

Discussion of
“Culling the Factor Zoo”

by Gurdip Bakshi, Timothy Christensen, John Crosby, and Xiaohui Gao

Yuhan Song

HEC Montréal

NFA 2026 Annual Meeting

Selecting the Best Factors from the Zoo

The number of factors keeps growing

- ▶ **1964:** one, the market (Sharpe, 1964, JF).
- ▶ **1993, 2015, 2018:** three, five, then six (Fama and French, 1993, JFE, 2015, JFE, 2018, JFE).
- ▶ **Today:** more than 400 published factors (Harvey, Liu, and Zhu, 2016, RFS; Jensen, Kelly, and Pedersen, 2023, JF).

Why not try every subset?

K	k	Subsets to compare
64	5	7 million
64	7	553 million
500	5	255 billion
500	7	1.5 quadrillion

Yet we still want a **small** model: easy to interpret and useful for evaluating asset managers.

This paper's question

Given K candidate factors, which k of them form the *best* model?

Mixed-Integer Optimization Finds the Best Subset

An SDF $n_t \geq 0$ prices asset i if $E[n_t R_{i,t}] = 1$. The paper solves **one** program:

$$\begin{array}{ll} \min_{\eta, \chi, n} & \frac{1}{T} \sum_t |n_t - f'_t \eta|^q \text{ ----- distance: model SDF vs. admissible SDF} \\ \text{s.t.} & 1 - \bar{c} \leq \frac{1}{T} \sum_t n_t R_{i,t} \leq 1 + \bar{c}, \quad i = 1, \dots, 203 \text{ -- admissible: prices within spread} \\ & 1/\sqrt{b^{\text{glr}}} \leq n_t \leq \sqrt{b^{\text{glr}}} \text{ ----- admissible: regularized rewards-for-risk} \\ & 1/\sqrt{b^{\text{glr}}} \leq f'_t \eta \leq \sqrt{b^{\text{glr}}}, \quad \frac{1}{T} \sum_t f'_t \eta = 1 \text{ ----- the model SDF is bounded too} \\ & \chi_j \in \{0, 1\}, \quad -B\chi_j \leq \eta_j \leq B\chi_j \text{ ----- } \chi_j = 0 \text{ forces } \eta_j = 0: \text{ factor } j \text{ is out} \\ & \sum_j \chi_j \leq k_{\max} \text{ ----- at most } k_{\max} \text{ factors} \end{array}$$

- **Distance:** squared ($q=2$), as in Hansen and Jagannathan (1997, JF), or absolute ($q=1$).
- **Bounds:** $b^{\text{glr}} = 9$ in all empirical results, i.e. $1/3 \leq n_t \leq 3$, a Sharpe ratio cap of 4.
- **Why a single solve suffices:** the switches χ_j and the loadings η are both unknowns. A mixed-integer solver searches subsets and loadings together, ruling out groups of subsets without evaluating them (Bertsimas, King, and Mazumder, 2016, AoS).

What They Find: A Sparse Model Wins Out of Sample

Preferred 7-factor model ($q=2$): SMB, CMA, $Q_{\text{expected growth}}$, MKT*, SMB*, MOM $_{t-6:t-1}$, SIEQ^{Banz}

Model	OOS alpha (%/yr) vs.			OOS
	FF5	BS6	FF5+MOM	Sharpe
7 factors	5.6	5.4	3.3	0.76
6 factors	4.6	4.5	2.5	0.71
5 factors	4.0	3.2	1.5	0.59
All 64 factors (dense)	1.7	1.3	-0.3	0.46

- **Sparse beats dense.** No evidence that the dense model outperforms.
- **A limit on k .** When k_{max} goes to 10–11 factors, many subsets are equally good, so there is no single best subset.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper points to drawbacks of LASSO and Bayesian methods:
 - ▶ LASSO is scale-sensitive and “makes important mistakes” (Feng, Giglio, and Xiu, 2020, JF).
 - ▶ Bayesian approaches do not guarantee a positive SDF and assume Gaussian likelihoods.
- ▶ The empirical comparisons focus on **fixed** benchmark models (FF5, BS6, FF5+MOM) and its own dense model.
- ▶ Open questions:
 - ▶ How large are these drawbacks in practice?
 - ▶ How do these methods perform on the same zoo?
 - ▶ Showing other **selection** methods on the same zoo would make the case even stronger.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper points to drawbacks of LASSO and Bayesian methods:
 - ▶ LASSO is scale-sensitive and “makes important mistakes” (Feng, Giglio, and Xiu, 2020, JF).
 - ▶ Bayesian approaches do not guarantee a positive SDF and assume Gaussian likelihoods.
- ▶ The empirical comparisons focus on **fixed** benchmark models (FF5, BS6, FF5+MOM) and its own dense model.
- ▶ Open questions:
 - ▶ How large are these drawbacks in practice?
 - ▶ How do these methods perform on the same zoo?
 - ▶ Showing other **selection** methods on the same zoo would make the case even stronger.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper points to drawbacks of LASSO and Bayesian methods:
 - ▶ LASSO is scale-sensitive and “makes important mistakes” (Feng, Giglio, and Xiu, 2020, JF).
 - ▶ Bayesian approaches do not guarantee a positive SDF and assume Gaussian likelihoods.
- ▶ The empirical comparisons focus on **fixed** benchmark models (FF5, BS6, FF5+MOM) and its own dense model.
- ▶ Open questions:
 - ▶ How large are these drawbacks in practice?
 - ▶ How do these methods perform on the same zoo?
 - ▶ Showing other **selection** methods on the same zoo would make the case even stronger.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper devotes more than a page of the introduction to the debate on the virtue of complexity (VoC) (Kelly, Malamud, and Zhou, 2024, JF).
- ▶ But
 - ▶ The VoC applies mainly when the number of regressors **exceeds** the number of observations.
 - ▶ This regime is hard to test with the paper's linear model. Kelly, Malamud, and Zhou (2024, JF) also document that a linear model performs poorly out of sample as the number of regressors approaches the number of observations.
 - ▶ The SDF is a **linear combination** of the factors, so by design it does not capture **nonlinear** relationships between the factors and the SDF.
- ▶ So
 - ▶ In my view, the paper and the VoC address somewhat different settings.
 - ▶ The paper's contribution stands on its own, and a much shorter VoC discussion might keep the focus on it.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper devotes more than a page of the introduction to the debate on the virtue of complexity (VoC) (Kelly, Malamud, and Zhou, 2024, JF).
- ▶ But
 - ▶ The VoC applies mainly when the number of regressors **exceeds** the number of observations.
 - ▶ This regime is hard to test with the paper's linear model. Kelly, Malamud, and Zhou (2024, JF) also document that a linear model performs poorly out of sample as the number of regressors approaches the number of observations.
 - ▶ The SDF is a **linear combination** of the factors, so by design it does not capture **nonlinear** relationships between the factors and the SDF.
- ▶ So
 - ▶ In my view, the paper and the VoC address somewhat different settings.
 - ▶ The paper's contribution stands on its own, and a much shorter VoC discussion might keep the focus on it.

Comment 1: Further Evidence on the Horse Race

- ▶ The paper devotes more than a page of the introduction to the debate on the virtue of complexity (VoC) (Kelly, Malamud, and Zhou, 2024, JF).
- ▶ But
 - ▶ The VoC applies mainly when the number of regressors **exceeds** the number of observations.
 - ▶ This regime is hard to test with the paper's linear model. Kelly, Malamud, and Zhou (2024, JF) also document that a linear model performs poorly out of sample as the number of regressors approaches the number of observations.
 - ▶ The SDF is a **linear combination** of the factors, so by design it does not capture **nonlinear** relationships between the factors and the SDF.
- ▶ So
 - ▶ In my view, the paper and the VoC address somewhat different settings.
 - ▶ The paper's contribution stands on its own, and a much shorter VoC discussion might keep the focus on it.

Comment 1: Further Evidence on the Horse Race

- ▶ The dense model is regularized “economically”: the gain-loss bound and the spread band replace a ridge penalty (Kozak, Nagel, and Santosh, 2020, JFE).
- ▶ But these bounds act on the SDF’s **values**, not on the loadings. Bounding $f_t'\eta$ caps $\eta'\Sigma_f\eta$.
 - ▶ Loadings in high-variance directions are limited, but loadings in low-variance directions, typically highly correlated factors, are not.
- ▶ The dense model includes many more highly correlated factors, so it is more likely to produce large loadings and to overfit in sample.
- ▶ A dense model with directly regularized loadings, not only bounded SDF values, would give a tighter comparison.

Comment 1: Further Evidence on the Horse Race

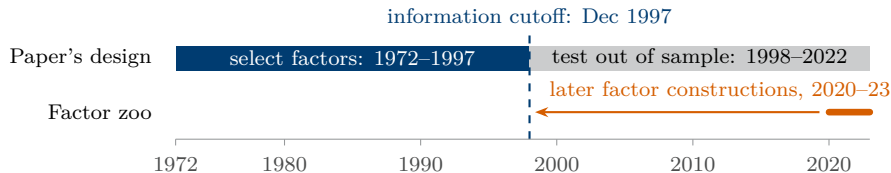
- ▶ The dense model is regularized “economically”: the gain-loss bound and the spread band replace a ridge penalty (Kozak, Nagel, and Santosh, 2020, JFE).
- ▶ But these bounds act on the SDF’s **values**, not on the loadings. Bounding $f_t'\eta$ caps $\eta'\Sigma_f\eta$.
 - ▶ Loadings in high-variance directions are limited, but loadings in low-variance directions, typically highly correlated factors, are not.
- ▶ The dense model includes many more highly correlated factors, so it is more likely to produce large loadings and to overfit in sample.
- ▶ A dense model with directly regularized loadings, not only bounded SDF values, would give a tighter comparison.

Comment 1: Further Evidence on the Horse Race

- ▶ The dense model is regularized “economically”: the gain-loss bound and the spread band replace a ridge penalty (Kozak, Nagel, and Santosh, 2020, JFE).
- ▶ But these bounds act on the SDF’s **values**, not on the loadings. Bounding $f_t'\eta$ caps $\eta'\Sigma_f\eta$.
 - ▶ Loadings in high-variance directions are limited, but loadings in low-variance directions, typically highly correlated factors, are not.
- ▶ The dense model includes many more highly correlated factors, so it is more likely to produce large loadings and to overfit in sample.
- ▶ A dense model with directly regularized loadings, not only bounded SDF values, would give a tighter comparison.

Comment 2: Possible Look-Ahead Bias in the Factor Zoo

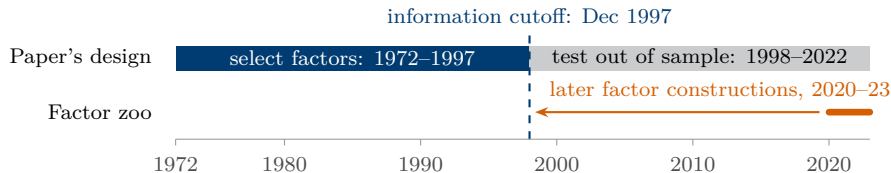
- ▶ The paper selects factors on 1972–1997 and tests on 1998–2022.
- ▶ But $Q_{\text{expected growth}}$ (Hou, Xue, and Zhang, 2020, RFS), MKT* and SMB* (Daniel et al., 2020, RFS) were **proposed after 1997**, and the 47 factors of Jensen, Kelly, and Pedersen (2023, JF) were **compiled in 2023**, using data that cover the test period.



- ▶ The zoo may partly reflect what worked in 1998–2022. Selecting among 64 recent candidates may therefore have an advantage over a **fixed** 5- or 6-factor benchmark.
- ▶ It would be valuable to re-select factors on rolling or expanding windows, and to restrict the zoo to factors published **before** the out-of-sample period.
- ▶ This can be also addressed by comparing with other selection methods on the same zoo, then there would be no information difference across methods.

Comment 2: Possible Look-Ahead Bias in the Factor Zoo

- ▶ The paper selects factors on 1972–1997 and tests on 1998–2022.
- ▶ But $Q_{\text{expected growth}}$ (Hou, Xue, and Zhang, 2020, RFS), MKT* and SMB* (Daniel et al., 2020, RFS) were **proposed after 1997**, and the 47 factors of Jensen, Kelly, and Pedersen (2023, JF) were **compiled in 2023**, using data that cover the test period.



- ▶ The zoo may partly reflect what worked in 1998–2022. Selecting among 64 recent candidates may therefore have an advantage over a **fixed** 5- or 6-factor benchmark.
- ▶ It would be valuable to re-select factors on rolling or expanding windows, and to restrict the zoo to factors published **before** the out-of-sample period.
- ▶ This can be also addressed by comparing with other selection methods on the same zoo, then there would be no information difference across methods.

Comment 3: Interpreting the SDF Loadings

SDF Loadings from Table IV

Model	Panel A: MIQP ($q=2$)		Panel B: MILP ($q=1$)	
	SMB	SIEQ ^{Banz}	SMB	SIEQ ^{Banz}
7-f	-16.4	+16.8	-21.8	+19.6
6-f	-28.8	+24.6	-28.8	+24.6
5-f	-21.0	+18.8	-21.3	+19.0

- ▶ The loadings on SMB and SIEQ^{Banz} are **opposite in sign** and **similar** in magnitude.
- ▶ The paper reports $\rho(\text{SMB}^*, \text{SMB}) \leq 0.8$ and $\rho(\text{SMB}^*, \text{SIEQ}^{\text{Banz}}) \leq 0.75$, but what is the correlation between SMB and SIEQ^{Banz}?
- ▶ I would expect it to be fairly high, say above 0.6, since **both factors capture size**.
- ▶ Large offsetting loadings on two highly correlated factors would benefit from more economic interpretation.

Comment 3: Interpreting the SDF Loadings

SDF Loadings from Table IV

Model	Panel A: MIQP ($q=2$)		Panel B: MILP ($q=1$)	
	SMB	SIEQ ^{Banz}	SMB	SIEQ ^{Banz}
7-f	-16.4	+16.8	-21.8	+19.6
6-f	-28.8	+24.6	-28.8	+24.6
5-f	-21.0	+18.8	-21.3	+19.0

- ▶ The loadings on SMB and SIEQ^{Banz} are **opposite in sign** and **similar** in magnitude.
- ▶ The paper reports $\rho(\text{SMB}^*, \text{SMB}) \leq 0.8$ and $\rho(\text{SMB}^*, \text{SIEQ}^{\text{Banz}}) \leq 0.75$, but what is the correlation between SMB and SIEQ^{Banz}?
- ▶ I would expect it to be fairly high, say above 0.6, since **both factors capture size**.
- ▶ Large offsetting loadings on two highly correlated factors would benefit from more economic interpretation.

Comment 3: Interpreting the SDF Loadings

SDF Loadings from Table IV

Panel A: MIQP ($q=2$)								
Model	constant	SMB	CMA	MKT*	QMJ	MOM $_{t-6:t-1}$	SIEQ ^{Banz}	Sharpe Ratio
6-f	+1.2	-28.8	-12.7	-7.4	-6.3	-4.2	+24.6	0.71
Panel B: MILP ($q=1$)								
Model	constant	SMB	CMA	MKT*	QMJ	MOM $_{t-6:t-1}$	SIEQ ^{Banz}	Sharpe Ratio
6-f	+1.2	-28.8	-12.7	-7.4	-6.3	-4.2	+24.6	0.71

- ▶ The 6-factor loadings are **identical** for $q=1$ and $q=2$.
- ▶ This may reflect robust loadings, but could it be that the bounds $1/3 \leq f'_t \eta \leq 3$ bind?
- ▶ If so, it would be helpful to report how often the bounds bind and how the loadings respond when the bounds change.

Comment 3: Interpreting the SDF Loadings

SDF Loadings from Table IV

Panel A: MIQP ($q=2$)								
Model	constant	SMB	CMA	MKT*	QMJ	MOM $_{t-6:t-1}$	SIEQ ^{Banz}	Sharpe Ratio
6-f	+1.2	-28.8	-12.7	-7.4	-6.3	-4.2	+24.6	0.71
Panel B: MILP ($q=1$)								
Model	constant	SMB	CMA	MKT*	QMJ	MOM $_{t-6:t-1}$	SIEQ ^{Banz}	Sharpe Ratio
6-f	+1.2	-28.8	-12.7	-7.4	-6.3	-4.2	+24.6	0.71

- ▶ The 6-factor loadings are **identical** for $q=1$ and $q=2$.
- ▶ This may reflect robust loadings, but could it be that the bounds $1/3 \leq f'_t \eta \leq 3$ bind?
- ▶ If so, it would be helpful to report how often the bounds bind and how the loadings respond when the bounds change.

Comment 4: Accounting Trading Costs in the OOS Alphas

- ▶ The paper builds the bid-ask spread into the SDF estimation, which makes the SDF more realistic.
- ▶ The OOS alphas against the benchmarks do not yet account for the cost of trading the MVE portfolios.
- ▶ Trading costs likely **differ** across portfolios, so the OOS alphas could change after costs.
- ▶ It would be helpful to charge the spread on the OOS MVE portfolio returns.

Comment 4: Accounting Trading Costs in the OOS Alphas

- ▶ The paper builds the bid-ask spread into the SDF estimation, which makes the SDF more realistic.
- ▶ The OOS alphas against the benchmarks do not yet account for the cost of trading the MVE portfolios.
- ▶ Trading costs likely **differ** across portfolios, so the OOS alphas could change after costs.
- ▶ It would be helpful to charge the spread on the OOS MVE portfolio returns.

Summary

- ▶ **A very nice paper on an important question, with a solid methodology and clear empirical results.**
- ▶ It would benefit from:
 - ▶ Comparing with other **selection** methods on the same zoo.
 - ▶ Restricting the zoo to factors available **before** the test period.
 - ▶ Providing more economic interpretation of the SDF loadings.
 - ▶ Accounting for trading costs in the OOS alphas.
- ▶ Looking forward to the next version!

References I

- Bertsimas, Dimitris, Angela King, and Rahul Mazumder, 2016, Best Subset Selection via a Modern Optimization Lens, *The Annals of Statistics* 44, 813–852.
- Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2020, The Cross-Section of Risk and Returns, *The Review of Financial Studies* 33, 1927–1979.
- Fama, Eugene F., and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 2015, A Five-Factor Asset Pricing Model, *Journal of Financial Economics* 116, 1–22.
- Fama, Eugene F., and Kenneth R. French, 2018, Choosing Factors, *Journal of Financial Economics* 128, 234–252.
- Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the Factor Zoo: A Test of New Factors, *The Journal of Finance* 75, 1327–1370.
- Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing Specification Errors in Stochastic Discount Factor Models, *The Journal of Finance* 52, 557–590.
- Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the Cross-Section of Expected Returns, *The Review of Financial Studies* 29, 5–68.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating Anomalies, *The Review of Financial Studies* 33, 2019–2133.
- Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is There a Replication Crisis in Finance?, *The Journal of Finance* 78, 2465–2518.
- Kelly, Bryan, Semyon Malamud, and Kangying Zhou, 2024, The Virtue of Complexity in Return Prediction, *The Journal of Finance* 79, 459–503.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the Cross-Section, *Journal of Financial Economics* 135, 271–292.
- Sharpe, William F., 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *The Journal of Finance* 19, 425–442.