Characteristics Are Covariances: A Unified Model of Risk and Return

Bryan T. Kelly & Seth Pruitt & Yinan Su

Journal of Financial Economics. 2019.Dec.

吕漫妮

2020年3月7日

Contents

the Euler equation for investment returns

$$E_t[m_{t+1}r_{i,t+1}] = 0 \quad \Leftrightarrow \quad E_t[r_{i,t+1}] = \underbrace{\frac{Cov_t(m_{t+1}, r_{i,t+1})}{Var_t(m_{t+1})}}_{\beta_{i,t}} \underbrace{\left(-\frac{Var_t(m_{t+1})}{E_t[m_{t+1}]}\right)}_{\lambda_t}.$$

cov(X,Y) = E(XY) - E(X)E(Y)

系统风险暴露，因子风险价格

Discount factor is linear in factors.

Why different assets earn different average returns?

$$r_{i,t+1} = \alpha_{i,t} + \beta'_{i,t} f_{t+1} + \epsilon_{i,t+1}$$

The factors and loadings are unobservable, how?

Treats these factors as observable by the econometrician, and estimates betas and alphas via regression e.g. Fama and French (1993).

Treats risk factors as latent and use factor analytic techniques, such as PCA, to estimate the factors and betas, e.g. Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986).

Models stock returns as a function of many characteristics at once

Instrumented Principal Components Analysis (IPCA) allows factor loadings to partially depend on observable asset characteristics that serve as instrumental variables.

Introduction

➢Central Motivation:

To build a model and estimator that admits the possibility that characteristics line up with average returns because they proxy for loadings on common risk factors.

➢Features

1.Excellent ability to jointly evaluate large numbers of characteristic predictors with minimal computational burden.

2.Has no presumption that the researcher can correctly specify the exact factors a priori.

3.Nest traditional, pre-specified factors within a more general IPCA specification.

Model and Estimation

➢ General Model Specification:

The general IPCA model specification for an excess return $r_{i,t+1}$ is

→ 动态因子载荷

K维潜在因子向量

$1 \times L$ $L \times K$

包含了可观测到的资产特征 → 工具向量

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \quad (3)$$

映射（大量特征→α, β）

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

误差部分（与工具变量不相关）

- Instrumenting the estimation of latent factor loadings with observable characteristics allows additional data to shape the factor model (PCA solely from returns data)
- Incorporating time-varying instruments makes it possible to estimate dynamic factor loadings, which is valuable when one seeks a model of conditional return behavior.

Model and Estimation

- General Model Specification:

The general IPCA model specification for an excess return $r_{i,t+1}$ is

动态因子载荷

K维潜在因子向量

L×K

1×L

包含了可观测的资产特征 → 工具向量

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \quad (3)$$

映射（大量特征 → α, β）

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

正交部分（与工具变量不相关）

- Estimation of gamma_beta amounts to finding a few linear combinations of candidate characteristics that best describe the loading structure (isolates the signal and averages out the noise)(a global mapping shared by all stocks in all periods).
- Gamma_beta also allows us to confront the challenge of migrating assets when standard response approach becomes infeasible because several characteristics is required for an adequate description of an asset's identity (migration in the asset's identity is tracked through its betas).

Model and Estimation

➢ General Model Specification:

动态因子载荷

The general IPCA model specification for an excess return $r_{i,t+1}$ is

K维潜在因子向量

1×L　　　　　　　　　　　　　　　L×K

包含了可观测的资产特征 → 工具向量

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \tag{3}$$

映射(大量特征→α,β)

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

误差部分(与工具变量不相关)

- Null hypothesis: Characteristics don't proxy for alpha→restrict gamma_alpha to zero
- The structure of alpha_i,t is a linear combination of instruments
- IPCA estimates it by finding the linear combination of characteristics (with weights given by gamma_alpha) that best describes conditional expected returns while controlling for the role of characteristics in factor risk exposure.

Model and Estimation

➢ General Model Specification:

动态因子载荷

The general IPCA model specification for an excess return $r_{i,t+1}$ is

K维潜在因子向量

1×L

包含了可观测的资产特征 → 工具向量

L×K

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \quad (3)$$

映射（大量特征 → α, β）

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

误差部分（与工具变量不相关）

- “Risk”, and systemic risk in particular, refers to correlated fluctuations in asset returns arising from the common return factors f.
- “Anomaly”, it simply will refer to a non-zero intercept alpha.
- The number of factors, K, is small, imposing a view that the empirical content of an asset pricing model is its parsimony in describing sources of systematic risk.
- Let the number of instruments, L, to be potentially large.

Model and Estimation

➢ Restricted Model - Specification:

- Kelly, Pruitt, and Su (2017) derive the IPCA estimator, prove IPCA consistently estimates model parameters and latent factors as the number of assets and the time dimension simultaneously grow large, as long as factors and residuals satisfy weak regularity conditions.
- Specification for a return： $r_{i,t+1} = z_{i,t}' \Gamma_\beta f_{t+1} + \epsilon^*_{i,t+1}$

where $\epsilon^*_{i,t+1} = \epsilon_{i,t+1} + \nu_{\alpha,i,t} + \nu_{\beta,i,t} f_{t+1}$ is a composite error.[14]

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

where r_t+1 is an N×1 vector of individual firm returns, Z_t is the N×L matrix that stacks the characteristics of each firm, epsilon stacks individual firm residuals. This restriction maintains that characteristics explain expected returns only insofar as they proxy for systematic risk exposures.

Model and Estimation

➢ Restricted Model - Derivation:

- Estimation objective: minimize the sum of squared composite model errors:

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

- The values of f_t+1 and gamma_beta satisfy the first-order conditions:

$$\hat{f}_{t+1} = \boxed{\left(\hat{\Gamma}'_\beta Z'_t Z_t \hat{\Gamma}_\beta\right)^{-1} \hat{\Gamma}'_\beta Z'_t} r_{t+1}, \quad \forall t \tag{6}$$

$$\rightarrow [(\hat{\Gamma}_\beta' Z_t')(Z_t \hat{\Gamma}_\beta)]^{-1} \hat{\Gamma}_\beta' Z_t' = (Z_t' \hat{\Gamma}_\beta)^{-1} (\hat{\Gamma}_\beta' Z_t')^{-1} \hat{\Gamma}_\beta' Z_t' = (Z_t' \hat{\Gamma}_\beta)^{-1} \Rightarrow \beta^{-1}$$

and

向量化函数　张量积

$$\mathrm{vec}(\hat{\Gamma}'_\beta) = \left(\sum_{t=1}^{T-1} Z'_t Z_t \otimes \hat{f}_{t+1} \hat{f}'_{t+1}\right)^{-1} \left(\sum_{t=1}^{T-1} \left[Z_t \otimes \hat{f}'_{t+1}\right]' r_{t+1}\right). \tag{7}$$

It has no closed-form solution and must be solved numerically (via alternating least squares in a matter of seconds even for high dimension systems).
Impose some assumptions to indentify gamma_beta and f (given that a non-singular K-dimensional rotation matrix R).

Model and Estimation

➢ Restricted Model - A Managed Portfolio Interpretation:

- Static factor model such as: $r_t = \beta f_t + \epsilon_t$
- Objective function in the static case is: $\min_{\beta,F} \sum_{t=1}^{T} (r_t - \beta f_t)' (r_t - \beta f_t)$
- The first-order condition for f_t: $f_t = (\beta'\beta)^{-1} \beta' r_t$
- Objective function for beta: $\max_{\beta} \operatorname{tr}\left(\sum_t (\beta'\beta)^{-1} \beta' r_t r_t' \beta\right)$
- Substitute the IPCA first-order condition into the original objective:

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

Model and Estimation

➢ Restricted Model - A Managed Portfolio Interpretation:

- Define *x* as L×1 vector: $x_{t+1} = \frac{Z_t' r_{t+1}}{N_{t+1}}$ is time t + 1 realization of returns on a set of L managed portfolios, and T×L matrix $X = [x_1, ..., x_T]'$

Prior empirical work shows that there tends to be a high degree of common variation in anomaly portfolios. IPCA recognizes this and estimates factors and loadings by focusing on the common variation in X. It estimates factors as the K linear combinations of X's columns, or "portfolios of portfolios".

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

$$\boxed{Z_t' Z_t} \leftarrow \boxed{T^{-1} \sum_t Z_t' Z_t}$$

1.gamma_beta equals to the first K eigenvectors of the sample second moment matrix of managed portfolio returns, $X'X = \sum_t x_t x_t'$
2.the estimates of f_t+1 would be the first K principal components of the managed portfolio panel

starting guess to initialize the numerical optimization

Model and Estimation

- Restricted Model – Compared with PCA:
- Factors that the explained covariation is re-weighted across time and across assets to emphasize observations associated with the most informative instruments. $\boxed{Z_t' Z_t}$
- Restricted Model – a generalization of period-by-period cross section regressions (Fama and MacBeth (1973) and Rosenberg (1974)):
- Let K = L
- When K < L, choose a reduced-rank set of regressors——the K < L combinations of characteristics that best fit the cross section regression.

Model and Estimation

➢ Unrestricted Model - Specification & Derivation :

- Intercepts are a linear combination of instruments with weights defined by the L × 1 parameter vector gamma_alpha: $r_{i,t+1} = z'_{i,t}\Gamma_\alpha + z'_{i,t}\Gamma_\beta f_{t+1} + \epsilon^*_{i,t+1}$ i.e.

  $r_{i,t+1} = z'_{i,t}\tilde{\Gamma}\tilde{f}_{t+1} + \epsilon^*_{i,t+1}$, where $\tilde{\Gamma} \equiv [\Gamma_\alpha, \Gamma_\beta]$ and $\tilde{f}_{t+1} \equiv [1, f'_{t+1}]'$.

- The first-order conditions:

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

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

Impose identification assumption $\Gamma'_\alpha\Gamma_\beta = 0_{1\times K}$ to identify gamma_alpha and gamma_beta. Of the total return predictability possessed by the instruments, only the orthogonal residual left unexplained by factor loadings is assigned to the intercept (allow risk loadings to explain as much of assets‘ mean returns as possible).

Asset Pricing Tests

➢ Testing the Zero Alpha:

- Test the zero alpha condition that distinguishes the restricted and unrestricted models.
- Model equation: $r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}$
- The null hypothesis: $H_0 : \Gamma_\alpha = \mathbf{0}_{L \times 1}$
- The Wald-type test statistic: $W_\alpha = \hat{\Gamma}'_\alpha \hat{\Gamma}_\alpha$
- Bootstrap proceeding:

$$\boxed{x_{t+1} = Z_t' r_{t+1} = (Z_t' Z_t)\Gamma_\alpha + (Z_t' Z_t)\Gamma_\beta f_{t+1} + Z_t' \epsilon^*_{t+1}}$$

1. Estimate the unrestricted model and retain the estimated parameters and the residuals fitted value.

2. For b = 1 to 1000, we generate the bth bootstrap sample of returns.

3. Re-estimate the unrestricted model, record the estimated test statistic, and calculate a p-value as the fraction of bootstrapped statistics $\tilde{W}^b_\alpha$ that exceed the one from the actual data

$$\boxed{\hat{\Gamma}_\alpha, \hat{\Gamma}_\beta, \text{ and } \{\hat{f}_t\}_{t=1}^T}$$

$$\downarrow$$

$$\boxed{\{\hat{d}_t\}_{t=1}^T}$$

$$\downarrow$$

$$\boxed{\tilde{x}^b_{t+1} = (Z_t' Z_t)\hat{\Gamma}_\beta \hat{f}_{t+1} + \tilde{d}^b_{t+1}, \quad \tilde{d}^b_{t+1} = q^b_{1,t+1} \hat{d}_{q^b_{2,t+1}}}$$

$$\downarrow$$

$$\boxed{\tilde{W}^b_\alpha = \tilde{\Gamma}^{b\prime}_\alpha \tilde{\Gamma}^b_\alpha}$$

Asset Pricing Tests

➢ Testing Observable Factor Models Versus IPCA:

- Whether observable factors (such as the FF5 model) significantly improve the model's description of the panel of asset returns while controlling for IPCA factors?
- Model: $r_{i,t+1} = \beta_{i,t} f_{t+1} + \delta_{i,t} g_{t+1} + \epsilon_{i,t+1}$ with $\delta_{i,t} = z'_{i,t}\Gamma_\delta + \nu_{\delta,i,t}$

  i.e. $r_{i,t+1} = z'_{i,t}\tilde{\Gamma}\tilde{f}_{t+1} + \epsilon^*_{i,t+1}$, where $\tilde{\Gamma} \equiv [\Gamma_\beta, \Gamma_\delta]$ and $\tilde{f}_{t+1} \equiv [f'_{t+1}, g'_{t+1}]'$
- The first-order conditions:

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

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

- The hypotheses: $H_0 : \Gamma_\delta = \mathbf{0}_{L\times M}$ vs. $H_1 : \Gamma_\delta \neq \mathbf{0}_{L\times M}$
- The Wald-type test statistic: $W_\delta = \text{vec}(\hat{\Gamma}_\delta)'\text{vec}(\hat{\Gamma}_\delta)$
- Bootstrap proceeding is the same as above.

Asset Pricing Tests

➢ Testing Instrument Significance:

- Test the incremental significance of an individual characteristic or set of characteristics while simultaneously controlling for all other characteristics.
- Partition the parameter matrix as: $\Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,L}]'$
- Hypotheses $H_0 : \Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,l-1}\ ,\ \mathbf{0}_{K\times 1}\ ,\gamma_{\beta,l+1}\ ,\ \dots\ ,\gamma_{\beta,L}]'$ vs. $H_1 : \Gamma_\beta = [\gamma_{\beta,1}\ ,\ \dots\ ,\gamma_{\beta,L}]'$
- The Wald-type statistic: $W_{\beta,l} = \hat{\gamma}'_{\beta,l}\hat{\gamma}_{\beta,l}$
- Bootstrap proceeding is the same as above.

Empirical Findings

- Data:

- Stock returns and characteristics sample: Freyberger, Neuhierl, and Weber (2017), begins in July 1962, ends in May 2014, and includes 12,813 firms. For each firm we have 36 characteristics.
- We restrict attention to i, t observations for which all 36 characteristics are non-missing, yielding 1,403,544 stock-month observations for 12,813 unique stocks.
- We cross-sectionally transform instruments period-by-period. In particular, we calculate stocks' ranks for each characteristic, then divide ranks by the number of non-missing observations and subtract 0.5. This maps characteristics into the [-0.5,+0.5] interval and focuses on their ordering as opposed to magnitude(results are qualitatively unchanged with no characteristic transformation).

Empirical Findings

➢ Measure Model Performance - In-sample Estimation:

- Describing the common variation in realized returns:

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

It includes the explained variation due to contemporaneous factor realizations and dynamic factor exposures, aggregated over all assets and time periods.

- Describe risk compensation:

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

It measures the accuracy of model-implied conditional expected returns.

Without further model structure, IPCA does not separately identify risk price dynamics. Hence, we hold estimated risk prices constant and predictive information enters return forecasts only through the instrumented loadings.

➢Model Performance:

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

➢Comparison with Existing Models:

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

➢Whether the inclusion of observable factors in the matched individual stock sample?

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

➢Comparison with Existing Models-Other Observable Factors:

Table IV
Other Observable Factors

Note. The table repeats the analysis of Table II for the Hou, Xue, and Zhang (HXZ, 2015), Stambaugh and Yuan (SY, 2017), and Barillas and Shanken (BS, 2018) factor models. Parameter counts in Panels C and D include latent factor realizations.

Test Assets	Statistic	Observable Factors HXZ	SY	BS
		Panel A: No instruments		
r_t	Total R^2	20.5	19.7	23.7
	Pred. R^2	0.18	0.37	0.14
	N_p	45804	45808	68706
		Panel B: With instruments		
r_t	Total R^2	14.7	14.4	15.6
	Pred. R^2	0.32	0.41	0.38
	N_p	148	148	222
		Panel C: With instruments and one IPCA factor		
r_t	Total R^2	16.9	17.0	17.5
	Pred. R^2	0.53	0.52	0.55
	N_p	747	784	821
		Panel D: With instruments and five IPCA factors		
r_t	Total R^2	18.8	18.8	18.9
	Pred. R^2	0.68	0.68	0.71
	N_p	3143	3328	3217

➢Out-of-sample Fits:

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

Recursive backward-looking estimation:

In every month t ≥120, we use all data through t to estimate the IPCA model and denote the parameter estimate $\hat{\Gamma}_{\beta,t}$, calculate the out-of-sample realized factor return at t + 1 as $\hat{f}_{t+1,t} = \left(\hat{\Gamma}'_{\beta,t} Z'_t Z_t \hat{\Gamma}_{\beta,t}\right)^{-1} \hat{\Gamma}'_{\beta,t} Z'_t r_{t+1}.$

The out-of-sample total R2 compares r_{t+1} to $Z_t \hat{\Gamma}_{\beta,t} \hat{f}_{t+1,t}$ and x_{t+1} to $Z'_t Z_t \hat{\Gamma}_{\beta,t} \hat{f}_{t+1,t}$
The out-of-sample predictive R2 is defined analogously, replacing $\hat{f}_{t+1,t}$ with the factor mean through t.

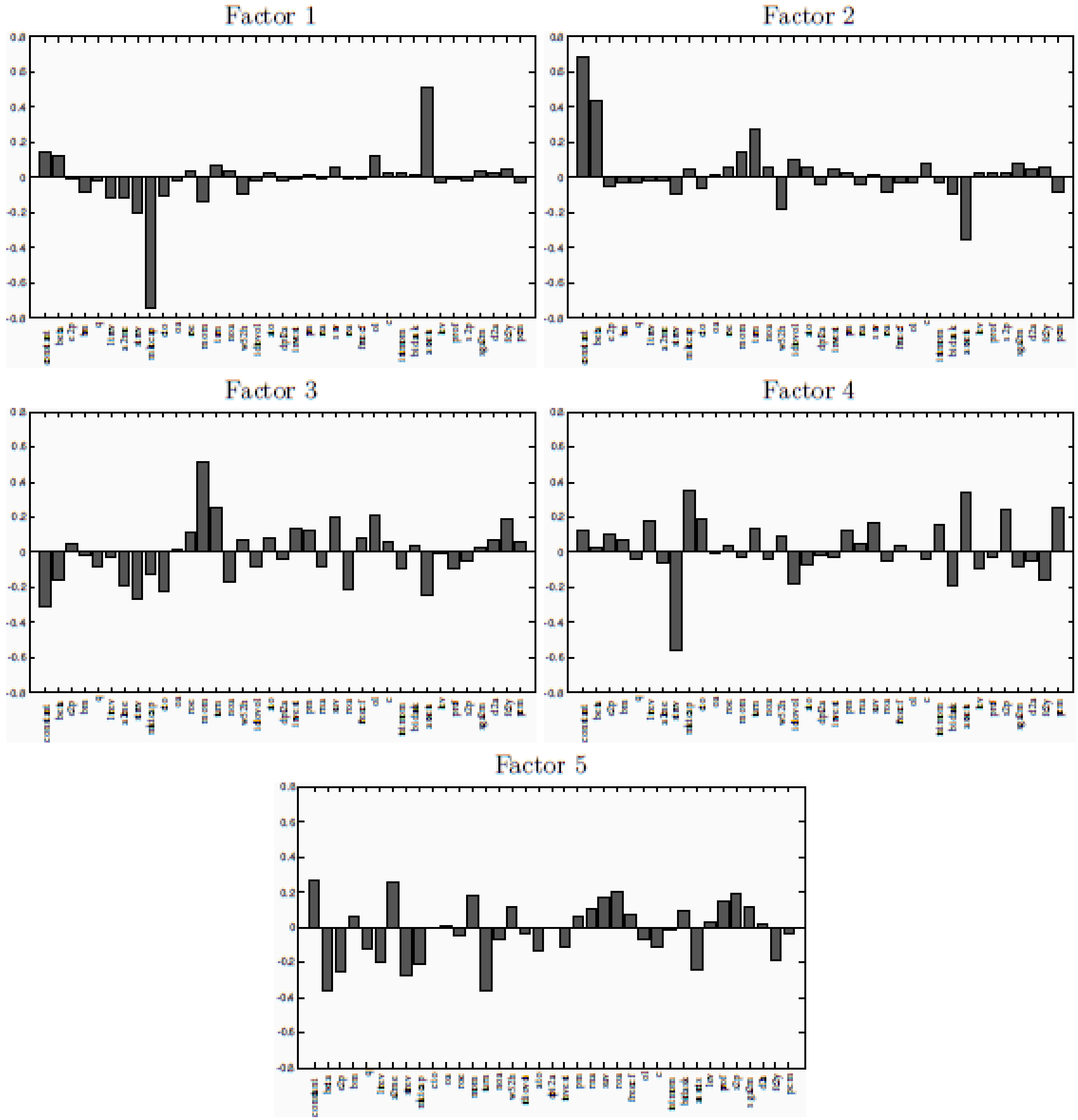
➢Unconditional Mean-Variance Efficiency:

- Zero intercepts in a factor pricing model are equivalent to multivariate mean-variance efficiency of the factors.
- Ours is a conditional asset pricing model——factor loadings are parameterized functions of conditioning instruments. It is attractive because it often maps directly to decisions of investors.
- Conditional models are more econometrically challenging than unconditional models because they typically require the researcher to estimate a dynamic model and take a stand on investors‘ conditioning information, leading the factor model literature to focus predominantly on unconditional estimation and testing.

➢Interpreting IPCA Factors:

Figure 2: Γ_β Coefficient Estimates

Note. The figure reports each column of the estimated Γ_β coefficient matrix from the $K = 5$ IPCA specification.



➢Large Versus 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

➢Annual Returns:

Table X
Annual Returns

Note. The table repeats the analysis of Table I using annual rather than monthly returns.

Test Assets	Statistic	K 1	2	3	4	5	6
		Panel A: In-sample					
r_t	Total R^2	15.8	18.4	19.4	20.2	20.6	21.0
	Pred. R^2	2.89	3.05	3.09	3.04	3.04	3.12
x_t	Total R^2	88.5	94.0	96.8	97.9	98.6	99.0
	Pred. R^2	15.7	16.2	16.4	16.2	16.2	16.2
		Panel B: Out-of-sample					
r_t	Total R^2	15.0	17.4	18.5	19.2	19.6	20.0
	Pred. R^2	3.05	3.32	3.40	3.23	3.30	3.26
x_t	Total R^2	87.8	94.1	96.7	98.1	98.6	98.9
	Pred. R^2	19.0	19.8	19.8	19.3	19.5	19.3

➢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	

➢Does a small subset of characteristics produce a factor model with similar explanatory power?

Table XII
IPCA Fits Excluding Insignificant Instruments

Note. IPCA percentage R^2 at the individual stock level including only the 10 characteristics from Table XI that are significant at the 1% level.

		K					
		1	2	3	4	5	6
Total R^2	$\Gamma_\alpha = \mathbf{0}$	14.6	16.1	17.1	17.7	18.2	18.4
	$\Gamma_\alpha \neq \mathbf{0}$	15.0	16.4	17.4	18.0	18.2	18.4
Pred. R^2	$\Gamma_\alpha = \mathbf{0}$	0.34	0.34	0.41	0.43	0.61	0.62
	$\Gamma_\alpha \neq \mathbf{0}$	0.67	0.66	0.66	0.66	0.65	0.65

➢Gauge stability

	Baseline	Large	Small	$K = 4$	$K = 6$	Annual
mktcap	**	**	**	**	**	**
assets	**	**	**	**	**	**
beta	**	**	**	**	**	**
strev	**		**		**	
mom	**	**	*		**	*
turn	**		**	**	**	
w52h	**	*	**	*	**	
a2me						
cto				*	*	
ol		*				
ltrev	**	**	**	**	**	**
fc2y		**			**	**
roa						
suv	**		**		**	
pcm	*	**			**	
idiovol	**	*			*	*
s2p						
bm						
bidask						
intmom	*			*	*	
roe						
sga2m		*				
ato	*		*	*	*	
e2p			*			
c						
noa						*
rna						
invest						
prof						
pm		**			**	*
d2a		*			*	
dpi2a						*
q						
freecf						
lev						*
oa						

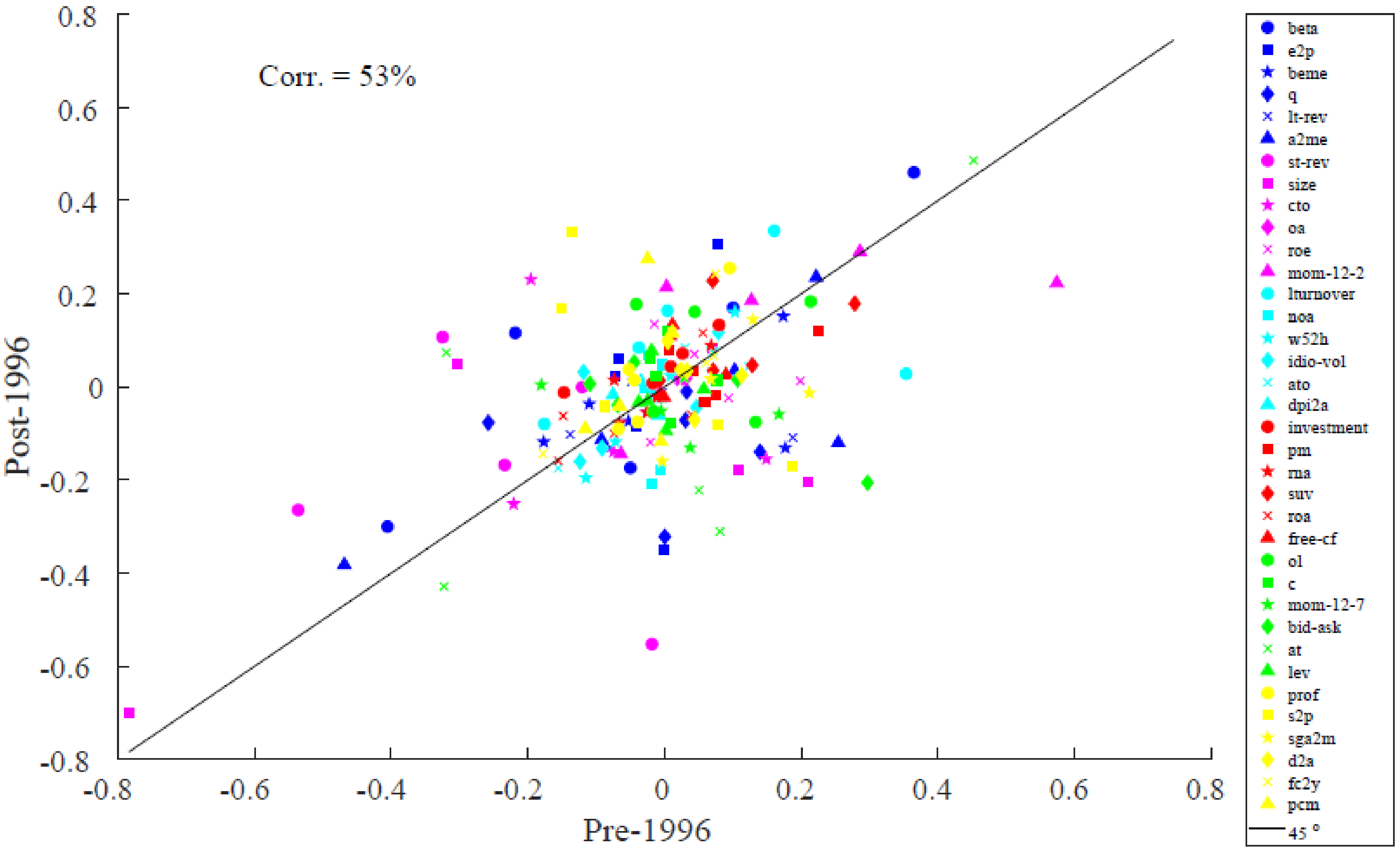
➢Static or Dynamic Loadings?

- Characteristics vary both in the cross section and over time. Do these two dimensions aid IPCA's estimation of the factor model in different ways?

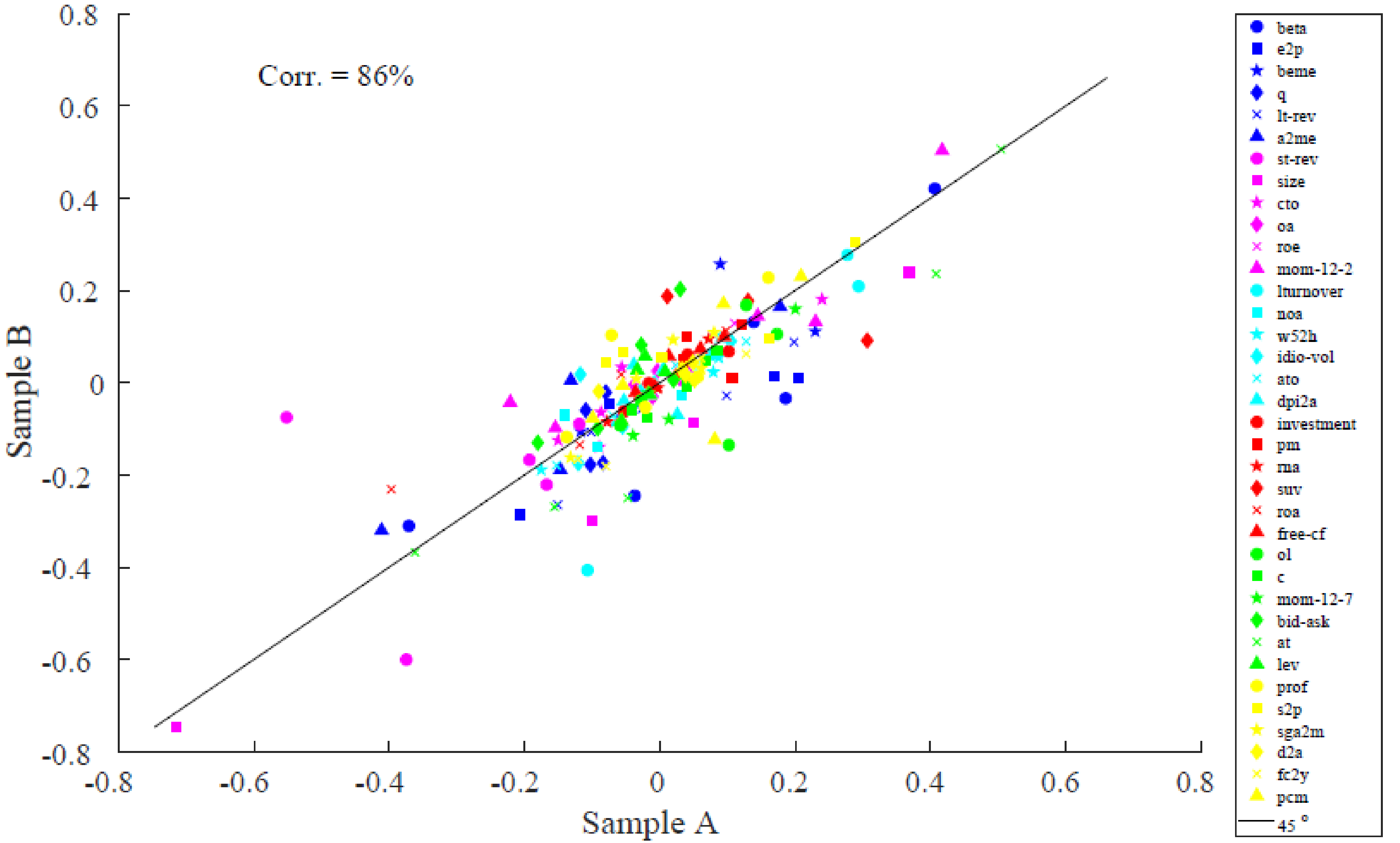
$$z_{i,t} = [1,\ \bar{c}'_{i,t},\ (c_{i,t} - \bar{c}_{i,t})']', \quad \text{where} \quad \bar{c}_{i,t} \equiv \frac{1}{t}\sum_{\tau=1}^{t} c_{i,\tau}$$

Estimation Sample	Fit Sample			
	Total R^2		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

➢Split Samples



➢Split Samples



Thanks!