Assignment 5.1, 5.2 and 5.3: Computing Distance to Default

Submission Details

- Submit through Canvas
- Use R Markdown for Assignments 5.1 - 5.3
- There are three parts to this assignment with different due dates
- Note, Method 3 takes a lot more time. Please get started early
  - Method 1 and 2: Due on Oct 17
  - Method 3: Due on Oct 24
- You have to submit ONLY
  - RMD file
  - Output in PDF format (Hide Code)
  - You don't need to submit any datasets

Computing Distance Default: Three Approaches

Method -1: Naive Computation

$$DD_{naive} \equiv \frac{log(E + F/F) + (r_{it-1} - \text{Naive}\sigma_V^2/2)T}{\text{Naive}\sigma_V \sqrt{T}}$$

where

$$\text{Naive}\sigma_V = \frac{E}{E+F}\sigma_E + \frac{F}{E+F}(0.05 + 0.25 * \sigma_E)$$

r_{it-1} is the firm's stock return over the previous year

- explore the results using naive $\sigma_D = (0.05 + 0.5 * \sigma_E)$
- explore the results using naive $\sigma_D = (0.25 * \sigma_E)$

Method -2: Directly Solving for the Unknowns

Based on the Black-Scholes formula, value of the equity is

$$E = V\mathcal{N}(d_1) - e^{-rT}F\mathcal{N}(d_2)$$

where

- E is the market value of the firm's equity,
- F is the face value of the firm's debt,
- r is the instantaneous risk-free rate,
- $\mathcal{N}(.)$ is the cumulative standard normal distribution function,
- $$d_1 = \frac{log(V/F) + (r + \sigma_V^2/2)T}{\sigma_V\sqrt{T}}$$
- $$d_2 = d_1 - \sigma_V\sqrt{T}$$

In this model, the second equation, using an application of Ito's lemma and the fact that $\frac{\partial E}{\partial V} = \mathcal{N}(d_1)$, links the volatility of the firm value and the volatility of the equity.

$$\sigma_E = \frac{V}{E}\mathcal{N}(d_1)\sigma_V$$

The unknowns in these two equations are

- the firm value V and
- the asset volatility σ_V.

The known quantities are

- equity value E,
- face value of debt or the default boundary F,
- risk-free interest rate r,
- time to maturity T.

Two nonlinear equations and two unknowns, we can directly solve for

$$V, \sigma_V$$

Method -3: Iterative Method

Alternately, as in KMV model, we can iteratively solve for V, σ_V,

- by starting with an initial value of σ_V,
- using the equity option equation to solve for asset value V for the sample period,
- construct the time-series of asset value and use this to compute the an estimate of σ_V.
- This process is repeated till the value of σ_V converges.
- Note: You are required to do the estimation for each year as of Jan 1. that means, one has to compute, say for Jan 1989, $sigma_V$ iteratively based on Jan 1 1988- Dec 31, 1988 stock data.

Computing Distance to Default Assignment: Method 1 and Method 2

Compute Distance to Default (DD) and Probability of Default (PD) for the sample of firms in the CRSP-COMPUSTAT intersection for the period 1970-2015. Use Method-1 (naive method) and Method-2 (direct solving) to compute DD. In all cases assume T=1 year.

1. Intersect COMPUSTAT and CRSP data (details given later in the assignment)
2. Compute DD and PD for these firms for each year (as of Jan 1 of each year)
3. Tabulate the following for the two DD and PD measures
    - Number of observations
    - Mean,
    - 25^{th}, 50^{th} and 75^{th} percentiles and
    - standard deviation
    - Minimum and Maximum of the variable
4. Compute the correlations between these two measures of DD and PD
5. Plot the mean, 25^{th}, 50^{th} and 75^{th} percentiles of DD for the two measures across time (for each year). Do you see any trends?
6. Get NBER recession data (`https://research.stlouisfed.org/fred2/series/USREC`). Compute the descriptive statistics for NBER recession = 1 and NBER recession = 0. Plot DD and PD with the recession data. What do you notice?
7. Get Moody's BAA-Fed Fund Spread(`https://research.stlouisfed.org/fred2/series/BAAFFM`). Plot DD and PD along with BAA spread. Reflect and comment.
8. Get Cleveland Financial Stress Index (`https://research.stlouisfed.org/fred2/series/CFSI`). Its not available for the entire time period. Plot DD and PD along with stress index. Reflect and comment.

Computing Distance to Default Assignment: Method 3

Compute Distance to Default (DD) and Probability of Default (PD) for the sample of firms in the CRSP-COMPUSTAT intersection for the period 1970-2015. Use Method-3 (iterative approach) to compute DD. Assume T=1 year.

- Intersect COMPUSTAT and CRSP data (details given later in the assignment)
- Compute DD and PD for these firms for each year (as of Jan 1 of each year)
- Tabulate the following for the DD and PD measure
  - Number of observations
  - Mean,
  - 25^{th}, 50^{th} and 75^{th} percentiles and
  - standard deviation
  - Minimum and Maximum of the variable
- Compute the correlations between Method 2 (iterative) with Method 1 (naive) and Method 2 (direct) measures of DD and PD
- Plot the mean, 25^{th}, 50^{th} and 75^{th} percetiles of DD for the iterative method along with the two measures (in the first part of the assignment) across time (for each year)

Data Extraction

- In order to reduce the time demands for the assignment, I have downloaded the required dataset and placed it on QCF servers
- You need to use the following two datasets
  - CRSP: dataset is DSF.SAS7BDAT (Same as used in Assignment 3).
  - COMPUSTAT FUNDA.SAS7BDAT (make sure to first subset and include only the variables required and with the appropriate filters)
  - You can use read sas function from library(haven) R package to import SAS7BDAT format dataset
- In addition, download risk free interest rate (3 month TB yields) from FRED, say `https://research.stlouisfed.org/fred2/series/DTB3` and save it as DAILYFED
- Combine the two datasets FUNDA and DSF based on CUSIP YEAR. Details below
- The risk-free interest rate is in another dataset called DAILYFED. Keep only one observation in January (so that the data is also at YEAR level) and combine with (FUNDA + DSF) based on YEAR

Details of the datasets and variables required for the assignment

- FUNDA (COMPUSTAT) data.
  - These are useful R packages for data manipulation: library(dplyr), library(lubridate)
  - Filters: filter(INDFMT == 'INDL' & DATAFMT == 'STD' & POPSRC == 'D' & FIC == 'USA' & CONSOL == 'C' & YEAR >= 1970 & YEAR <= 2017)
  - Variables required: CUSIP (CUSIP = substr(CUSIP,1,8)), YEAR (YEAR = year(DATADATE)), DLC (DLC = DLC*1000000), DLTT (DLTT = DLTT*1000000)

- Compute Face value of debt (F) as DLC+ 0.5*DLTT
- After this step you should have CUSIP, YEAR and Default Boundary (F) for each firm-year

- DSF (CRSP) data
  - variables required: CUSIP, DATE, PRC, SHROUT (SHROUT = SHROUT*1000), RET
  - You can construct YEAR variable as YEAR = YEAR(DATE);
  - market capitalization E = ABS(PRC) * SHROUT
  - The data is DAILY. You need to first compute the standard deviation of equity returns (RET) based on the last one year.
  - You can use the group by and summarise function to compute the cumulative annual annual return and standard deviation for each firm (CUSIP) for each YEAR

    ```
    DSF %>% group_by(CUSIP,YEAR) %>%
    summarise(annret = exp(sum(log(1+RET))), sigmae =
    sd(RET)*sqrt(250), E = first(E))
    ```

  - the above code collapses the DAILY data to ANNUAL data
  - Note: for the first part of the assignment you would need only ANNUAL data from this point on (once the daily standard deviation is computed)
  - After the above steps you should have a dataset with CUSIP, YEAR, this year's cumulative return (ANNRET), this year's annualized daily standard deviation (SIGMAE) and this year's equity (E)
  - Note that you would need to LAG the cumulative annual return and equity standard deviation. An easy way to do would be to say YEAR = YEAR+1 so for example, standard deviation computed from returns of 2008 is assigned to observations in 2009.

- Linking DSF (CRSP) and FUNDA (COMPUSTAT)
  - Use CUSIP which is the unique firm identifier in both the datasets

  - Make sure that the FUNDA data is lagged appropriately (so that it is available to the market at the time of estimation).
  - So after lagging, you can merge with DSF data using CUSIP and YEAR
  - After this step you should have CUSIP, YEAR, last year's cumulative return (ANNRET), last year's annualized daily standard deviation (SIGMAE), last year's Default Boundary (F) and last year's equity (E)
  - This data should be sufficient to compute DD using the Naive and Direct Solving approach

- DAILYFED data with risk-free interest rate: keep DATE and DTB3.
  - Compute continuosly compounded risk-free interest rate as r = log(1+DTB3/100)
  - Compute YEAR = YEAR(DATE)
  - group_by(YEAR) %>% summarise(r = first(r)) so that only the first observation for each YEAR is present
  - Now you have this data risk-free interest rate also at ANNUAL level

- Now combine DAILYFED which is at YEAR level with the combined (FUNDA + DSF) based on YEAR

R commands that may be useful for the Data Analysis

As I mentioned in the class, R provides multiple methods to perform any given task. Some of the following commands may be useful

- setwd()
- getwd()
- glm()
- rootSolve package
- append()

- sink()
- order()
- ggplot()+ geom_histogram()
- min()
- max()
- mean()
- sd(x,na.rm=TRUE)
- quantile(column,c(.25,.50,.75))
- print()
- print(paste())
- cor()

Data Analysis

Steps in the assignment

1. Use / Create SAS datasets from QCF servers / FRED (FUNDA, DSF and DAILYFED)
2. Keep only the variables that you require.
3. Lag variables appropriately
4. Create the required variables in each data set based on the instructions given before
5. Merge the datasets based on the instructions given above
6. Compute the required statistics
    - Use rootSolve package to directly solve for DD (two unknowns, two equations)

- Use sink() to direct output to a file and use it again to return output to terminal
- Use appropriate commands to compute the necessary descriptive statistics (min(), max(), sd(x, na.rm=TRUE), quantile(column, c(.25, .50, .75)), mean()).
- Save the results in PDF format. I don't want to see hundreds of pages of output.
- Check the R Console to see if there are any errors in your code
- Upload the R Markdown code and results (in PDF form) to canvas