# Sample Description

## 1. Main description

I combine several datasets to clean this small sample. The return data is from CRSP, and the characteristics are obtained from the SAS code of Green, Hand, and Zhang (2017) (GHZ). I also download FF3 factors and market equity breakpoints from Kenneth R. French’s website to merge to this sample. All characteristics excluding dummy and score variables are cross-sectionally winsorized at the 1st and 99th percentiles.

## 2. Filters

(1) Stocks in NYSE, AMEX, and NASDAQ ()

(2) Common shares ()

(3) Delete the observations which cannot match to GHZ characteristics

(4) Non-missing return

(5) Non-missing lagged market equity (ME)

(6) Exclude microcap stocks (lagged ME less than NYSE 20% breakpoint)

(7) Sample period: from 2000/01 to 2019/12

## 3. Other information

(1) Stock returns are adjusted for delisting. Missing delisting returns are filled following Shumway (1997), Shumway and Warther (1999).

(2) The downloaded FF3 factors (ME breakpoints) are recorded in percent (in millions).

(3) GHZ’s SAS code is downloaded from

<https://sites.google.com/site/jeremiahrgreenacctg/home>.

I simply extend the sample period to 2019 and re-run the code to obtain the characteristics.