burde her heller brukte en modell som kan håndtere heltall direkte.)
- **Histogram og QQplott:** Histogrammet tyder på at residualene har en litt assymetrisk fordeling sammenlignet med normalfordelingen og QQ-plottet viser at residualene avviker fra normalfordelingen i høyre hale. Totalt sett er ikke dette et katastrofalt avvik fra normalfordelingen.
- **Autokorrelasjonsplott:** Autokorrelasjonsplottet viser at det er autokorrelasjon mellom residualene ved flere lag. Y_t er en tidsrekke som kan være autokorrelert.

  (Ikke nødvendig for full pott: Figuren tyder på at det autoregressive leddet ϕY_{t-1} og de andre to forklaringsvariablene ikke ikke klarer å plukke opp autokorrelasjonen i Y_t. Dette kan tyde på både at smitte-effekten strekker seg lengre tilbake i tid og at de to andre forklaringsvariablene har en forsinket effekt.)

  Samlet sett bør en være litt skeptisk til denne modellen.

(c) Formel for konfidensintervall for en regresjonskoeffisient β er $\widehat{\beta} \pm t_{1-\alpha/2} \cdot S(\widehat{\beta})$, der $S(\widehat{\beta})$ er det estimerte standardavviket til koeffisientestimatet. I dette tilfellet har vi såpass mange observasjoner at vi kan sette t-kvantilen $t_{1-0.05/2}$ til 1.96 for et 95% konfidensitervall, og får

$$[-0.035 \pm 1.96 \cdot 0.013] = [\mathbf{-0.0605, -0.0095}].$$

(d) Nei, i utgangspunktet ikke. For å predikere Y_{t+1} med denne modellen må vi også vite verdien av $\texttt{GDP}_{t+1}$ og $\texttt{SP500vol}_{t+1}$, altså verdien av forklaringsvariablene neste måned. Man måtte i såfall hatt tilgang til egne prediksjoner av disse.

(e) Interaksjonsleddene tillater at effekten av de to risikofaktorene kommer an på om det er en måned med høy finansiel risiko eller ikke. Det virker ikke å være signifikant forskjell på den systematisk risikoen i måneder med normal og høy risiko siden interaksjonsleddene mellom `riskperiod` og de to variablene `GDP` og `SP500vol` ikke er signifikante. Det kan derimot se ut til at det er en større smitteeffekt i måneder med høy risiko siden interaksjonsleddet mellom `riskperiod` og Y_{t-1} er signifikant positivt. Vi ser at $\phi = 0.119$ i normale måneder, mens $\phi = 0.119 + 0.165 = 0.284$ i risikomåneder.

(f)

$$\begin{aligned}
E(Y_t) &= E\left(\beta_0 + \phi Y_{t-1} + \beta_1 X_t + \epsilon_t\right) \\
&= \beta_0 + \phi E(Y_{t-1}) + \beta_1 E(X_t) + 0 \\
&= \beta_0 + \phi E(\beta_0 + \phi Y_{t-2} + \beta_1 X_{t-1} + \epsilon_{t-1}) + \beta_1 E(X_t) \\
&= \beta_0 + \phi\beta_0 + \beta_1 E(X_t) + \phi\beta_1 E(X_{t-1}) + \phi^2 E(Y_{t-2}) \\
&= (1+\phi)\beta_0 + (1+\phi)\beta_1\mu_x + \phi^2 E(Y_{t-2}) \quad \text{siden} \quad E(X_t) = E(X_{t-1}) = \mu_x \\
&= (1+\phi+\phi^2)\beta_0 + (1+\phi+\phi^2)\beta_1\mu_x + \phi^3 E(Y_{t-3}) \\
&\vdots \\
&= (1+\phi+\cdots+\phi^{k-1})\beta_0 + (1+\phi+\cdots+\phi^{k-1})\beta_1\mu_x + \phi^k E(Y_{t-k}) \\
&\vdots \\
&= \frac{\beta_0}{1-\phi} + \frac{\beta_1\mu_x}{1-\phi} + 0, \quad \text{siden} \quad |\phi| < 1 \\
&= \frac{\beta_0 + \beta_1\mu_x}{1-\phi}
\end{aligned}$$

Formelen over viser hvordan X_t påvirker Y_t over tid siden $E(Y_t)$ kan ses på som gjennomsnittet til Y_t i det lange løp. Dersom $\beta_1 \neq 0$ forskyves $E(Y_t)$ med $\beta_1\mu_x/(1-\phi)$ fra hva som hadde vært $E(Y_t)$ dersom $\beta_1 = 0$.

INFO (ikke nødvendig for full pott): Generelt når man har autoregressive ledd som ϕY_{t-1} med i regresjonsmodellen kan man ikke tolke β_1 på vanlig måte. Man kan i prinsippet si at β_1 er endringen i $E(Y_t|Y_{t-1}, X_t)$ når X_t øker med en enhet, men dette er ikke særlig intuitivt og tolkningen over er kanskje lettere å forstå.

Oppgave 3

(a) La $p = P(\text{konkurs} \mid \texttt{re}, \texttt{ebit}, \texttt{me}, \texttt{s})$. Den estimerte modellen er da:

$$p = \frac{\exp(z)}{1+\exp(z)}, \quad z = \widehat{\beta}_0 + \widehat{\beta}_1\texttt{re} + \widehat{\beta}_2\texttt{ebit} + \widehat{\beta}_3\texttt{me} + \widehat{\beta}_4\texttt{s} = -5.248 - 0.962\texttt{re} + 39.604\texttt{ebit} + 2.724\texttt{me} - 2.375\texttt{s}.$$

Tolkning av `re`: Dersom de andre forklaringsvariablene er uendret vil èn enhets økning i variabelen `re` være assosiert med at oddsen for at en selskap går konkurs avtar med en multiplikativ faktor $e^{-0.962} = 0.382$ (en reduksjon på ca. 62%). Det kan se ut til at selskaper som kan vise til god historisk fortjeneste har lavere sjanse for å gå konkurs.

(b) Vi regner først ut den lineære komponenten:

$$z = \widehat{\beta}_0 + \widehat{\beta}_1\texttt{re} + \widehat{\beta}_2\texttt{ebit} + \widehat{\beta}_3\texttt{me} + \widehat{\beta}_4\texttt{s} = -5.248 - 0.962 \cdot 0.4 + 39.604 \cdot 0.05 + 2.724 \cdot 0.74 - 2.375 \cdot 0.1 = -1.874.$$

Da angir modellen følgende sannsynligheten for å at dette selskapet skal gå konkurs:

$$p = \exp(-1.874)/(1+\exp(-1.874)) = 0.133$$

og siden $0.133 > \delta = 0.1$ klassifiserer vi dette selskapet til å gå konkurs; $\hat{Y} = 1$.

(c) Siden investeringsselskapet har lav toleranse for risiko er man veldig opptatt av å styre unna de selskapene som går konkurs. Man bør derfor velge en terskelverdi som gir høy sensitivitet, altså $\delta = 0.05$. Ut fra figuren vil vi da ha nærmere 90% sannsynlighet for å klassifisere riktig selskaper som kommer til å gå konkurs. Dette vil gå på bekostning av at vi "går glipp av" mange selskaper som faktisk ikke går konkurs, siden sannsynligheten for å klassifisere selskaper som ikke går konkurs til å gå konkurs (1-spesifisitet) er høyere for denne terskelverdien sammenlignet med de andre terskelverdiene.