Firm Characteristics and Expected Stock Returns

Authors:
Yufeng Han
Ai He
David E. Rapach
Guofu Zhou

August 6, 2019

Presented by
Ziqi Tan
FIN500R Section 3
Sept 5^{th}, 2019

Agenda

- Introduction & Motivation
- Discuss the Work & Results
- Conclusion of this Paper
- My Thoughts on the Improvements

Firm Characteristics and Expected Stock Returns

62 Pages • Posted: 13 Jun 2018 • Last revised: 8 Aug 2019

Yufeng Han
University of North Carolina (UNC) at Charlotte - Finance

Ai He
University of South Carolina - Department of Finance

David Rapach
Saint Louis University; Washington University in St. Louis

Guofu Zhou
Washington University in St. Louis - John M. Olin Business School; China Academy of Financial Research (CAFR)

Date Written: August 6, 2019

Abstract

Complementing the widely used conventional multiple regression approach — which can suffer from overfitting with a large number of predictors — we propose a combination Lasso (C-Lasso) approach to improve out-of-sample forecasts of cross-sectional expected stock returns via shrinkage. Using 99 firm characteristics and an out-of-sample period spanning more than four decades, an approach that blends conventional and C-Lasso forecasts delivers unbiased estimates of the cross-sectional dispersion in expected returns. Similarly, combining spread portfolios formed from conventional and C-Lasso forecasts generates substantial performance gains. Our results indicate that more characteristics matter for cross-sectional expected returns than previously believed, due to time-varying characteristic premia.

Keywords: Cross-sectional expected stock returns, Characteristic premia, Forecast combination, Lasso, Forecast encompassing, Fama-MacBeth regression

JEL Classification: G11, G14

Suggested Citation:

Han, Yufeng and He, Ai and Rapach, David and Zhou, Guofu, Firm Characteristics and Expected Stock Returns (August 6, 2019). Available at SSRN: https://ssrn.com/abstract=3185335 or http://dx.doi.org/10.2139/ssrn.3185335

Introduction & Motivation

- Some Terminologies
- Define the problem
- How this Paper Addressed it
- Discuss the Work & Results
- Conclusion of this Paper
- My Thoughts on the Improvements

Terminologies

- Firm Characteristics
- Cross-sectional Expected Stock Returns

Conventional Multiple Regression Approach

$$r_{i,t} = a_t + \sum_{j=1}^{J} b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \qquad (1)$$

Where i is the individual stock,

$z_{i,j,t}$ is the month-t value for the jth characteristic for stock i.

The month (t + 1) cross-sectional return forecasts are given by

$$\hat{r}_{i,t+1|t} = \hat{a}_t + \sum_{j=1}^{J} \hat{b}_{j,t} z_{i,j,t} \qquad (2)$$

Where $\hat{a}_t, and\ \hat{b}_{j,t}$ are the OLS or WLS estimates of $a_t, and\ \ b_{j,t}$.

Unconventional Multiple Regression Approach Proposed by this Paper

Apply a robust forecast combination approach using machine learning tools to perform both shrinkage and variable selection in regression models with a large number of explanatory variables.

Introduction & Motivation

Define the Problem

- When we use conventional forecasts that rely on ordinary or weighted least squares to estimate high-dimensional linear regressions, the predictive ability of firm characteristics for US stock returns declines substantially after 2003.

High-dimensional linear regression → (OLS/WLS) → The conventional forecasts overstate the cross-sectional dispersion in expected returns

↓

An indication of overfitting → A substantive decline in predictive ability

↓ The question becomes

How to utilize information from the entire set of firm characteristics but in a manner that guards against overfitting?

Motivation to Tackle this Problem

- During the past decade, while the alpha generated from minimum volatility factor persists, factors such as value and growth did not have a satisfying performance. This further threats active managers because of the rising doubts on whether active managers can deliver after-fee alpha by actively selecting stocks.
- Increasing factor zoo
- A growing literatures employs machine leaning methods, including the Lasso, in finance.

How this Paper Addressed the Overfitting Problem – Comparison

Benchmark

Out-of-sample forecasts using a conventional multiple regression approach

$$r_{i,t} = a_t + \sum_{j=1}^{J} b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \qquad (1)$$

$$\hat{r}_{i,t+1|t} = \hat{a}_t + \sum_{j=1}^{J} \hat{b}_{j,t} z_{i,j,t} \qquad (2)$$

Competing Model - Combination Estimation

Produce return forecasts by first fitting a series of cross-sectional univariate regressions, each of which includes an individual firm characteristic as a predictor variable

Then pool the cross-sectional return forecasts corresponding to the individual characteristics (shrinkage strategy to guard against overfitting)

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

Step 1

For month t, we first estimate a series of cross-sectional univariate regressions, **relates returns to an individual characteristic:**

$$r_{i,t} = a_{j,t} + b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, I_t; \; j = 1, \ldots, J_{t-1}$$

ri,t the month t return for stock i
zi,j,t−1 the jth firm characteristic for stock in month (t − 1)
It the number of stocks available in quarter t
Jt−1 the number of characteristics available at the end of quarter t − 1

Input:
return for stock i in month t, firm characteristic for stock i in month (t-1)

Output:
j intercepts and j betas

Step 2

Construct month (t + 1) return forecasts for each stock based on each characteristic:

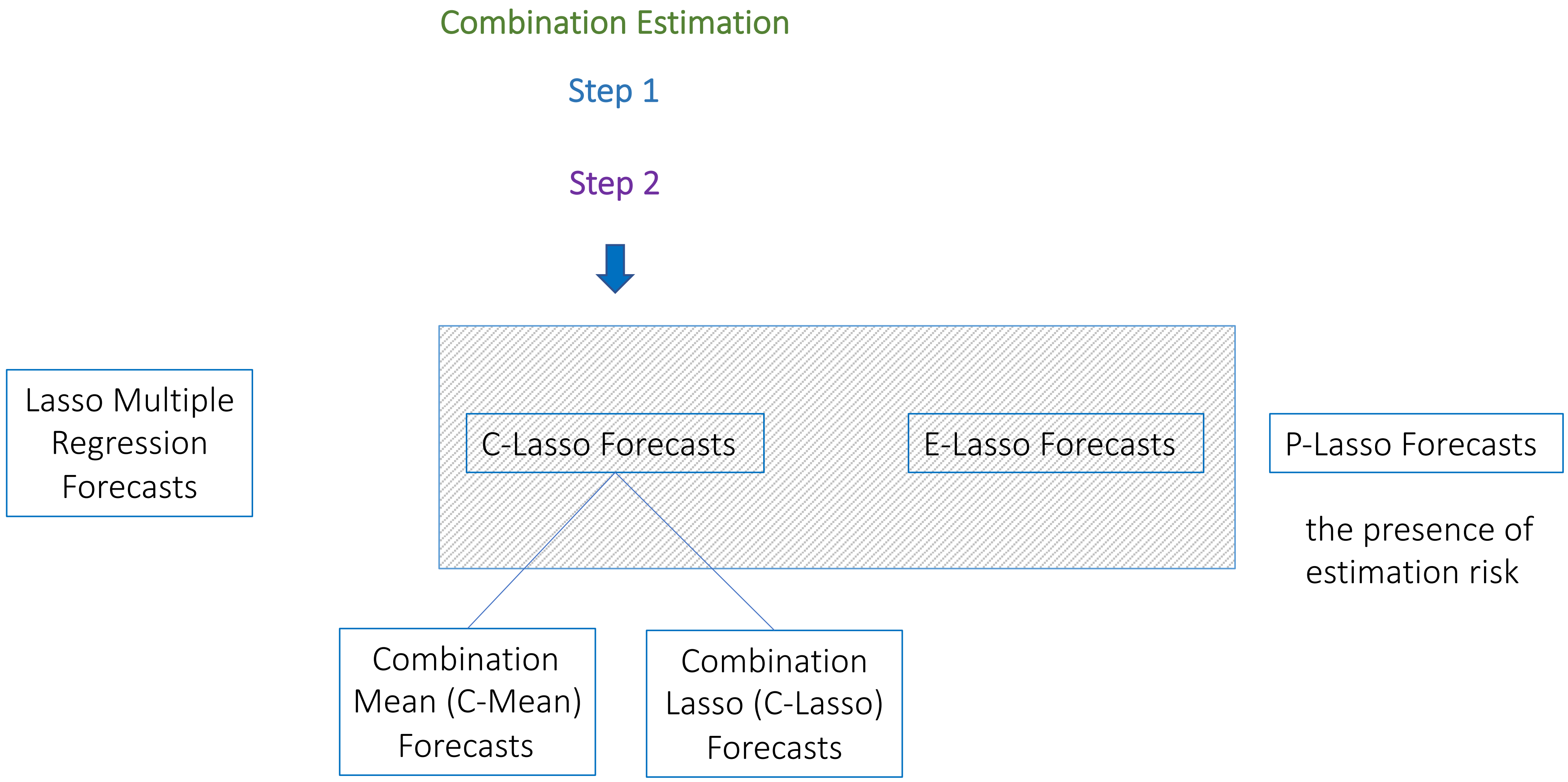
$$\hat{r}^{(j)}_{i,t+1|t} = \hat{a}_{j,t} + \hat{b}_{j,t} z_{i,j,t} \quad \text{for } i = 1, \ldots, I_{t+1}; \; j = 1, \ldots, J_t$$

Where aˆj,t and ˆbj,t come from Step 1

Input:
aˆj,t and ˆbj,t (intercept and beta from Step 1), the jth firm characteristic for stock i in month t

Output:
return forecasts for stock i
for characteristic j in month (t+1)

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow



How this Paper Addressed the Overfitting Problem – Summary

	Train: Regressor (X)	Train: Regressand (Y)	Train: Assign weights?	Predict: X	Predict: Coefficient(s)	Predict: Y hat
Lasso Multiple Regression Forecasts	Factor exposure to all characteristics the month (t-1) value for the jth characteristic for stock i	Realized return the month t realized return for stock i	Lasso	Factor exposure to all characteristics the month t value for the jth characteristic for stock i	Weights assigned to each factor in the training model in month(t-1)	
C-Lasso Forecasts: C-Mean Forecasts	Factor exposure to each characteristic the month (t-1) value for the jth characteristic for stock i	Realized return the month t realized return for stock i	OLS	Return Forecasts from simple linear regression the return forecasts based on the individual characteristic	Simple average	
C-Lasso Forecasts: C-Lasso Forecasts	Same as C-Mean	Same as C-Mean	Same as C-Mean	Same as C-Mean	Weights assigned to each y hat in the training model in month(t-1)	
E-Lasso Forecasts	Blend the conventional and C-Lasso forecasts to improve the statistical accuracy of cross-sectional return forecasts					
P-Lasso Forecasts	Blend the conventional and C-Lasso forecasts to improve the investment performance				The presence of estimation risk	

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

Lasso Multiple Regression Forecasts

Step 3

Instead of estimating Equation (1) via conventional OLS or WLS, we use the following objective function:

$$\arg\min_{a_t \in \mathbb{R},\, \mathbf{b}_t \in \mathbb{R}^J} \left\{ \frac{1}{2I_t} \sum_{i=1}^{I_t} w_{i,t} \left[r_{i,t} - \left(a_t + \sum_{j=1}^{J} b_{j,t} z_{i,j,t-1} \right) \right]^2 + \lambda_t \|\mathbf{b}_t\|_1 \right\},$$

where

$$\mathbf{b}_t = \begin{bmatrix} b_{1,t} & \dots & b_{J,t} \end{bmatrix}',$$

The Lasso multiple regression forecasts are given by

$$\hat{r}^{\text{Lasso}}_{i,t+1|t} = \hat{a}^{\text{Lasso}}_t + \sum_{j=1}^{J} \hat{b}^{\text{Lasso}}_{j,t} z_{i,j,t},$$

for $i = 1, \dots, I_{t+1}$, where $\hat{a}^{\text{Lasso}}_t$ and $\hat{b}^{\text{Lasso}}_{j,t}$ are the unweighted or weighted Lasso estimates of a_t and $b_{j,t}$, respectively, for $j = 1, \dots, J$

Input:
Same as conventional approach

Output:
Lasso multiple regression forecasts for stock i

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

C-Mean Forecasts

Step 3

Compute a simple combination forecast of ri,t+1 by taking the arithmetic mean (or trimmed mean) of the individual forecasts:

$$\hat{r}^{\text{Mean}}_{i,t+1|t} = \frac{1}{J_t}\sum_{j=1}^{J_t} \hat{r}^{(j)}_{i,t+1|t} \quad \text{for } i = 1,\ldots,I_{t+1}$$

Input:
return forecast for stock i for firm characteristic j in month (t+1) from Step 2

Output:
Simple average return forecast for stock i

C-Mean Forecasts 2.0

Step 3

$$\hat{r}^{\text{Mean}}_{i,t+1|t} = \bar{r}_t + \frac{1}{J_t}\sum_{j=1}^{J_t} \hat{b}_{j,t}(z_{i,j,t} - \bar{z}_{j,t}) \quad \text{for } i = 1,\ldots,I_{t+1},$$

where

$$\bar{r}_t = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t} r_{i,t},$$

$$\bar{z}_{j,t} = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t} z_{i,j,t-1},$$

Input:
aˆj,t and ˆbj,t (intercept and beta from Step 1), the jth firm characteristic for stock i in month t

Output:
Adjusted return forecasts for stock i for characteristic j in month (t+1)

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

C-Lasso Forecasts

Step 3

Improve combination forecasts in a time-series context - use the Lasso to refine the cross-sectional C-Mean forecasts

Consider the following cross-sectional version of a multiple regression for month t involving the univariate regression forecasts:

$$r_{i,t} = \xi_t + \sum_{j=1}^{J} \phi_{j,t} \hat{r}^{(j)}_{i,t|t-1} + \varepsilon_{i,t} \qquad (3)$$

We estimate Equation (3) using the Lasso objective function:

$$\underset{\xi_t \in \mathbb{R},\, \boldsymbol{\phi}_t \in \mathbb{R}^J_{\geq 0}}{\arg\min} \left\{ \frac{1}{2I_t} \sum_{i=1}^{I_t} w_{i,t} \left[r_{i,t} - \left(\xi_t + \sum_{j=1}^{J} \phi_{j,t} \hat{r}^{(j)}_{i,t|t-1} \right) \right]^2 + \lambda_t \|\boldsymbol{\phi}_t\|_1 \right\},$$

where

$$\boldsymbol{\phi}_t = \begin{bmatrix} \phi_{1,t} & \dots & \phi_{J,t} \end{bmatrix}'.$$

Input:
return forecast for stock i for firm characteristic j in month (t+1) from Step 2

Output:
Next slide

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

C-Lasso Forecasts

Step 3 (Continued)

$$r_{i,t} = \xi_t + \sum_{j=1}^{J} \phi_{j,t} \hat{r}^{(j)}_{i,t|t-1} + \varepsilon_{i,t} \qquad (3)$$

Let $\hat{\mathcal{M}}_t \subseteq \{1, \ldots, J\}$ denote the index set of cross-sectional univariate regression forecasts selected by the Lasso in Equation (3). The C-Lasso forecasts are given by

$$\begin{aligned} \hat{r}^{\text{C-Lasso}}_{i,t+1|t} &= \frac{1}{|\hat{\mathcal{M}}_t|} \sum_{j \in \hat{\mathcal{M}}_t} \hat{r}^{(j)}_{i,t+1|t} \\ &= \frac{1}{|\hat{\mathcal{M}}_t|} \sum_{j \in \hat{\mathcal{M}}_t} \left[\bar{r}_t + \hat{d}_{j,t}(z_{i,j,t} - \bar{z}_{j,t-1}) \right] \\ &= \bar{r}_t + \sum_{j \in \hat{\mathcal{M}}_t} \frac{1}{|\hat{\mathcal{M}}_t|} \hat{d}_{j,t}(z_{i,j,t} - \bar{z}_{j,t-1}) \end{aligned}$$

Input:
return forecast for stock i for firm characteristic j in month (t+1) from Step 2

Output:
C-Lasso forecasts for stock i for characteristic j in month (t+1)

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

E-Lasso Forecasts

Step 3

E-Lasso blends the conventional and C-Lasso forecasts, are given by

$$\hat{r}^{\text{E-Lasso}}_{i,t+1|t} = \left(1 - \hat{\theta}_t\right)\tilde{r}_{i,t+1|t} + \hat{\theta}_t \hat{r}^{\text{C-Lasso}}_{i,t+1|t},$$

for $i = 1, \ldots, I_{t+1}$, where $\hat{\theta}_t$ is the OLS or WLS estimate of θ_t

where

$$\tilde{\theta}_t = \frac{1}{M} \sum_{m=0}^{M-1} \hat{\theta}_{t-m}.$$

Note: in the paper, the author expects "moderate" values of M corresponding to two to four years to be most effective.

Input:
Conventional forecasts and C-Lasso forecasts

Output:
E-Lasso forecasts for stock i
for characteristic j in month (t+1)

How this Paper Addressed the Overfitting Problem – A Step-by-step Workflow

Combination Estimation

P-Lasso Forecasts

Step 3

P-Lasso blends the weights for the decile spread portfolios based on the conventional and C-Lasso forecasts to improve investment performance.

Specifically, let ω1,t+1 and ω2,t+1 denote the It+1-dimensional vectors of month-(t + 1) weights for the spread portfolios based on the conventional and C-Lasso forecasts. We construct a P-Lasso allocation whose weights are given by

$$\omega_{P,t+1} = (1 - \rho_t^{\text{MV}})\omega_{1,t+1} + \rho_t^{\text{MV}}\omega_{2,t+1},$$

where

$$\rho_t^{\text{MV}} = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_{12}}{\hat{\sigma}_1^2 - 2\hat{\sigma}_{12} + \hat{\sigma}_2^2},$$

$\hat{\sigma}_1^2$ ($\hat{\sigma}_2^2$) is the sample variance for the spread portfolio based on the conventional (C-Lasso) forecasts, and $\hat{\sigma}_{12}$ is the sample covariance for the spread portfolio returns. In computing ρ_t^{MV}, we estimate the sample variances and covariance using data through month t.

Input:
Conventional forecasts and C-Lasso forecasts

Output:
P-Lasso forecasts for stock i
for characteristic j in month (t+1)

Discuss the Work & Results

A brief description of its work

- Range: 1965:01–2018:06
- Investment horizon: monthly
- Number of firm characteristics: 99
- Data transformation: winsorization
- 4 cases of portfolio construction:
  - Value weighting for all stocks (VW-All)
  - Equal weighting for large stocks (EW-Large)
  - Equal weighting excluding micro-cap stocks (EW-ExMicro)
  - Equal weighting for all stocks (EW-All)
- 6 cases of out-of-sample return forecasts
  - Conventional
  - Lasso Multiple Regression
  - C-Mean
  - C-Lasso
  - E-Lasso
  - P-Lasso
- 2 test methods to analyze the forecasts
  - Predictive slopes
  - Forecast encompassing tests
- 4 competing cases that have test results
  - Conventional Forecasts
  - C-Lasso Forecasts
  - E-Lasso Forecasts
  - P-Lasso Forecasts

Discuss the Work & Results

How the overfitting problem is addressed

Forecasting Approach	Overfitting Problem	
Conventional Forecasts	When J is large (a high-dimensional model), the cross-sectional return forecasts are susceptible to overfitting. This concern is exacerbated when forecasting stock returns, as the noise component in returns is inherently sizable.	4
Lasso Multiple Regression Forecasts	Like the conventional regression forecasts—the Lasso multiple regression forecasts are typically characterized by significant overfitting. Thus, it insufficiently shrinks the coefficient estimates	4
C-Mean Forecasts	It makes two adjustments: (i) it replaces the OLS or WLS multiple regression slope coefficient estimates with their univariate counterparts; (ii) it shrinks the magnitude of each slope coefficient by the factor 1/J, which has the effect of strongly shrinking the forecast to the cross-sectional mean return.	3
C-Lasso Forecasts	It incorporates both the generally strong shrinkage property of the C-Mean forecasts and the ability of the Lasso to select relevant predictor variables.	2
E-Lasso Forecasts/P-Lasso Forecasts	We can improve overall out-of-sample performance by pooling the conventional and C-Lasso forecasts. The encompassing framework provides a method for optimally pooling the conventional and C-Lasso forecasts.	1

Discuss the Work & Results

Suitability of each forecasting approach

Forecasting Approach	When it Shines	Ranking of robustness
Conventional Forecasts	Performs relatively well when characteristic premia are fairly stable	4
Lasso Multiple Regression Forecasts	n.a.	4
C-Mean Forecasts	The strong shrinkage property of forecast combination works to stabilize the forecasts by making them significantly less volatile. Forecast stabilization helps to improve out-of-sample performance in environments with a low signal-to-noise ratio	2
C-Lasso Forecasts	Smoothing the univariate coefficient estimates over time when forming the combination forecasts tends to make the cross-sectional return forecasts too conservative.(con) It is likely to prove especially useful for tracking cross-sectional expected returns when characteristic premia are time varying/vary substantially over time	2
E-Lasso Forecasts/P-Lasso Forecasts	A flexible shrinkage strategy Allows the data to inform the degree of shrinkage	1

- Introduction & Motivation

- Discuss the Work & Results
  - A brief description
  - How the overfitting problem is addressed
  - Suitability of each forecasting approach
  - Partial test results

- Conclusion of this Paper

- My Thoughts on the Improvements

Discuss the Work & Results

Partial test results

Table 4: Forecast encompassing tests

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\hat{\theta}$		$1-\hat{\theta}$		$\hat{\theta}$		$1-\hat{\theta}$	
Out-of-sample period	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.
	Panel A: VW-All				*Panel B: EW-Large*			
1975:01–2018:06	0.61	**7.21**	0.39	**4.67**	0.77	**13.17**	0.23	**3.97**
1975:01–1984:12	0.59	**5.78**	0.41	**3.98**	0.67	**10.48**	0.33	**5.19**
1985:01–1994:12	0.85	**5.11**	0.15	0.92	0.89	**14.55**	0.11	**1.75**
1995:01–2004:12	0.17	0.85	0.83	**4.17**	0.58	**3.19**	0.42	**2.32**
2005:01–2018:06	0.76	**5.02**	0.24	1.55	0.89	**21.56**	0.11	**2.67**
	Panel C: EW-ExMicro				*Panel D: EW-All*			
1975:01–2018:06	0.61	**7.31**	0.39	**4.59**	0.41	**5.85**	0.59	**8.57**
1975:01–1984:12	0.44	**4.70**	0.56	**5.94**	0.31	**4.58**	0.69	**10.39**
1985:01–1994:12	0.59	**6.77**	0.41	**4.70**	0.18	**2.12**	0.82	**9.88**
1995:01–2004:12	0.49	**2.20**	0.51	**2.29**	0.31	**1.81**	0.69	**3.98**
2005:01–2018:06	0.85	**7.34**	0.15	1.28	0.72	**16.23**	0.28	**6.39**

(2) (6) C-Lasso, (4) (8) Conventional

Conclusion of this Paper

- By overcoming the overfitting problem that can plague conventional multiple regression forecasts, methods in this paper indicate that a larger number of firm characteristics are relevant for explaining cross-sectional expected stock returns than previously believed.
- Nearly all of the 99 characteristics that this paper considers are relevant a good portion of the time, while approximately 20 to 30 are relevant on average at a given point in time. These results are consistent with time-varying characteristic premia, which is particularly important around business-cycle recessions.
- The C-Lasso approach accommodates time-varying characteristic premia in a manner that guards against overfitting.
- The E-Lasso approach optimally blends conventional multiple regression forecasts with the C-Lasso forecasts, and compared to peers, E-Lasso forecasts in this paper appear to provide the best out-of-sample estimates to date of the cross-sectional dispersion in expected returns.
- The P-Lasso approach, similarly to blending the conventional and C-Lasso forecasts to improve the statistical accuracy of cross-sectional return forecasts, blends spread portfolios formed from the conventional and substantially enhances performance in the form of higher Sharpe ratios.

- Key takeaways:

a. By fixing the overfitting problem, a larger number of firm characteristics are relevant for explaining cross-sectional expected stock returns even after 2003.
b. We expect conventional multiple regression forecasts to perform relatively well when characteristic premia are fairly stable, while the C-Lasso forecasts will likely perform better when premia vary substantially over time.
c. We can interpret the E-Lasso and P-Lasso approach as flexible shrinkage strategy. By estimating the weights on two forecasts, we allow the data/sample variances and covariance to inform the degree of shrinkage.
d. The P-Lasso allocations produce significantly positive average returns and better Sharpe ratios for all cases and all samples.

Thoughts on the Improvements

- Challenges and doubts may encounter and proposals to solve it

  Data-related:

  How to deal with missing value? – Fill it with the universe median/leave it blank

  Some firm characteristics are only available on quarterly basis - Not sure yet

  The number of available factors may lead to biased estimation in early years (before 2006) – focus on back-testing results in recent 10 years

  Model-related:

  Lasso may choose non of the variables for some month t – use combination estimation

  Lasso can generate extremely large coefficients – Try Adaptive Lasso

  The coefficients generated by Lasso are random even for the same data set – Use iteration to tone the parameters

- Thoughts on data transformation

  Instead of winsorization, use log transformation and Box-cox transformation

- Consider non-linear relationship

  Apply non-linear models such as Random Forest and Neutral Network

- May consider to design an algorithm, self-adjusted to the degree of time-varying

  Instead of minimizing MSE/MSFE at one point of time, minimizing MSE over the last certain of periods