Backtesting Project

A-B-G-I

May 4, 2018

- Factors
  - 12 JKKL Factors
  - The Factors We Picked
- Model
  - Data/Factors
  - Results:
- Correlation
- Improvements?
  - Graphs and fitted stats
- What are we doing next?
  - ML/ RF modeling

Factors

12 JKKL Factors

Number	Factors	Category	Descriptions	Effect
1	RETP	Momentum	Cumulative Market Adjusted Return for the Preceding 6 Months	pos
2	RETP2	Momentum	Cumulative Market Adjusted Return for the 2nd Preceding 6 Months	pos
3	TURN	Trading Volume	Average Daily Volume Turnover	neg
4	SIZE	Size	Market Cap (Natural Log)	neg
5	FREV	Earning Surpirse	Analyst earnings forecast revisions to price	pos
6	LTG	Growth	Long-term growth forecast	neg
7	SUE	Earning Surprise	Standardized unexpected earnings	pos
8	SG	Growth	Sales Growth	neg
9	TA	Earning Quality	Total Accruals to total assets	neg
10	CAPEX	Growth	Capital expenditures to total assets	neg
11	BP	Growth	Book to Price	pos
12	EP	Growth	Earnings to Price	pos

The Factors We Picked

Number	Factors	Descriptions	Research

Number	Factors	Descriptions	Research
13	DP	Historically, there has a been a positive relation between Dividend/Price or Dividend Yield and future returns	Litzenberger and Ramaswamy (1982)
14	Volume	Firms with larger amounts of volume subsequently have lower furture returns	Ang et al. (2006)
15	total Q	Total Q is a new proxy for Tobin's Q. Tobin's Q is traditionally Market Equity value + Market value of liabilities divided by equity book value + liabilities book value. Total Q includes intangible capital in the denominator.	Peters and Taylor (2016)
16	Off Balance Sheet Asset(OffBS)	what degree intangible capital is kept off of or not listed on the balance sheet	Peters and Taylor (2016)
17	M-Score	Attempts to encapsulate likelihood of firm-level earnings manipulation. This factor uses eight sub-factors calculated with compustat data; additionally, Beneish finds that firms with a score greater than -1.78 are more often than not earnings manipulators	Beneish's paper

Model

Data/Factors

In-Sample 1985-1998

Out-sample 1985-2013

Here's what the Datatable looks like:

```
head(Data,5)
```

```
##    year permno ind  yearlyRet        retp       ret2p      turn  lagsize
## 1: 1993  10010  24 -0.4153792  0.04635837 -0.33497324 23.488430 11.25170
## 2: 1993  10074  20  0.5494987 -0.13162166  0.40484643 57.700921 10.55641
## 3: 1993  10138  35  0.2592031  0.15833998 -0.01834576 53.869122 13.33725
## 4: 1993  10138  35  0.2592031  0.15833998 -0.01834576 53.869122 13.33725
## 5: 1993  10147  22  1.7789487  0.63323234  0.57894882  5.163551 13.57594
##           FREV     LTG        SUE       SG          ta      CAPEX
## 1: -0.04944709 28.9375 -1.6160340 1.352598  0.03519428 0.06290010
## 2:  0.00000000 28.9375 -0.3994654 1.419220 -0.04426388 0.12948991
## 3:  0.00000000 13.2500  0.5658795 1.244849 -0.06458850 0.04420138
## 4:  0.00000000 13.2500  0.5658795 1.244849 -0.06458850 0.04420138
## 5:  0.00000000 28.9375  1.1973422 1.515659  0.12660722 0.10516545
##           bp          EP   VOL      q_tot K_int_offBS          DP
## 1: 0.2784134 -0.03462688 12407  0.9440737   33.051460 0.000000000
## 2: 0.2093319  0.06666672   853  1.2400050    5.678446 0.004545455
## 3: 0.2780287  0.06231292 11920 10.6024800    0.000000 0.004482759
## 4: 0.2780287  0.06231292 11920 10.6024800    0.000000 0.000000000
## 5: 0.2712372  0.04047619 44671  7.8852210  199.970300 0.000000000
```

Results:

```
## Simple return: 13.7745046230993
```

```
## Annulized raw return: 12.1075815461762
```

```
## Geometric average of annulized raw: 12.7213471657791
```

```
## Geometric average of annulized excess returns: 11.0542713167365
```

```
## Monthly sharpe ratio: 0.824325805586877
```

```
## CAPM
```

```
##
## Call:
## lm(formula = r_excess ~ `Mkt-RF`, data = Res)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6666 -2.3837  0.0694  2.4514 14.4634
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.15994    0.31431    3.69  0.00031 ***
## `Mkt-RF`    -0.36211    0.06806   -5.32  3.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.91 on 154 degrees of freedom
## Multiple R-squared:  0.1553, Adjusted R-squared:  0.1498
## F-statistic:  28.3 on 1 and 154 DF,  p-value: 3.597e-07
```

```
## FF 3 factor, time series regression
```

```
##
## Call:
## lm(formula = r_excess ~ `Mkt-RF` + SMB + HML, data = Res)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.168 -2.050 -0.096  2.259 15.546
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.13244    0.31178   3.632 0.000384 ***
## `Mkt-RF`    -0.32599    0.07124  -4.576 9.79e-06 ***
## SMB         -0.18750    0.12793  -1.466 0.144814
## HML          0.31195    0.11006   2.834 0.005216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.81 on 152 degrees of freedom
## Multiple R-squared:  0.2083, Adjusted R-squared:  0.1927
## F-statistic: 13.33 on 3 and 152 DF,  p-value: 9.031e-08
```

```
## Carhart 4 factor
```

```
##
## Call:
## lm(formula = r_excess ~ `Mkt-RF` + SMB + HML + Mom, data = Res)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1821 -1.9129 -0.0666  1.9665 16.3721
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.06983    0.29944   3.573 0.000474 ***
## `Mkt-RF`    -0.19311    0.07683  -2.514 0.013003 *
## SMB         -0.16647    0.12280  -1.356 0.177245
## HML          0.25796    0.10650   2.422 0.016615 *
## Mom          0.22740    0.06015   3.780 0.000225 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.653 on 151 degrees of freedom
## Multiple R-squared:  0.2767, Adjusted R-squared:  0.2576
## F-statistic: 14.44 on 4 and 151 DF,  p-value: 5.214e-10
```

```
## Monthly information ratio under Fama French 3 factor model: 0.300166650697856
```

```
## Annualized FF3 IR: 1.03980777949293
```

```
## Monthly information ratio under Corhart 4 factor model: 0.296688571496051
```

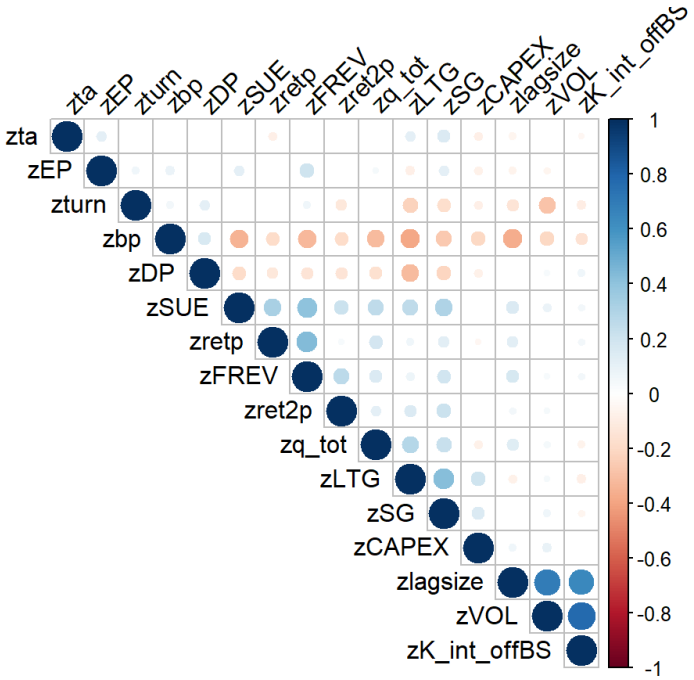
```
## Annualized C4 IR: 1.02775935971238
```

Correlation

Correlation Matrix

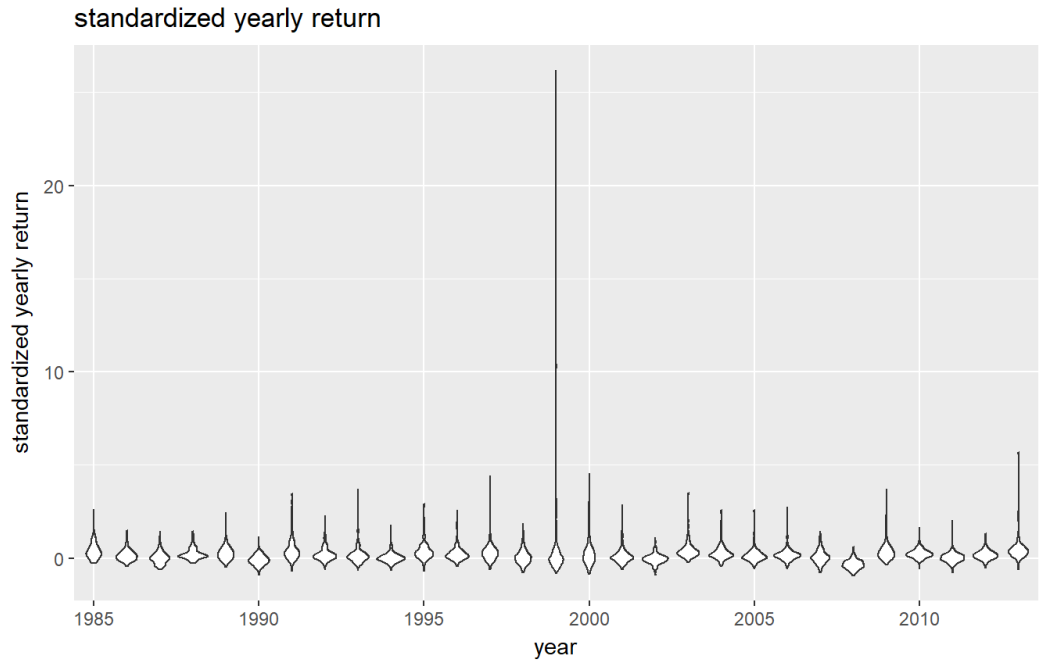
	zretp	zret2p	zturn	zlagsize	zFREV	zLTG	zSUE	zSG	zta	zCAPEX	zbp	zEP	zVOL	zq_tot	zK_int_offBS	zDP
zretp	1															
zret2p	0.04	1														
zturn	0.01	-0.12	1													
zlagsize	0.13	0.05	-0.15	1												
zFREV	0.43	0.27	0.07	0.17	1											
zLTG	0.07	0.15	-0.22	-0.07	0.07	1										
zSUE	0.32	0.22	0.02	0.16	0.4	0.25	1									
zSG	0.12	0.22	-0.18	0	0.2	0.43	0.3	1								
zta	-0.08	0	-0.02	-0.06	0	0.11	0	0.16	1							
zCAPEX	-0.05	0.02	-0.08	0.07	0.01	0.21	0.02	0.16	-0.08	1						
zbp	-0.18	-0.19	0.06	-0.37	-0.33	-0.39	-0.34	-0.27	-0.04	-0.2	1					
zEP	0.01	0.02	0.06	-0.07	0.2	-0.09	0.11	0.12	0.12	-0.08	0.09	1				
zVOL	0.02	0.04	-0.29	0.7	0.04	0.04	0.08	0.06	-0.03	0.09	-0.2	-0.06	1			
zq_tot	0.19	0.12	0.01	0.14	0.16	0.28	0.25	0.23	-0.01	-0.07	-0.32	0.04	0.05	1		
zK_int_offBS	0.05	0.01	-0.09	0.64	0.05	-0.09	0.05	-0.05	-0.04	0.02	-0.15	-0.02	0.77	-0.07	1	
zDP	-0.12	-0.15	0.11	0	-0.15	-0.31	-0.18	-0.22	-0.04	-0.07	0.17	0.03	0.04	-0.16	0.07	1

Corrgram

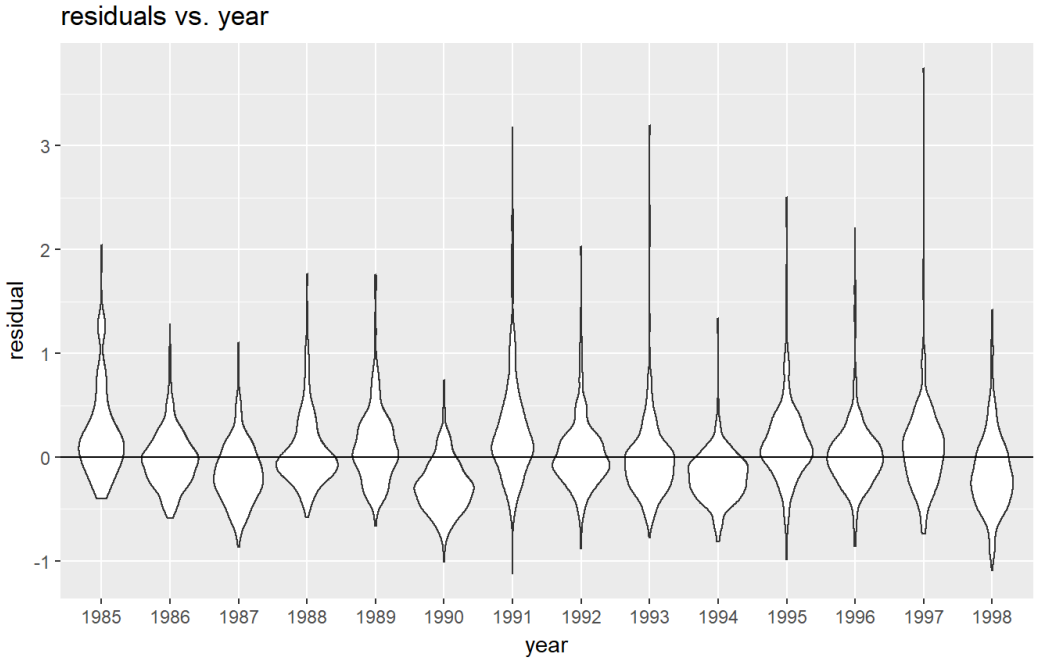


Improvements?

Graphs and fitted stats



term	estimate	std.error	statistic	p.value
(Intercept)	-8.4193153	2.9302435	-2.8732477	0.0040814
year	0.0043250	0.0014711	2.9400459	0.0032980
zretp	0.0282914	0.0066914	4.2280050	0.0000240
zret2p	0.0150185	0.0062253	2.4125054	0.0158822
zturn	0.0002136	0.0064600	0.0330723	0.9736184
zlagsize	-0.0479979	0.0093157	-5.1523547	0.0000003
zFREV	0.0092893	0.0073189	1.2692160	0.2044282
zLTG	-0.0264486	0.0075392	-3.5081253	0.0004556
zSUE	-0.0124399	0.0068979	-1.8034411	0.0713844
zSG	0.0005290	0.0068700	0.0769964	0.9386297
zta	-0.0299439	0.0059741	-5.0122540	0.0000006
zCAPEX	-0.0092563	0.0061737	-1.4993136	0.1338609
zbp	0.0402549	0.0075579	5.3261728	0.0000001
zEP	-0.0163874	0.0061765	-2.6531845	0.0080010
zDP	-0.0828134	0.0062132	-13.3285435	0.0000000
zVOL	-0.0026452	0.0108538	-0.2437145	0.8074628
zq_tot	0.1023892	0.0065961	15.5226415	0.0000000
zK_int_offBS	0.0617827	0.0098131	6.2959456	0.0000000



What are we doing next?

ML/ RF modeling

Dividing the training and testing randomly 80-20

```
set.seed(1234)
  test_df <- final_df %>%
  group_by(Direction) %>%
  sample_frac(.2) %>%
  ungroup()

training_df <- final_df %>%
  anti_join(test_df, by="RegionID")

training_df
```

Trainng the rf model

```
rf <- randomForest(Direction~., data=training_df %>% select(-RegionID), ntree = 100)
rf
```

10-fold Crossvalidation: Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it.

Note: In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

```
result_df <- createFolds(final_df$Direction, k=10) %>%
  purrr::imap(function(test_indices, fold_number) {
    train_df <- final_df %>%
      select(-RegionID) %>%
      slice(-test_indices)

    test_df <- final_df %>%
      select(-RegionID) %>%
      slice(test_indices)

    # fit the two models
    rf <- randomForest(Direction~., data=train_df, ntree=100)

    dt <- randomForest(Direction~., data=train_df, ntree=10)
}) %>%
  purrr::reduce(bind_rows)
result_df

result_df %>%
  mutate(error_rf = observed_label != predicted_label_rf,
         error_dt = observed_label != predicted_label_dt) %>%
  group_by(fold) %>%
  summarize(rf = mean(error_rf), dt = mean(error_dt)) %>%
  tidyr::gather(model, error, -fold) %>%
  lm(error~model, data=.) %>%
  broom::tidy()
```