Test Assets and Weak Factors

Stefano Giglio, Dacheng Xiu and Dake Zhang

Yale SOM, Chicago Booth, Chicago Booth

TEDS and Nanyang Technological University

October 21, 2021

Introduction

- Estimation and testing of **risk premia of factors** is central in testing economic mechanisms of asset pricing theories
  - Typically, implications are about **nontradable** factors
  - E.g.: intermediary capital, liquidity, consumption
- **Interpretation:** how much are investors willing to pay to hedge that risk, and **only** that risk (holding all other risks constant)
- **How do we estimate risk premia for a factor g_t?**
  1. Take a set of test assets
  2. Build a hedging portfolio for the factor of interest using those assets
  3. Risk premium is the average excess return of that portfolio

Existing literature mainly focus on 2; Little work on 1.

A preview of the model

- Linear factor model with p factors v_t (zero-mean) for $N \times 1$ excess returns r_t

$$r_t = \beta\gamma + \beta v_t + u_t$$

- γ are the factor risk premia for v_t
- β are the risk exposures
- Factor of interest g_t, which could be (one of) the factor v_t.
- The parameter of interest is the risk premium of g_t. If g_t is (an entry of) v_t, then its risk premium is (corresponding) γ.

Introduction

- In theory, easy to estimate: standard **two-pass cross-sectional regression**

$$\widehat{\gamma} = (\widehat{\beta}^{\intercal}\widehat{\beta})^{-1}\widehat{\beta}^{\intercal}\bar{r}.$$

- Two pitfalls have been emphasized in the literature: both **very** important in applications

1. **Omitted factor bias:** what are the right risk factors to control for?
    - Often theory is too stylized; we know many factors drive returns in reality

2. **Weak factor bias:** what if our test assets appear only weakly related to g_t? i.e., $\beta \approx 0$
    - If **any** of the factors are weak, inference on **all** factors is contaminated, even when g_t is strong!
    - Intuitively, this is due to errors-in-variables.

Introduction

- We propose a procedure to estimate risk premia and conduct inference that addresses **both problems**
- Our method integrates:
  - **PCA** to address the omitted factor problem (as in Giglio and Xiu (2021, JPE))
  - **Supervised asset selection** to address the weak factor problem
- Key insight: tackle the issue of **weak factors** via its connection to **test assets.**
  - Standard view: strength of factors is a property of the factors
  - Our view: strength of factors is a **property of the test assets**: any factor could be strong or weak
  - Selecting assets appropriately can make a weak factor effectively strong!

Which test assets?

Which test assets are typically used when estimating risk premia?

- **"Standard"** sets: e.g., Fama-French 25
  - Many risk factors are not reflected in this cross-section
  - They are **weak** and their risk premia cannot be strongly identified
- **"Large"** datasets, sorted by many characteristics
  - Contain exposure to many risk factors
  - But again many of them are potentially **weak**: reflected only in a small subsection
  - Not sufficient to solve the "weak factor" issue
- **"Targeted"**, beta-sorted portfolios
  - Univariate sorts: omitted factors

SPCA: intuition

Two key insights behind our procedure:

1. We can use the factor g_t to help **select** particularly informative assets, and discard the rest
2. We can extract (from returns of the selected portfolios) **latent factors** that help control for the omitted factors, and isolate the risk premium of the factor of interest

Dealing with weak factors in a latent factor model by itself is an open problem.

Literature Review

- Kan and Zhang (1999, JF) and Kleibergen (2009, JoE): Weak factors lead to distorted inference.
- Lettau and Pelger (2020, JoE, RFS): propose rp-PCA.
- Freyaldenhoven (2019) proposes an estimator of the number of "weak" factors.
- Pesaran and Smith (2019): FM estimator converges more slowly as factors become weaker.
- Bailey, Kapetanios, and Pesaran (2020, JAE) propose a measure of factor strength.
- Anatolyev and Mikusheva (2018, JoE): propose four-splitting estimator.

- Bai and Ng (2008, JoE) propose hard thresholding, Lasso, and elastic net to drop some data prior to PCA for forecasting.
- Huang, Jiang, Li, Tong, and Zhou (2021, MS) propose scaled-PCA to improve forecasting performance.

- Bair and Tibshirani (2004, Plos Biology) and Bair, Hastie, Paul, and Tibishirani (2006, JASA) propose SPCA in a restrictive setting in which screening is only done once.

Outline

1. Setup: linear factor model
2. Thee-pass estimator
3. Weak factors
4. How asset selection solves the weak factor problem: supervised PCA
5. Simulations
6. Empirical results

1. The general model

More generally, we will consider this model:

$$r_t = \beta\gamma + \beta v_t + u_t, \quad g_t = \xi + \eta v_t + z_t$$

- **Risk premium** of g_t is $\eta\gamma$: the object of interest
- Econometricians observe r_t and g_t.
- v_t is latent – this means we allow for omitted factors.
- z_t is measurement error.

Notation:

- For any time series of vectors $\{a_t\}_{t=1}^T$, we denote $\bar{a} = \frac{1}{T}\sum_{t=1}^T a_t$.
- In addition, we write $\bar{a}_t = a_t - \bar{a}$.
- We use the capital letter A to denote the matrix $(a_1 : a_2 : \ldots : a_T)$.
- We write $\bar{A} = A - \bar{a}\iota_T^\intercal$, where we use ι_k to denote a $k \times 1$ vector of 1s.

2. A Three-pass Approach to Estimating Risk Premia

$$r_t = \beta H^{-1} H \gamma + \beta H^{-1} H v_t + u_t$$
$$g_t = \xi + \eta H^{-1} H v_t + z_t$$

Giglio and Xiu (2021, JPE) propose a **three-pass estimator** to obtain $\eta\gamma$.

1. Extract latent factors of returns $\widehat{v}_t = H v_t$ via PCA
2. Use cross-sectional regression to estimate latent factor risk premia $\widehat{\gamma} = H\gamma$
3. Regress g_t on the latent factors $\widehat{v}_t$ via time-series regression to obtain $\widehat{\eta} = \eta H^{-1}$

Risk premium of g_t is estimated as $\widehat{\eta}\widehat{\gamma} = \eta H^{-1} H \gamma = \eta\gamma$

3. Weak factors

- The three-pass procedure requires that all latent factors are **strong**
  - If some of the factors v_t are **weak** in the panel of returns, PCA might not recover it
  - Leads to a bias in risk premium

3. Weak factors

- To illustrate the issue, we start with a 1 latent-factor model (without measurement error z_t):
$$r_t = \beta\gamma + \beta v_t + u_t, \quad g_t = \xi + \eta v_t$$
- An sufficient condition for consistency with respect to $\eta\gamma$:
$$N/(||\beta||^2 T) \to B = 0.$$
- If it converges to $B > 0$, then
$$\widehat{\gamma}_g^{PCA} \xrightarrow{p} (1 + B)^{-1}\eta\gamma.$$
- Note: it's **not sufficient to have many assets**: adding assets with small beta with respect to g_t actually hurts!

3. Alternative procedures

- The three-pass approach is equivalent to conducting a PCA regression of g_t on r_t.
- What if we use PLS or Ridge?

$$\text{PLS:} \quad \widehat{\gamma}_g^{PLS} = ||\bar{G}\bar{R}^\intercal \bar{R}||^{-2} \bar{G}\bar{R}^\intercal \bar{R}\bar{G}^\intercal \bar{G}\bar{R}^\intercal \bar{r};$$

$$\text{Ridge:} \quad \widehat{\gamma}_g^{Ridge} = \bar{G}\bar{R}^\intercal \left(\bar{R}\bar{R}^\intercal + \mu \mathbb{I}_N\right)^{-1} \bar{r}.$$

- The same issue remains: if $N/(||\beta||^2 T) \to B \geq 0$,

$$\widehat{\gamma}_g^{PLS} \xrightarrow{p} (1 + B)^{-1} \eta \gamma.$$

Similarly, if, in addition, $\mu/(||\beta||^2 T) \to D \geq 0$,

$$\widehat{\gamma}_g^{Ridge} \xrightarrow{p} (1 + B + D)^{-1} \eta \gamma.$$

3. Alternative procedures

- Lettau and Pelger (2020, JoE and RFS) propose rp-PCA to estimate **factors** in an AP model.
  - Instead of PCA on $T^{-1}\bar{R}\bar{R}^{\intercal}$, they suggest applying PCA on $T^{-1}RR^{\intercal}+\mu\bar{r}\bar{r}^{\intercal}$, where μ is a tuning parameter.
  - This is appealing in that the information on the mean is used since $r_t=\beta\gamma+\beta v_t+u_t$.
  - They prove that in the case of strong factors, this procedure could be more efficient than PCA to recover **factors**.
  - In the case of extremely weak factors, i.e., $||\beta||=O_p(1)$, however, no estimators can be consistent. Their procedure seems to produce a higher correlation with the true factors than PCA.
- For **risk premia** estimation, we prove, if $N/(||\beta||^2T)\to B\geq 0$, for any $\mu>-1$

$$\widehat{\gamma}_g^{rpPCA}\xrightarrow{p} w(1+B)^{-1}\eta\gamma+(1-w)\eta(\gamma+\gamma^{-1}B),$$

where

$$w=\frac{2+2B}{1+2B+\sqrt{(1-a)^2+4(1+\mu)\gamma+a}},\qquad a=(1+\mu)(\gamma^2+B)-B.$$

4. Supervised PCA

1. **Supervised selection:** Compute the (univariate) correlation of factor g_t with all returns r_t, and select only assets with sufficiently high correlation (top $q \times N$ assets sorted by correlation, where q is a tuning parameter)
   $\widehat{I} \subset [N]$: $\widehat{I} = \left\{ i \,\middle|\, T^{-1} |\bar{R}_{[i]} \bar{G}^\intercal| \geq c_q \right\}$, where c_q is the $(1-q)$-quantile of $\left\{ T^{-1} |\bar{R}_{[i]} \bar{G}^\intercal| \right\}_{i \in [N]}$.

2. **PCA:** Extract latent factor(s) from this subset

- Key assumption: there exists a subset $I_0 \subset [N]$ such that $\left\| \beta_{[I_0]} \right\| \asymp \sqrt{N_0}$ where $N_0 = |I_0| \to \infty$.
- Asymptotic guarantee: in the one factor case, suppose $\log N / T \to 0$, for any choice of q such that $qN/N_0 \to 0$ and $qN \to \infty$, we have

$$\widehat{\gamma}_g^{SPCA} \xrightarrow{p} \eta \gamma.$$

4. Supervised PCA: challenges in the multivariate case

- In general, PCA estimator of risk premia becomes inconsistent if $N/(\lambda_{\min}(\beta^\intercal\beta)T) \nrightarrow 0$.
- In a multi-factor model, even if all factors are strong by themselves, a related problem arises when some of the factors' exposures are highly correlated.
  - **Example 1.** Suppose the loading matrix of R is

$$\beta = \left[\begin{array}{c|c} \beta_{11} & \beta_{12} \\ \hline \beta_{21} & \beta_{22} \end{array}\right],$$

    where β_{11} and β_{12} are $N_0 \times 1$ vectors and β_{21} and β_{22} are $(N - N_0) \times 1$ vectors.
  - If $\beta_{21} = \beta_{22}$, then even if either factor is strong, the same "weak factor" issue arises.

4. Supervised PCA: challenges in the multivariate case

- If we only do **selection and PCA only once,** sometimes **too few** assets are screened out.
  - **Example 2.** Suppose the loading matrix of R is
    $$\beta = \left[\begin{array}{c|c} \beta_{11} & \beta_{12} \\ \hline \beta_{21} & 0 \end{array}\right],$$
  - The loading matrix of g_t is $\eta = [1, 1]$: $g_t = v_{1t} + v_{2t}$, so we need to recover both factors to get $\eta\gamma$
  - If N_0/N is small, v_{2t} is weak
  - If $\beta_{21} \neq 0$, **all** assets are correlated with g_t and thus the selection will not zoom in on N_0!
  - The factor extracted is strong, and the weak factor will be missed.

4. Supervised PCA: challenges in the multivariate case

- If we only do **selection and PCA only once,** sometimes **too many** assets are screened out.
  - **Example 3.** Suppose the loading matrix of R is

$$\beta = \left[\begin{array}{c|c} \beta_{11} & \beta_{11} \\ \hline 0 & \beta_{22} \end{array}\right],$$

  - The loading matrix of G is $\eta = [1,0]$, so $g_1 = v_{1t}$
  - Both factors are strong but loadings are highly correlated
  - After the screening, only the first half will remain
  - But the first half is now a 1-factor model with factor $v_{1t} + v_{2t}$, so v_{1t} cannot be identified.

4. Supervised PCA with Projection and Iteration

1. **Supervised selection**
2. **PCA** Extract one latent factor $\widehat{V}_{(1)}$ from this subset, estimate $\widehat{\beta}_{(1)}$, $\widehat{\eta}_{(1)}$, and estimate risk premia $\widehat{\gamma}_{(1)}$ only using the selected set
3. **Projection:** Compute residuals of g_t and (all) R_t with respect to this factor
4. **... iterate:** Compute the (univariate) correlation of the g_t residual with the R_t residuals; select assets with sufficiently high correlation
   - Extract latent factor $\widehat{V}_{(2)}$, $\widehat{\beta}_{(2)}$, $\widehat{\eta}_{(2)}$, and $\widehat{\gamma}_{(2)}$ from this subset... build residuals... repeat. Stop at the number of factors $\widehat{p} = k - 1$, (if $c_q^{(k)} < c$, for some threshold c.)
5. Once all factor risk premia are collected, $\widehat{\gamma}_g^{SPCA} = \sum_{k=1}^{\widehat{p}} \widehat{\eta}_{(k)} \widehat{\gamma}_{(k)}$.

Note: the resulting factors are orthogonal, betas are only estimated for small cross sections.

4. Supervised PCA: Consistency and Inference

- Iterative screening and projection approach addresses the problems in these examples.
- Consistency: $\widehat{\gamma}_g^{SPCA} \xrightarrow{p} \eta\gamma$ if
  - all factors are strong in some subset I_0: $\lambda_{\min}(\beta_{[I_0]}^\intercal \beta_{[I_0]}) \asymp N_0$.
  - $\log(NT)(N_0^{-1} + T^{-1}) \to 0$, $c^{-1}(\log NT)^{1/2}(q^{-1/2}N^{-1/2} + T^{-1/2}) \to 0, \quad c, qN/N_0 \to 0$,
- We have CLT results:

$$\sqrt{T}\left(\widehat{\gamma}_g^{SPCA} - \eta\gamma\right) \xrightarrow{d} N(0, \Phi),$$

  if $T^{-1/2}N_0 \to \infty$, $q^{-1}N^{-1}T^{1/2} \to 0$, $\lambda_{\min}(\eta^\intercal\eta) \gtrsim 1$.
- Φ can be estimated by a standard Newey-West-type estimator.

4. Supervised PCA

SPCA gives us:

- **Consistency** in all cases
- **Asymptotic inference** with one additional assumption: $\lambda_{\min}(\eta^\intercal \eta) \gtrsim 1$
  - Recall that we need to identify both v, weak and strong, and η
  - Potential issue for inference (not for consistency): **low** η, e.g., $\eta \sim T^{-1/3}$
  - Assume that this is obviated by the presence of other factors in G.

Our paper also provides theoretical results based on Lasso and Ridge, by first estimating the SDF, then recover its correlation with the factor of interest.

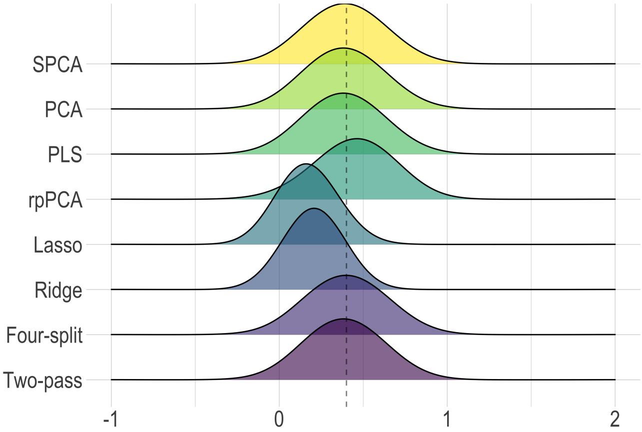
5. Simulation

- 4-factor DGP: RmRf, SMB, HML and one potentially weak factor, V.
- The first three factors are calibrated to match the empirical data.
- V is a potentially weak factor:

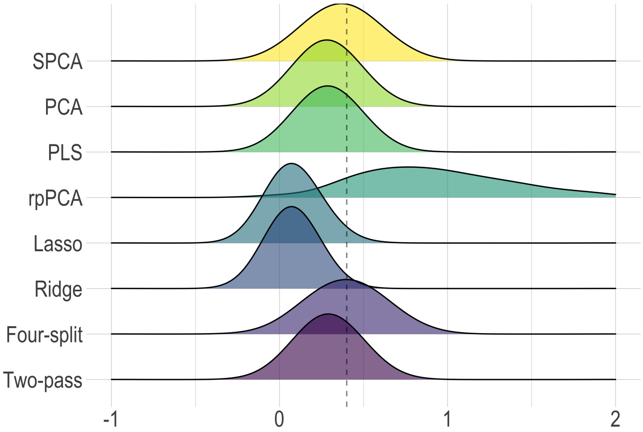
$$\beta_{i,V} \overset{i.i.d.}{\sim} aN(0,1)+(1-a)N(0,0.1^2)$$

- $a = 0.05,\ 0.5$
- $N = 2000,\ T = 120$
- We compare all relevant estimators, including four-split by Anatoyev and Mikusheva (2021, JoE)

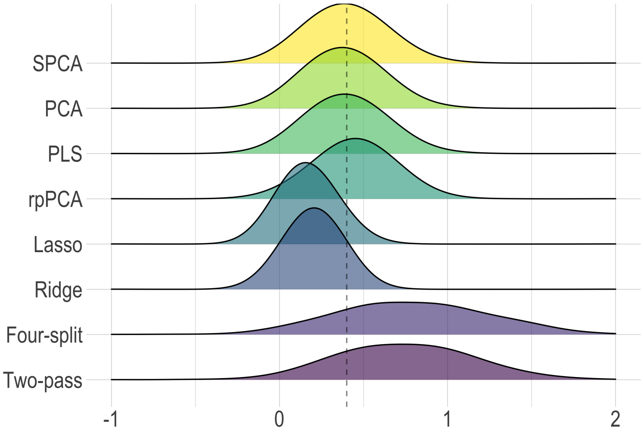
5. Simulation: benchmark case, $a = 0.5$, no measurement error



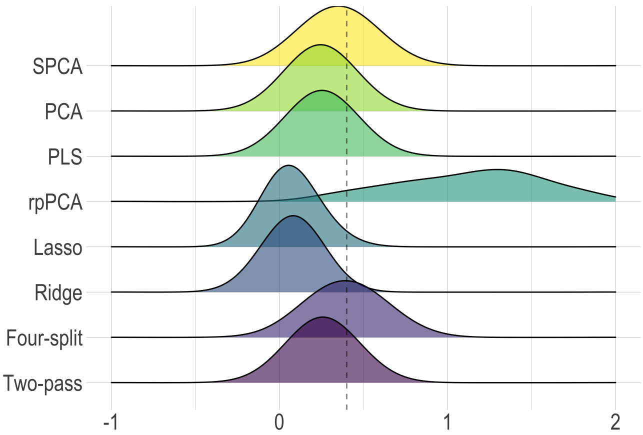
5. Simulation: $a = 0.05$



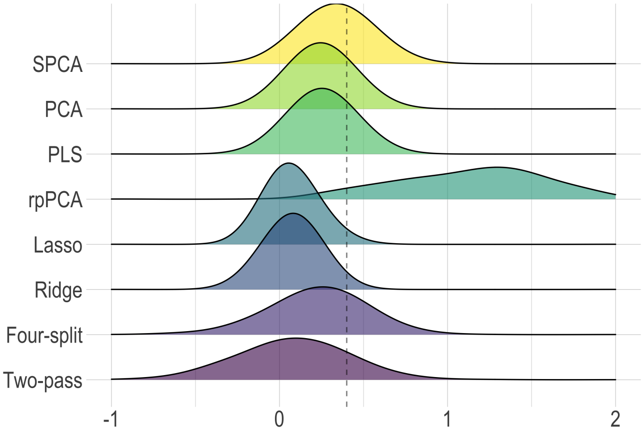
5. Simulation: $a = 0.5$ but with measurement error



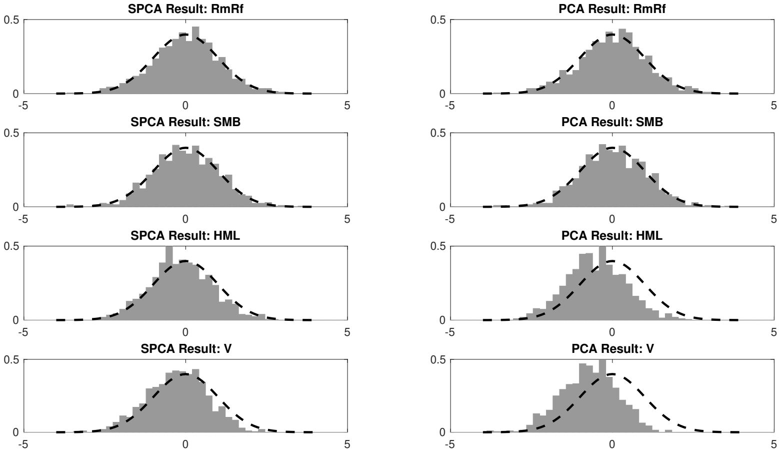
5. Simulation: highly correlated β



5. Simulation: with HML factor missing



5. Simulation: CLT, highly correlated β

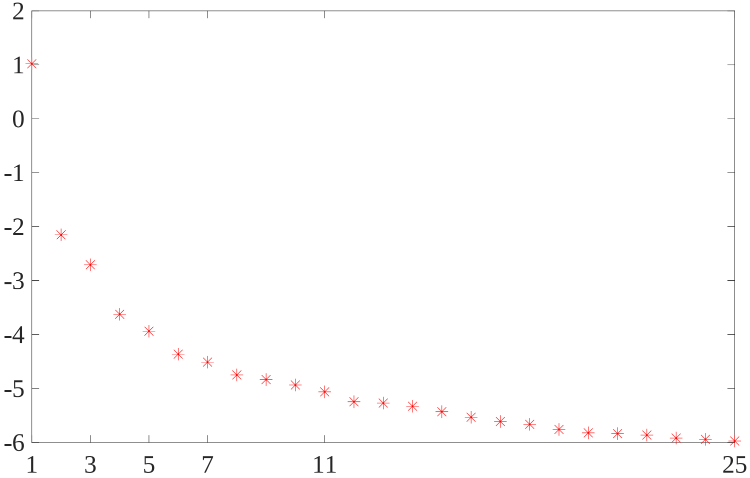


6. Empirical evidence

- Data: Chen and Zimmermann (2020) data for the universe of returns
  - Portfolios sorted by 210 characteristics, 1976-2019, 901 portfolios
  - 49 industry portfolios
- Factors:
  - Tradable: Market, SMB, HML, RMW, CMA, MOM, BAB, QMJ
  - Nontradable: Liquidity, Intermediary capital, IP, Macro PCs (LN 2010), the term spread; the credit spread; the unemployment rate; two sentiment indexes, one from Huang et al. (2015, RFS) and one from Baker and Wurgler (2006, JF); oil price growth AR(1) innovations; and consumption growth AR(1) innovations.

6. Empirical evidence

- Latent factors. Log eigenvalues:



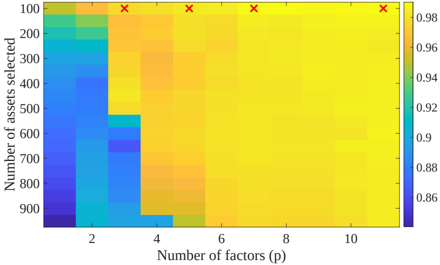
6. Empirical evidence

- Split the sample in half. Estimation and tuning parameter in the first half, evaluation in the second half.
- 100 times 3-fold CV, pick the parameter to maximize validation sample R^2 for the target factors for the implied mimicking portfolio (as in the simulations)
- Pick median value for the tuning parameter, use it OOS to evaluate the model

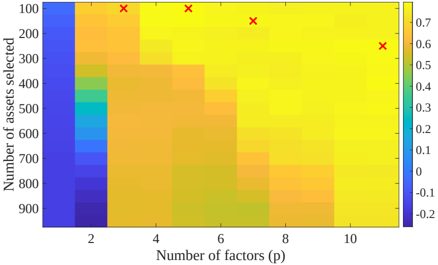
6. Empirical evidence

	Avg. ret. (train.)	Avg. ret. (eval.)	3 Latent Factors			5 Latent Factors			7 Latent Factors			11 Latent Factors			
			RP	# Assets	R^2	RP	# Assets	R^2	RP	# Assets	R^2	RP	# Assets	R^2	Stderrs for RP
Market	74	62	68	100	0.98	70	100	0.98	72	100	0.99	74	100	0.99	26
HML	39	-7	50	100	0.70	37	100	0.79	39	150	0.78	44	250	0.79	18
SMB	12	25	15	100	0.82	5	100	0.85	10	100	0.85	10	100	0.85	18
RMW	37	28	-8	100	-0.18	40	100	0.56	33	100	0.61	27	150	0.66	9
CMA	26	19	36	250	0.41	40	100	0.55	27	200	0.55	31	350	0.53	11
Momentum	91	30	67	100	0.79	86	100	0.87	102	100	0.87	101	100	0.88	23
BAB	126	56	112	100	0.43	120	100	0.38	112	150	0.35	128	150	0.45	20
QMJ	41	39	-9	100	0.43	28	100	0.81	31	100	0.80	36	150	0.78	10
Liquidity			70	550	0.01	85	650	0.02	83	700	0.04	95	900	0.03	25
Intermed. Cap.			112	100	0.59	101	100	0.56	121	150	0.55	116	350	0.52	41
IP growth			-4	950	-0.01	-4	950	-0.02	-5	950	-0.03	-2	950	0.00	3
LN 1			225	550	-0.28	202	650	-0.19	150	700	-0.11	54	950	-0.12	146
LN 2			-70	950	-0.05	-79	950	-0.12	-24	950	-0.16	-29	950	-0.17	82
LN 3			96	400	0.03	86	650	0.06	16	700	0.06	-21	850	0.05	78
Consumption			2	950	-0.01	3	950	0.00	3	950	-0.01	2	950	-0.01	2
Fin. Unc.			-61	350	-0.08	-48	750	0.00	-40	850	0.09	-41	950	0.10	17
Real Unc.			-6	950	0.05	-7	950	0.04	-9	950	0.04	-11	950	0.06	12
Macro Unc.			-7	950	0.08	-10	950	0.08	-10	950	0.08	-16	950	0.09	10
Term			229	950	-0.11	81	950	-0.36	-57	950	-0.54	262	950	-0.59	372
Credit			41	950	-0.03	62	950	-0.03	41	950	-0.02	-43	950	-0.03	77
Unempl.			65	950	0.00	109	950	-0.01	112	950	-0.01	110	950	0.00	108
Sentiment HJTZ			-24	950	0.01	-27	950	-0.03	-18	950	-0.06	-40	950	-0.07	76
Sentiment BW			57	950	0.00	64	950	0.00	50	950	0.01	16	950	-0.02	71
Oil			-37	950	-0.05	-62	950	-0.02	-42	950	-0.03	-20	950	-0.02	41

6. Empirical evidence

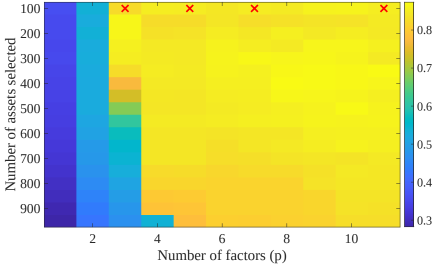


(a) Market

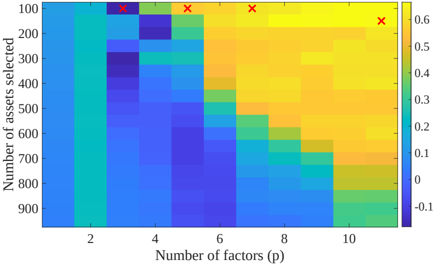


(b) HML

6. Empirical evidence

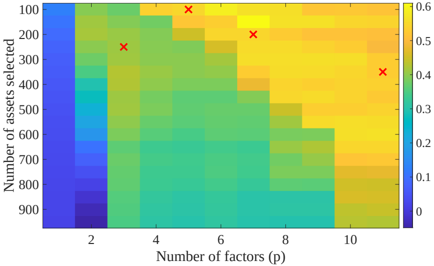


(c) SMB

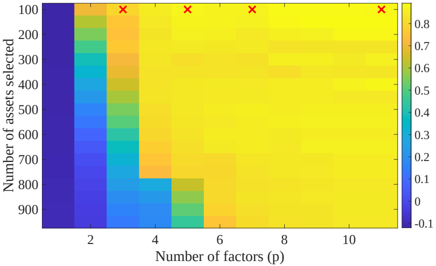


(d) RMW

6. Empirical evidence

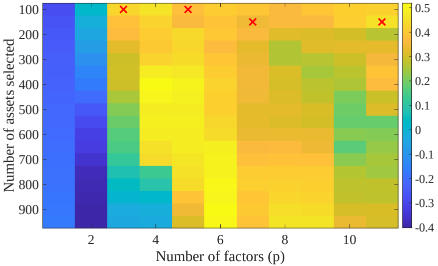


(e) CMA

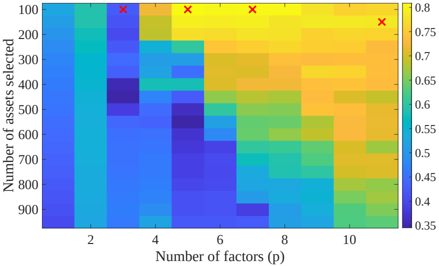


(f) Momentum

6. Empirical evidence

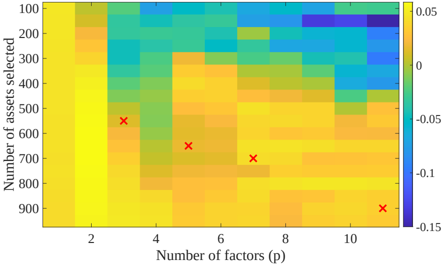


(g) BAB

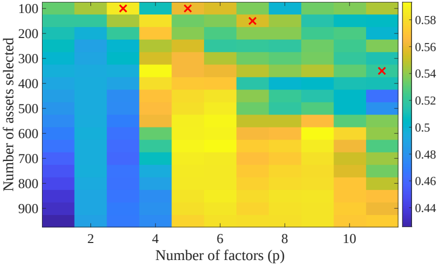


(h) QMJ

6. Empirical evidence

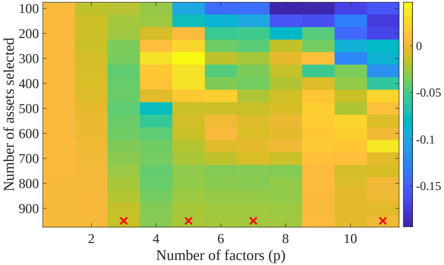


(i) Liquidity

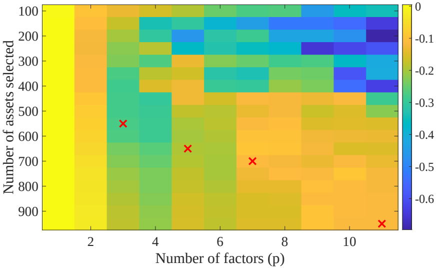


(j) Intermediary Capital

6. Empirical evidence

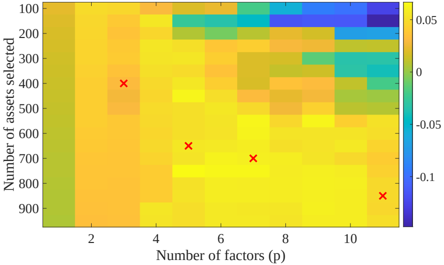


(k) IP Growth

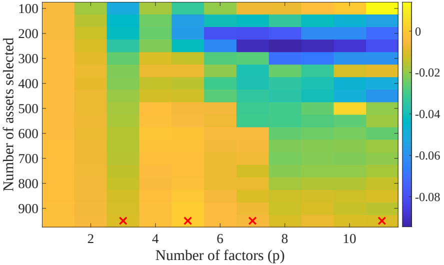


(l) LN #1

6. Empirical evidence

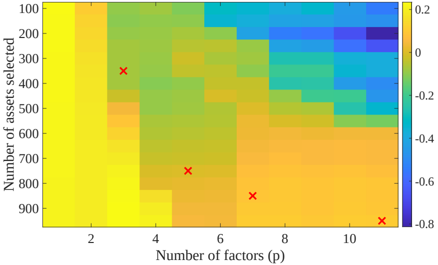


(m) LN #3

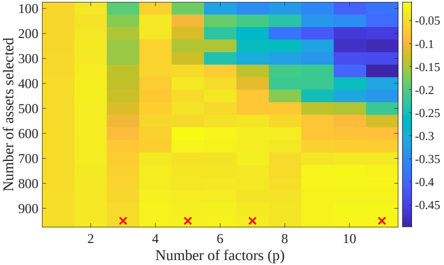


(n) Consumption

6. Empirical evidence



(o) Fin. Unc.



(p) Oil

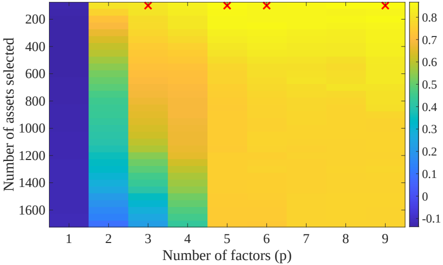
6. Empirical evidence: asset selection

	Factor #1		Factor #2		Factor #3	
	Asset	Corr	Asset	Corr	Asset	Corr
Mom	IntMom09	0.44	Mom12mOffSeason02	0.79	Mom12m08	0.64
	IntMom10	0.4	Mom12mOffSeason03	0.76	BMdec05	0.63
	MomVol10	0.37	Size01	0.74	IntMom03	0.63
	MomVol09	0.36	ResidualMomentum01	0.73	SP05	0.62
	IntMom08	0.36	ResidualMomentum02	0.73	ShareIss5Y05	0.62
	Mom12m10	0.36	NumEarnIncrease01	0.72	BookLeverage02	0.62
	FirmAgeMom05	0.35	ShareIss5Y01	0.7	cfp05	0.61
	Mom12mOffSeason10	0.34	MomVol03	0.69	BMdec04	0.61
	Mom12mOffSeason09	0.33	CompEquIss01	0.68	ShareIss1Y05	0.6
	Mom12m09	0.33	Mom12m03	0.68	LRreversal04	0.6
RMW	Industry:Gold	0.27	OperProf05	0.54	OperProfRD01	0.53
	MomOffSeason10	0.27	OperProfRD09	0.53	RoE01	0.47
	AccrualsBM02	0.27	CBOperProf09	0.5	GP01	0.45
	DelEqu05	0.27	RoE05	0.49	CBOperProf02	0.45
	LRreversal05	0.27	CBOperProf10	0.49	DolVol01	0.44
	roaq01	0.26	Leverage02	0.49	OperProfRD02	0.44
	AssetGrowth10	0.26	OperProfRD08	0.49	CBOperProf01	0.43
	DolVol05	0.25	realestate03	0.49	OperProf01	0.41
	ChEQ05	0.25	GP05	0.49	RoE02	0.4
	Price05	0.25	GP04	0.48	VolMkt02	0.4

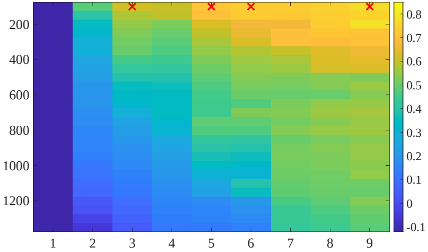
6. Empirical evidence: asset selection

	Factor #1		Factor #2		Factor #3	
	Asset	Corr	Asset	Corr	Asset	Corr
Liq.	InvGrowth06	0.47	InvGrowth06	0.28	InvGrowth06	0.3
	NetPayoutYield07	0.47	BetaFP09	0.26	DolVol01	0.27
	PayoutYield05	0.46	EntMult06	0.25	XFIN08	0.26
	PayoutYield07	0.46	NetPayoutYield07	0.24	MeanRankRevGrowth01	0.26
	BetaFP03	0.46	PayoutYield07	0.24	BetaFP03	0.25
	DelLTI02	0.46	PayoutYield05	0.24	ShortInterest01	0.25
	IntanBM03	0.46	cfp04	0.23	BetaFP09	0.24
	EntMult06	0.46	BetaFP10	0.23	EntMult06	0.24
	VolMkt04	0.46	XFIN08	0.23	PayoutYield07	0.24
	PayoutYield06	0.46	ShortInterest01	0.22	ChEQ04	0.23
Interm.	Industry:Banks	0.9	Industry:banks	0.76	Industry:banks	0.7
	Industry:Fin	0.84	Industry:Fin	0.56	Industry:Fin	0.47
	IntMom05	0.8	DelEqu02	0.46	DebtIssuance02	0.38
	EquityDuration04	0.8	grcapx3y02	0.44	NOA10	0.36
	IdioVolAHT05	0.8	OScore02	0.43	ChAssetTurnover04	0.35
	IdioVol3F05	0.79	GrLTNOA10	0.43	HerfAsset05	0.35
	MaxRet08	0.79	ChAssetTurnover04	0.43	ShareRepurchase01	0.35
	Illiquidity01	0.79	IntMom05	0.43	HerfBE05	0.35
	IdioRisk05	0.79	IdioVolAHT05	0.42	DelEqu05	0.32
	CBOperProf03	0.78	Tax01	0.42	Beta05	0.32

6. Empirical evidence: varying the universe of test assets

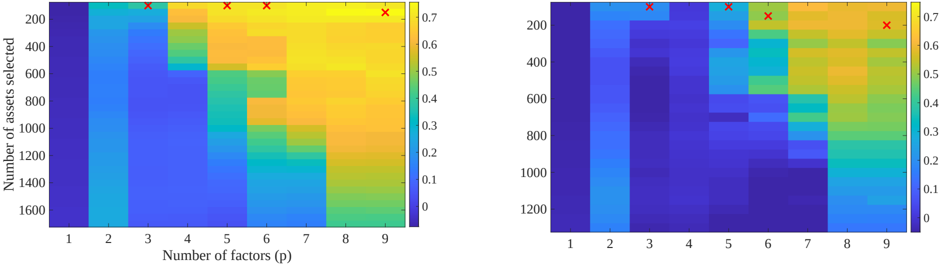


(q) Momentum



(r) Momentum w/o momentum test assets

6. Empirical evidence: varying the universe of test assets



(s) Profitability

(t) Profitability w/o profitability test assets

6. Additional applications

1. Performance evaluation of funds: Buffett's alpha
   - Measure Berkshire Hathaway alpha relative to latent factor model
   - Frazzini et al. (2013): alpha goes away when benchmarking to FF4 + BAB + QMJ
   - SPCA: alpha goes away when $p = 6$
   - No need to take a stand on the factors
   - Selected assets: idiosyncratic volatility sorts, profitability, leverage.

2. De-noise factors
   - Use SPCA to remove measurement error from FF5

	CZ		HXZ	
	no zero-beta	w/ zero-beta	no zero-beta	w/ zero-beta
FF5	19	19	12	12
Daniel et al. (2020)	39	17	35	13
SPCA (5 factors)	20	20	10	10
SPCA (7 factors)	17	17	11	11
SPCA (11 factors)	17	17	11	11

Conclusions

- SPCA gives a **robust** way to estimate risk premia for any nontradable factor
  - **Selection** step strengthens the factor of interest in the cross-section
  - Estimation of **latent factors** removes **omitted factor bias** concern
  - Alternation of the two helps pick **all** factors, strong and weak
- Combines insights from decades of asset pricing research, with a new approach to the weak factor issue
- When we are worried about weak factors, the solution is not to throw away the factor: it is to properly select the assets around that factor