Schrödinger's Sparsity in the Cross Section of Stock Returns

Doron Avramov[1], Guanhao Feng[2], Jingyu He[2] and **Shuhua Xiao**[2]

May 18, 2025 Peking University

Ninth PKU-NUS Annual International Conference on Quantitative Finance and Economics

[1]Reichman University [2]City University of Hong Kong

Sparsity

A central challenge in modern statistics: **addressing high-dimensional problems**

Sparse modeling

- **selection** for **sparse** models: L_1 penalty
- *Usually, researchers assume that the underlying signal is sparse, and advanced methods are designed to recover such signals effectively.*

Empirical asset pricing:

- Feng et al. (2020) and Bryzgalova et al. (2023):
  Evidence of sparsity in factor risk prices within cross-sectional regressions
- *Assumption: the cross section of returns is driven by a limited number of factors.*

Sparsity

A central challenge in modern statistics: **addressing high-dimensional problems**

Dense modeling

- **Shrinkage**: L_2 penalty

Empirical asset pricing:

- Kozak, Nagel, and Santosh (2020):

  A characteristics-sparse SDF cannot explain the cross section of returns.
- Kozak and Nagel (2023): Factors derived from characteristics through sorting / characteristic weighting / OLS cross-sectional regression slopes do not span the SDF unless a large number of characteristics are used simultaneously.
- Shen and Xiu (2025): When signals are weak, ridge regression outperforms Lasso for prediction. Equivalently, the predictive model might not be sparse.

Sparsity

- Addressing high-dimensional problems is a central challenge in modern statistics.
- Statisticians have developed lots of tools:
  - **Shrinkage**: L_2 penalty.
  - Selection for **sparse** models: L_1 penalty.
- Asset pricing
  - Sparse modeling
  - Dense modeling
- These modeling outcomes are often artifacts of the imposed prior.
- A less frequently explored question arises:

Are asset pricing models inherently sparse?

Motivation: Illusion of Sparsity

Giannone, Lenza, and Primiceri (2021) (GLP) propose a Bayesian sparse model that parametrizes the level of sparsity

- Link L_1 and L_2: **no assumption**, but posterior.

- They examine various types of economic data
  - Macro / Finance / Micro

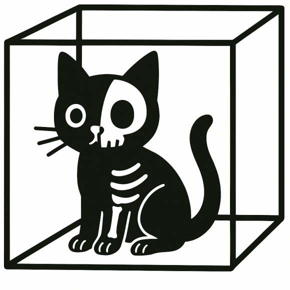
- Findings: the posterior distribution does not typically concentrate on a single sparse model.

$\Rightarrow$ This phenomenon highlights an illusion of sparsity in economic data.

- They did not emphasize factors.

Motivation: Schrödinger's Sparsity

Existing approaches: require researchers to commit *ex ante* to either a sparse (selection) or dense (shrinkage) specification prior to examining the cross section and adhere to that assumption throughout the modeling process.

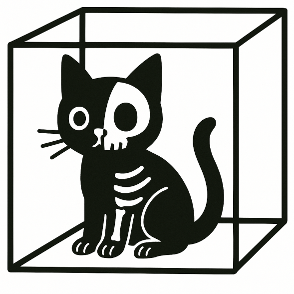


Schrödinger's cat

- We cannot determine whether the cat is alive or dead until we open the box.

Motivation: Schrödinger's Sparsity

Existing approaches: require researchers to commit *ex ante* to either a sparse (selection) or dense (shrinkage) specification prior to examining the cross section and adhere to that assumption throughout the modeling process.



Schrödinger's cat

- We cannot determine whether the cat is alive or dead until we open the box.
- We cannot determine whether the model is **sparse or dense** until we "**open the box**."

Research Questions

We investigate whether asset pricing models are sparse within the conditional latent factor structure of IPCA.

Following the idea of Giannone, Lenza, and Primiceri (2021) in examining sparsity levels, we study the **sparsity of characteristics** in the conditional latent factor model of Kelly, Pruitt, and Su (2019), which introduces observable characteristics as instruments for loadings on latent factors.

Contribution

Methodology Innovations

We propose a novel Bayesian sparse conditional (latent) factor model.

- We permit sparsity levels to be freely estimated or fixed exogenously.
- We separate the sparsity of alphas from that of betas.
- We incorporate observable traded factors alongside latent ones.
  - estimate conditional versions of well-known models
  - recover unspanned components

Empirical Findings

- $\cdots$

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics

- When sparsity is imposed exogenously:

  Highest performance $\sim$ the imposed level aligns with the endogenous level selected by the posterior

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics

- When sparsity is imposed exogenously:

  Highest performance $\sim$ the imposed level aligns with the endogenous level selected by the posterior

- Mispricing is typically sparser than factor loadings.

  Complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa.

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics
- When sparsity is imposed exogenously:

  Highest performance $\sim$ the imposed level aligns with the endogenous level selected by the posterior
- Mispricing is typically sparser than factor loadings.

  Complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa.
- Sparsity varies across test asset sets.

  Fama–French 25 portfolios $\sim$ Sparse models

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics

- When sparsity is imposed exogenously:

  Highest performance $\sim$ the imposed level aligns with the endogenous level selected by the posterior

- Mispricing is typically sparser than factor loadings.

  Complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa.

- Sparsity varies across test asset sets.

  Fama–French 25 portfolios $\sim$ Sparse models

- Sparsity is time-varying. Models become more sparse during recessions.

Key Results

- Best-performing models are neither extremely sparse nor fully dense.

  $\sim$ A substantial yet selective set of characteristics

- When sparsity is imposed exogenously:

  Highest performance $\sim$ the imposed level aligns with the endogenous level selected by the posterior

- Mispricing is typically sparser than factor loadings.

  Complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa.

- Sparsity varies across test asset sets.

  Fama–French 25 portfolios $\sim$ Sparse models

- Sparsity is time-varying. Models become more sparse during recessions.

- Models that combine observable and latent factors outperform those that use either component alone.

Full Model

$$
\begin{aligned}
r_{i,t} &= \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) + \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})\mathbf{f}_t + \epsilon_{i,t} \qquad (1) \\
\text{where} \quad \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) &= \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1} \\
\boldsymbol{\beta}(\mathbf{Z}_{i,t-1}) &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(\mathbf{I}_K \otimes \mathbf{Z}_{i,t-1}), \quad \epsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right)
\end{aligned}
$$

- $r_{i,t}$: return of asset i at time t
- $\mathbf{f}_t$: K latent factors
- $\mathbf{Z}_{i,t-1}$: vector, L firm characteristics for asset i at time $t-1$

We assume independent spike-and-slab priors on the regression coefficient Giannone, Lenza, and Primiceri (2021).

Core Notation: q

Spike-and-slab prior, a Bayesian variable selection prior.

$$P(\beta \neq 0) = q, \quad P(\beta = 0) = 1 - P(\beta \neq 0) = 1 - q.$$

$$\beta = \begin{cases} \mathcal{N}\left(0, \gamma^2\right) \text{ with prob } q & \text{The regressor is chosen. } \sim L_2 \text{ penalty} \\ 0 \text{ with prob } 1 - q & \text{The regressor is not chosen. } \sim L_1 \text{ penalty} \end{cases}$$

Spike-and-slab prior, a Bayesian variable selection prior.

$$P(\beta \neq 0) = q, \quad P(\beta = 0) = 1 - P(\beta \neq 0) = 1 - q.$$

$$\beta = \begin{cases} \mathcal{N}\left(0, \gamma^2\right) \text{ with prob } q & \text{The regressor is chosen. } \sim L_2 \text{ penalty} \\ 0 \text{ with prob } 1 - q & \text{The regressor is not chosen. } \sim L_1 \text{ penalty} \end{cases}$$

- Traditional spike-and-slab prior: q is a specific value.
- Giannone et al., 2021: q has its prior so that we can sample q.
  - These priors probabilistically interpolate between variable selection and shrinkage, allowing the degree of sparsity to be estimated from the data.
- **Prior settings of $q \neq$ precise control of sparsity levels!**

Sparse BayesIPCA Model

$$r_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}] + \epsilon_{i,t}.$$

Independent spike-and-slab priors on the regression coefficient (GLP)

- **Global prior**:

  The same sparsity level of mispricing (alpha) and factor loadings (beta)

$$[\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1] \overset{iid}{\sim} \begin{cases} \mathcal{N}\left(0, \gamma^2\right) & \text{with prob } q \\ 0 & \text{with prob } 1 - q \end{cases}$$

$$q \sim \text{Beta}(a_q, b_q),$$

$$\gamma^2 \sim \mathrm{IG}(A/2, B/2)$$

$$\alpha_0, \boldsymbol{\beta}_0 \overset{iid}{\sim} \mathcal{N}\left(0, \xi^2\right), \quad \xi^2 \sim \mathrm{IG}(C/2, D/2)$$

Sparse BayesIPCA Model

$$r_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}] + \epsilon_{i,t}.$$

Independent spike-and-slab priors on the regression coefficient (GLP)

- **Separate priors**:
  Different sparsity levels of mispricing (alpha) and factor loadings (beta)

$$\boldsymbol{\alpha}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}\left(0, \gamma_\alpha^2\right) & \text{with prob } q_\alpha \\ 0 & \text{with prob } 1 - q_\alpha \end{cases}, \quad \boldsymbol{\beta}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}\left(0, \gamma_\beta^2\right) & \text{with prob } q_\beta \\ 0 & \text{with prob } 1 - q_\beta \end{cases}$$

$q_\alpha \sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha}),$ $\quad$ $q_\beta \sim \text{Beta}(a_{q_\beta}, b_{q_\beta}),$

$\gamma_\alpha^2 \sim \mathrm{IG}(A_\alpha/2, B_\alpha/2),$ $\quad$ $\gamma_\beta^2 \sim \mathrm{IG}(A_\beta/2, B_\beta/2),$

$$r_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}] + \epsilon_{i,t}.$$

We design joint priors to directly control the sparsity level (i.e., control the number of selected characteristics).

M restricts the number of characteristics driving alpha (beta).

- **(Global) joint prior**:

$$(\tau_1, \tau_2, \cdots, \tau_L) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbf{I}\left(\sum_{i=1}^{L} \tau_i = M\right)$$

- **(Separate) joint priors**:

$$(\tau_1^{\alpha}, \tau_2^{\alpha}, \cdots, \tau_L^{\alpha}) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbf{I}\left(\sum_{i=1}^{L} \tau_i^{\alpha} = M_{\alpha}\right)$$

$$(\tau_1^{\beta_k}, \tau_2^{\beta_k}, \cdots, \tau_L^{\beta_k}) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbf{I}\left(\sum_{i=1}^{L} \tau_i^{\beta} = M_{\beta}\right)$$

Empirical Findings

(i) Sparsity for P-Tree 100 Test Assets

(ii) Large Sets of Test Assets

- Heterogeneous Roles of Characteristics

(iii) Time-varying Sparsity

- Dynamic Roles of Characteristics

(iv) Resurrecting Conditional Observable Factors Model

Dataset

Main test assets:

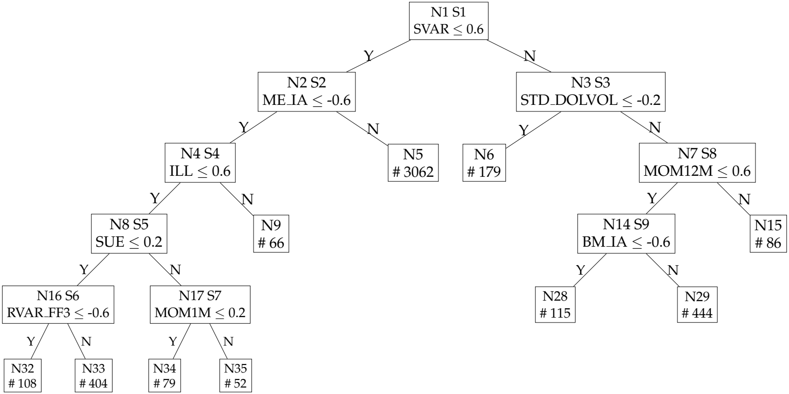
- P-Tree (Cong et al., 2025, JFE) test assets, from Jan-1990 to Dec-2024, monthly.
  - Constructed based on monthly observations of U.S. stocks from 1980 to 2024.
  - 20 $\mathbf{Z}_{i,t}$ firm characteristics.

Other test assets:

- 25 ME/BM portfolios (FF25), 61 long-short portfolios for each characteristic (LS61), 357 bivariate-sorted portfolios (Bi357).
- 500 stocks with the highest and 500 stocks with the lowest average market equity (Big ind500 / Small ind500).

(i) Sparsity for P-Tree 100 Test Assets

Figure 1: Panel Tree from 1980 to 1989



(i) Sparsity for P-Tree 100 Test Assets

Table 1: Model Performance under Global Sparse Priors

		CSR^2			TP.Sp		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.							
	0.1	29.37	**43.66**	$\boxed{\textbf{55.57}}$	0.35	1.36	0.92
q prior mean	0.5	29.54	43.63	54.79	0.35	1.44	0.92
	0.9	**29.71**	43.62	53.89	**0.35**	**1.50**	**0.95**
Panel B: Fixed # selected chars.							
	2	25.44	**52.49**	**51.02**	**0.44**	**1.11**	0.48
M	10	29.53	38.32	41.51	0.35	0.87	**1.12**
	18	27.48	39.31	42.02	0.33	0.55	0.95
Panel C: No sparsity							
M	20	**29.92**	36.88	45.23	0.35	0.57	0.95

Benchmark: CAPM.

q prior mean is 0.1. $K = 5 \sim M_\alpha = 1, M_\beta = 9$.

(i) Sparsity for P-Tree 100 Test Assets

Table 2: Model Performance under Separate Sparse Priors on Alphas and Betas

		CSR^2			TP. Sp		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.							
(q_α prior mean, q_β prior mean)	0.1,0.1	29.17	44.09	**59.20**	0.34	0.75	0.71
	0.5,0.1	29.37	43.27	58.47	0.35	0.77	0.79
	0.9,0.1	29.41	43.54	58.00	0.35	1.14	0.68
	0.1,0.5	29.29	43.53	57.82	0.34	0.75	1.00
	0.5,0.5	29.48	42.49	56.84	0.35	1.01	1.14
	0.9,0.5	29.53	43.65	54.94	0.35	1.17	0.92
	0.1,0.9	29.48	**45.11**	58.72	0.34	0.99	0.77
	0.5,0.9	29.64	42.48	56.84	0.35	1.00	**1.14**
	0.9,0.9	**29.73**	44.13	56.69	**0.35**	**1.27**	0.90
Panel B: Fixed # selected chars.							
(M_α, M_β)	2,2	25.44	49.34	48.39	**0.44**	**1.10**	0.95
	10,2	27.98	51.07	50.10	0.37	0.57	0.87
	18,2	25.17	47.01	38.00	0.32	0.79	0.68
	2,10	28.85	51.17	56.83	0.42	0.60	0.87
	10,10	29.59	37.87	41.20	0.35	0.89	0.97
	18,10	27.19	40.97	39.03	0.32	0.47	0.88
	2,18	29.81	**54.91**	**56.99**	0.43	0.65	1.13
	10,18	**29.88**	34.24	51.26	0.36	1.01	**1.22**
	18,18	27.46	39.30	42.11	0.33	0.53	0.94

Benchmark: CAPM.

(i) Sparsity for P-Tree 100 Test Assets

- **Unrestricted # selected chars**:
  - Global prior:

    q prior mean is 0.1. $K = 5 \sim M_\alpha = 1, M_\beta = 9$.
  - Separate priors:

    Both prior means of q_α and q_β are 0.1. $K = 5 \sim M_\alpha = 1, M_\beta = 10$.
- **Fix # selected chars**:
  - Global prior: $K = 5 \sim M_\alpha = 2, M_\beta = 18$
  - Separate priors: $K = 5 \sim M_\alpha = 2, M_\beta = 18$.

(i) Sparsity for P-Tree 100 Test Assets

- **Unrestricted # selected chars**:
  - Global prior:

    q prior mean is 0.1. $K = 5 \sim M_\alpha = 1, M_\beta = 9$.
  - Separate priors:

    Both prior means of q_α and q_β are 0.1. $K = 5 \sim M_\alpha = 1, M_\beta = 10$.
- **Fix # selected chars**:
  - Global prior: $K = 5 \sim M_\alpha = 2, M_\beta = 18$
  - Separate priors: $K = 5 \sim M_\alpha = 2, M_\beta = 18$.

- Best-performing models are neither extremely sparse nor fully dense.
- # chars driving factor loading (beta) exceeds that of those driving mispricing (alpha).
- When sparsity is imposed exogenously, model performance is highest when the imposed level aligns with the endogenous level selected by the posterior.

(ii) Large Sets of Test Assets

Table 3: Sparsity for Different Test Assets

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

- Sparsity levels vary across different types of test assets.

  E.g., FF25 sparser.

(ii) Large Sets of Test Assets

Table 3: Sparsity for Different Test Assets

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

- Panel A: Within the same category of test assets, a larger number of assets generally requires more characteristics.

(ii) Large Sets of Test Assets

Table 3: Sparsity for Different Test Assets

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

- Panel B: Among test assets of the same type and size, those that are harder to explain tend to require more characteristics to capture mispricing.

(ii) Large Sets of Test Assets

Table 3: Sparsity for Different Test Assets

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

- Panel B: Complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa.

Table 3: Sparsity for Different Test Assets

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

- Panel C: There is substantial variation in the sparsity levels across commonly used test assets.

(iii) Time-varying Sparsity

Table 4: Time Variation Analysis: Sparsity in Regimes

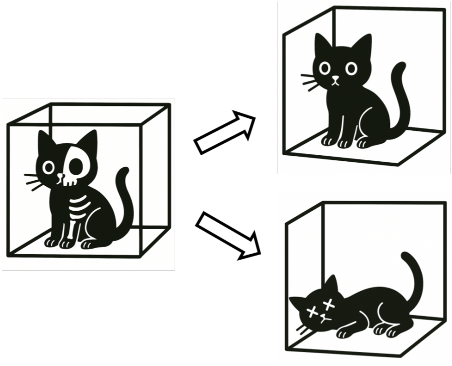
	Different periods					
	Regime1	Regime2	Regime3	Normal	Recession	Full
Panel A: Global prior						
q	0.37	0.41	0.42	0.47	0.42	0.48
Panel B: Separate priors						
q_α	0.30	0.29	0.23	0.27	0.24	0.31
q_β	0.42	0.46	0.56	0.54	0.53	0.59

- Settings of time periods:
  - Follow breakpoints in Smith and Timmermann (2021) to split time periods. (July 1998 and June 2010)
  - Define recession periods based on the Sahm Rule, totaling 88 months.
- Asset pricing models tend to be sparser during recessions.

Schrödinger's Sparsity

Sparsity levels vary across both **cross-sectional** and **time-series** dimensions.

$\Leftarrow$ i) Type and number of test assets; ii) Time periods / Macro conditions



Assuming the asset pricing model to be either sparse or dense a priori may be inappropriate.

Empirical Findings

(i) Sparsity for P-Tree 100 Test Assets

(ii) Large Sets of Test Assets

- Heterogeneous Roles of Characteristics

(iii) Time-varying Sparsity

- Dynamic Roles of Characteristics

(iv) **Resurrecting Conditional Observable Factors Model**

Model with Observable and Latent Factors

- In the conditional observable factor model, alpha and beta can be (sparse) functions of high-dimensional characteristics.
- Augmenting latent factors helps recover unspanned components in observable factor models.

Model with Observable and Latent Factors

- In the conditional observable factor model, alpha and beta can be (sparse) functions of high-dimensional characteristics.
- Augmenting latent factors helps recover unspanned components in observable factor models.

$$
\begin{aligned}
r_{i,t} &= \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) + \boldsymbol{\beta}(\mathbf{Z}_{i,t-1}) \underbrace{\left[\tilde{\mathbf{f}}_t, \mathbf{f}_t\right]}_{\mathbf{F}_t} + \epsilon_{i,t} \\
&= \underbrace{\alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1}}_{\text{mispricing}} + \underbrace{\boldsymbol{\beta}_0 \tilde{\mathbf{f}}_t + \boldsymbol{\beta}_1 [\tilde{\mathbf{f}}_t \otimes \mathbf{Z}_{i,t-1}]}_{\text{obs. factors, conditional beta}} + \underbrace{\boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}]}_{\text{latent factors, dynamic loadings}} + \epsilon_{i,t}.
\end{aligned}
$$

(iv) Resurrecting Conditional Observable Factors Model

Table 5: Augmented Observable Factor Models

	CSR^2	TP.Sp	(q_α, q_β)	$\beta_{0,\text{MKT}}$	$\boldsymbol{\alpha}$ RMSE
Panel A: only obs					
MKT	14.93	0.57	0.45,0.63	1.15	0.0032
FF5	50.38	1.13	0.26,0.61	1.07	0.0014
Panel B: only latent					
LF1	29.48	0.35	0.49,0.53	/	0.0036
LF5	56.81	1.13	0.23,0.34	/	0.0011
Panel C: obs + latent					
MKT+LF1	53.87	0.87	0.31,0.65	1.14	0.0015
MKT+LF5	56.45	1.39	0.24,0.46	0.98	0.0007
FF5+LF1	50.55	1.23	0.33,0.65	1.06	0.0012
FF5+LF5	60.33	1.53	0.18,0.42	0.95	0.0001
Panel D: uncond. model					
MKT	/	0.57	/	1.19	0.0060
FF5	49.25	1.13	/	1.09	0.0042

Benchmark: CAPM.

- Panel A v.s. Panel C: Jointly considering both observable and latent factors helps mitigate model misspecification.
  - $\beta_{0,\text{MKT}}$: be closed to 1 after introducing latent factors.
  - $\boldsymbol{\alpha}$ RMSE: decreases after introducing latent factors.

(iv) Resurrecting Conditional Observable Factors Model

Table 5: Augmented Observable Factor Models

	CSR^2	TP.Sp	(q_α, q_β)	$\beta_{0,\mathrm{MKT}}$	$\boldsymbol{\alpha}$ RMSE
Panel A: only obs					
MKT	14.93	0.57	0.45,0.63	1.15	0.0032
FF5	50.38	1.13	0.26,0.61	1.07	0.0014
Panel B: only latent					
LF1	29.48	0.35	0.49,0.53	/	0.0036
LF5	56.81	1.13	0.23,0.34	/	0.0011
Panel C: obs + latent					
MKT+LF1	53.87	0.87	0.31,0.65	1.14	0.0015
MKT+LF5	56.45	1.39	0.24,0.46	0.98	0.0007
FF5+LF1	50.55	1.23	0.33,0.65	1.06	0.0012
FF5+LF5	60.33	1.53	0.18,0.42	0.95	0.0001
Panel D: uncond. model					
MKT	/	0.57	/	1.19	0.0060
FF5	49.25	1.13	/	1.09	0.0042

Benchmark: CAPM.

- Panel A v.s. Panel D: The conditional factor model outperforms the unconditional model in cross-sectional explanatory power.

(iv) Resurrecting Conditional Observable Factors Model

Figure 2: Characteristics Importance in Alphas and Betas across Different Models

	α_1	$\beta_{1,MKT}$
α_0 , $\beta_{0,MKT}$	−0.0014	1.1507
ABR	0.0000	−0.0411
ACC	−0.0002	0.0020
ADM	0.0000	−0.0090
AGR	−0.0000	0.0383
BASPREAD	−0.0020	−0.0002
BETA	−0.0016	0.1464
BM	0.0000	0.0359
CFP	0.0006	−0.0405
EP	0.0011	−0.0595
ME	0.0000	−0.0001
MOM1M	0.0016	−0.0159
MOM12M	0.0017	−0.0394
NI	0.0000	−0.0576
OP	0.0000	0.0007
RDM	0.0013	0.0005
ROE	0.0000	0.0003
SEAS1A	0.0000	0.0393
SP	0.0019	−0.0307
SUE	0.0018	−0.0274
SVAR	−0.0001	0.2309

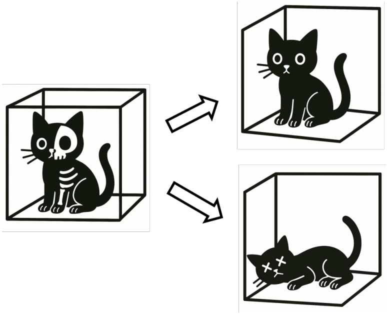
(a) MKT

	α_1	β_{1,LF_1}	$\beta_{1,MKT}$
α_0 , β_{0,LF_1} , $\beta_{0,MKT}$	0.0000	−0.5604	1.1409
ABR	0.0000	−0.0022	−0.0156
ACC	−0.0000	−0.0546	0.0015
ADM	0.0003	0.0017	−0.0210
AGR	0.0000	0.0115	0.0384
BASPREAD	−0.0001	−0.1224	−0.0032
BETA	−0.0010	−0.0945	0.1278
BM	0.0001	0.1802	−0.0016
CFP	−0.0000	−0.0111	−0.0303
EP	0.0001	0.1403	−0.0448
ME	−0.0018	0.3740	−0.0136
MOM1M	0.0000	−0.0029	−0.0357
MOM12M	0.0014	0.0365	−0.0192
NI	0.0000	0.0429	−0.0305
OP	0.0001	0.2001	−0.0001
RDM	0.0000	0.0003	−0.0001
ROE	0.0000	−0.0003	0.0008
SEAS1A	−0.0000	0.0410	0.0255
SP	0.0001	0.2445	−0.0006
SUE	0.0014	0.1628	−0.0330
SVAR	−0.0003	−0.5751	0.2263

(b) MKT + LF1

Summary

- An important research problem: **Are the asset pricing models sparse**?
  - Schrödinger's Sparsity
- A new approach, the BayesIPCA Model, combines the **Bayesian framework of factor estimation** and the **characteristics-based model (IPCA)**.
  - An important extension for considering the **spike-and-slab prior** while estimating the conditional (latent) factor model.
- By avoiding pre-specified assumptions on sparsity or density, our approach **endogenously** determines whether the model is sparse or dense.



Summary

- An important research problem: **Are the asset pricing models sparse**?
  - Schrödinger's Sparsity
- A new approach, the BayesIPCA Model, combines the **Bayesian framework of factor estimation** and the **characteristics-based model (IPCA)**.
  - An important extension for considering the **spike-and-slab prior** while estimating the conditional (latent) factor model.
- By avoiding pre-specified assumptions on sparsity or density, our approach **endogenously** determines whether the model is sparse or dense.
- Based on our method, we can:
  - Identify the global / separate sparsity levels of the asset-pricing model
  - Investigate the characteristics that drive mispricing and factor loadings, and assess their relative importance
  - Resurrect the conditional observable factors model

Thank you!

Technical details

Evaluation Measures

$$\mathrm{CSR}^2 = 1 - \frac{\sum_{i=1}^{N}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t} - \widehat{r}_{i,t})\right)^2}{\sum_{i=1}^{N}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t} - \beta_i \mathrm{MktRF}_t)\right)^2},$$

where $\widehat{r}_{i,t} = \widehat{\boldsymbol{\beta}}(\mathbf{z}_{i,t-1})\mathbf{F}_t$.

Why cross-sectional R^2?

- Sharpe ratio of the factor-efficient portfolio (Investment)
- Cross-sectional R^2 (Asset pricing)

IPCA factors generated by portfolios have much lower Sharpe ratios than their individual stock counterparts.

CS R^2 is difficult to calculate for the unbalanced individual stock return panel.

$\Longrightarrow$ BK proposes using Total R^2, which is directly related to the objectives of IPCA but does not measure traditional pricing errors.

Review: Bayesian APT (Arbitrage Pricing Theory) Factor Model

Geweke and Zhou (1996)

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\epsilon}_t$$

- $\mathbf{r}_t = (r_{1,t}, \cdots, r_{N,t})$: a vector of returns of N asset at time t
- $\boldsymbol{\alpha} = \mathbb{E}[\mathbf{r}_t]$, the expected return on asset.
- "pervasive" factor assumptions:

  $\mathbb{E}[\mathbf{f}_t] = 0$, $\mathbb{E}[\mathbf{f}_t\mathbf{f}_t'] = \mathbf{I}$, $\mathbb{E}(\boldsymbol{\epsilon}_t \mid \mathbf{f}_t) = 0$, $\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' \mid \mathbf{f}_t] = \boldsymbol{\Sigma}$.
- Gibb sampler, draw $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.
- $\mathbf{f}_t$ and $\mathbf{r}_t$ are jointly normally distributed.

  Draw $\mathbf{f}$ conditional on $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and the data:

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{r}_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \boldsymbol{\beta}' \\ \boldsymbol{\beta} & \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma} \end{pmatrix} \right].$$

$$\mathbb{E}(\mathbf{f}_t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) = \boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1}(\mathbf{r}_t - \boldsymbol{\alpha}),$$

$$\mathrm{Cov}(\mathbf{f}_t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) = \mathbf{I} - \boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1}\boldsymbol{\beta}.$$

Review: IPCA

Kelly, Pruitt, and Su (2019)

$$r_{i,t} = \mathbf{z}'_{i,t-1}\Gamma_\alpha + \mathbf{z}'_{i,t-1}\Gamma_\beta \mathbf{f}_t + \epsilon_{i,t}$$

$$r_{i,t} = \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) + \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})\mathbf{f}_t + \epsilon_{i,t}$$

$$\text{where} \quad \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) = \mathbf{Z}'_{i,t-1}\Gamma_\alpha = \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1}$$

$$\boldsymbol{\beta}(\mathbf{Z}_{i,t-1}) = \mathbf{Z}'_{i,t-1}\Gamma_\beta = \boldsymbol{\beta}_1(\mathbf{I}_K \otimes \mathbf{Z}_{i,t-1})$$

- Estimate of $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$ and $\mathbf{f}_t$ by optimization:

$$\min_{\Gamma_\beta, \Gamma_\alpha, \mathbf{f}} \sum_{t=1}^{T} \left(\mathbf{r}_t - \mathbf{Z}_{t-1}\Gamma_\beta \mathbf{f}_t - \mathbf{Z}_{t-1}\Gamma_\alpha\right)' \left(\mathbf{r}_t - \mathbf{Z}_{t-1}\Gamma_\beta \mathbf{f}_t - \mathbf{Z}_{t-1}\Gamma_\alpha\right).$$

- Method: Alternating Least Square (ALS)
- Some conclusions:
  - Dynamic betas (parameterized functions of observable characteristics)
  - Accept $\boldsymbol{\alpha}_1 = \mathbf{0}$ ($\Gamma_\alpha = \mathbf{0}$).

Alpha Tests in Different Models

		# α_0 and $\alpha_{1,i} \neq 0$			p-value		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.							
q prior mean	0.1	10	5	1	0	0	0
	0.5	10	5	1	0	0	0
	0.9	10	5	1	0	0	0
Panel B: Fixed # selected chars.							
M	2	4	2	2	0	0	0
	10	14	4	3	0	0	0
	18	14	12	9	0	0	0
	20	21	18	16	0	0	0

Number of Selected Characteristics in Different Models

		M_α			M_β		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Global prior							
q prior mean	0.1	10	5	1	10	11	9
	0.5	10	5	2	10	11	9
	0.9	10	5	1	11	11	9
Panel B: Separate priors							
(q_α prior mean, q_β prior mean)	0.1,0.1	10	5	1	10	11	10
	0.5,0.1	10	5	1	10	11	10
	0.9,0.1	10	5	1	10	11	10
	0.1,0.5	10	4	1	10	12	10
	0.5,0.5	10	4	2	10	12	14
	0.9,0.5	10	5	2	10	11	10
	0.1,0.9	10	5	1	11	11	11
	0.5,0.9	10	4	2	11	12	14
	0.9,0.9	10	5	2	11	11	14

(ii) Time-varying Sparsity: Dynamic Roles of Characteristics

Figure 3: Changing Roles of Characteristics in Regimes

	Regime1	Regime2	Regime3	Normal	Recession
α_0	0.0001	0.0001	0.0000	0.0000	0.0000
ABR	0.0000	0.0000	0.0000	0.0000	0.0000
ACC	−0.0000	−0.0001	0.0000	−0.0000	−0.0003
ADM	0.0001	0.0000	0.0000	0.0000	0.0000
AGR	0.0000	0.0001	0.0000	−0.0001	0.0000
BASPREAD	−0.0011	−0.0004	−0.0019	−0.0019	0.0000
BETA	−0.0001	0.0001	−0.0000	−0.0001	−0.0001
BM	0.0003	0.0000	0.0000	0.0002	−0.0002
CFP	0.0014	0.0035	−0.0000	0.0000	0.0000
EP	0.0000	0.0001	−0.0000	0.0000	−0.0000
ME	−0.0001	−0.0028	0.0000	−0.0002	−0.0003
MOM1M	−0.0000	0.0003	0.0000	−0.0001	0.0000
MOM12M	0.0020	0.0008	0.0000	0.0009	0.0001
NI	−0.0002	0.0000	0.0000	0.0000	0.0000
OP	0.0000	0.0000	−0.0000	0.0000	0.0025
RDM	0.0001	−0.0000	0.0000	0.0003	0.0000
ROE	0.0000	0.0000	0.0000	0.0000	0.0000
SEAS1A	0.0010	0.0000	0.0000	0.0000	−0.0000
SP	0.0004	−0.0001	0.0000	0.0000	0.0000
SUE	0.0022	0.0029	0.0019	0.0018	0.0030
SVAR	0.0004	−0.0006	0.0000	0.0001	−0.0000

(a) α_1

	Regime1	Regime2	Regime3	Normal	Recession
β_{0,LF_1}	0.9389	0.9540	0.9456	−0.9139	−0.4336
ABR	0.0000	−0.0010	−0.0001	0.0000	0.0006
ACC	0.0013	0.0000	0.0367	0.0000	−0.0005
ADM	0.0132	0.0000	−0.0083	0.0000	0.0031
AGR	0.0000	0.0106	−0.0050	−0.0118	−0.0092
BASPREAD	0.0000	−0.0001	0.0000	0.0000	0.0043
BETA	0.2323	0.1944	0.2348	−0.2881	−0.0748
BM	−0.0478	0.0011	0.0751	0.0670	0.0950
CFP	−0.0591	−0.0258	−0.0046	0.0623	−0.0098
EP	0.0033	−0.0002	0.0000	0.0001	−0.0033
ME	−0.1544	−0.0610	−0.0984	0.1276	−0.1319
MOM1M	−0.0830	−0.0887	−0.0614	0.0389	0.0027
MOM12M	0.0001	−0.0457	−0.0714	0.0381	−0.0534
NI	0.0000	−0.0000	0.0034	−0.0001	0.0000
OP	0.0000	0.0000	0.0107	0.0265	−0.0010
RDM	0.0543	0.0536	−0.0070	−0.0395	−0.0020
ROE	0.0000	0.0000	0.0000	0.0000	0.0028
SEAS1A	0.0000	0.0000	0.0000	0.0000	0.0072
SP	0.0000	0.0107	0.0000	0.0326	0.0041
SUE	−0.0061	0.0000	0.0000	0.0000	0.0065
SVAR	0.1452	0.1791	0.1466	−0.2245	0.0786

(b) β_{1,LF_1}

(iii) Large Sets of Test Assets: Heterogeneous Roles of Characteristics

Figure 4: Heterogeneous Characteristics in Test Assets (mispricing)

	PT100	PT200	PT400	Big500	Small500	FF25	LS61	Bi357
α_0	0.0001	−0.0000	0.0000	−0.0001	−0.0017	0.0002	0.0014	0.0001
ABR	0.0001	0.0012	0.0013	0.0003	0.0056	0.0000	0.0013	0.0009
ACC	−0.0004	−0.0010	−0.0010	−0.0033	−0.0054	0.0000	−0.0004	−0.0010
ADM	0.0001	0.0000	0.0000	−0.0000	−0.0001	0.0000	0.0000	−0.0004
AGR	−0.0000	−0.0001	−0.0001	−0.0002	0.0000	−0.0000	−0.0001	0.0000
BASPREAD	−0.0022	−0.0021	−0.0018	−0.0000	0.0011	−0.0000	0.0004	−0.0002
BETA	−0.0001	−0.0000	−0.0001	−0.0012	0.0001	−0.0001	0.0004	−0.0001
BM	0.0001	0.0001	0.0002	0.0003	−0.0005	−0.0000	0.0001	0.0001
CFP	0.0001	0.0004	0.0008	0.0001	0.0005	−0.0000	−0.0001	−0.0001
EP	0.0000	0.0001	0.0004	0.0002	0.0146	−0.0000	−0.0002	0.0007
ME	−0.0008	−0.0004	−0.0010	−0.0054	−0.0128	−0.0000	−0.0003	0.0001
MOM1M	−0.0000	−0.0001	−0.0001	−0.0032	−0.0114	0.0000	−0.0001	−0.0001
MOM12M	0.0015	0.0013	0.0010	0.0010	0.0070	0.0000	0.0010	0.0004
NI	−0.0000	−0.0003	−0.0010	0.0000	−0.0086	−0.0023	0.0000	−0.0010
OP	0.0000	0.0000	−0.0000	0.0001	−0.0014	0.0000	−0.0000	0.0006
RDM	−0.0000	0.0002	0.0004	0.0019	0.0042	0.0000	0.0003	0.0007
ROE	0.0000	0.0001	0.0000	0.0005	0.0014	0.0000	0.0000	0.0000
SEAS1A	0.0000	0.0000	−0.0000	0.0002	0.0012	0.0000	0.0002	0.0000
SP	0.0001	0.0004	0.0002	−0.0000	−0.0008	−0.0000	−0.0000	0.0001
SUE	0.0020	0.0013	0.0012	0.0018	0.0278	0.0000	0.0012	0.0012
SVAR	−0.0001	0.0003	0.0002	−0.0004	−0.0001	−0.0002	0.0005	0.0001

Color scale: 0, 0.2, 0.4, 0.6, 0.8, 1

(iii) Large Sets of Test Assets: Heterogeneous Roles of Characteristics

Figure 5: Heterogeneous Characteristics in Test Assets (factor loading)

	PT100	PT200	PT400	Big500	Small500	FF25	LS61	Bi357
β_{0,LF_1}	−0.9678	−0.9493	0.9794	0.0563	0.7208	0.9388	0.2298	0.9620
ABR	0.0397	0.0213	−0.0223	−0.0024	0.0001	−0.0339	0.0016	−0.0131
ACC	0.0000	−0.0080	0.0017	0.0028	0.0000	−0.0137	−0.0125	0.0136
ADM	0.0000	0.0000	−0.0053	−0.0019	−0.0001	0.0000	−0.0025	0.0231
AGR	−0.0112	−0.0111	0.0233	0.0007	0.0000	0.0000	−0.0043	0.0026
BASPREAD	0.0000	0.0310	−0.0279	0.0000	−0.0257	0.0498	−0.2295	0.1048
BETA	−0.2197	−0.2061	0.1784	0.0159	0.0704	0.2418	−0.2342	0.1416
BM	−0.0154	0.0088	−0.0024	0.0008	−0.0624	0.0458	0.0788	0.0024
CFP	0.0176	0.0435	0.0001	−0.0035	−0.0034	0.0000	0.0680	−0.0262
EP	0.0000	0.0000	0.0252	−0.0005	−0.0507	0.0000	0.0296	0.0205
ME	0.0400	0.0734	0.0354	−0.0046	−0.0668	0.0000	0.0664	−0.0225
MOM1M	0.0520	0.0547	−0.0120	−0.0075	−0.1227	−0.0167	0.0631	−0.0503
MOM12M	0.0312	0.0432	0.0120	−0.0029	−0.1273	0.0349	−0.0034	−0.0178
NI	0.0000	0.0000	0.0006	−0.0027	0.1076	0.0660	−0.0047	0.0326
OP	−0.0177	−0.0049	0.0208	0.0023	−0.0400	0.0000	0.0332	0.0051
RDM	−0.0140	−0.0294	0.0117	0.0016	0.1062	−0.0017	−0.0993	0.0498
ROE	0.0000	0.0000	0.0002	−0.0022	0.0126	0.0000	0.0110	−0.0003
SEAS1A	0.0000	0.0000	0.0175	0.0014	0.0119	0.0000	−0.0133	0.0192
SP	−0.0279	−0.0083	0.0383	0.0054	0.1034	0.0564	0.0384	0.0327
SUE	0.0000	0.0008	−0.0054	−0.0000	0.0482	0.0000	0.0455	−0.0180
SVAR	−0.0781	−0.2028	0.0010	0.0118	0.2578	0.1086	−0.2839	0.1737

Color scale: 0, 0.2, 0.4, 0.6, 0.8, 1

References

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance 78*(1), 487–557.

Cong, L., G. Feng, J. He, and X. He (2025). Growing the efficient frontier on panel trees. *Journal of Financial Economics 167*, 104024.

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

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies 9*(2), 557–587.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S. and S. Nagel (2023). When do cross-sectional asset pricing factors span the stochastic discount factor? Technical report, National Bureau of Economic Research.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Shen, Z. and D. Xiu (2025). Can machines learn weak signals? Technical report, University of Chicago.

Smith, S. C. and A. Timmermann (2021). Break risk. *Review of Financial Studies 34*(4), 2045–2100.