Ensuring sparsity in a high-dimensional environment with correlated regressors: A penalization approach

Hugo Finizola Stellet[1]

[1] Puc-Rio

July 2023

Outline

Factor-based asset pricing

Factor-based models have been relevant since the CAPM proposed the market factor (see Sharpe (1964) and Lintner (1965)).

- Fama and French (1993, 2015) incorporated additional factors, updating the benchmark for market efficiency.

Factor-based asset pricing

Factor-based models have been relevant since the CAPM proposed the market factor (see Sharpe (1964) and Lintner (1965)).

- Fama and French (1993, 2015) incorporated additional factors, updating the benchmark for market efficiency.

However, researchers have continuously claimed the discovery of new pricing factors, often referred to as ***pricing anomalies***:

> Defined by Brennan and Xia (2001) as *"a statistically significant difference between the realized average returns associated with certain characteristics of securities."*

The *zoo of factors*

However, an overwhelming number of anomalies have been claimed as significant pricing factors over the past few years:

- Hou et al. (2020) even compiled a data library with ***447 anomaly factors*** from the literature.

The *zoo of factors*

However, an overwhelming number of anomalies have been claimed as significant pricing factors over the past few years:

- Hou et al. (2020) even compiled a data library with ***447 anomaly factors*** from the literature.

Cochrane (2011) characterized this situation as a ***factor zoo***, advocating against the proliferation of factors.

- Nevertheless, it is hard to imagine that all of these factors are ***jointly*** relevant in pricing the average cross-section of returns.

The *zoo of factors*

However, an overwhelming number of anomalies have been claimed as significant pricing factors over the past few years:

- Hou et al. (2020) even compiled a data library with ***447 anomaly factors*** from the literature.

Cochrane (2011) characterized this situation as a ***factor zoo***, advocating against the proliferation of factors.

- Nevertheless, it is hard to imagine that all of these factors are ***jointly*** relevant in pricing the average cross-section of returns.

This raises the question: What factors are ***truly relevant*** for predicting the cross-section of asset returns?

Time-varying factors and sparsity

Factors' relevancy may not remain constant throughout the time series. Sun (2023) divides his sample into two periods, revealing substantial variations in the selected factors between these periods.

- A framework that allows for ***exploring the time-varying relevance*** of factors could be of significant interest.

Time-varying factors and sparsity

Factors' relevancy may not remain constant throughout the time series. Sun (2023) divides his sample into two periods, revealing substantial variations in the selected factors between these periods.

- A framework that allows for ***exploring the time-varying relevance*** of factors could be of significant interest.

Asset pricing models have implicitly made a ***silent sparsity assumption***, typically considering only a handful of factors.

- It might be advantageous to ***impose a similar degree of sparsity*** in machine learning-based factor selection models.

Research objectives

Scan the factor zoo in a ***time-varying*** fashion while enabling factor selection and return predictions over distinct periods:

- We propose variations in the usual approach while conducting true out-of-sample exercises and assessing predictability against a more coherent benchmark.

Research objectives

Scan the factor zoo in a ***time-varying*** fashion while enabling factor selection and return predictions over distinct periods:

- We propose variations in the usual approach while conducting true out-of-sample exercises and assessing predictability against a more coherent benchmark.

Explore different methodologies for ***guaranteeing sparsity*** in shrinkage regressions:

- Usual criteria for defining penalization parameter(s) may be more permissive than ideal, and a more drastic dimension-reducing approach could be appropriate.

Preview of results

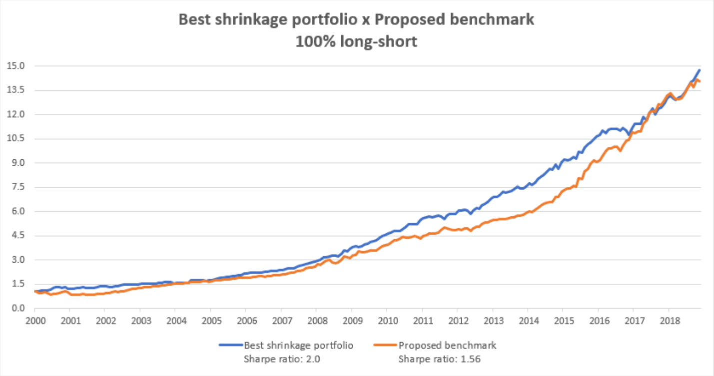
Our findings indicate that:

- Ensuring some level of ***sparsity is desirable***, as factor selection performs better when a smaller number of factors is used;
- Factor-asset ***correlation should be measured over a shorter period***, closer to the time of return prediction, and;
- ***Predictability*** improves when the ***time series*** used to estimate selected factors' β's is ***longer***.

Preview of results

Considering non-scaled long-short strategies, our best portfolio obtained about the ***same return*** as the proposed strict benchmark.

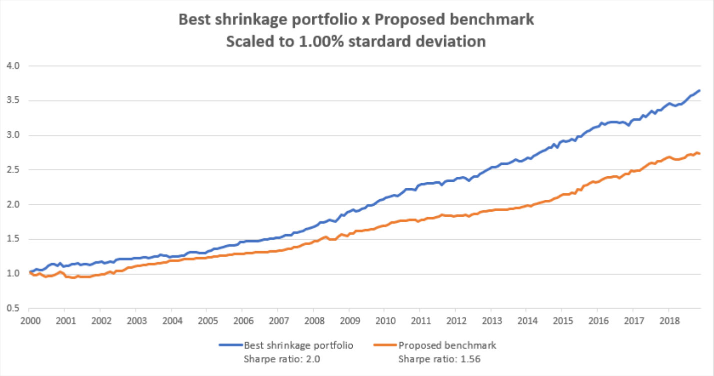
- Experiencing ***less volatility***.



Preview of results

A visually fair comparison is obtained by scaling the returns to the same standard deviation (1.00%).

- The shrinkage portfolio appears to ***outclass the benchmark***.



Literature

Machine-learning-driven asset pricing

Horowitz (2016), Huang et al. (2010), Chinco et al. (2019), Hiraki and Sun (2022)

- Apply machine learning techniques to factor selection.

Literature

Machine-learning-driven asset pricing

Horowitz (2016), Huang et al. (2010), Chinco et al. (2019), Hiraki and Sun (2022)

- Apply machine learning techniques to factor selection.

Shrinkage regressions applied to the SDF model

Feng et al. (2020), Sun (2023), Freyberger et al. (2020)

- Establish a framework that separates factor selection and returns prediction periods while enabling a truly out-of-sample exercise;
- Provide insights into the optimal time frames for evaluating each aspect of the problem.

Literature

Penalization parameter(s) of shrinkage regressions

Schwarz (1978), Akaike (1974), Stone (1974)

- Propose a criteria focused on the number of selected regressors to determine the penalization parameter(s).

Literature

Penalization parameter(s) of shrinkage regressions

Schwarz (1978), Akaike (1974), Stone (1974)

- Propose a criteria focused on the number of selected regressors to determine the penalization parameter(s).

Benchmark for accepting factors' significance

Feng et al. (2020), Fama and French (2015, 1993)

- Apply the out-of-sample methodology to establish a stricter benchmark for verifying the results' significance.

Stochastic discount factor model

We employ the Stochastic Discount Factor (SDF) methodology, as described in Cochrane (2009). The SDF (m) is denoted as:

$$m := r_0^{-1}[1 - b'(f - \mathbb{E}[f])], \tag{1}$$

where r_0^{-1} is a constant zero-beta rate, f is a $K \times 1$ vector of factor returns, and b is the $K \times 1$ vector of SDF's coefficients.

SDF pricing error

For the model to hold, any admissible value for ***m must satisfy the fundamental asset pricing equation***: $\mathbb{E}[Rm] = 0$ - where R is a $N \times 1$ vector of asset returns.

- However, this equality ***may not hold when m is unknown*** beforehand and is instead estimated from some model.

SDF pricing error

For the model to hold, any admissible value for ***m must satisfy the fundamental asset pricing equation***: $\mathbb{E}[Rm] = 0$ - where R is a $N \times 1$ vector of asset returns.

- However, this equality ***may not hold when m is unknown*** beforehand and is instead estimated from some model.

Taking the natural definition of the pricing error $e(b)$ and denoting the *unknown* SDF as $m(b)$, ***the pricing error*** can be written as:

$$
\begin{aligned}
e(b) = \mathbb{E}[Rm(b)] &= \mathbb{E}[R]\,\mathbb{E}[m(b)] + \mathbb{C}\text{ov}[R, m(b)] \\
&= r_0^{-1}\,\mathbb{E}[R]\,\mathbb{E}[1 - b'(f - \mathbb{E}[f])] + r_0^{-1}\,\mathbb{C}\text{ov}[R, 1 - b'(f - \mathbb{E}[f])] \\
&= r_0^{-1}(\mathbb{E}[R] - \mathbb{C}\text{ov}[R, f]b) \\
&= r_0^{-1}(\mu_R - Cb)
\end{aligned}
\tag{2}
$$

SDF pricing error

Taking the pricing error quadratic form as

$$Q(b) = e(b)'We(b), \tag{3}$$

where W is an appropriate $N \times N$ weighting matrix, it is possible to ***estimate risk prices b*** by minimizing their quadratic error $Q(b)$:

$$\begin{aligned} \hat{b} &= \arg\min_b Q(b) \\ &= \arg\min_b [(\mu_R - Cb)'W(\mu_R - Cb)] \end{aligned} \tag{4}$$

Optimization equation

Therefore,

$$\hat{\hat{b}} = (\hat{C}'\hat{W}\hat{C})^{-1}\hat{C}'\hat{W}\hat{\mu}_R,^{1} \tag{5}$$

where $\hat{\mu}_f = (1/T)\sum_{t=1}^{T} f_t$, $\hat{\mu}_R = (1/T)\sum_{t=1}^{T} R_t$, and $\hat{C} = \hat{\mathbb{C}}\mathrm{ov}(R, f) = (1/T)\sum_{t=1}^{T}(R_t - \hat{\mu}_R)(f_t - \hat{\mu}_f)'$.

- Notice that $\hat{\hat{b}}$ is an ***empirical estimate*** of $\hat{b}$, which makes use of ***sample estimates*** $\hat{C}$ and $\hat{\mu}_R$.

[1] As a constant, r_0 can be dropped from the optimization.

Optimization equation

Therefore,

$$\hat{\hat{b}} = (\hat{C}'\hat{W}\hat{C})^{-1}\hat{C}'\hat{W}\hat{\mu}_R,^1 \tag{5}$$

where $\hat{\mu}_f = (1/T)\sum_{t=1}^{T} f_t$, $\hat{\mu}_R = (1/T)\sum_{t=1}^{T} R_t$, and $\hat{C} = \hat{\mathbb{C}}\mathrm{ov}(R, f) = (1/T)\sum_{t=1}^{T}(R_t - \hat{\mu}_R)(f_t - \hat{\mu}_f)'$.

- Notice that $\hat{\hat{b}}$ is an ***empirical estimate*** of $\hat{b}$, which makes use of ***sample estimates*** $\hat{C}$ and $\hat{\mu}_R$.

As test assets are both abundant and have economic interpretations, Ludvigson (2013) suggests setting W as the identity, therefore:

$$\hat{\hat{b}} = \arg\min_{b}[(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] \tag{6}$$

[1] As a constant, r_0 can be dropped from the optimization.

Achieving sparsity

Regularization and ***dimension-reducing techniques*** have emerged as possible solutions to scan the zoo of factors for relevant ones.

Shrinkage estimators are usually implemented by adding a penalty term ($\Omega(b)$) to Equation 6:

$$\hat{\hat{b}} = \arg\min_b [(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] + \Omega(b) \tag{7}$$

Achieving sparsity

Regularization and ***dimension-reducing techniques*** have emerged as possible solutions to scan the zoo of factors for relevant ones.

Shrinkage estimators are usually implemented by adding a penalty term ($\Omega(b)$) to Equation 6:

$$\hat{\hat{b}} = \arg\min_{b} [(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] + \Omega(b) \tag{7}$$

The definition of $\Omega(b)$ depends on the specific technique. Some commonly used definitions are:

- LASSO: $\Omega(b)_{\text{LASSO}} = \lambda \sum_{j=1}^{K} |b_j|$
- Elastic Net: $\Omega(b)_{\text{eNet}} = \lambda \sum_{j=1}^{K} [(1-\alpha)|b_j|^2 + \alpha |b_j|]$
- Adaptive LASSO: $\Omega(b)_{\text{A-LASSO}} = \lambda \sum_{j=1}^{K} \hat{w}_j |b_j|$

Achieving sparsity

Setting the parameter(s) of the ***penalty term*** is a task for researchers. Literature suggests some methodologies for accomplishing this, like:

- Cross-Validation (CV)
- Information Criteria:
  - Bayesian Information Criterion (BIC)
  - Akaike Information Criterion (AIC)

As suggested in Freyberger et al. (2020), we believe that the ***BIC*** methodology is more adequate for our purposes.

Underlying sparsity assumption

The asset pricing literature has ***implicitly relied on*** the concept of ***sparsity*** by comparing new factors against models with a small set of control variables, as discussed in Feng et al. (2020).

> Those benchmark models represent a ***selection of a few factors*** from the vast array of potentially relevant factors for explaining cross-sectional expected returns.

Underlying sparsity assumption

The asset pricing literature has ***implicitly relied on*** the concept of ***sparsity*** by comparing new factors against models with a small set of control variables, as discussed in Feng et al. (2020).

> Those benchmark models represent a ***selection of a few factors*** from the vast array of potentially relevant factors for explaining cross-sectional expected returns.

However, what distinguishes this type of sparsity from a ***machine-learning***-based approach?

- Distinctively from traditional asset pricing models, the former ***does not have biases*** toward any specific explanatory factor.

Forcing sparsity

What if we want to consider only the ***top n most relevant factors***?

- One possibility is to ***rank the magnitude of*** the $\hat{\hat{b}}$'s ***coefficients*** and consider only the n largest as relevant - see Sun (2023).

Forcing sparsity

What if we want to consider only the ***top n most relevant factors***?

- One possibility is to ***rank the magnitude of*** the $\hat{\hat{b}}$'s ***coefficients*** and consider only the n largest as relevant - see Sun (2023).

However, when some regressors exhibit significant correlation, potentially relevant factors may have their $\hat{\hat{b}}$ magnitudes diminished (even disregarded) due to the presence of correlated factors.

- Top n ranked factors approach might overlook ***the effect of correlated factors***, which already tend to have their $\hat{\hat{b}}$'s values reduced.

Forcing sparsity

Sun (2023) tackles this issue by altering the definition of $\Omega(b)$ in Equation 7 and utilizing the OWL estimator (Figueiredo and Nowak, 2016).

- However, in the out-of-sample analysis, he still only considers the top n factors, disregarding other selected factors.

Forcing sparsity

Sun (2023) tackles this issue by altering the definition of $\Omega(b)$ in Equation 7 and utilizing the OWL estimator (Figueiredo and Nowak, 2016).

- However, in the out-of-sample analysis, he still only considers the top n factors, disregarding other selected factors.

Moreover, when arbitrarily disregarding selected factors, we may not be fully leveraging the ***properties of the chosen shrinkage regression***.

- Therefore, it might not be appropriate to eliminate any selected factor, and considering ***another criterion*** for properly setting the ***penalization parameter*** may be a better option.

Ensuring sparsity

What if, instead of adopting one of the existing criteria, which usually focuses on minimizing prediction errors, we decide to *choose the penalization parameter to **ensure** a predetermined number of factors*?

Ensuring sparsity

What if, instead of adopting one of the existing criteria, which usually focuses on minimizing prediction errors, we decide to *choose the penalization parameter to* ***ensure*** *a predetermined number of factors*?

- By ***fixing*** the number of selected factors n, we aim to ***preserve regression properties*** while achieving a model that is as ***shrunk*** as desired.

Ensuring sparsity

The idea is to ***dynamically set the penalization parameter*** (λ) to ensure that *n* factors survive the shrinking process.

λ will be set as follows:

- Start with a reasonable initial value for λ and run the regression to determine the number of selected factors.
  - If the number of selected factors is too small, decrease λ.
  - If the number of selected factors is too large, increase λ.
- Repeat this process until the regression returns the desired number of factors.[2]

[2] The rate at which λ changes should be adjusted accordingly as the number of selected factors approaches *n*.

Time frame(s) of interest

Choosing the appropriate time frame is crucial to avoid potential biases and ensure the validity of findings:

- For a ***time-invariant*** model, it is advisable to consider the ***longest*** possible ***time series*** of returns (see Feng et al., 2020).
- However, if we aim for ***out-of-sample predictability***, it may be interesting to allow for time-varying selection, where the ***relevancy of factors may change*** over time.

Out-of-sample analysis

Out-of-sample (OOS) analysis of asset pricing models typically involves a ***long-short*** strategy based on the predicted returns of assets.

We propose a slight modification to the methodology employed in Freyberger et al. (2020), which consists of, in a ***rolling-windows setup***:

[3] Similarly to a post-LASSO estimator (Belloni and Chernozhukov, 2013).

Out-of-sample analysis

Out-of-sample (OOS) analysis of asset pricing models typically involves a ***long-short*** strategy based on the predicted returns of assets.

We propose a slight modification to the methodology employed in Freyberger et al. (2020), which consists of, in a ***rolling-windows setup***:

- Utilizing the model's selected factors to ***predict assets'*** 1-month-ahead ***returns***. Predictions are calculated using ordinary least squares (OLS) regression;[3]

[3] Similarly to a post-LASSO estimator (Belloni and Chernozhukov, 2013).

Out-of-sample analysis

Out-of-sample (OOS) analysis of asset pricing models typically involves a ***long-short*** strategy based on the predicted returns of assets.

We propose a slight modification to the methodology employed in Freyberger et al. (2020), which consists of, in a ***rolling-windows setup***:

- Utilizing the model's selected factors to ***predict assets'*** 1-month-ahead ***returns***. Predictions are calculated using ordinary least squares (OLS) regression;[3]
- ***Establishing*** a long (short) ***position*** on assets with the highest (lowest) predicted returns, and verifying the resulting ***Sharpe ratio***.

[3] Similarly to a post-LASSO estimator (Belloni and Chernozhukov, 2013).

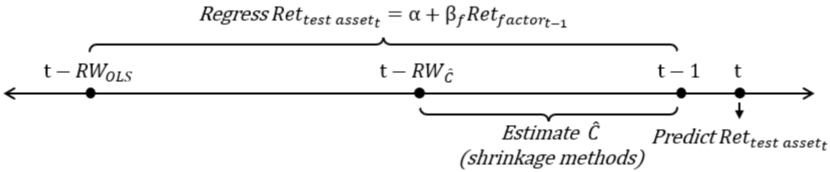
Time-varying selection

Our approach differs from Freyberger et al. (2020) as ***we separate the periods used for factor selection and returns prediction***, considering two distinct rolling window parameters:

Time-varying selection

Our approach differs from Freyberger et al. (2020) as ***we separate the periods used for factor selection and returns prediction***, considering two distinct rolling window parameters:

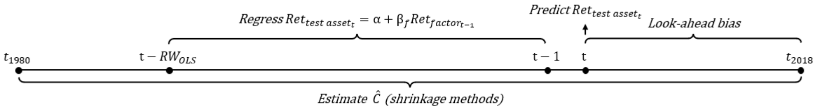
- $RW_{\hat{C}}$: Rolling window length for calculating $\hat{C}$, the estimate of the factor covariance matrix - and selecting relevant factors;
- RW_{OLS}: Rolling window length for regressing chosen factors' returns onto the returns of the test assets.



Full panel selection

Out-of-sample analysis in large panels can be challenging, as selecting factors using all available data introduces ***look-ahead bias*** into results.

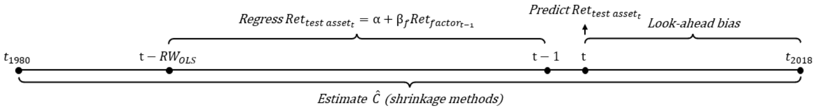
- Thereafter, ***relevant factors must be re-estimated every period*** prior to predicting assets' returns.



Full panel selection

Out-of-sample analysis in large panels can be challenging, as selecting factors using all available data introduces ***look-ahead bias*** into results.

- Thereafter, ***relevant factors must be re-estimated every period*** prior to predicting assets' returns.



Factor selection in shorter panels also provides insights into how asset returns' cross-section ***has been*** behaving ***near*** prediction period:

It also enables the study of how relevant factors for pricing the average cross-section of returns ***change*** over time.

Sources

Stock data from ***January 1980 to December 2018*** is obtained from the Center for Research in Security Prices (***CRSP***) and ***Compustat*** databases.[4]

The ***80 pricing anomalies*** are the same as in Sun (2023).

Market excess return and risk-free rates are obtained from ***Kenneth French's online data library***.

Micro stocks are defined as stocks from firms that, at the given time, have a market capitalization ***smaller than the 20th percentile of NYSE's stocks***.

[4] Acknowledgment to Chuanping Sun for providing the data.

Anomaly factors

The characteristic-based factor portfolios are constructed as the ***spread return between the top and bottom decile portfolios*** with respect to each firm characteristic.

- In this approach, all factors are calculated in a ***high-minus-low*** manner, which distinguishes them from usual characteristics.[5]

[5] This is simply to ease the calculations and will only impact the signal of the factor's $\hat{\hat{b}}$'s, not their magnitude.

Anomaly factors

The characteristic-based factor portfolios are constructed as the ***spread return between the top and bottom decile portfolios*** with respect to each firm characteristic.

- In this approach, all factors are calculated in a ***high-minus-low*** manner, which distinguishes them from usual characteristics.[5]

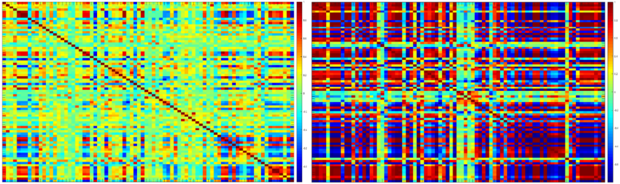
To ensure comparability between factors' coefficients, all ***anomaly factors are demeaned and scaled*** to have the same standard deviation as the market factor.

[5] This is simply to ease the calculations and will only impact the signal of the factor's $\hat{\hat{b}}$'s, not their magnitude.

A tour through the zoo

Sun (2023) examines the correlation between factors, comparing measurements using time series and factor loadings[6] approaches.

- In the time series measurement, 16% of the anomaly factors exhibit correlations greater than 0.5 (absolute values);
- The factor loading approach shows ***significantly higher correlations***, indicating that applying the standard Fama-MacBeth procedure could lead to ***multicollinearity issues***.



(a) Factor correlation measured by time series (b) Factor correlation measured by factor loadings

[6] The coefficients of explanatory variables on the second stage F-MB regression

Test assets

The choice between using individual stocks or sorted portfolios as test assets is a topic of ongoing debate in the literature.

- While ***individual stocks*** may seem appealing, they tend to be ***noisy*** and introduce errors in variables - especially when the regression is performed on estimated variables;
- Consequently, several articles[7] advocate for using sorted portfolios as test assets.

[7] See Fama and French (2008), Hou et al. (2015), and Feng et al. (2020)

Test assets

The choice between using individual stocks or sorted portfolios as test assets is a topic of ongoing debate in the literature.

- While ***individual stocks*** may seem appealing, they tend to be ***noisy*** and introduce errors in variables - especially when the regression is performed on estimated variables;
- Consequently, several articles[7] advocate for using sorted portfolios as test assets.

In line with Feng et al. (2020), we construct ***bi-variate sorted test portfolios*** by crossing size with every other available characteristic, resulting in a total of 1,896 diversified portfolios as our set of test assets.

[7] See Fama and French (2008), Hou et al. (2015), and Feng et al. (2020)

Empirical analysis

We applied the proposed framework to scan the zoo of factors and reduce its dimensionality. We:

- Considered Elastic Net (eNet), LASSO, and Adaptive LASSO (A-LASSO) as candidate ***shrinkage techniques***;
- Conducted tests for four distinct ***periods*** - 60, 120, 180, and 240 months - for both $RW_{\hat{C}}$ and RW_{OLS}, and;
- Explored three possibilities for setting the ***penalization parameter***: through BIC (using all or top n selected factors) and fixing the number of selected regressors.

Shrinking the zoo

When setting the penalization parameter (λ) through BIC and ***considering all selected factors***, we obtained the following Sharpe ratios:

$RW_{\hat{C}}$	60			120			180			240		
RW_{OLS}	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
60	0.60	0.71	0.24	0.51	0.82	0.74	0.72	0.74	0.41	0.74	0.91	0.43
120	0.82	0.88	0.59	0.69	0.86	0.94	0.91	0.94	0.71	0.82	0.90	0.54
180	0.94	1.02	0.54	0.76	1.06	0.90	1.03	1.05	0.87	0.81	0.81	0.66
240	0.89	1.14	0.43	0.69	0.95	0.86	0.95	0.91	0.76	0.82	0.92	0.60

Shrinking the zoo

When setting the penalization parameter (λ) through BIC and ***considering all selected factors***, we obtained the following Sharpe ratios:

$RW_{\hat{C}}$	60			120			180			240		
RW_{OLS}	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
60	0.60	0.71	0.24	0.51	0.82	0.74	0.72	0.74	0.41	0.74	0.91	0.43
120	0.82	0.88	0.59	0.69	0.86	0.94	0.91	0.94	0.71	0.82	0.90	0.54
180	0.94	1.02	0.54	0.76	1.06	0.90	1.03	1.05	0.87	0.81	0.81	0.66
240	0.89	1.14	0.43	0.69	0.95	0.86	0.95	0.91	0.76	0.82	0.92	0.60

On one hand, given the less-labored application, these results are ***not particularly impressive***.

On the other hand, they can ***provide valuable insights*** into the impact of sparsity in factor selection.

Time-varying selection - # of factors

Let's examine the restrictiveness of factor selection in this time-varying manner, taking a look at the ***average number of factors selected***:

$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Average # of factors	15.6	9.4	22.4	17.3	9.2	20.2	17.6	10.7	21.5	16.6	10.9	20.1

Time-varying selection - # of factors

Let's examine the restrictiveness of factor selection in this time-varying manner, taking a look at the ***average number of factors selected***:

$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Average # of factors	15.6	9.4	22.4	17.3	9.2	20.2	17.6	10.7	21.5	16.6	10.9	20.1

The table above highlights that ***performance improves with fewer factors*** considered.

- Therefore, imposing a ***stricter degree of sparsity*** could potentially lead to better predictability.

Guaranteeing sparsity

The results presented so far did not prioritize guaranteeing sparsity:

- ***Unnecessary factors*** may introduce more ***noise*** than desirable.

Guaranteeing sparsity

The results presented so far did not prioritize guaranteeing sparsity:

- ***Unnecessary factors*** may introduce more ***noise*** than desirable.

Traditional models often make underlying sparsity assumptions:

- It should be possible to achieve ***similar sparsity*** levels through shrinkage regressions ***without any prior*** expectations.

Guaranteeing sparsity

The results presented so far did not prioritize guaranteeing sparsity:

- ***Unnecessary factors*** may introduce more ***noise*** than desirable.

Traditional models often make underlying sparsity assumptions:

- It should be possible to achieve ***similar sparsity*** levels through shrinkage regressions ***without any prior*** expectations.

We proceed to examine the out-of-sample results for the two approaches to guaranteeing sparsity:

- ***Forcingly*** considering only the top n selected factors.
- Deliberately setting the penalization parameter of the shrinkage regression to ***ensure*** the selection of a fixed n factors.

Forcing sparsity - Top *n* factors results

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.31	1.25	1.30	1.25	1.23	0.98	1.00	1.12	1.22	0.98	1.15	1.11
Top 5	1.18	1.06	1.20	1.09	1.14	1.22	0.81	0.93	1.01	1.03	1.11	1.06
Top 10	0.59	0.64	0.65	0.63	0.97	0.87	0.86	0.82	0.75	0.83	1.00	0.64
RW_{OLS}	120											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.27	1.25	1.16	1.36	1.37	1.12	1.09	1.24	1.30	0.95	1.05	1.02
Top 5	1.22	1.12	1.23	1.26	1.36	1.32	0.86	1.05	1.06	0.84	0.86	1.08
Top 10	0.93	0.86	0.80	0.79	1.10	1.15	0.94	1.00	0.98	0.87	0.91	0.68
RW_{OLS}	180											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.50	1.38	1.22	1.46	1.56	1.22	1.27	1.39	1.45	1.13	1.20	1.12
Top 5	1.37	1.26	1.13	1.37	1.50	1.46	1.02	1.22	1.22	0.94	0.91	1.23
Top 10	1.16	1.01	0.71	0.83	1.27	1.15	1.09	1.15	1.12	0.92	0.95	0.79
RW_{OLS}	240											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.54	1.28	1.26	1.39	1.46	1.13	1.30	1.42	1.33	1.20	1.29	1.12
Top 5	1.34	1.20	1.24	1.32	1.45	1.17	1.08	1.16	1.04	1.07	0.98	1.26
Top 10	1.08	1.10	0.85	0.77	1.13	1.01	0.98	1.12	1.00	0.88	1.04	0.85

Forcing sparsity - Top *n* factors results

Despite potentially jeopardizing the regression's properties, top *n* methodology sensibly ***improves*** overall ***Sharpe ratios***, revealing that:

- ***Higher*** number of factors (n) ***compromises*** performance;
- ***Shorter*** periods for covariances estimations ($RW_{\hat{C}}$) ***increase*** Sharpe;
- ***Longer*** periods for predicting returns (RW_{OLS}) ***enhance*** results.

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
All	0.60	0.71	0.24	0.51	0.82	0.74	0.72	0.74	0.41	0.74	0.91	0.43
Top 3	1.31	1.25	1.30	1.25	1.23	0.98	1.00	1.12	1.22	0.98	1.15	1.11
Top 5	1.18	1.06	1.20	1.09	1.14	1.22	0.81	0.93	1.01	1.03	1.11	1.06
Top 10	0.59	0.64	0.65	0.63	0.97	0.87	0.86	0.82	0.75	0.83	1.00	0.64
RW_{OLS}	180											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
All	0.94	1.02	0.54	0.76	1.06	0.90	1.03	1.05	0.87	0.81	0.81	0.66
Top 3	1.50	1.38	1.22	1.46	1.56	1.22	1.27	1.39	1.45	1.13	1.20	1.12
Top 5	1.37	1.26	1.13	1.37	1.50	1.46	1.02	1.22	1.22	0.94	0.91	1.23
Top 10	1.16	1.01	0.71	0.83	1.27	1.15	1.09	1.15	1.12	0.92	0.95	0.79

Ensuring sparsity - Fixed *n* factors results

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.29	1.38	1.22	1.17	1.12	1.20	1.30	1.24	1.33	1.34	1.23	0.97
Fixed 5	1.11	1.06	1.11	0.91	1.13	1.10	0.79	0.85	1.00	1.00	1.09	0.97
Fixed 10	0.81	0.62	0.74	1.00	0.86	0.78	0.95	0.85	0.92	0.79	0.54	0.54

RW_{OLS}	120											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.53	1.56	1.47	1.21	1.25	1.38	1.39	1.26	1.38	1.34	1.26	0.89
Fixed 5	1.37	1.35	1.41	0.97	1.35	1.19	0.91	0.98	1.11	0.87	1.04	1.13
Fixed 10	0.88	0.91	1.08	1.08	0.97	1.02	0.87	0.87	1.08	0.90	0.73	0.70

RW_{OLS}	180											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.70	1.91	1.67	1.30	1.31	1.56	1.56	1.41	1.50	1.51	1.45	0.93
Fixed 5	1.57	1.72	1.43	1.04	1.46	1.38	1.05	1.10	1.25	1.01	1.23	1.19
Fixed 10	1.13	1.04	1.01	1.12	1.00	1.05	0.95	1.07	1.19	0.91	0.82	0.91

RW_{OLS}	240											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.74	2.00	1.57	1.48	1.33	1.39	1.52	1.49	1.44	1.61	1.58	0.99
Fixed 5	1.56	1.74	1.55	1.05	1.45	1.15	1.11	1.19	1.19	0.96	1.24	1.13
Fixed 10	1.24	1.29	1.05	0.97	0.91	0.99	0.89	0.98	1.10	0.82	0.83	0.88

Ensuring sparsity - Fixed *n* factors results

Fixing n ***further improves*** the ***OOS Sharpe ratios***, particularly when:

- n is smaller, $RW_{\hat{C}}$ is shorter, and RW_{OLS} is longer.

The results indicate that ***adjusting penalty factors*** accordingly yields even ***better predictability*** than considering only the top n factors.

- *Without compromising regressions' properties.*

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
All	0.60	0.71	0.24	0.51	0.82	0.74	0.72	0.74	0.41	0.74	0.91	0.43
Fixed 3	1.29	1.38	1.22	1.17	1.12	1.20	1.30	1.24	1.33	1.34	1.23	0.97
Fixed 5	1.11	1.06	1.11	0.91	1.13	1.10	0.79	0.85	1.00	1.00	1.09	0.97
Fixed 10	0.81	0.62	0.74	1.00	0.86	0.78	0.95	0.85	0.92	0.79	0.54	0.54
RW_{OLS}	240											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
All	0.89	1.14	0.43	0.69	0.95	0.86	0.95	0.91	0.76	0.82	0.92	0.60
Fixed 3	1.74	2.00	1.57	1.48	1.33	1.39	1.52	1.49	1.44	1.61	1.58	0.99
Fixed 5	1.56	1.74	1.55	1.05	1.45	1.15	1.11	1.19	1.19	0.96	1.24	1.13
Fixed 10	1.24	1.29	1.05	0.97	0.91	0.99	0.89	0.98	1.10	0.82	0.83	0.88

Comparing bananas with bananas

On the one hand, practitioners often consider an annualized ***Sharpe ratio above one***[8] as interesting as a ”rule of thumb” for good results.

- On the other hand, in rigorous scientific research, ***we should not rely on ”rules of thumb”***, and;
- Considered anomalies were found significant, suggesting that ***abnormal returns could be a byproduct*** of test assets' relevance.

[8] Gross of trading and slippage costs.

Comparing bananas with bananas

On the one hand, practitioners often consider an annualized ***Sharpe ratio above one***[8] as interesting as a "rule of thumb" for good results.

- On the other hand, in rigorous scientific research, ***we should not rely on "rules of thumb"***, and;
- Considered anomalies were found significant, suggesting that ***abnormal returns could be a byproduct*** of test assets' relevance.

Therefore, solely observing Sharpe ratios may raise doubts about the results. To ensure a ***fair comparison***, we will benchmark our results against a ***stricter and more appropriate benchmark***.

This will allow us to compare bananas with bananas and draw more meaningful conclusions.

[8] Gross of trading and slippage costs.

Farming fair bananas

To ensure a fair comparison, we faced our results with a benchmark based on the classic 3-factor model:[9]

- We generated out-of-sample results ***considering the classic factors*** as relevant and ***constructing hedge portfolios*** accordingly.

The Sharpe ratios obtained were:

RW_{OLS}	60	120	180	240
3-factor Fama-French	1.38	1.57	1.64	1.49

[9]See Fama and French (1993).

Bananas can be better than bananas

RW_{OLS}	60	120	180	240
3-factors Fama-French	1.38	1.57	1.64	1.49

Cross-examining our results with the benchmark based on the 3-factor Fama-French model, we draw some conclusions:

- Using neither all selected factors nor the top *n* factors outperformed the Fama-French-sorted portfolios.

Bananas can be better than bananas

RW_{OLS}	60	120	180	240
3-factors Fama-French	1.38	1.57	1.64	1.49

Cross-examining our results with the benchmark based on the 3-factor Fama-French model, we draw some conclusions:

- Using neither all selected factors nor the top n factors outperformed the Fama-French-sorted portfolios.
- However, the ***fixed*** n ***factors methodology*** demonstrated the ***potential to beat this strict benchmark***:
  - Notably, ***shorter*** periods for ***factor selection*** and ***longer*** periods for ***predicting test asset returns*** yielded better results.

The golden bananas

Portfolios that were able to beat the benchmark share some common characteristics:

- Fixed n (with $n \in \{3, 5\}$), $RW_{\hat{C}} = 60$, and $RW_{OLS} \in \{180, 240\}$.

Regression	Methodology	$RW_{\hat{C}}$	RW_{OLS}	SR
LASSO	Fixed 3	60	240	2.00
LASSO	Fixed 3	60	180	1.91
eNet	Fixed 3	60	240	1.74
LASSO	Fixed 5	60	240	1.74
LASSO	Fixed 5	60	180	1.72
eNet	Fixed 3	60	180	1.70
A-LASSO	Fixed 3	60	180	1.67

Shorter periods for **covariance** estimations, **longer** periods for **predicting** test asset returns, and **higher sparsity** appear desirable for factor shrinkage success.

Conclusion

The SDF model should be used with caution in extensive panels:

- Potential ***look-ahead bias*** and ***data availability*** issues may jeopardize some studies.

Conclusion

The SDF model should be used with caution in extensive panels:

- Potential ***look-ahead bias*** and ***data availability*** issues may jeopardize some studies.

Factor models are not well-behaved with a high number of factors:

- Researchers need to ***ensure a certain level of sparsity***, which cannot be achieved through standard penalization criteria.

Conclusion

The SDF model should be used with caution in extensive panels:

- Potential ***look-ahead bias*** and ***data availability*** issues may jeopardize some studies.

Factor models are not well-behaved with a high number of factors:

- Researchers need to ***ensure a certain level of sparsity***, which cannot be achieved through standard penalization criteria.

Adjusting penalization parameter(s) yields favorable results:

- It ***outperforms*** the approach of ***considering only high-magnitude*** factors, while still ***maintaining regression's properties***.

Conclusion

Regressing factors' over assets' returns should consider longer periods:

- ***Returns' predictions*** are more accurate when based on ***longer time series***.

Conclusion

Regressing factors' over assets' returns should consider longer periods:

- ***Returns' predictions*** are more accurate when based on ***longer time series***.

Factor screening should be done over shorter periods:

- The ***covariance*** between factors and asset returns should be ***captured over shorter time periods***, closer to the prediction period.

Next steps

Some improvements are still in the production pipeline, such as:

- Incorporation of ***trading costs***;
- Inclusion of ***additional metrics*** for evaluating portfolio performance (asymmetry, kurtosis, drawdowns);
- ***Simulation*** of OOS results where n factors are chosen randomly.

Next steps

Some improvements are still in the production pipeline, such as:

- Incorporation of ***trading costs***;
- Inclusion of ***additional metrics*** for evaluating portfolio performance (asymmetry, kurtosis, drawdowns);
- ***Simulation*** of OOS results where n factors are chosen randomly.

Furthermore, our research agenda could be extended to investigate the phenomenon from different angles, such as:

- Studying how the ***nature of relevant factors changes over time***, allowing a deeper understanding of asset pricing dynamic nature;
- Exploring the ***shorter period for covariance*** (and factors' relevance) estimation, considering daily - or even intra-day - data.

References I

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

A. Belloni and V. Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 2013.

M. J. Brennan and Y. Xia. Assessing asset pricing anomalies. *The Review of Financial Studies*, 14(4):905–942, 2001.

A. Chinco, A. D. Clark-Joseph, and M. Ye. Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492, 2019.

J. H. Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

J. H. Cochrane. Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108, 2011.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

References II

E. F. Fama and K. R. French. Dissecting anomalies. *The journal of finance*, 63(4):1653–1678, 2008.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

G. Feng, S. Giglio, and D. Xiu. Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370, 2020.

M. Figueiredo and R. Nowak. Ordered weighted l1 regularized regression with strongly correlated covariates: Theoretical aspects. In *Artificial Intelligence and Statistics*, pages 930–938. PMLR, 2016.

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

K. Hiraki and C. Sun. A toolkit for exploiting contemporaneous stock correlations. *Journal of Empirical Finance*, 65:99–124, 2022.

References III

J. L. Horowitz. Variable selection and estimation in high-dimensional models. *Canadian Journal of Economics*, 48(2):389–407, 2016.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

K. Hou, C. Xue, and L. Zhang. Replicating anomalies. *The Review of financial studies*, 33(5):2019–2133, 2020.

J. Huang, J. L. Horowitz, and F. Wei. Variable selection in nonparametric additive models. *Annals of Statistics*, 38:2282–2313, 2010.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965. ISSN 00346535, 15309142. URL `http://www.jstor.org/stable/1924119`.

S. C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. *Handbook of the Economics of Finance*, 2:799–906, 2013.

References IV

G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133, 1974.

C. Sun. Factor correlation and the cross section of asset returns: a correlation-robust approach. *Working paper*, 2023.