Replication for 'Bond Risk Premiums with Machine Learning'

YoungHae Kim

January 15, 2021

1. Data

- Monthly yield-data : Liu, Wu(2020), Reconstructuring the yield curve
  - Annualized continuously-compounded zero coupon yield in percentage points
  - Extract → date : 1971.09 ~ 2019.12, maturity : 1month ~ 120month
  - `https://sites.google.com/view/jingcynthiawu/yield-data`
- Monthly macro-data : McCracken, Ng(2015)
  - Use 'current.csv' data → extract 1971.09 ~ 2019.12
  - `https://research.stlouisfed.org/econ/mccracken/fred-databases/`
  - In above site, 135 variables are described in 'Appendix_table_update' file. But now 7 variables are omitted → 128 variables
  - Omitted variables : NAMPI (Group1), NAPMEI (Group2), NAPM, NAPMNOI, NAPMSDI, NAPMII (Group4), NAPMPRI (Group7)
- Construct forward-rate, excess-return from yield-data
  - Define the zero-coupon yield at t with a maturity of n as $y_t(n)$ $(t = \frac{1}{12}, \frac{2}{12}, ...48\frac{3}{12})$ $(n = \frac{1}{12}, \frac{2}{12}, ...10)$
  - The price of the n-year discount bond at time t relates to the zero-coupon yield : $log(P_t(n)) = -ny_t(n)$
  - The forward rate with maturity n at time t is dened as the return for a loan starting at $t+n-1$ and maturing at $t+n$: $f_t(n) = log(P_t(n-1)) - log(P_t(n))$
  - The excess return : $rx_{t+1}(n) = log(P_{t+1}(n-1)) - log(P_t(n)) - y_t(1)$