HW 5

Kaiyi Li

kl2538

N18881575

This R notebook is created by Kaiyi Li for volatility class homework 5 at 10:27 AM Friday, March 20, 2020.

Github link: https://github.com/LiKaiy1/volatility/blob/master/HW5/HW5_Kaiyi_Li.pdf

1.

Before anything started, we read the dataset spd.csv

```
spd <- read.csv('spd.csv')
spd <- data.frame(
  year = spd$year.t,
  Dt = spd$X..defaults..Dt.,
  Nt = spd$X..obligors..Nt.
)
spd
```

a)

$$\hat{p} = \frac{1}{T}\sum_{t=1}^{T}\frac{D_t}{N_t}$$

```
spd = cbind(spd,pt = spd$Dt/spd$Nt)
spd
```

```
average_default_prob <- sum(spd$pt)/length(spd$year)
average_default_prob
```

Therefore, the average default probability would be **0.001144022**.

b)

Given

$$
\begin{gathered}
A_i = w_i Z + \sqrt{1-w_i^2}\varepsilon_i \\
cov(\varepsilon_i, \varepsilon_j) = 0 \\
cov(Z, \varepsilon_j) = 0 \\
\forall i
\end{gathered}
$$

Find $\rho_{i,j}^{asset}$:

For $E(A_i), E(A_j)$, they can be calculated as

$$
\begin{gathered}
E(A_i) = w_i E(Z) + \sqrt{1-w_i^2} E(\varepsilon_i) \\
E(A_j) = w_j E(Z) + \sqrt{1-w_j^2} E(\varepsilon_j)
\end{gathered}
$$

For the covariance, we have

$$
\begin{gathered}
Cov(A_i, A_j) = E[(A_i - E(A_i))(A_j - E(A_j))] \\
= E[(w_i(Z - E(Z)) + \sqrt{1-w_i^2}(\varepsilon_i - E(\varepsilon_i)))(w_j(Z - E(Z)) + \sqrt{1-w_j^2}(\varepsilon_j - E(\varepsilon_j)))] \\
= E[w_i w_j (Z - E(Z))^2 + w_i \sqrt{1-w_j^2}(Z - E(Z))(\varepsilon_j - E(\varepsilon_j)) + \\
w_j \sqrt{1-w_i^2}(Z - E(Z))(\varepsilon_i - E(\varepsilon_i)) + \sqrt{1-w_i^2}\sqrt{1-w_j^2}(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))] \\
= E[w_i w_j Var(Z) + w_i \sqrt{1-w_j^2} Cov(Z, \varepsilon_j) + \\
w_j \sqrt{1-w_i^2} Cov(Z, \varepsilon_i) + \sqrt{1-w_i^2}\sqrt{1-w_j^2} Cov(\varepsilon_i, \varepsilon_j)] \\
\text{Given we have} \\
Cov(\varepsilon_i, \varepsilon_j) = 0, Cov(Z, \varepsilon_j) = 0, Cov(Z, \varepsilon_i) = 0 \\
\text{So,} \\
Cov(A_i, A_j) = w_i w_j Var(Z)
\end{gathered}
$$

Eventually, we have the correlation as $\rho_{i,j}^{asset}$, which is calculated as

$$
\begin{aligned}
\rho_{i,j}^{asset} &= \frac{Cov(A_i, A_j)}{\sigma_{A_i}\sigma_{A_j}} \\
&= \frac{w_i w_j Var(Z)}{\sigma_{A_i}\sigma_{A_j}}
\end{aligned}
$$

Note that if we suggest A, Z, ε all follows normal distribution $\mathcal{N}(0,1)$, we would have

$$
\rho_{i,j}^{asset} = \frac{Cov(A_i, A_j)}{1 * 1} = w_i w_j
$$

c)

The default probability for all obligors, p, and the default correlation becames

$$\rho_{ij} = \frac{p_i p_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}}$$
Given that
$$p_i = p_j = p$$
$$\rho_{ij} = \frac{p}{1-p}$$

However, the asset correlation, $\rho_{i,j}^{asset}$ has nothing to do with default probability, and therefore should remain the same.

But with all the companies have the same default probabilities, $w_i = w_j = w$. And then the asset correlation can be written as

$$\rho_{i,j}^{asset} = \frac{w^2 Var(Z)}{\sigma_{A_i}\sigma_{A_j}}$$

Or as we assume normality for all variables, we would have $\rho_{i,j}^{asset} = w^2$.

d)

$$\hat{p}_{2t} = \frac{D_t(D_t - 1)}{N_t(N_t - 1)}$$

```
spd = cbind(spd,p2 = (spd$Dt*(spd$Dt-1))/(spd$Nt*(spd$Nt-1)))
spd
```

```
average_joint_default_prob <- sum(spd$p2)/length(spd$year)
average_joint_default_prob
```

Therefore, the average probability for joint defaults over T years would be **0.000002199**.

e)

The correlation then is

$$\begin{aligned}\rho_{i,j}^{asset} &= \frac{Cov(A_i, A_j)}{\sigma_{A_i}\sigma_{A_j}} \\ &= \frac{w_i w_j Var(Z)}{\sigma_{A_i}\sigma_{A_j}} \\ &= w_i w_j\end{aligned}$$

f)

$$p_{ij} = \Phi_2(d_i, d_j, \rho_{ij}^{asset})$$

First, we find out the default threshold d:

```
d <- qnorm(sum(spd$pt)/length(spd$year))
d
```

The default threshold is -3.050049. The joint probability of default is estimated to be 0.000002199.

Find the asset correlation which makes joint probability of default same as our estimate.

```
library(rootSolve)
library(pbivnorm)
f <- function(rho){pbivnorm(d,d,rho) - average_joint_default_prob}
uniroot(f,c(0,1))$root
```

The asset correlation ρ_{ij}^{asset} is **0.4885019**.

2.

```
lgd <- as.data.frame(read.csv('lgd.csv'))
lgd
```

Question a and b are done in excel using VLOOKUP.

a)

```
lgd_modified <- read.csv("lgd_modified.csv",header = TRUE)
tail(lgd_modified$LGD_A)
```

The last five value of LGD_A are 0.538, 0.538, 0.365, 0.538, 0.538.

b)

```
head(lgd_modified$I_DEF)
```

The first five value of I_DEF are 1.415, 1.415, 1.183, 2.353, 2.353.

c)

```
# LGD <- lgd_modified$LGD
# LEV <- lgd_modified$LEV
# LGD_A <- lgd_modified$LGD_A
# I_DEF <- lgd_modified$I_DEF
regression <- lm(LGD~LEV+LGD_A+I_DEF,data = lgd_modified)
summary(regression)
```

The summary of this regression is

```
Call:
lm(formula = LGD ~ LEV + LGD_A + I_DEF, data = lgd_modified)

Residuals:
     Min       1Q   Median       3Q      Max
-0.67598 -0.19928  0.05206  0.24367  0.44643

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  0.57015    0.10195   5.593 4.89e-08 ***
LEV          0.20414    0.06983   2.923  0.00372 **
LGD_A       -0.14364    0.16216  -0.886  0.37642
I_DEF        0.02212    0.01280   1.729  0.08480 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2832 on 312 degrees of freedom
Multiple R-squared:  0.04593, Adjusted R-squared:  0.03675
F-statistic: 5.006 on 3 and 312 DF,  p-value: 0.0021
```

The estimated model is (has cluster errors)

$$L\hat{G}D = 0.57015 + 0.20414LEV - 0.14364LGD\_A + 0.02212I\_DEF$$

d)

```
new_data <- as.data.frame(
  cbind(
  LEV = c(0.607452818682007),
  LGD_A = c(0.365),
  I_DEF = c(3.783))
)
# new_data
predict(regression,newdata=new_data)
```

The predicted result is 0.7254252.

e)

```
LGD <- lgd_modified$LGD
a=mean(LGD)/var(LGD)*(mean(LGD)*(1-mean(LGD))-var(LGD))
b=(1-mean(LGD))/var(LGD)*(mean(LGD)*(1-mean(LGD))-var(LGD))
# u=pbeta(LGD, a,b)
# hist(u)
```

a is 1.125016, b is 0.6023947.

f)

```
lgd_t <- as.data.frame(
  cbind(
    lgd_modified,
    TLGD=qnorm(pbeta(LGD, a,b))
  )
)
# lgd_t
regression_normal <- lm(TLGD~LEV+LGD_A+I_DEF,data = lgd_t)
summary(regression_normal)
```

```
Call:
lm(formula = TLGD ~ LEV + LGD_A + I_DEF, data = lgd_t)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.78209 -0.56487 -0.01416  0.70799  1.84399 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) -0.30625    0.33393  -0.917  0.35979   
LEV          0.59802    0.22872   2.615  0.00937 **
LGD_A       -0.36898    0.53115  -0.695  0.48777   
I_DEF        0.06540    0.04191   1.560  0.11969   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9277 on 312 degrees of freedom
Multiple R-squared:  0.0365,	Adjusted R-squared:  0.02723 
F-statistic: 3.939 on 3 and 312 DF,  p-value: 0.008811
```

$$\hat{TLGD} = -0.30625 + 0.59802LEV - 0.36898LGD\_A + 0.06540I\_DEF$$

g)

```
#predict
tlgd_pred <- predict(regression_normal,newdata=new_data)
#transform original scale
bLGD=qbeta(pnorm(tlgd_pred),a,b)
bLGD
```

The predicted LGD is 0.7789802.