Schrödinger's Sparsity in the Cross Section of Stock Returns

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

Dec 15, 2025

Global AI Finance Research Conference

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

High-dimensional AP has two *different* modeling choices and assumptions.

- **Sparse modeling: L_1 penalty, Lasso regression**
  - Feng, Giglio, and Xiu (JF 2020), Freybergr, Neuhierl, and Weber (RFS 2020), and Bybee, Kelly, and Su (RFS 2023)
- **Dense modeling: L_2 penalty, Ridge regression**
  - Kozak, Nagel, and Santosh (JFE 2020) and Kozak and Nagel (WP 2023) — SDF requires a large number of characteristics.

Empirical findings frequently mirror prior assumptions instead of revealing the true nature of data.

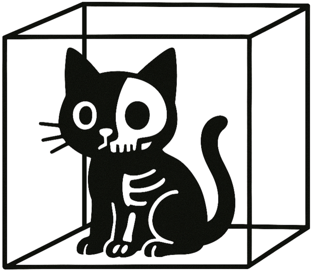
Giannone, Lenza, and Primiceri (ECTA 2021) (GLP2021) develop a Bayesian sparse model that learns sparsity levels in linear regression.

- Test six high-dimensional datasets (Macro/Finance/Micro); Find the posterior distribution rarely concentrates on a single sparse model.

$\Rightarrow$ *illusion of sparsity*

Can sparsity be treated not as an assumption, but as an inferred property of the data?

- GLP2021 links L_1 and L_2: **no prespecified assumption**, but posterior learning for the unknown proportion of non-zero coefficients.



Schrödinger's cat

- A cat, entangled with a quantum system, remains in a superposition of alive and dead states until observed.
- The nature of AP models — sparse or dense — are in a state of superposition until empirical data is observed.

We examine the sparsity of Asset Pricing models within the conditional latent factor framework of IPCA with potentially mispricing.

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

- $\mathbf{f}_t$: K latent factors (can include observable factors).
- $\mathbf{Z}_{i,t-1}$: L characteristics.

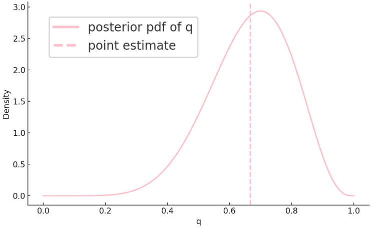
Research Questions

- Built on IPCA (Kelly, Pruitt, and Su, JFE 2019; Chen, Roussanov, and Wang, WP 2023) and Bayesian unconditional latent factor model (Geweke and Zhou, RFS 1996).

  — A New Perspective: Probability of char sparsity

- Our focus is on the char-driven betas and potentially mispricing.

  — why Bayes?



- Allow sparsity prob. to be data-inferred or exogenously fixed, enabling model estimation without / with sparsity assumptions.

Model

Model Setting

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

- $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) = \mathbf{0} \Rightarrow$ Risk-based pricing model / factor model
  - Mapping $\mathbf{Z}_{i,t-1} \mapsto \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$ encodes systematic risk exposure
  - **Hypo**: Factor structure is both sufficient and complete for spanning the cross section of $\mathbb{E}[r_{i,t}]$
- $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) \neq \mathbf{0} \Rightarrow$ Data-generating process for expected returns
  - Additional characteristic-driven components in expected returns are needed beyond any risk-based factor representation
  - **Hypo**: Factor structure is one component of a forecasting model

Let $d = 1$ or 0 denote selected or not selected, the spike and slab prior on β is

$$\beta \mid d \sim d\mathcal{N}\left(0, \xi_1^2\sigma^2\right) + (1-d)\mathcal{N}\left(0, \xi_0^2\sigma^2\right)$$

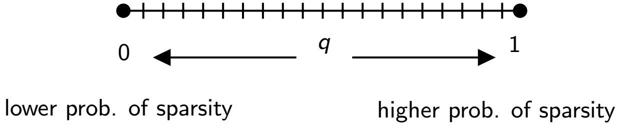
$$P(d=0) = 1 - P(d=1) = q$$

Hence, when ξ_1 is related large and ξ_0 shrinks to zero:

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

Spike-and-Slab Prior: Endogenous q

- Standard spike-and-slab prior: q is a specific value.
- GLP2021: q has its prior so that one can sample: $q \sim \mathrm{Beta}(a, b)$
  - These priors probabilistically balance variable selection and shrinkage.



- **Prior settings of $q \neq$ precise control of sparsity levels!**

Prior: Learning Sparsity Probability

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

- Independent spike-and-slab priors on $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$
- **Separate priors**: different sparsity levels of alpha and beta.

$$[\boldsymbol{\alpha}_1]_l \sim \begin{cases} \mathcal{N}(0, \gamma_\alpha^2) & \text{if } d_l^\alpha = 1 \\ 0 & \text{if } d_l^\alpha = 0 \end{cases} \qquad [\boldsymbol{\beta}_1]_l \sim \begin{cases} \mathcal{N}(0, \gamma_\beta^2) & \text{if } d_l^\beta = 1 \\ 0 & \text{if } d_l^\beta = 0 \end{cases}$$

$$d_l^\alpha \sim \text{Bernoulli}(1 - q_\alpha) \qquad d_l^\beta \sim \text{Bernoulli}(1 - q_\beta)$$

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

$$\gamma_\alpha^2 \sim \mathcal{IG}\left(A_{\gamma_\alpha}/2, B_{\gamma_\alpha}/2\right) \qquad \gamma_\beta^2 \sim \mathcal{IG}\left(A_{\gamma_\beta}/2, B_{\gamma_\beta}/2\right)$$

- Higher posterior mean of q_α (or q_β), higher prob. of sparsity.

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

- Directly control the sparsity level (i.e., control # selected char.).

  M_α and M_β restrict the number of char. driving alpha and beta.

- **(Separate) joint priors**:

$$(d_1^\alpha, d_2^\alpha, \cdots, d_L^\alpha) \sim \left[\prod_{l=1}^{L} \text{Bernoulli}(1 - q_\alpha)\right] \times \mathbf{I}\left(\sum_{l=1}^{L} d_l = M_\alpha\right),$$

$$(d_1^\beta, d_2^\beta, \cdots, d_L^\beta) \sim \left[\prod_{l=1}^{L} \text{Bernoulli}(1 - q_\beta)\right] \times \mathbf{I}\left(\sum_{l=1}^{L} d_l = M_\beta\right).$$

- Larger M_α (or M_β), lower **sparsity level**.

Schrödinger's Sparsity

Data

20 characteristics:

- Categories for frictions, momentum, investment, intangibles, value-versus-growth, and profitability.

Main test assets:

- P-Tree (Cong, Feng, He, and He, JFE 2025) test assets (1990-2024)
  - Sequential decreasing alphas by boosted trees
  - Constructed based on the past sample (1980-1989)

Other test assets:

- 25 ME/BM portfolios (ME/BM25), 360 bivariate-sorted portfolios (Bi360), and 610 univariate-sorted portfolios (Uni610).
- 500 stocks with the 1st-500th and 501st-1000th average market equity (Big ind500 / Small ind500).

(i) Learning Sparsity

Table 1: Performance for Various Models

		CSR^2			SR			INS q_β and $\widehat{M}_\beta$		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Learning Sparsity										
	0.9	14.7	63.1	68.3	0.49	0.38	0.92	0.59 (11)	0.57 (12)	0.60 (11)
q_β prior mean	0.5	14.6	61.9	68.0	0.49	0.51	0.90	0.43 (11)	0.44 (12)	0.47 (11)
	0.1	14.5	62.8	68.9	0.49	0.54	0.98	0.24 (13)	0.30 (12)	0.32 (11)
Panel B: Fixed Sparsity Level										
	2	13.6	63.1	62.7	0.50	0.23	0.63	/	/	/
M_β	10	13.8	62.6	64.9	0.49	0.60	0.66	/	/	/
	18	14.9	64.0	66.2	0.49	0.54	0.55	/	/	/
Panel C: No Sparsity										
M_β	20	14.4	62.8	65.4	0.49	0.45	0.45	/	/	/
Panel D: IPCA										
M_β	20	17.8	61.7	70.8	0.33	0.50	0.74	/	/	/

- Models are not very sparse, nor dense
- Learn rather than impose sparsity in conditional asset pricing models

(ii) Test Assets and Sparsity

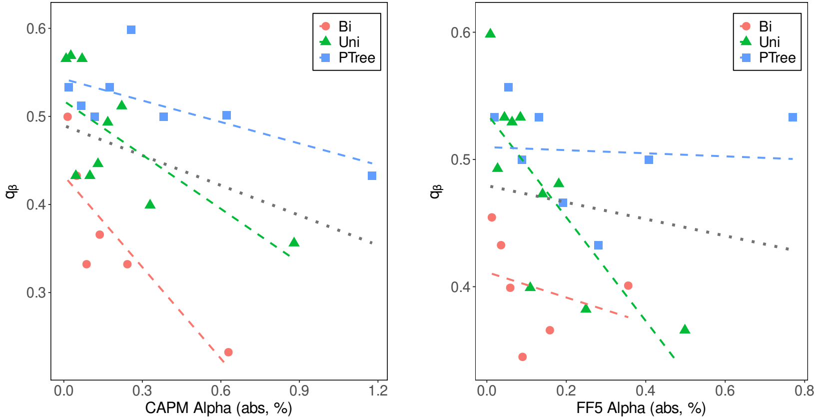
Table 2: Sparsity for Different Test Assets

	CSR^2	SR	q_β	$\widehat{M}_\beta$
Panel A: P-Tree				
100	59.6	1.12	0.50	10
200	69.4	0.68	0.37	14
400	63.3	1.01	0.26	17
Panel B: Ind. Stock				
Big500	46.9	1.54	0.30	16
Small500	30.1	4.16	0.42	12
Panel C: Others				
ME/BM25	53.6	0.82	0.50	10
Bi360	71.6	1.15	0.21	19
Uni610	66.1	0.87	0.23	18

- Sparsity levels change across different types of test assets.
- Panels A, C: Assets that are more difficult to price require more chars.
- Panel B: Effect of potential mispricing.

Pricing Difficulty versus Sparsity

Figure 1: Sparsity and Pricing Difficulty for Different Test Assets



Sparsity is linked to pricing difficulties of test assets.

(iii) Macro Regimes and Sparsity

Table 3: Sparsity in Structural Breaks / Business Cycles

	CSR^2	SR	q_β	$\widehat{M}_\beta$
Panel A: Sequential segmentation				
Regime1	52.9	1.40	0.53	9
Regime2	36.6	0.74	0.53	9
Regime3	68.9	0.53	0.50	10
Panel B: Macro-driven segmentation				
Normal	61.7	0.83	0.49	10
Recession	21.9	1.01	0.56	8
Panel C: Full period				
Whole	52.1	0.73	0.46	11

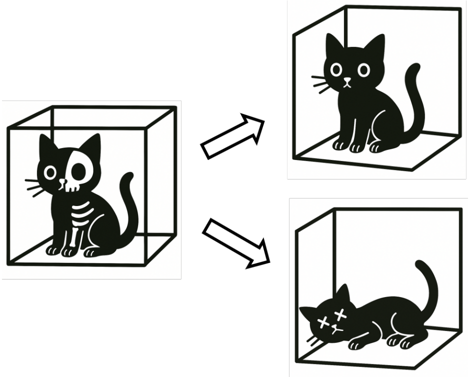
- Settings of time periods:
  - Breakpoints in Smith and Timmermann (RFS 2021): July 1998 and June 2010.
  - Define recession periods based on the Sahm Rule (88 months).
- AP models tend to be sparser during recessions.

  ⇒ Macro conditions dominate.

Schrödinger's Sparsity Everywhere!

Sparsity Prob. change across both **cross-sectional** and **time-series** dimensions.

⇒ i) Test assets / Pricing difficulty; ii) Time periods / Macro conditions



Learning sparsity prob, instead of assuming AP model to be

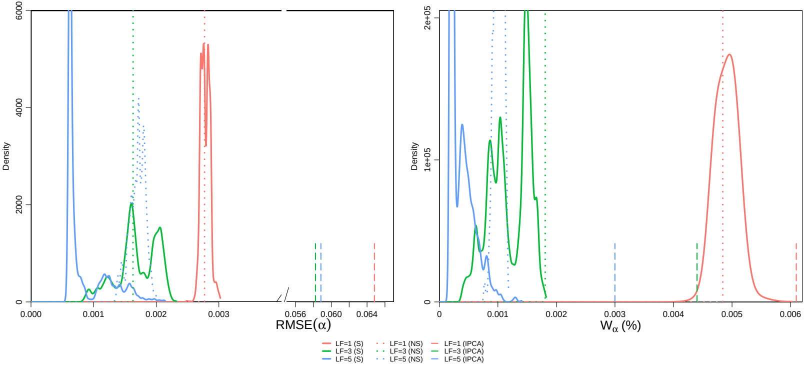
either sparse or dense ex ante

Learning Sparsity with Mispricing

(i) Mispricing Test

$$\widehat{\boldsymbol{\alpha}}_{it}^{(g)} = \widehat{\alpha}_0^{(g)} + \widehat{\boldsymbol{\alpha}}_1^{(g)\top} \mathbf{Z}_{i,t-1}$$

- Scale of the coefficient vector: $W_\alpha^{(g)} = \widehat{\Gamma}_\alpha^{(g)\prime} \widehat{\Gamma}_\alpha^{(g)}$, where $\Gamma_\alpha = [\alpha_0, \boldsymbol{\alpha}_1]$
- Scale of the implied mispricing: $\widehat{\boldsymbol{\alpha}}^{(g)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\alpha}}_{it}^{(g)}\right)^2}$



(a) Value / density of RMSE($\boldsymbol{\alpha}$) (b) Value / density of $W_\alpha(\%)$

(ii) Investment Performance

Table 4: Forecast-Implied Investment Performance (Sharpe Ratio) for Various Models

		Sign-adj. Value-Weighted			Sign-adj. Equal-Weighted			Forecast-Weighted		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Learning Sparsity										
(q_α,q_β) prior mean	0.5,0.5	0.59	-0.14	**0.83**	0.42	-0.01	**0.76**	0.46	0.05	**0.78**
		INS posterior of q_α, q_β and $\widehat{M}_\alpha, \widehat{M}_\beta$								
		$K=1$: (0.57,0.44) and (7,11); $K=3$: (0.68,0.43) and (5,12); $K=5$: (0.80,0.44) and (0,12)								
Panel B: Fixed Sparsity Level										
	2,2	0.59	0.70	0.63	0.43	0.46	0.43	0.42	0.52	0.48
	10,2	0.58	0.59	-0.61	0.42	0.42	-0.42	0.40	0.39	-0.39
	18,2	0.57	0.33	-0.50	0.43	0.14	-0.37	0.40	0.12	-0.31
	2,10	0.68	0.35	0.73	0.48	0.11	0.55	0.50	0.12	0.56
(M_α, M_β)	10,10	0.63	0.73	0.61	0.43	0.46	0.42	0.47	0.49	0.39
	18,10	0.65	0.74	0.61	0.44	0.53	0.42	0.47	0.48	0.47
	2,18	0.70	0.09	0.71	0.52	-0.04	0.57	0.53	0.03	0.59
	10,18	0.68	0.47	0.61	0.49	0.21	0.42	0.51	0.18	0.38
	18,18	0.68	0.41	0.16	0.49	0.18	-0.12	0.51	0.17	-0.14
Panel C: No Sparsity										
(M_α, M_β)	20	0.67	0.72	0.74	0.46	0.48	0.51	0.51	0.46	0.54
Panel D: IPCA										
(M_α, M_β)	20	0.66	0.66	0.74	0.52	0.48	0.56	0.55	0.53	0.57

(iii) Risk and Mispricing

Table 5: Performance for Various Models with Mispricing

		CSR^2_{adj}			Pure-alpha SR			Alpha long-short SR		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Learning Sparsity										
(q_α, q_β) prior mean	0.5,0.5	13.3	63.9	69.3	0.50	0.73	0.86	0.81	0.76	1.04
Panel B: Fixed Sparsity Level										
	2,2	13.3	64.0	63.7	0.04	0.46	0.46	0.55	0.60	0.48
	10,2	13.6	63.6	63.7	0.55	0.86	0.86	0.92	0.94	0.88
	18,2	13.9	63.8	63.7	0.53	0.75	0.64	1.00	0.80	0.75
	2,10	13.2	61.6	66.6	0.04	0.41	0.54	0.54	0.14	0.93
(M_α, M_β)	10,10	12.8	62.5	63.9	0.54	0.76	0.64	0.85	0.77	1.00
	18,10	13.4	59.5	65.9	0.54	0.37	0.40	0.97	0.69	0.79
	2,18	13.4	62.0	65.8	0.01	0.45	0.10	0.53	0.41	0.44
	10,18	13.9	61.0	67.3	0.54	0.58	0.56	0.85	0.83	0.92
	18,18	12.9	61.2	65.0	0.54	0.56	0.56	0.97	0.87	0.90
Panel C: No Sparsity										
(M_α, M_β)	20	12.2	58.9	68.4	0.49	0.46	0.30	0.87	0.75	0.67
Panel D: IPCA										
(M_α, M_β)	20	16.4	59.2	69.3	0.67	0.56	0.43	0.81	0.74	0.77

▸ Go to metrics

Summary

- QUESTION: **How can researchers determine model assumptions before examining the data?**

  ⇒ Schrödinger's Sparsity: the true state remains unknowable until observed

  - treating sparsity as a probabilistic property rather than a binary assumption

- A new approach, a flexible Bayesian framework
  - Utilizing the independent / joint spike-and-slab priors
  - **Endogenously** determine whether the model is sparse or dense, without imposing prior assumptions on sparsity or density
  - **Exogenously** control the sparsity level of the model

- Empirical findings:
  - Best models lie between the extremes of full sparsity and full density
  - Learning sparsity matters
  - Cross section: Sparsity probability is linked to the pricing difficulty of test assets
  - Time series: Sparsity depends on macro states and increases during recessions

(iii) Risk and Mispricing (measurement)**

- Adjusted cross-sectional R^2
  - Non-traded factor:

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

  - A traded (realized) factor proxy:

$$\mathbf{f}_t^{\text{traded}} = \boldsymbol{\beta}^\top \left(\boldsymbol{\beta}\boldsymbol{\beta}^\top + \boldsymbol{\Sigma}\right)^{-1}\mathbf{r}_t$$

  - The fitted return obtained from the risk-exposure channel ($\tilde{r}_{i,t}$):

$$\widetilde{r}_{i,t} = \widehat{\boldsymbol{\beta}}_0^\top \mathbf{f}_t^{\text{traded}} + \widehat{\boldsymbol{\beta}}_1^\top \left(\mathbf{f}_t^{\text{traded}} \otimes \mathbf{Z}_{i,t-1}\right)$$

- Alpha strategies
  - Pure-alpha strategy

$$\boldsymbol{w}_{t-1}^{\text{PA}} = \widetilde{\mathbf{Z}}_{t-1}\left(\widetilde{\mathbf{Z}}_{t-1}^\top \widetilde{\mathbf{Z}}_{t-1}\right)^{-1}\widehat{\Gamma}_\alpha, \text{ where } \widetilde{\mathbf{Z}}_{t-1} = [\mathbf{1}, \mathbf{Z}_{t-1}]$$

$$R_t^{\text{PA}} = (\boldsymbol{w}_{t-1}^{\text{PA}})^\top \left(\mathbf{r}_t - \widehat{\boldsymbol{\beta}}_0^\top \mathbf{f}_t^{\text{traded}} - \widehat{\boldsymbol{\beta}}_1^\top \left(\mathbf{f}_t^{\text{traded}} \otimes \mathbf{Z}_{t-1}\right)\right).$$

  - Alpha long-short strategy

$$\boldsymbol{w}_{t-1}^{\text{LS}} = \widetilde{\mathbf{Z}}_{t-1}\widehat{\Gamma}_\alpha - \text{mean}(\widetilde{\mathbf{Z}}_{t-1}\widehat{\Gamma}_\alpha),$$

Back

Observable Factors and Sparsity

In the conditional model, beta are functions of char.

- $\mathbf{f}^L$: Latent factor

$$r_{i,t} = \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})^\top \mathbf{f}_t^L + \epsilon_{i,t} = \underbrace{\boldsymbol{\beta}_0^\top \mathbf{f}_t^L + \boldsymbol{\beta}_1^\top [\mathbf{f}_t^L \otimes \mathbf{Z}_{i,t-1}]}_{\text{latent factors, conditional beta}} + \epsilon_{i,t}$$

- $\mathbf{f}^O$: Pre-specified factor

$$r_{i,t} = \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})^\top \mathbf{f}_t^O + \epsilon_{i,t} = \underbrace{\boldsymbol{\beta}_0^\top \mathbf{f}_t^O + \boldsymbol{\beta}_1^\top [\mathbf{f}_t^O \otimes \mathbf{Z}_{i,t-1}]}_{\text{obs. factors, conditional beta}} + \epsilon_{i,t}$$

Replacing latent factors with pre-specified factors

- Interpretation of "sparsity"
- Persistence of sparsity patterns

(iv) Prespecified Factors and Sparsity

- Obs: Market factor; Fama-French five factors (FF5)
- Latent but prespecified: Five factors estimated via IPCA (IPCA5)

Figure 3: Sparsity across Factors and Test Assets