Problem Set 2 Solution

Denis Mokanov

April 15, 2019

```
library(data.table)
library(ggplot2)
library(lfe)
library(foreign)
library(stargazer)
rm(list = ls())
```

Question 1: On marginal significance and trading strategy improvements

1. Using the Fama-MacBeth regression approach, what are the average return, standard deviation and Sharpe ratio of the trading strategy implied by using only an intercept and lnInv on the right hand side in the regressions?

```
# Import data and set as data.table
StockRetAcct_DT = as.data.table(read.dta("StockRetAcct_insample.dta"))

# create excess returns in levels
StockRetAcct_DT[, `:=`(ExRet, exp(lnAnnRet) - exp(lnRf))]

# Hindsight is 20/20, flip the sign on lnInv to get positive returns
StockRetAcct_DT[, `:=`(neg_lnInv, -1 * lnInv)]

# Fama-MacBeth Regressions
port_ret = StockRetAcct_DT[, .(lambda = felm(ExRet ~ neg_lnInv, na.action = na.omit)$coef[2]),
    by = year]
fm_output = list(MeanReturn = mean(port_ret$lambda), StdReturn = sqrt(var(port_ret$lambda)),
    SR_Return = mean(port_ret$lambda)/sqrt(var(port_ret$lambda)), tstat_MeanRet = sqrt(1 +
        2014 - 1980) * mean(port_ret$lambda)/sqrt(var(port_ret$lambda)))
fm_output
```

```
## $MeanReturn
## [1] 0.08679146
##
## $StdReturn
## [1] 0.1486441
##
## $SR_Return
## [1] 0.5838877
##
## $tstat_MeanRet
## [1] 3.454326
```

2. What is the analytical expression for the portfolio weights in this case?

$$w_{i,t-1} = \frac{1}{N_t} \frac{\ln(Inv_{i,t-1}) - \mathbb{E}_i\left[\ln(Inv_{i,t-1})\right]}{\text{var}_i\left(\ln(Inv_{i,t-1})\right)}$$

3. You worry that there is industry-related noise associated with the characteristic lnInv and want to clean up your trading strategy with the goal of reducing exposure to unpriced industry risks. What regressions to you run? Report mean, standard deviation, and Sharpe ratio of the 'cleaned-up' trading strategy.

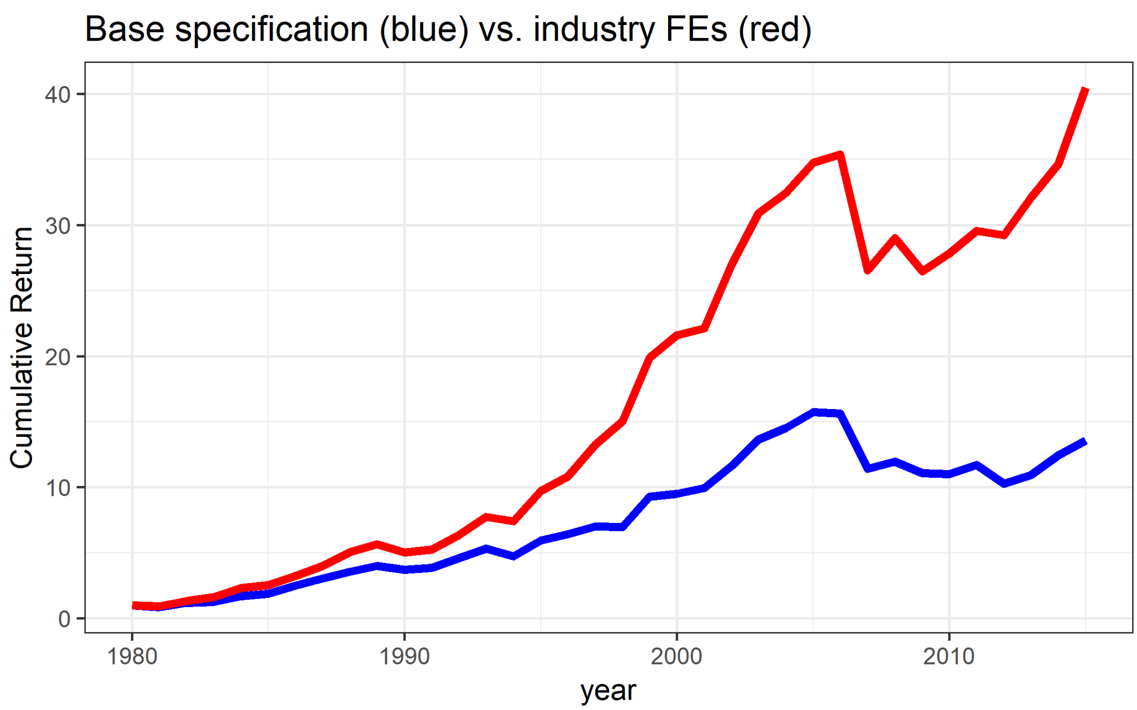
```
# Fama-MacBeth Regressions
port_ret_ind_FE = StockRetAcct_DT[, .(lambda = felm(ExRet ~ neg_lnInv | ff_ind |
    0 | 0, na.action = na.omit)$coef[1]), by = year]
fm_output_ind_FE = list(MeanReturn = mean(port_ret_ind_FE$lambda), StdReturn = sqrt(var(port_ret_ind_FE
    SR_Return = mean(port_ret_ind_FE$lambda)/sqrt(var(port_ret_ind_FE$lambda)),
    tstat_MeanRet = sqrt(1 + 2014 - 1980) * mean(port_ret_ind_FE$lambda)/sqrt(var(port_ret_ind_FE$lambda
fm_output_ind_FE
```

```
## $MeanReturn
## [1] 0.08257762
##
## $StdReturn
## [1] 0.1019642
##
## $SR_Return
## [1] 0.8098685
##
## $tstat_MeanRet
## [1] 4.791247
```

4. As in the class notes, plot the cumulative returns to the simple and the 'cleaned-up' trading strategies based on your new signal, lnInv. Make sure both trading strategies result in portfolios with a 15% return standard deviation.

```
# For both strategies
for (i in c("", "_ind_FE")) {
    # Scale returns to have a 15% standard deviation
    assign(paste0("scaled_ret", i), get(paste0("port_ret", i))$lambda * 0.15/sqrt(var(get(paste0("port_
        i))$lambda)))
    assign(paste0("Cum_ret", i), 0)
    for (j in 1:35) {
        # Calculate cumulative returns
        assign(paste0("Cum_ret", i), c(get(paste0("Cum_ret", i)), get(paste0("Cum_ret",
            i))[j] + log(1 + get(paste0("scaled_ret", i))[j])))
    }
}

# Plot cumulative returns
qplot(c(1980:2015), exp(Cum_ret), geom = "line", xlab = "year", ylab = "Cumulative Return",
    color = I("blue"), size = I(1.5), main = "Base specification (blue) vs. industry FEs (red)") +
    geom_line(aes(y = exp(Cum_ret_ind_FE)), color = I("red"), size = I(1.5)) +
    theme_bw()
```


Base specification (blue) vs. industry FEs (red)
Cumulative Return
40
30
20
10
0
1980
1990
2000
2010
year

Question 2: Predicting medium to long-run firm-level return variance

1. Try with and without industry and year fixed effects, with and without clustering of standard errors. Discuss which specification makes most sense to you. In particular, discuss the effect of a year fixed effect. What is the intuition for the impact of this fixed effect?

```
# Define next year's rv. Make sure next observation is actually next year.
setorder(StockRetAcct_DT, FirmID, year)
StockRetAcct_DT[, `:=`(next_rv, shift(rv, type = "lead")), by = FirmID]
StockRetAcct_DT[, `:=`(next_year, shift(year, type = "lead")), by = FirmID]
StockRetAcct_DT[next_year != (year + 1), `:=`(next_rv, NA)]

# Define previous year's return
StockRetAcct_DT[, `:=`(prev_ExRet, shift(ExRet)), by = FirmID]
StockRetAcct_DT[, `:=`(prev_year, shift(year)), by = FirmID]
StockRetAcct_DT[prev_year != (year - 1), `:=`(prev_ExRet, NA)]

# Run some panel regressions
r1 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet,
    data = StockRetAcct_DT, na.action = na.omit)
r2 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    0 | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
r3 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year, data = StockRetAcct_DT, na.action = na.omit)
r4 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
r5 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year + ff_ind, data = StockRetAcct_DT, na.action = na.omit)
r6 = felm(next_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year + ff_ind | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
stargazer(r1, r2, r3, r4, r5, r6, type = "text", report = "vc*t", add.lines = list(c("Year FE",
    "N", "N", "Y", "Y", "Y", "Y"), c("Ind FE", "N", "N", "N", "N", "Y", "Y"),
    c("Firm, Year Clustering", "N", "Y", "N", "Y", "N", "Y")), omit.stat = "ser")
```

```
##
## ===========================================================================================
##                                                 Dependent variable:
##                             ---------------------------------------------------------------
##                                                       next_rv
##                                 (1)        (2)         (3)        (4)         (5)        (6)
## -------------------------------------------------------------------------------------------
## rv                           0.426***   0.426**    0.554***   0.554***   0.516***    0.516***
##                         t = 103.445 t = 2.545  t = 146.543 t = 4.007  t = 132.257 t = 3.782
##
## lnBM                        -0.024***  -0.024***  -0.009***    -0.009    -0.004***    -0.004
##                         t = -25.707 t = -3.219 t = -13.553 t = -1.380 t = -5.270  t = -0.786
##
## lnProf                      -0.055***  -0.055***  -0.036***  -0.036***  -0.039***  -0.039***
##                         t = -17.379 t = -4.322 t = -16.184 t = -5.234 t = -17.280 t = -6.324
##
```

```
## lnLever                    -0.006***    -0.006    -0.005***    -0.005      0.0003      0.0003
##                          t = -6.045  t = -0.944 t = -8.080  t = -0.815  t = 0.337  t = 0.076
##
## lnIssue                    0.030***     0.030*    0.031***    0.031***   0.034***    0.034***
##                           t = 9.525  t = 1.745  t = 14.105  t = 3.124  t = 15.055  t = 3.200
##
## lnInv                      0.069***    0.069***   0.034***    0.034***   0.032***    0.032***
##                          t = 22.351  t = 3.026  t = 15.508  t = 2.589  t = 15.063  t = 2.712
##
## prev_ExRet                  -0.0004     -0.0004   0.019***     0.019     0.021***     0.021
##                          t = -0.260  t = -0.016 t = 16.829  t = 1.398  t = 19.018  t = 1.588
##
## Constant                   0.073***    0.073***
##                          t = 47.373  t = 3.352
##
## -------------------------------------------------------------------------------------------
## Year FE                       N           N           Y           Y           Y           Y
## Ind FE                        N           N           N           N           Y           Y
## Firm, Year Clustering         N           Y           N           Y           N           Y
## Observations               46,685      46,685      46,685      46,685      46,685      46,685
## R2                          0.268       0.268       0.653       0.653       0.661       0.661
## Adjusted R2                 0.268       0.268       0.652       0.652       0.660       0.660
## ===========================================================================================
## Note:                                                            *p<0.1; **p<0.05; ***p<0.01
```

I test four specifications - with and without industry and year fixed effects, and with and without firm and year clustering. We can see that clustering significantly decreases the t-statistic of all coefficients, and that fixed effects increase the predictive power of rv. This likely is due to the year fixed effects accounting for changes in systematic risk driving differences in observed realized variance across time.

2. Also try forecasting at the 5-year horizon (rv in 5 years). How do the results change? Can we predict return variance 5-years ahead? Is the 5-year lagged rv significant, or are other variables more important?

```
# Define five year forward rv
five_year_rv = copy(StockRetAcct_DT[, .(FirmID, year, rv)])
setorder(five_year_rv, FirmID, year)
five_year_rv[, `:=`(year, year - 5)]
setnames(five_year_rv, "rv", "next_5_rv")
StockRetAcct_DT = merge(StockRetAcct_DT, five_year_rv, by = c("FirmID", "year"),
    all.x = T)
setkey(StockRetAcct_DT)

# Run some panel regressions
r1 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet,
    data = StockRetAcct_DT, na.action = na.omit)
r2 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    0 | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
r3 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year, data = StockRetAcct_DT, na.action = na.omit)
r4 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
r5 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
    year + ff_ind, data = StockRetAcct_DT, na.action = na.omit)
r6 = felm(next_5_rv ~ rv + lnBM + lnProf + lnLever + lnIssue + lnInv + prev_ExRet |
```

```r
    year + ff_ind | 0 | year + FirmID, data = StockRetAcct_DT, na.action = na.omit)
stargazer(r1, r2, r3, r4, r5, r6, type = "text", report = "vc*t", add.lines = list(c("Year FE",
    "N", "N", "Y", "Y", "Y", "Y"), c("Ind FE", "N", "N", "N", "N", "Y", "Y"),
    c("Firm, Year Clustering", "N", "Y", "N", "Y", "N", "Y")), omit.stat = "ser")
```

```
##
## ==========================================================================================
##                                                Dependent variable:
##                          -----------------------------------------------------------------
##                                                    next_5_rv
##                              (1)        (2)         (3)        (4)         (5)        (6)
## ------------------------------------------------------------------------------------------
## rv                         -0.001     -0.001    0.217***    0.217***   0.156***    0.156**
##                          t = -0.273 t = -0.031 t = 41.127  t = 2.697  t = 28.931  t = 2.429
##
## lnBM                      -0.025***  -0.025***  -0.012***   -0.012*    -0.002**    -0.002
##                          t = -19.639 t = -3.270 t = -12.045 t = -1.796 t = -2.083 t = -0.456
##
## lnProf                    -0.079***  -0.079***  -0.036***  -0.036***  -0.039***  -0.039***
##                          t = -16.501 t = -2.898 t = -10.048 t = -3.157 t = -10.934 t = -3.597
##
## lnLever                   -0.007***   -0.007    -0.003***   -0.003     0.002**     0.002
##                          t = -5.339 t = -0.849 t = -2.886 t = -0.342  t = 2.015  t = 0.522
##
## lnIssue                   0.053***   0.053***   0.024***   0.024***   0.029***    0.029***
##                          t = 12.114 t = 4.141   t = 7.510  t = 3.333  t = 9.184  t = 4.643
##
## lnInv                     0.035***   0.035**    0.034***   0.034***   0.030***    0.030***
##                           t = 8.045 t = 2.183  t = 10.488 t = 2.899  t = 9.598  t = 3.141
##
## prev_ExRet                0.033***    0.033*    0.004**     0.004     0.007***     0.007
##                          t = 16.488 t = 1.711   t = 2.257  t = 0.425  t = 4.720  t = 0.975
##
## Constant                  0.133***   0.133***
##                          t = 64.125 t = 6.491
##
## ------------------------------------------------------------------------------------------
## Year FE                      N          N           Y          Y           Y          Y
## Ind FE                       N          N           N          N           Y          Y
## Firm, Year Clustering        N          Y           N          Y           N          Y
## Observations               31,220     31,220     31,220      31,220     31,220     31,220
## R2                          0.057      0.057      0.487      0.487       0.511      0.511
## Adjusted R2                 0.057      0.057      0.487      0.487       0.511      0.511
## ==========================================================================================
## Note:                                                      *p<0.1; **p<0.05; ***p<0.01
```

Regression coefficients are less significant when predicting realized variance at a five-year horizon. With firm and year clustering, and industry and year fixed effects, the t-stat of rv is only 2.429. While other variables have higher explanatory power, they are harder to motivate (lnIssue as an example).

3. What are the benefits of the panel approach, versus simply running one regression for each firm? What are the potential costs?

The primary issue with running a regression for each firm, is that the time-series for each firm is not long enough to get statistically significant estimates. The panel approach allows you to estimate covariates across many firms, allowing for more observations to be used in the regression. However, the panel specification will identify the aggregate affect of the covariates, which may not be directly applicable to any specific firm.