Data Description

Data source: Thomson Reuters Datastream, the Bloomberg Data License Back Office Product, CRI data base

The data comes from several sources and requires a large amount of preprocessing which was done when I was doing research on KMV model. Now, it's a 315MB csv file, with 262406 rows in total. It consists a large sample of U.S. public firms over the period from 1991 to 2012. Here only companies traded on NYSE, AMEX and Nasdaq are included. For each row, it consists the features of a certain company in a certain month

The following features are included in the dataset:

1. Level of DTD.
2. Trend of DTD.
3. Level of (Cash + Short term investments)/Total assets, abbreviated as CASH/TA.
4. Trend of CASH/TA.
5. Level of Net income/Total assets, abbreviated as NI/TA.
6. Trend of NI/TA.
7. Level of log (Firm market capitalization/ Economy's median market capitalization), abbreviated as SIZE.
8. Trend of SIZE.
9. Current value of (Market capitalization + Total liabilities)/Total asset, abbreviated as M/B.
10. Current value of SIGMA.
11. Treasury rate

Explanation:

1. DTD: firm's distance-to-default, which is a volatility adjusted leverage measure based on Merton (1974).
3. CASH/TA: ratio of the sum of cash and short-term investments to the total assets.
5. NI/TA: ratio of net income to the total assets.
7. SIZE: logarithm of the ratio of firm's market equity value to the average market equity value of the S&P500 firm.
9. M/B: market-to-book asset ratio.
10. SIGMA: 1-year volatility
11. Treasury rate: 3-month US Treasury bill rate.