What Firm Characteristics Drive US Stock Returns?

Yufeng Han[1] Ai He[2] David Rapach[3] Guofu Zhou[4]

[1]UNCC [2]Emory Univ [3]Saint Louis Univ [4]Washington Univ in St Louis

American Finance Association Annual Meeting

January 6, 2019

Predicting cross-sectional returns

- CS predictive regressions based on firm characteristics
  - Haugen & Baker (1996)
  - Hanna & Ready (2005)
  - Lewellen (2015)
  - Green, Hand & Zhang (2017)
- Plethora of plausible predictors
  - Model uncertainty/instability
- CS multiple regression forecast $\Rightarrow$ overfitting
- What we do $\Rightarrow$ forecast combination

Forecast combination

- CS univariate regression $\Rightarrow r_{i,t} = a_{j,t} + b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t}$
  - $r_{i,t} \Rightarrow$ return for firm i in month t
  - $z_{i,j,t-1} \Rightarrow$ characteristic j for firm i in month $t-1$
  - $i = 1, \ldots, I_t; \; j = 1, \ldots, J$
- Univariate forecast $\Rightarrow \hat{r}^{(j)}_{i,t+1|t} = \hat{a}_{j,t} + \hat{b}_{j,t} z_{i,j,t}$
  - $i = 1, \ldots, I_{t+1}; \; j = 1, \ldots, J$
  - $\hat{a}_{j,t}$ $(\hat{b}_{j,t}) \Rightarrow$ OLS/WLS estimate of $a_{j,t}$ $(b_{j,t})$
- Simple combination forecast $\Rightarrow \hat{r}^{\text{Mean}}_{i,t+1|t} = \frac{1}{J} \sum_{j=1}^{J} \hat{r}^{(j)}_{i,t+1|t}$

Forecast combination

- Refine forecast using ML (Diebold & Shin forthcoming)
- Granger & Ramanathan (1984) regression
  - $r_{i,t} = a_t^{\mathrm{GR}} + \sum_{j=1}^{J} b_{j,t}^{\mathrm{GR}} \hat{r}_{i,t|t-1}^{(j)} + \varepsilon_{i,t}$
    - Select regressors using ordinary/weighted LASSO/ENet
- LASSO/ENet combination forecast
  - Include individual forecasts selected by LASSO/ENet
- No look-ahead bias

Conventional approach

- CS multiple regression $\Rightarrow r_{i,t} = a_t^{\mathrm{MR}} + \sum_{j=1}^{J} b_{j,t}^{\mathrm{MR}} z_{i,j,t-1} + \varepsilon_{i,t}$
- Forecast $\Rightarrow \hat{r}_{i,t+1|t}^{\mathrm{MR}} = \bar{a}_t^{\mathrm{MR}} + \sum_{j=1}^{J} \bar{b}_{j,t}^{\mathrm{MR}} z_{i,j,t}$
  - $\bar{a}_t^{\mathrm{MR}} = \frac{1}{120} \sum_{s=0}^{119} \hat{a}_{t-s}^{\mathrm{MR}}$
  - $\bar{b}_{j,t}^{\mathrm{MR}} = \frac{1}{120} \sum_{s=0}^{119} \hat{b}_{j,t-s}^{\mathrm{MR}}$
- Combination vis-à-vis MR (Rapach, Strauss & Zhou 2010)
  - $\hat{r}_{i,t+1|t}^{\mathrm{Mean}} = \bar{r}_t + \frac{1}{J} \sum_{j=1}^{J} \hat{b}_{j,t}(z_{i,j,t} - \bar{z}_{j,t})$
    - Replace MR coefficients with univariate counterparts
    - Shrink forecast to cross-sectional mean

Fama-MacBeth predictive slope

- Lewellen (2015) approach
- CS regression $\Rightarrow r_{i,t} = a_t^{\text{FM}} + b_t^{\text{FM}} \hat{r}_{i,t|t-1} + \varepsilon_{i,t}$
  - $\hat{b}_t^{\text{FM}} \Rightarrow$ OLS/WLS estimate of b_t^{FM}
  - $R_t^2 \Rightarrow$ CS R^2 statistic
- Time-series averages
  - $\hat{b}^{\text{FM}} = \frac{1}{T} \sum_{t=1}^{T} \hat{b}_t^{\text{FM}}$
  - $R^2 = \frac{1}{T} \sum_{t=1}^{T} R_t^2$
- Test H_0: $b^{\text{FM}} \leqslant 0$ vs H_A: $b^{\text{FM}} > 0$

Fama-MacBeth predictive slope

- Dispersion of CS expected returns
  - $\hat{b}^{\text{FM}} = 1 \Rightarrow$ unbiased on average
    - Mincer-Zarnowitz regression
  - $\hat{b}^{\text{FM}} < 1 \Rightarrow$ overstates dispersion on average
    - Signals overfitting
  - $\hat{b}^{\text{FM}} > 1 \Rightarrow$ understates dispersion on average

Fama-MacBeth predictive slope

- Demeaned CS forecast error
  - $\hat{u}_{i,t|t-1} = (r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)$
  - Not concerned with forecasting $\bar{r}_t$ per se
- (CS VW) $\text{MSFE}_t = \frac{1}{I_t} \sum_{i=1}^{I_t} w_{i,t} \hat{u}^2_{i,t|t-1}$
- FM interpretation
  - $\text{MSFE}_t^{\text{Naive}} - \text{MSFE}_t = \left(2\hat{b}_t^{\text{FM}} - 1\right) \hat{\sigma}^2_{\hat{r},t}$
    - $\hat{b}^{\text{FM}} > 0.5 \Rightarrow \hat{r}_{i,t|t-1}$ outperforms naive forecast on average
  - $\text{MSFE}_t = \text{bias}_t^2 \hat{\sigma}_{\hat{r},t} + \left(1 - R_t^2\right) \hat{\sigma}^2_{r,t}$
    - $\text{bias}_t = \hat{b}_t^{\text{FM}} - 1$

Forecast encompassing

- Composite forecast $\Rightarrow$ naive & competing forecasts
  - $\hat{r}^*_{i,t|t-1} = (1-\theta_t)\bar{r}_{t-1} + \theta_t \hat{r}_{i,t|t-1}$
- (CS VW) $\text{MSFE}_t = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^*_{i,t|t-1}\right)^2$
- $\hat{b}^{\text{FM}}_t \Rightarrow$ estimate of optimal θ_t
  - $\hat{b}^{\text{FM}} \Rightarrow$ estimate of average optimal θ_t
- Encompassing test
  - H_0: $b^{\text{FM}} \leqslant 0 \Rightarrow$ naive encompasses competing on average
  - H_A: $b^{\text{FM}} > 0 \Rightarrow$ ~~naive encompasses competing on average~~

Forecast encompassing

- Composite forecast $\Rightarrow$ competing forecasts A & B
  - $\hat{r}^{\dagger}_{i,t|t-1} = (1-\eta_t)\hat{r}^{A}_{i,t|t-1} + \eta_t \hat{r}^{B}_{i,t|t-1}$
- (CS VW) $\text{MSFE}_t = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^{\dagger}_{i,t|t-1}\right)^2$
- CS regression $\Rightarrow$ $\hat{e}^{A}_{i,t|t-1} = a^{\dagger}_t + b^{\dagger}_t\left(\hat{e}^{A}_{i,t|t-1} - \hat{e}^{B}_{i,t|t-1}\right) + \varepsilon_{i,t}$
  - $\hat{e}^{k}_{i,t|t-1} = r_{i,t} - \hat{r}^{k}_{i,t|t-1}$ for $k = \text{A}, \text{B}$
  - $\hat{b}^{\dagger}_t \Rightarrow$ OLS/WLS estimate of $b^{\dagger}_t$
- $\hat{b}^{\dagger}_t \Rightarrow$ estimate of optimal η_t
  - $\hat{b}^{\dagger} = \frac{1}{T}\sum_{t=1}^{T}\hat{b}^{\dagger}_t \Rightarrow$ estimate of average optimal η_t

Forecast encompassing

- A encompasses B?
  - H_0: $b^{\dagger} \leqslant 0 \Rightarrow$ A encompasses B on average
  - H_A: $b^{\dagger} > 0 \Rightarrow$ ~~A encompasses B on average~~
- B encompasses A?
  - H_0: $1 - b^{\dagger} \leqslant 0 \Rightarrow$ B encompasses A on average
  - H_A: $1 - b^{\dagger} > 0 \Rightarrow$ ~~B encompasses A on average~~
- Compare info content of conventional/combination forecasts
  - A $\Rightarrow$ conventional
  - B $\Rightarrow$ combination

Spread portfolios

- Sort stocks based on $\hat{r}_{i,t+1|t}$ at end of month t
- Form decile portfolios
- 10-1 spread portfolio for month $t+1$
  - Go long (short) tenth (first) decile

Data

- Sample period ⇒ 1980:01–2017:12
- NYSE/AMEX/NASDAQ stocks with market value on CRSP
- Data for 94 firm characteristics
  - CRSP, Compustat, I/B/E/S
- 3 cases (following Green, Hand & Zhang 2017)
  - Value Weighted
  - Equal Weighted excl Microcap
  - Equal Weighted

FM regressions results—VW

Full out-of-sample period (1990:01–2017:12)

Method	$\hat{b}^{\text{FM}}$	t-stat	R^2
Conventional	0.31	3.10***	1.67%
Mean	1.56	2.43***	5.37%
Tr mean	2.23	2.44***	5.35%
LASSO	1.64	3.96***	4.14%
ENet	1.66	3.95***	4.14%

FM regressions results—VW

Pre-2003 out-of-sample period (1990:01–2002:12)

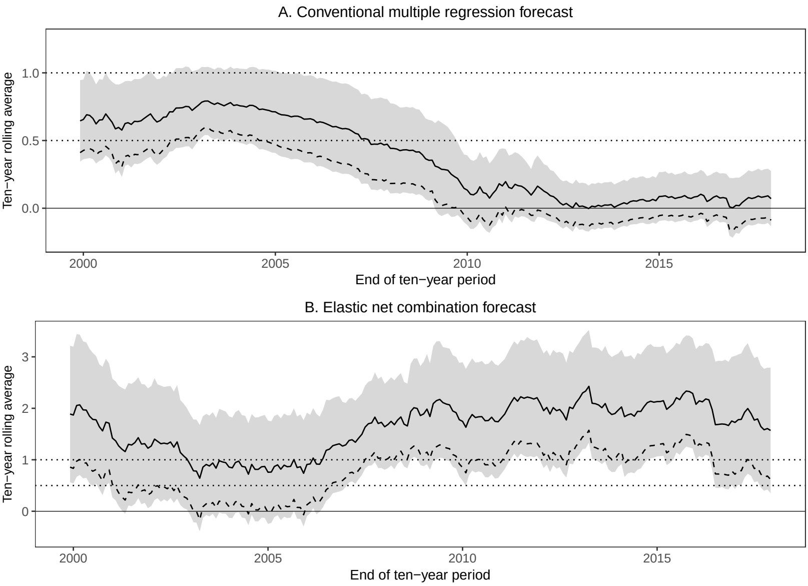
Method	$\hat{b}^{\mathrm{FM}}$	t-stat	R^2
Conventional	0.63	4.14***	2.11%
Mean	1.47	1.47*	6.13%
Tr mean	1.97	1.36*	6.24%
LASSO	1.73	2.72***	4.51%
ENet	1.76	2.69***	4.54%

FM regressions results—VW

Post-2003 out-of-sample period (2004:01–2017:12)

Method	$\hat{b}^{FM}$	t-stat	R^2
Conventional	0.05	0.58	1.24%
Mean	1.94	2.40***	4.74%
Tr mean	2.92	2.62***	4.60%
LASSO	1.55	2.70***	3.89%
ENet	1.58	2.74***	3.89%

10-year rolling estimates of FM slopes—VW



Encompassing test results—VW

Full out-of-sample period (1990:01–2017:12)

	$\hat{b}^{\dagger}$		$1-\hat{b}^{\dagger}$	
Method	Coef	t-stat	Coef	t-stat
Mean	0.70	6.62***	0.30	2.89***
Tr mean	0.70	6.62***	0.30	2.82***
LASSO	0.68	6.57***	0.32	3.14***
ENet	0.68	6.62***	0.32	3.12***

Encompassing test results—VW

Pre-2003 out-of-sample period (1990:01–2002:12)

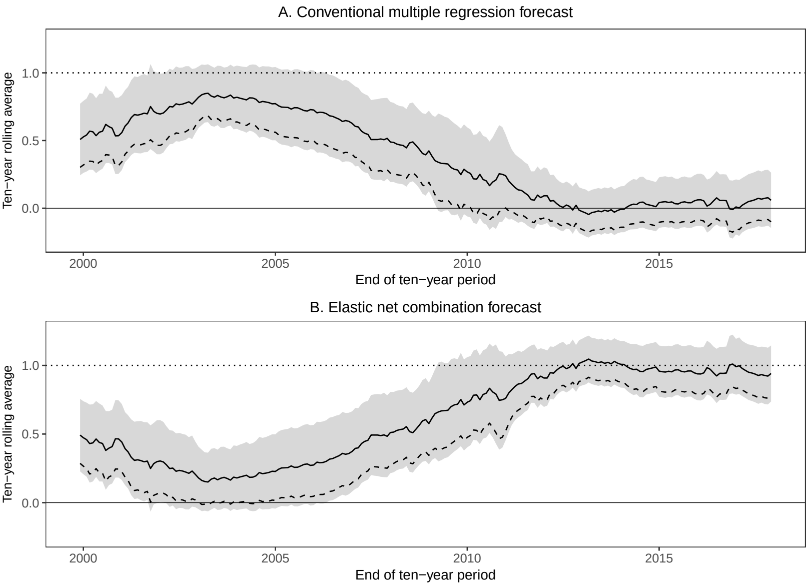
	$\hat{b}^{\dagger}$		$1-\hat{b}^{\dagger}$	
Method	Coef	t-stat	Coef	t-stat
Mean	0.36	2.35***	0.64	4.22***
Tr mean	0.36	2.34***	0.64	4.12***
LASSO	0.35	2.28**	0.65	4.22***
ENet	0.35	2.32**	0.65	4.23***

Encompassing test results—VW

Post-2003 out-of-sample period (2004:01–2017:12)

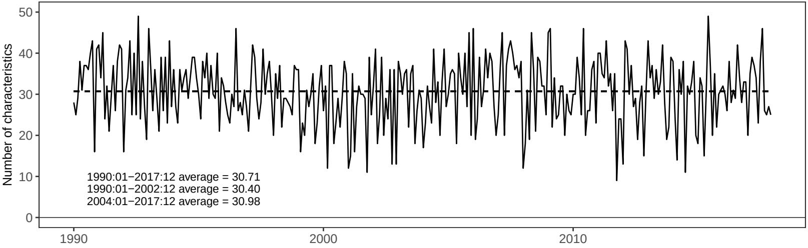
	$\hat{b}^{\dagger}$		$1 - \hat{b}^{\dagger}$	
Method	Coef	t-stat	Coef	t-stat
Mean	0.98	10.62***	0.02	0.18
Tr mean	0.99	10.73***	0.01	0.13
LASSO	0.95	10.86***	0.05	0.54
ENet	0.95	10.82***	0.05	0.51

10-year rolling estimates of encompassing coefs—VW

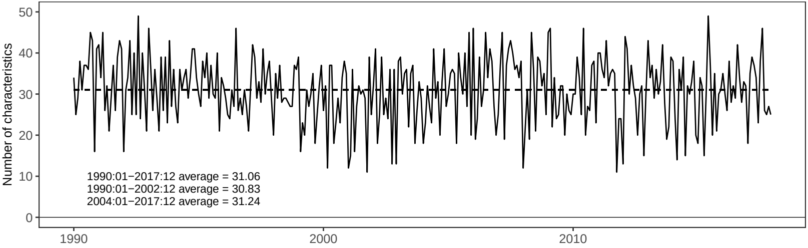


Number of selected characteristics—VW

A. Weighted LASSO



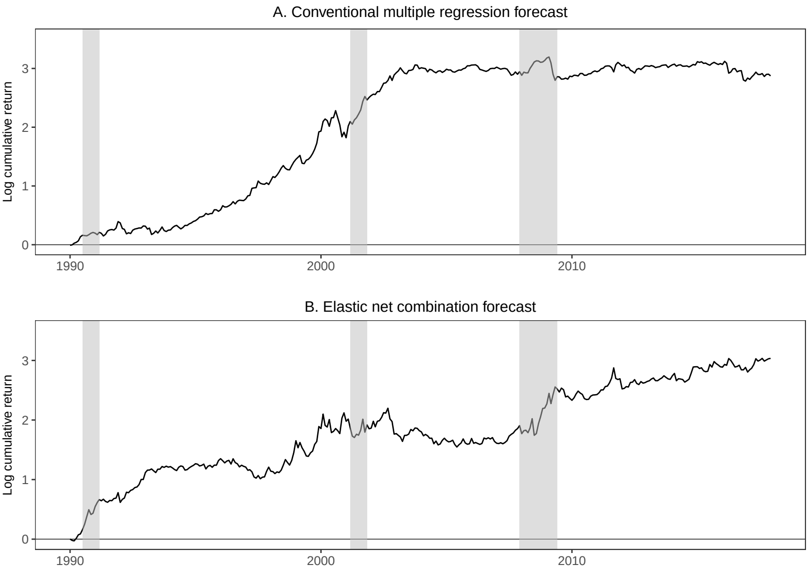
B. Weighted elastic net



ENet selection frequencies—VW

- Significant churn in characteristics selected over time
  - Nearly all $> 20\%$
  - Vast majority $> 30\%$
  - All $< 50\%$
- Top 9 ($> 40\%$)
  - `mom1m`, `tb`, `sin`, `nincr`, `salerec`
  - `chnanalyst`, `cinvest`, `chpmia`, `IPO`

Spread portfolio log cumulative returns—VW



Risk-adjusted average returns—VW

Carhart (1997) 4-factor model

	$\hat{\alpha}$		$\hat{\Delta}$	
Method	Coef	t-stat	Coef	t-stat
Conventional	1.41%	4.79***	−1.55%	−3.81***
Mean	1.56%	2.52**	−0.51%	−0.59
Tr mean	1.56%	2.60***	−0.60%	−0.74
LASSO	1.30%	2.37**	−0.21%	−0.27
ENet	1.31%	2.39**	−0.22%	−0.29

Risk-adjusted average returns—VW

Fama & French (2015) 5-factor model

	$\hat{\alpha}$		$\hat{\Delta}$	
Method	Coef	t-stat	Coef	t-stat
Conventional	1.99%	5.42***	−2.07%	−4.14***
Mean	1.27%	2.03**	−0.23%	−0.28
Tr mean	1.32%	2.19**	−0.37%	−0.45
LASSO	1.50%	2.65***	−0.32%	−0.42
ENet	1.50%	2.65***	−0.33%	−0.43

Risk-adjusted average returns—VW

Hou, Xue & Zhang (2015) *q*-factor model

	$\hat{\alpha}$		$\hat{\Delta}$	
Method	Coef	*t*-stat	Coef	*t*-stat
Conventional	1.57%	4.17***	−1.86%	−3.62***
Mean	1.60%	2.46**	−0.46%	−0.53
Tr mean	1.59%	2.55**	−0.57%	−0.67
LASSO	1.61%	2.72***	−0.49%	−0.61
ENet	1.62%	2.73***	−0.51%	−0.63

Take-aways

- Information in 94 firm characteristics useful over time
  - But need to process information so that we avoid overfitting
- Many characteristics matter (pre & post 2003)
- Economic implications
  - Characteristic premia vary over time
  - Sizable # of characteristics relevant at each point in time
- Keen challenge $\Rightarrow$ develop asset pricing models
  - Avoid overfitting
  - Facilitate economic interpretation