IPCA Model

FinanceHub, 16/Nov/2020

Presented by:
Helder Palaro

Instrumented Principal Component Analysis*

Bryan Kelly — Yale, AQR, NBER
Seth Pruitt — Arizona State University
Yinan Su — Johns Hopkins University

July 30, 2020

Theory, estimation, asymptotic results, examples.

Characteristics *Are* Covariances: A Unified Model of Risk and Return*

Bryan Kelly — Yale University, AQR Capital Management, and NBER
Seth Pruitt — Arizona State University
Yinan Su — Johns Hopkins University

October 15, 2018

Model cross section of returns. Statistical tests of alpha, observable factors and characteristics.

Asset pricing for equities

Common approaches

Approach 1: risk factors treated as observable. Sorted portfolios based on factors. Alphas and betas estimated via regression. Example: Fama and French (1993)

Approach 2: risk factors treated as latent. PCA and betas are estimated from panel of returns. Example: Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986).

IPCA Model

$$r_{i,t+1} = \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}$$

$$\beta_{i,t} = z'_{i,t} \Gamma_\beta + \nu_{\beta,i,t}$$

N assets *r*
K factors *f*
L characteristics *z*
T data points *t*

As in Approach 1, risk premia are still determined by exposures to risk factors *f*, but as in Approach 2, these factors are considered latent.

Asset characteristics *z* (P/E, P/B, etc) serve as instrumental variables to the time-varying conditional loadings β.

The mapping Γ from characteristics *z* to loadings β is fixed over time and across individuals.

Benefits of the IPCA model

1) Researcher does not need to specify the risk factors a priori.

2) Individual loadings to each risk factor is an unnecessary excess (NxK parameters). IPCA requires only how characteristics map into their factor loadings through Γ. (LxK + TxK parameters).

3) Because of (2), a large number of assets *N* or characteristic predictors *L* can be handled.

4) The PCA approach lacks the flexibility to incorporate other data beyond returns, and can only accommodate static loadings.

Benefits of the IPCA model (cont.)

5) The IPCA estimator converges at $N^{1/2}$ faster than the PCA estimates.

6) Stocks evolve over time, moving from growth to value, for example. Standard response is to dynamically form portfolios, which gets more difficult with many characteristics. IPCA provides a more elegant solution. Betas are parameterized as function of characteristics.

7) It easily handles missing data (unbalanced panels).

Model estimation

$$r_{i,t+1} = z'_{i,t}\Gamma_\beta f_{t+1} + \epsilon^*_{i,t+1}$$

$$r_{t+1} = Z_t \Gamma_\beta f_{t+1} + \epsilon^*_{t+1},$$

$$\min_{\Gamma_\beta, F} \sum_{t=1}^{T-1} (r_{t+1} - Z_t\Gamma_\beta f_{t+1})' (r_{t+1} - Z_t\Gamma_\beta f_{t+1}) .$$

The values of f_{t+1} and Γ_β that minimize (5) satisfy the first-order conditions

$$\hat{f}_{t+1} = \left(\hat{\Gamma}'_\beta Z'_t Z_t \hat{\Gamma}_\beta\right)^{-1} \hat{\Gamma}'_\beta Z'_t r_{t+1}, \quad \forall t$$

and

$$\text{vec}(\hat{\Gamma}'_\beta) = \left(\sum_{t=1}^{T-1} Z'_t Z_t \otimes \hat{f}_{t+1}\hat{f}'_{t+1}\right)^{-1} \left(\sum_{t=1}^{T-1} \left[Z_t \otimes \hat{f}'_{t+1}\right]' r_{t+1}\right) .$$

Model estimation (cont.)

The optimization does not admit an analytical solution in general, but is speedily solved numerically by an alternating least squares (ALS) algorithm. It iterates between minimizing over Γ while holding $\{f_t\}$ fixed, and minimizing over $\{f_t\}$ while holding Γ fixed, until convergence.

Managed portfolios

PCA: $r_t = \beta f_t + \varepsilon_t$

$$\min_{\beta,F} \sum_{t=1}^{T} (r_t - \beta f_t)' (r_t - \beta f_t)$$

$$f_t = (\beta'\beta)^{-1} \beta' r_t$$

$$\max_{\beta} \text{ tr} \left(\sum_t (\beta'\beta)^{-1} \beta' r_t r_t' \beta \right)$$

Solution for β is given by the first K eigenvectors of $\sum_t r_t r_t'$.

IPCA: $r_t = \beta_t f_t + \varepsilon_t$

$$\max_{\Gamma_\beta} \text{ tr} \left(\sum_{t=1}^{T-1} (\Gamma_\beta' Z_t' Z_t \Gamma_\beta)^{-1} \Gamma_\beta' Z_t' r_{t+1} r_{t+1}' Z_t \Gamma_\beta \right)$$

Solution for β_t is given by the first K eigenvectors of $X'X = \sum_t x_t x_t'$.
where $x_{t+1} = \frac{Z_t' r_{t+1}}{N_{t+1}}$ are the L managed portfolios based on each of the L characteristics.

Asymptotic results

Assuming that the characteristics are orthogonal to errors (plus some technical conditions), then the ALS estimators are consistent (they converge in probability to the true Γ and $\{f_t\}$ when N, T $\rightarrow \infty$).

Assuming some technical conditions about convergence in distribution of characteristics and risk factors, then the ALS estimators are asymptotically normal.

Unrestricted IPCA model

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1},$$

$$\alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \nu_{\alpha,i,t}, \quad \beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t}.$$

Expected returns depend on characteristics in a way that is not explained by exposure to systematic risk.

Statistical test: $H_0 : \Gamma_\alpha = \mathbf{0}_{L\times 1}$ $H_1 : \Gamma_\alpha \neq \mathbf{0}_{L\times 1}.$

The null hypothesis states that alphas are unassociated with characteristics. Although idiosyncratic mispricings are still allowed.

Test statistic (Wald): $W_\alpha = \hat{\Gamma}'_\alpha \hat{\Gamma}_\alpha.$

Inference is conducted by bootstrap.

Testing observable factors

$$r_{i,t+1} = \beta_{i,t} f_{t+1} + \delta_{i,t} g_{t+1} + \epsilon_{i,t+1}.$$

$$\beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t} \qquad \delta_{i,t} = z'_{i,t}\Gamma_\delta + \nu_{\delta,i,t},$$

M observable factors *g*.

Statistical test: $H_0 : \Gamma_\delta = \mathbf{0}_{L\times M} \quad \text{vs.} \quad H_1 : \Gamma_\delta \neq \mathbf{0}_{L\times M}.$

Test statistics (Wald) : $W_\delta = \text{vec}(\hat{\Gamma}_\delta)'\text{vec}(\hat{\Gamma}_\delta).$

If we reject the null hypothesis, it means that the observable factors *g* hold explanatory power for the returns above and beyond the baseline IPCA factors.

Inference is also done by bootstrap.

Testing characteristic significance

Describe by its L columns: $\Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,L}]'$,

Statistical test:

$$H_0 : \Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,l-1}\ ,\ \mathbf{0}_{K\times 1}\ ,\gamma_{\beta,l+1}\ ,\ \dots\ ,\gamma_{\beta,L}]' \quad \text{vs.} \quad H_1 : \Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,L}]'.$$

If we reject the null hypothesis, it means that the l^{th} characteristic contributes for describing the expected returns variation through the IPCA risk factors.

Test statistic: (Wald) $W_{\beta,l} = \hat{\gamma}'_{\beta,l}\hat{\gamma}_{\beta,l}.$

Again, inference is conducted by bootstrap.

Evaluating model fit

$$\text{Total } R^2 = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1})\right)^2}{\sum_{i,t} r_{i,t+1}^2}$$

Describes how well the systematic factor risk model explains the variance of the stock panel returns (contemporaneously).

$$\text{Predictive } R^2 = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda})\right)^2}{\sum_{i,t} r_{i,t+1}^2}$$

Describes how well the systematic factor risk model predicts the returns for the next period.

Data

- Sample from July/1962 to May/2014.
- It includes 12,813 stocks trading on Nyse, Amex or Nasdaq.
- Cross-sectionally calculate stock ranks for each characteristic to avoid outliers.
- List of 36 characteristics:

beta (beta), assets-to-market (a2me), total assets (assets), sales-to-assets (ato), book-to-market (bm), cash-to-short-term-investment (c), capital turnover (cto), capital intensity (d2a), ratio of change in PP&E to change in total assets (dpi2a), earnings-to-price (e2p), fixed costs-to-sales (fc2y), cash flow-to-book (freecf), idiosyncratic volatility with respect to the FF3 model (idiovol), investment (invest), leverage (lev), market capitalization (mktcap), turnover (turn), net operating assets (noa), operating accruals (oa), operating leverage (ol), price-to-cost margin (pcm), profit margin (pm), gross profitability (prof), Tobin's Q (q), price relative to its 52-week high (w52h), return on net operating assets (rna), return on assets (roa), return on equity (roe), momentum (mom), intermediate momentum (intmom), short-term reversal (strev), long-term reversal (ltrev), sales-to-price (s2p), SG&A-to-sales (sga2s), bid-ask spread (bidask), and unexplained volume (suv).

Results

Table I
IPCA Model Performance

Note. Panel A and B report total and predictive R^2 in percent for the restricted ($\Gamma_\alpha = \mathbf{0}$) and unrestricted ($\Gamma_\alpha \neq \mathbf{0}$) IPCA model. These are calculated with respect to either individual stocks (Panel A) or characteristic-managed portfolios (Panel B). Panel C reports bootstrapped p-values in percent for the test of $\Gamma_\alpha = \mathbf{0}$.

		K					
		1	2	3	4	5	6
		Panel A: Individual Stocks (r_t)					
Total R^2	$\Gamma_\alpha = \mathbf{0}$	14.8	16.4	17.4	18.0	18.6	18.9
	$\Gamma_\alpha \neq \mathbf{0}$	15.2	16.8	17.7	18.4	18.7	19.0
Pred. R^2	$\Gamma_\alpha = \mathbf{0}$	0.35	0.34	0.41	0.42	0.69	0.68
	$\Gamma_\alpha \neq \mathbf{0}$	0.76	0.75	0.75	0.74	0.74	0.72
		Panel B: Managed Portfolios (x_t)					
Total R^2	$\Gamma_\alpha = \mathbf{0}$	90.3	95.3	97.1	98.0	98.4	98.8
	$\Gamma_\alpha \neq \mathbf{0}$	90.8	95.7	97.3	98.2	98.6	98.9
Pred. R^2	$\Gamma_\alpha = \mathbf{0}$	2.01	2.00	2.10	2.13	2.41	2.39
	$\Gamma_\alpha \neq \mathbf{0}$	2.61	2.56	2.54	2.51	2.50	2.46
		Panel C: Asset Pricing Test					
W_α p-value		0.00	0.00	0.00	0.00	2.06	52.1

Benchmark models

K = 1 factor : CAPM

K = 3 factors : Fama-French (1993), market, size, value

K = 4 factors : Carhart (1997) FF3 + momentum

K = 5 factors : Fama-French (2015) FF3 + investment and profitability

K = 6 factors : FF5 + momentum

Results (2)

	β_t	β
Latent Factors	IPCA (A)	PCA (D)
Observ Factors	IPCA-observ (C)	FF (B)

Test Assets	Statistic	K = 1	3	4	5	6
		Panel A: IPCA				
r_t	Total R^2	14.9	17.6	18.2	18.7	19
	Pred. R^2	0.36	0.43	0.43	0.70	0.70
	N_p	636	1908	2544	3180	3816
x_t	Total R^2	90.3	97.1	98.0	98.4	98.8
	Pred. R^2	2.01	2.10	2.13	2.41	2.39
	N_p	636	1908	2544	3180	3816
		Panel B: Observable Factors (no instruments)				
r_t	Total R^2	11.9	18.9	20.9	21.9	23.7
	Pred. R^2	0.31	0.29	0.28	0.29	0.23
	N_p	11452	34356	45808	57260	68712
x_t	Total R^2	65.6	85.1	87.5	86.4	88.6
	Pred. R^2	1.67	2.07	1.98	2.06	1.96
	N_p	37	111	148	185	222
		Panel C: Observable Factors (with instruments)				
r_t	Total R^2	10.4	14.2	15.3	14.7	15.6
	Pred. R^2	0.27	0.37	0.33	0.38	0.34
	N_p	37	111	148	185	222
x_t	Total R^2	66.9	87.2	89.5	88.3	90.3
	Pred. R^2	1.63	2.07	1.96	2.06	1.96
	N_p	37	111	148	185	222
		Panel D: Principal Components				
r_t	Total R^2	16.8	26.2	29.0	31.5	33.8
	Pred. R^2	< 0	< 0	< 0	< 0	< 0
	N_p	12051	36153	48204	60255	72306
x_t	Total R^2	88.4	95.5	96.7	97.3	97.9
	Pred. R^2	2.02	2.13	2.17	2.20	2.22
	N_p	636	1908	2544	3180	3816

Results (3)

Table III
IPCA Fits Including Observable Factors

Note. Panels A and B report total and predictive R^2 from IPCA specifications with various numbers of latent factors K (corresponding to columns) while also controlling for observable factors according to equation (14). Rows labeled 0, 1, 4, and 6 correspond to no observable factors or the CAPM, FFC4, or FFC6 factors, respectively. Panel C reports tests of the incremental explanatory power of each observable factor model with respect to the IPCA model. In all specifications, both latent and observable factor loadings are instrumented with observable firm characteristics. R^2's and p-values are in percent.

Observ. Factors	K = 1	2	3	4	5	6
			Panel A: Total R^2			
0	14.8	16.4	17.4	18.0	18.6	18.9
1	15.8	16.8	17.5	18.1	18.6	18.9
4	17.3	17.9	18.3	18.6	18.8	19.1
6	17.5	18.0	18.4	18.7	18.9	19.1
			Panel B: Predictive R^2			
0	0.35	0.34	0.41	0.42	0.69	0.68
1	0.35	0.40	0.41	0.50	0.69	0.68
4	0.45	0.66	0.67	0.66	0.71	0.69
6	0.50	0.66	0.67	0.66	0.67	0.69
			Panel C: Individual Significance Test p-value			
MKT–RF	26.1	91.8	84.4	60.7	49.7	46.5
SMB	2.97	2.26	2.28	1.92	1.32	1.36
HML	2.72	1.34	29.6	62.0	60.7	61.2
RMW	0.92	6.70	11.4	9.10	13.0	14.7
CMA	11.9	10.5	9.02	7.08	14.3	13.9
MOM	0.00	0.00	0.00	0.68	1.82	36.2

Results (4) – Out-of-sample

Table V
Out-of-sample Fits

Note. The table reports out-of-sample total and predictive R^2 in percent with recursive estimation scheme.

Test Assets	Statistic	K 1	2	3	4	5	6
r_t	Total R^2	13.9	15.3	16.3	16.9	17.5	17.8
	Pred. R^2	0.34	0.33	0.55	0.61	0.60	0.60
x_t	Total R^2	89.5	94.8	96.4	97.4	98.2	98.6
	Pred. R^2	2.21	2.15	2.42	2.44	2.42	2.42

Expanding window to estimate , factors are updated by relation:

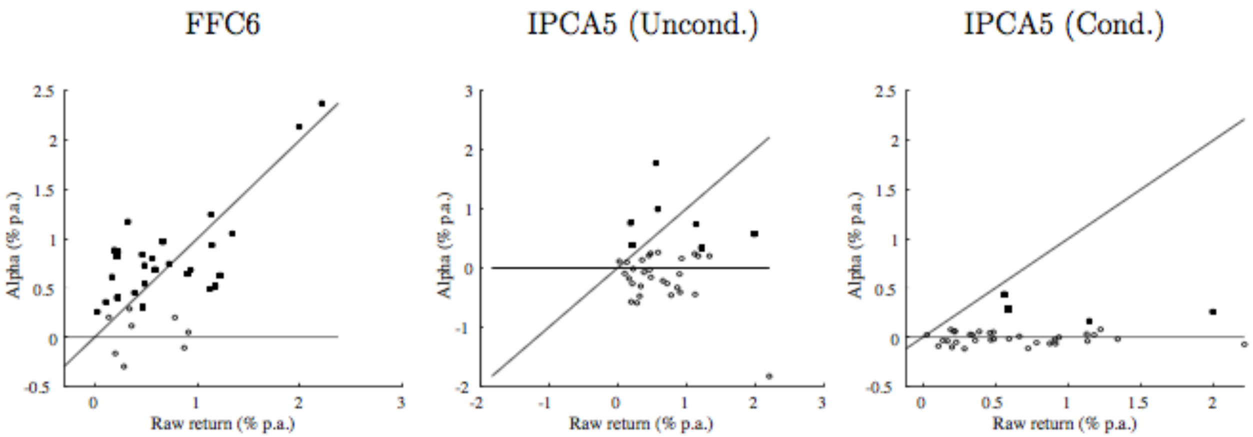
$$\hat{f}_{t+1,t} = \left(\hat{\Gamma}'_{\beta,t} Z'_t Z_t \hat{\Gamma}_{\beta,t}\right)^{-1} \hat{\Gamma}'_{\beta,t} Z'_t r_{t+1}$$

Efficiency – Intercept (Alpha)

Zero intercepts in a factor pricing model are equivalent to multivariate mean-variance efficiency of the factors.

Figure 1: Alphas of Characteristic-Managed Portfolios

Note. The left and middle panels report unconditional alphas for characteristic-managed portfolios (x_t), relative to FFC6 factors and five IPCA factors, respectively, estimated from time series regression. The right panel reports the time series averages of conditional alphas in the baseline five-factor IPCA model. Alphas are plotted against portfolios' raw average excess returns. Alphas with t-statistics in excess of 2.0 are shown with filled squares, while insignificant alphas are shown with unfilled circles.



Performance

Out-of-sample analysis
It does not consider costs, which can be a problem since IPCA tends to have higher turnover.

Table VI
Out-of-Sample Factor Portfolio Sharpe Ratios

Note. The table reports out-of-sample annualized Sharpe ratios for individual factors ("univariate") and for the mean-variance efficient portfolio of factors in each model ("tangency").

	K					
	1	2	3	4	5	6
			Panel A: IPCA			
Univariate	0.62	0.04	1.67	1.33	0.97	0.54
Tangency	0.62	0.62	2.49	3.09	3.89	4.05
			Panel B: Observable Factors			
Univariate	0.46	0.33	0.41	0.46	0.62	0.51
Tangency	0.46	0.51	0.78	1.01	1.29	1.37

Arbitrage portfolios

Uses the unrestricted version of IPCA:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1},$$

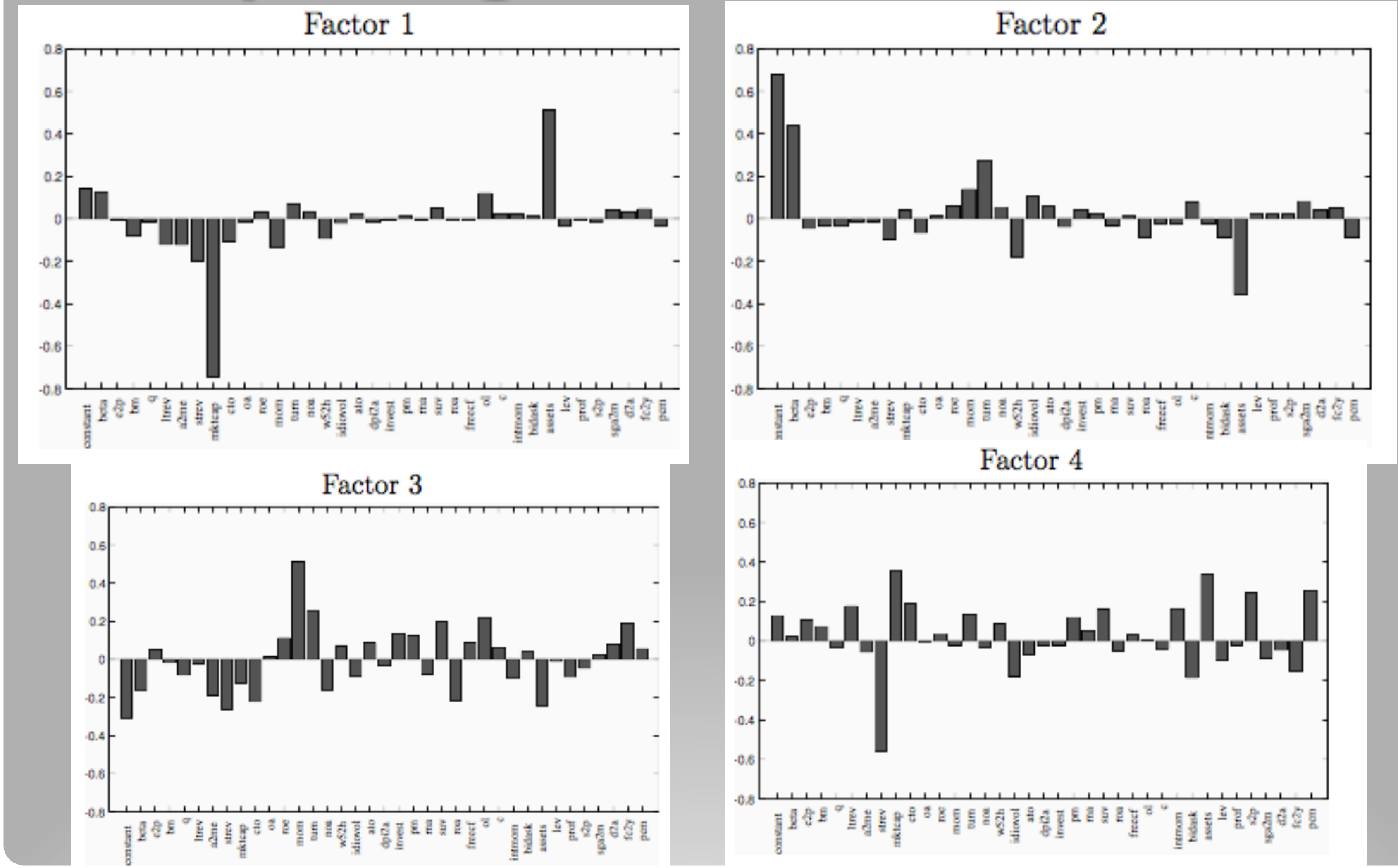
$$\alpha_{i,t} = z'_{i,t}\Gamma_{\alpha} + \nu_{\alpha,i,t}, \quad \beta_{i,t} = z'_{i,t}\Gamma_{\beta} + \nu_{\beta,i,t}.$$

Table VII
IPCA Pure-Alpha Portfolios

Note. The table reports out-of-sample annualized Sharpe ratios for a portfolio designed to exploit characteristic-based mispricing estimated from Γ_{α} in the unrestricted IPCA model.

	K					
	1	2	3	4	5	6
Sharpe Ratio	0.55	0.72	0.82	1.01	1.04	1.07

Interpreting IPCA Factors



Large and small stocks

Table VIII
IPCA Performance for Large versus Small Stocks

Note. Panel A and B report in-sample and out-of-sample total and predictive R^2 for subsamples of large and small stocks. We evaluate fits within each subsample using the same parameters (estimated from the unified sample of all stocks). All estimates use the restricted ($\Gamma_\alpha = \mathbf{0}$) IPCA specification.

		K					
		1	2	3	4	5	6
		Panel A: Large Stocks					
In-Sample	Total R^2	23.7	27.1	29.0	30.0	30.5	31.2
	Pred. R^2	0.32	0.31	0.40	0.43	0.56	0.53
Out-of-Sample	Total R^2	22.4	25.9	27.3	28.1	29.0	29.7
	Pred. R^2	0.40	0.32	0.46	0.52	0.41	0.39
		Panel B: Small Stocks					
In-Sample	Total R^2	14.7	15.8	17.0	17.5	18.1	18.3
	Pred. R^2	0.70	0.69	0.76	0.75	1.10	1.10
Out-of-Sample	Total R^2	14.7	15.7	16.9	17.5	17.9	18.2
	Pred. R^2	0.74	0.80	1.02	1.08	1.07	1.09

Table IX
Out-of-sample Tangency Sharpe Ratios, Large Versus Small Stocks

Note. The table repeats the analysis of Table VI for large and small stocks using parameters estimated separately in each subsample.

	K					
	1	2	3	4	5	6
Large	0.51	0.58	1.10	1.41	2.03	2.61
Small	0.64	1.25	2.76	2.82	4.15	4.19

Which characteristics matter?

Table XI
Individual Characteristic Contribution

Note. The table reports the contribution of each individual characteristic to overall model fit, defined as the reduction in total R^2 from setting all Γ_β elements pertaining to that characteristic to zero (in the restricted IPCA specification with $K = 5$). ** and * denote that a variable significantly improves the model at the 1% and 5% levels, respectively.

mktcap	2.84	**	roa	0.07		c	0.03
assets	1.64	**	suv	0.07	**	noa	0.03
beta	0.56	**	pcm	0.06	*	rna	0.02
strev	0.47	**	idiovol	0.05	**	invest	0.02
mom	0.33	**	s2p	0.05		prof	0.02
turn	0.31	**	bm	0.04		pm	0.02
w52h	0.14	**	bidask	0.04		d2a	0.01
a2me	0.14		intmom	0.03	*	dpi2a	0.01
cto	0.13		roe	0.03		q	0.01
ol	0.11		sga2m	0.03		freecf	0.01
ltrev	0.10	**	ato	0.03	*	lev	0.01
fc2y	0.08		e2p	0.03		oa	0.00

Model including only the 10 characteristics significant at 1% has similar performance to the full model.

Split Samples

Table XV
IPCA Cross-Validation for Split Samples

Note. The table reports total and predictive R^2 for 50-50 split samples using parameters estimated separately in each subsample. Rows correspond to the sample from which the Γ_β parameter is estimated and columns represent the sample in which fits are evaluated. In particular, when row and column labels differ, we are using fits in one sample (e.g., the first half) to cross-validate the reliability of parameters estimated in the other sample (e.g., the second half stocks). All estimates use the $K = 5$ restricted ($\Gamma_\alpha = \mathbf{0}$) IPCA specification. Panel A reports data split by time into first and second half of the sample and Panel B randomly splits the sample by CRSP permno.

Estimation Sample	Fit Sample: Total R^2		Fit Sample: Predictive R^2	
A. Time Split				
	Pre-1996	Post-1996	Pre-1996	Post-1996
Pre-1996	18.8	17.9	0.80	0.60
Post-1996	18.0	18.7	0.69	0.67
B. Random Split				
	A	B	A	B
A	18.3	18.4	0.69	0.68
B	17.8	18.8	0.67	0.68

Conclusion

- IPCA treats characteristics as instrumental variables for estimating dynamic loadings on latent factors.
- The estimator is easy to work as standard PCA, but it allows the research to bring information beyond returns.
- By estimating latent factors instead of a pre-specified observable factors, IPCA successfully describes the variation of stock returns and risk compensation. The model does that parsimoniously (low dimension).
-The model outperforms leading observable factor models, in-sample and out-of-sample
- Only a subset of stock characteristics are responsible for IPCA empirical success.
- The authors introduced a set of statistical asset pricing tests. When researches encounters a new anomaly characteristics, they can test whether it contributes as a risk factor loading or as an anomaly alpha.