

OTDK-dolgozat

2025

Neumann János Egyetem

**THE ROLE OF FIRM CHARACTERISTICS IN EXPLAINING STOCK
RETURNS FOR DIFFERENT INDUSTRIES**

**A VÁLLALATI KARAKTERISZTIKÁK SZEREPE A
RÉSZVÉNYHOZAMOK MAGYARÁZATÁBAN A KÜLÖNBÖZŐ
IPARÁGAK ESETÉBEN**

KANYÁRI BENCE PÉTER
NEMZETKÖZI GAZDASÁG ÉS GAZDÁLKODÁS MSC

Konzulens: Nagy Olivér

Kézirat lezárásának dátuma: 2024.10.26.

ABSTRACT

This paper examines whether different firm characteristics drive asset prices in various industries in the US stock market. The study uses the Instrumented Principal Component Analysis (IPCA), which utilizes observable firm characteristics as instruments to estimate latent factor loadings. The IPCA is estimated separately for each industry group, and its performance is analyzed. The paper also determines the importance of firm characteristics in explaining stock returns in each industry group. The data set used in the analysis includes 12,813 firms across 11 industry groups from July 1962 to May 2014 and consists of 36 distinct firm characteristics. The study finds that estimating IPCA by industry groups improves its ability to describe the common variation of returns, confirming the homogeneity of stocks within the same industry. However, it demonstrates varying effectiveness in explaining the differences in average returns. The research highlights the importance of market capitalization and total assets across all industries but also demonstrates that the importance of other firm characteristics varies across different industry groups. Overall, the findings emphasize the role of industry context in asset pricing and contribute to a better understanding of how firm characteristics influence expected returns.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. METHODOLOGY	4
2.1. THE IPCA MODEL	4
2.1.1. <i>The managed portfolios interpretation</i>	5
2.2. MEASURING PERFORMANCE	5
2.3. CHARACTERISTICS IMPORTANCE.....	7
3. EMPIRICAL RESULTS	8
3.1. DATA.....	8
3.2. IPCA PERFORMANCE	12
3.2.1. <i>In-sample performance</i>	12
3.2.2. <i>Out-of-sample performance</i>	15
3.2.3. <i>Parameter stability</i>	19
3.3. WHICH CHARACTERISTICS MATTER?.....	20
3.3.1. <i>Robustness</i>	22
4. CONCLUSION.....	25
REFERENCES.....	26

LIST OF TABLES AND FIGURES

Table 1: List of Characteristics.....	9
Table 2: Summary Table of Characteristics.....	10
Table 3: Restrictive model's Total R^2	13
Table 4: Unrestrictive model's Total R^2	13
Table 5: Restrictive model's Predictive R^2	14
Table 6: Unrestrictive model's Predictive R^2	15
Table 7: Restrictive model's Out-of-sample Total R^2	16
Table 8: Restrictive model's Out-of-sample Predictive R^2	16
Table 9: Out-of-sample Annualized Univariate Sharpe Ratios	18
Table 10: Out-of-sample Annualized Tangency Sharpe Ratios.....	18
Figure 1: Parameter Stability	19
Figure 2: Characteristics Importance	20
Figure 3: Top 10 Characteristics by Variable Importance.....	21
Figure 4: Correlation of Characteristic Importance	22
Figure 5: Time Series Average Characteristics Importance.....	23
Figure 6: Top 10 Characteristics by Time Series Average of Variable Importance	24

1. INTRODUCTION

The major challenge in asset pricing is understanding why different assets yield different returns. According to the arbitrage pricing theory (Chamberlain, 1983; Ross, 1976), the cross-section of expected returns should be explained by exposure to systematic risk factors. Many studies have used firm characteristics, such as size or momentum, to proxy exposure to these non-diversifiable systemic risk factors (see, for instance, Carhart, 1997; Fama & French, 1993, 2015). As a result, a range of firm characteristics has been identified that can predict stock returns (Goyal, 2012; Green et al., 2017; Harvey et al., 2016; Mclean & Pontiff, 2016), leading to a "zoo" of new variables (Cochrane, 2011). However, the challenge lies in determining which characteristics are truly driving expected returns and which are merely noise.

The Instrumented Principal Component Analysis (IPCA) introduced by Kelly et al. (2019), allows us to determine which characteristics are most important in explaining expected returns. The IPCA is a latent factor model that estimates factor loadings using a large number of observable firm characteristics as instruments. Therefore, the application of IPCA has several advantages. First, due to the direct observability of the characteristics used as instruments, latent factors can be estimated more efficiently. Second, conditional betas incorporate firm characteristics, allowing firms to switch identities. Third, and most importantly, we can determine to what extent each firm characteristic can explain the variation in expected returns.

In my paper, I look into whether asset prices are driven by different characteristics in various industries in the US stock market. Additionally, I analyze the performance of the IPCA model by industry when its parameters are estimated separately for each industry group. The IPCA is not capable of directly handling data that is structured hierarchically. Therefore, when applied to all stocks, it can only capture broad, market-wide risk drivers and may overlook more subtle, industry-specific risk factors. However, if IPCA is estimated separately for each industry group, it can more accurately capture these industry-specific risk factors. As a result, it can be more effective in describing the drivers of stock returns.

In theory, an industry refers to a group of firms that produce homogenous products or close substitutes (Chou et al., 2012). Therefore, firms within the same industry tend to have higher return comovements due to their shared common fundamentals. If this is true, the IPCA should better describe the common variation of expected returns when estimated separately in each industry group. However, It is important to note that a good description of the common variation in returns does not necessarily imply a good description of risk compensation. Finally, as firm characteristics vary widely across industries (Chou et al., 2012), by examining the importance of firm characteristics in explaining expected returns, we can expect that there are characteristics that have explanatory power only for certain industries.

Empirical evidence suggests that industry plays an important role in describing stock returns. Moskowitz & Grinblatt (1999) found that individual stock momentum is mostly driven by industry momentum and that stocks within an industry tend to be more highly correlated than stocks in different industries. Hou & Robinson (2006) also discovered that firms in

industries with high concentration tend to earn lower returns, even when controlling for size, book-to-market ratios, and momentum. Chan et al. (2007) state that the comovement in returns is stronger for large-cap stocks belonging to the same industry than it is for small-cap stocks within the same industry. Chou et al. (2012) highlighted that industry returns include important risk premiums that cannot be fully explained by size, book-to-market ratios, and momentum. Overall, previous results suggest that there are unique industry-related patterns that drive asset prices, so the performance of the IPCA model and the importance of characteristics can differ across industry groups.

IPCA outperforms competing factor models in jointly explaining realized variation in individual returns and differences in average returns. Most factor models either use pre-selected characteristics to identify factors (e.g. Fama & French, 1993, 2015) or treat them as latent (e.g. Connor & Korajczyk, 1988). At the same time, the IPCA allows factors to be treated as latent while estimating factor exposures using observable characteristics as instruments. Thus, using both latent factors and observable characteristics, IPCA is more efficient in describing stock returns. This result has already been stated in several studies. Kelly et al. (2019) examined the performance of IPCA in the US stock market. They found that IPCA outperforms other factor models and successfully describes the riskiness of stock returns and risk compensation. Windmüller (2022) applied the IPCA for global stock returns and has shown that IPCA is also effective for non-US stocks and that it performs better in emerging markets than in developed markets. Finally, Ma et al. (2023) found that the IPCA is particularly effective and performs better than competing factor models in the Chinese stock market.

Kelly et al. (2019) also examine the importance of firm characteristics in their work and found that the two most important characteristics that can explain stock returns in the US stock market are market capitalization and total assets. Furthermore, among the 36 characteristics they analyzed, only ten are significant. They propose a test using the bootstrap procedure to decide which characteristic is significant in explaining returns. However, I do not use this test in my paper, as it is extremely computationally intensive. I define the importance of characteristics in a manner similar to Kelly et al. (2019) and Ma et al. (2023), specifically as the decrease in the explained portion of the common variation in individual stock returns that is attributable solely to the given characteristic.

I analyzed 12,813 unique firms of 11 industry groups between July 1962 and May 2014. I used the same database as Freyberger et al. (2017) and Kelly et al. (2019), that contains 36 distinct firm characteristics for each firm. These characteristics are constructed from COMPUSTAT, and the returns are from the CRSP (Freyberger et al., 2017). I have classified each firm into industry groups based on their Standard Industrial Classification (SIC) codes, which I downloaded from the CRSP. I used the 12 industry classifications available on Kenneth French's website, and I combined the Energy and Utilities industry groups to ensure that I have a sufficient number of firms in all industry groups.

First, I examined the performance of the IPCA by industry groups, and I found that the model can better describe the common variation of individual stock returns when it is estimated separately in each industry group. However, it is less effective in describing the differences in

average returns on average. I also analyzed the out-of-sample factor returns and found that the model produces a lower Sharp ratio in each industry group when estimated separately compared to the case when the model is estimated without applying the industry breakdown. These results are in line with my expectations and confirm that there are larger comovements among individual returns within the same industry. Additionally, these results are consistent with the theory that stocks belonging to the same industry are more homogeneous, and therefore, their exposure to systematic risk factors is similar. This allows investors to construct a less diversified portfolio, resulting in a lower Sharp ratio for the mean-variance efficient portfolio.

The results indicate that market capitalization and total assets are among the most important characteristics for all industry groups. However, the importance of different firm characteristics varies across industry groups. Moreover, on average, only three out of the five most important and five out of the ten most important characteristics are shared between the two cases when the model is estimated separately by industry group and when it is estimated without industry breakdown. Taking all industry groups together, the five most important characteristics are market capitalization, total assets, sales-to-price, book-to-market equity, and assets-to-market equity.

This paper is organized as follows. In Section 2, I describe the IPCA model. Section 3 contains the data description and the empirical results, and Section 4 concludes.

2. METHODOLOGY

In my paper, I use the instrumented principal component analysis (IPCA), which was first introduced by Kelly et al. (2017) and then extended to a cross-section of stock returns by Kelly et al. (2019). IPCA is a latent factor model that considers the factor loadings partially dependent on the characteristics. This means the characteristics are treated as instrumental variables of the latent conditional loadings. Thus, by including instruments, we can improve estimates of factors and loadings without correctly specifying exact factors a priori. In addition, IPCA allows for dynamic loadings instead of static loadings used in traditional methods. Overall, the IPCA introduces a new approach to latent factor analysis in asset pricing literature, which helps us better understand the relationship between stock returns and their characteristics.

2.1. THE IPCA MODEL

IPCA formally assumes that a conditional, linear multifactor model generates excess return $r_{i,t+1}$:

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

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

where $f_{t+1}: \mathbb{R}^{K \times 1}$ is the latent factor realizations, $\beta_{i,t}: \mathbb{R}^{1 \times K}$ denotes the corresponding asset-specific factor loadings, $\alpha_{i,t}$ is an asset-specific intercept, $\epsilon_{i,t+1}$ is the residual return of asset i at time $t + 1$ and K denotes the number of latent factors. Additionally, both $\alpha_{i,t}$ and $\beta_{i,t}$ depend linearly on the firm-level characteristics $z_{i,t}: \mathbb{R}^{L \times 1}$, $\Gamma_{\alpha}: \mathbb{R}^{L \times 1}$ and $\Gamma_{\beta}: \mathbb{R}^{L \times K}$ denote their corresponding characteristic coefficients, respectively, and L denote the number of characteristics. $v_{\alpha,i,t}$ and $v_{\beta,i,t}$ denote the error terms in the alpha and beta estimates, which are uncorrelated with the asset characteristics.

We can differentiate between two types of IPCA models based on the presence or absence of an anomalous intercept. If the relationship between the expected returns and characteristics of an asset is determined solely by compensating for exposure to latent risk factors, then IPCA identifies these latent factors and their loadings without considering any anomaly intercept ($\Gamma_{\alpha} = 0_{L \times 1}$). This is known as the restricted model, where the characteristics only serve as proxies for systemic risk exposures related to the latent factors in the pricing equation. In this particular case, Equation 1 can be simplified as follows:

$$r_{i,t+1} = z'_{i,t}\Gamma_{\beta}f_{t+1} + \epsilon_{i,t+1}^*, \quad (2)$$

where $\epsilon_{i,t+1}^* = \epsilon_{i,t+1} + v_{\alpha,i,t} + v_{\beta,i,t}f_{t+1}$ represents the composite error, which captures both the idiosyncratic firm-level shocks ($\epsilon_{i,t+1}$) and the deviation from true factor model parameters ($v_{\alpha,i,t} + v_{\beta,i,t}f_{t+1}$).

On the other hand, if there is a possibility that characteristics explain a portion of the return without encompassing systemic risk, the IPCA model displays an anomalous feature, meaning it has a nonzero intercept. This is the unrestricted model, which assumes that

intercepts are a linear combination of instruments. Therefore, we can rewrite Equation 1 as follows:

$$r_{i,t+1} = z'_{i,t}\Gamma_\alpha + z'_{i,t}\Gamma_\beta f_{t+1} + \epsilon_{i,t+1}^*. \quad (3)$$

The model is estimated using the alternative least squares method, as minimizing the objective function can only be done numerically. However, the numerical problem can be solved relatively quickly as the exact solution can be approximated well, as explained in Section 2.1.1. To obtain an exact solution for IPCA, additional identification assumptions are required, like most latent factor models. Therefore, in line with the work of Kelly et al. (2019) and Ma et al., (2023), it is assumed that $\Gamma_\beta' \Gamma_\beta = \mathbb{I}_K$ for both the restricted and unrestricted models. In addition to this, for the unrestricted model, $\Gamma_\alpha' \Gamma_\beta = 0_{1 \times K}$ is also imposed to achieve econometric identification.

2.1.1. The managed portfolios interpretation

To develop intuition for the IPCA factor and loading estimators, it is helpful to write the static counterpart of IPCA without any instruments. A static model can be constructed similarly as in (Connor & Korajczyk, 1988):

$$x_{t+1} = \beta f_{t+1} + \varepsilon_{t+1}, \quad (4)$$

$$x_{t+1} = \frac{Z'_t r_{t+1}}{N_{t+1}},$$

where $x_{t+1}: \mathbb{R}^{L \times 1}$ is the characteristic-managed portfolios, $\beta: \mathbb{R}^{L \times K}$ is the loadings, $f_{t+1}: \mathbb{R}^{K \times 1}$ is the latent factors and $\varepsilon_{t+1}: \mathbb{R}^{L \times 1}$ is the errors at time $t + 1$. Furthermore, $r_{t+1}: \mathbb{R}^{N_{t+1} \times 1}$ is the individual returns, N_{t+1} is the number of individual stocks at time $t + 1$ and $Z_t: \mathbb{R}^{N_{t+1} \times L}$ is the stacks of each firm characteristics (z_t) at time t . We can see x_{t+1} represents the returns on a set of L managed portfolios at time $t + 1$. It is important to note that the calculation of x_{t+1} only requires individual stock returns for two consecutive periods, so unbalanced data are not a concern.

Equation 4 represents the characteristic-managed portfolio interpretation of the IPCA, introduced in (Kelly et al., 2019). They showed that, as long as $Z'_t Z_t$ is not too volatile, estimating Equation 4 (using PCA) can provide a good approximation of the estimate of Equation 2. However, the role of $Z'_t Z_t$ in the difference between the two estimates highlights that, in the IPCA, the weights assigned to each instrument's contribution are adjusted to emphasize observations associated with the most informative instruments rather than being static across time and assets.

2.2. MEASURING PERFORMANCE

Following Gu et al. (2021) and Kelly et al. (2019), I utilize two types of R^2 statistics to evaluate the performance of the IPCA model in two distinct ways. The first one is the total R^2 , which satisfies the condition that a successful asset pricing model should be capable of accurately describing the common variation in realized excess returns. It quantifies the portion of return

variance accounted for by the dynamic nature of conditional loadings and current factor realizations aggregated over all assets and time periods. For individual stocks, it summarizes how well the systematic factor risk explains the realized riskiness in the individual stocks panel, and we define it as

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

The second metric is the predictive R^2 , which is derived from the condition that a successful asset pricing model should be able to describe the differences in average returns across assets. It represents the fraction of realized excess return variation explained by the model's description of conditional expected returns. We define it for individual stocks as

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

where $\hat{\lambda}$ is the unconditional time-series factor mean of $\hat{f}_{t+1}$.

IPCA distinguishes itself by focusing on the evolution of factor loadings over time rather than relying on static factor loadings like many other models. However, a limitation arises from the assumption that the risk prices remain consistent across different sections and times. IPCA cannot independently assess risk price dynamics; it presumes risk prices to be fixed. It implies that any predictive value for future returns is incorporated solely through adjustments in the factor loadings based on available instruments. Consequently, when applying the $\Gamma_\alpha = 0_{L \times 1}$ restriction, the predictive R^2 metric reflecting the model's ability to describe risk compensation solely through exposure to systematic risk (Kelly et al., 2019).

We can determine the total and the predictive R^2 in an out-of-sample manner. To do this, I use recursive backward-looking estimation and track the post-estimation performance of the restricted IPCA model. I estimate the restricted model every month t , where I have observations for at least 120 months similar to Kelly et al. (2019) and Lewellen (2015), using only observations up to t . If $\hat{\Gamma}_{\beta,t}$ denotes the backward-looking parameter estimate, the out-of-sample realized factor return at $t + 1$, based on Kelly et al. (2019), can be calculated as

$$\hat{f}_{t+1,t} = \left(\hat{\Gamma}'_{\beta,t} Z'_t Z_t \hat{\Gamma}_{\beta,t} \right)^{-1} \hat{\Gamma}'_{\beta,t} Z'_t r_{t+1}. \quad (5)$$

Then, the out-of-sample total R^2 is defined as

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

Finally, the out-of-sample predictive R^2 is defined as

$$\text{Out-of-sample predictive } R^2 = 1 - \frac{\sum_{i,t} (r_{i,t+1} - z'_{i,t} \hat{\Gamma}_{\beta,t} \hat{\lambda}_t)^2}{\sum_{i,t} r_{i,t+1}^2},$$

where $\hat{\lambda}_t$ the factor means through t .

2.3. CHARACTERISTICS IMPORTANCE

The IPCA also allows us to determine which characteristics are important in explaining stock returns. To do this, we start from the restricted model described in Equation 2, and express Γ_{β} in the following way:

$$\Gamma_{\beta} = [\gamma_{\beta,1}, \dots, \gamma_{\beta,L}]',$$

where $\gamma_{\beta,l}: \mathbb{R}^{K \times 1}$ maps characteristic l to loadings on the K latent factors. If the l^{th} characteristic has no impact on the model performance, then it cannot affect any of the loadings. Therefore, the entire l^{th} row of Γ_{β} must be zero. To evaluate the importance of the l^{th} characteristic (l^{th} element of $z_{i,t}$), we run two versions of the restricted model - one with the l^{th} row of Γ_{β} set to zero ($\gamma_{\beta,l} = 0$), and another where it is allowed to be non-zero. We then, following Ma et al. (2023), compare the distance between the two estimates and calculate the importance of the given characteristic as the reduction in the total R^2 .

For estimating the IPCA model, I used the Python IPCA estimation class published alongside Kelly et al. (2017) study, which is available on Seth Priutt's website¹.

¹ <https://www.dropbox.com/s/wte544zzcd03wvs/ipca.py?dl=0>

3. EMPIRICAL RESULTS

3.1. DATA

I use the same database as Kelly et al. (2019) utilized by Freyberger et al. (2017). The database contains the stock returns and characteristics data for 12,813 firms from July 1962 to May 2014, taken from the CRSP and COMPUSTAT databases (Freyberger et al., 2017). The characteristics are already lagged, and each firm has 36 different characteristics for each month. I calculate excess returns using the monthly risk-free rates available on the Kenneth French's website². Moreover, I only consider observations where all 36 characteristics are present, resulting in a final sample of 1,403,544 observations for the 12,813 unique stocks. Table 1 provides a description of each characteristic.

I have classified each firm into 11 industry groups based on their Standard Industrial Classification (SIC) codes. For each firm-month pair, I obtained the SIC codes from the CRSP database and then classified them based on the 12 industry group classifications available on Kenneth French's website³. However, I combined the Energy and Utilities industry groups, resulting in only 11 industry groups. This approach was necessary to ensure that there were sufficient firms and observations in each group. The 11 industry groups are the following: Consumer Nondurables (NoDur); Consumer Durables (Durbl); Manufacturing (Manuf); Oil, Gas, and Coal Extraction and Products, and Utilities (Enrgy + Utils); Chemicals and Allied Products (Chems); Business Equipment (BusEq); Telephone and Television Transmission (Telcm); Wholesale, Retail, and Some Services (Shops); Healthcare, Medical Equipment, and Drugs (Hlth); Finance (Money); and Other (Other). Table 2 reports summary statistics for various firm characteristics of each industry group. I calculate all statistics annually and then average them over time. The Total refers to the aggregate case where I consider all observations and do not apply any industry breakdown, similar to Kelly (2019).

² https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

³ https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html

Table 1: List of Characteristics

The table reports the characteristics and a brief description of each, based on Freyberger et al. (2017). The notation column contains the notation of the characteristics used in the paper.

Characteristic	Notation	Description	Reference
assets-to-market equity	a2me	Total assets over market capitalization as of December t-1	Bhandari (1988)
total assets	assets	Total assets	Gandhi & Lustig (2015)
sales-to-assets	ato	Net sales over lagged net operating assets	Soliman (2008)
market beta	beta	Product of correlations between the excess return of stock i and the market excess return and the ratio of volatilities	Frazzini & Pedersen (2014)
bid-ask spread	bidask	Average daily bid-ask spread in the previous months	Chung & Zhang (2014)
book-to-market equity	bm	Ratio of book value of equity to market value of equity	Davis et al. (2000); Rosenberg et al. (1985)
cash-to-assets	c	Ratio of cash and short-term investments to total assets	Palazzo (2012)
capital turnover	cto	Ratio of net sales to lagged total assets	Haugen & Baker (1996)
capital intensity	d2a	Ratio of depreciation and amortization to total assets	Gorodnichenko & Weber (2016)
change in property, plant, and equipment	dpi2a	Changes in property, plants, and equipment and inventory over lagged total assets	Lyandres et al. (2008)
earnings-to-price	e2p	Ratio of income before extraordinary items to the market capitalization as of December t-1	Basu (1983)
fixed costs-to-sales	fc2y	Ratio of selling, general, and administrative expenses, research and development expenses, and advertising expenses to net sales	D'Acutto et al. (2018)
cash flow-to-book	freecf	Ratio of net income, depreciation and amortization, less change in working capital, and capital expenditure over the book-value of equity	Hou et al. (2011)
idiosyncratic volatility	idiovol	Standard deviation of the residuals from a regression of excess returns on the Fama and French (1993) three-factor model	Ang et al. (2006)
intermediate momentum	intmom	Cumulative return from 12 months before the return prediction to seven months before	Novy-Marx (2012)
investment	invest	Percentage year-on-year growth rate in total assets	Cooper et al. (2008)
leverage	lev	Ratio of long-term debt and debt in current liabilities to the sum of long-term debt, debt in current liabilities, and stockholders' equity	Lewellen (2015)
long-term reversal	ltrev	Cumulative return from 36 months before the return prediction to 13 months before	De Bondt & Thaler (1985)
market capitalization	mkcap	Product of the price and shares outstanding	Banz (1981)
momentum	mom	Cumulative return from 12 months before the return prediction to two months before	Fama & French (2016)
net operating assets	noa	Difference between operating assets minus operating liabilities scaled by lagged total assets	Hirshleifer et al. (2004)
operating accruals	oa	Changes in non-cash working capital minus depreciation scaled by lagged total assets	Sloan (1996)
operating leverage	ol	Sum of cost of goods sold and selling, general, and administrative expenses over total assets	Novy-Marx (2011)
price-to-cost margin	pcm	Difference between net sales and costs of goods sold divided by net sales	Bustamante & Donangelo (2017)
profit margin	pm	Operating income after depreciation over net sales	Soliman (2008)
profitability	prof	Gross profitability divided by the book value of equity	Ball et al. (2015)
Tobin's Q	q	Total assets, the market value of equity minus cash and short-term investments, minus deferred taxes scaled by total assets	
return on net operating assets	rna	Ratio of operating income after depreciation to lagged net operating assets	Soliman (2008)
return-on-assets	roa	Income before extraordinary items to lagged total assets	Balakrishnan et al. (2010)
return-on-equity	roe	Income before extraordinary items to lagged book-value of equity	Haugen & Baker (1996)
sales-to-price	s2p	Ratio of net sales to the market capitalization as of December	Lewellen (2015)
sales, general and administrative costs-to-sales	sga2s	Ratio of selling, general and administrative expenses to net sales	
short-term reversal	strev	Lagged one-month return	Jegadeesh (1990)
standard unexplained volume	suv	Difference between actual volume and predicted volume in the previous month	Garfinkel (2009)
turnover	turn	Last month's volume over shares outstanding	Datar et al. (1998)
closeness to 52-week high	w52h	The ratio of stock price at the end of the previous calendar month and the previous 52 week high price	George & Hwang (2004)

Table 2: Summary Table of Characteristics

The table reports the mean and the time-series standard deviation of the 36 characteristics, excess returns (ret) and the number of firms (n) by industry groups and for all observations (Total). Total assets (assets) is in thousands dollar and market capitalization (mktcap) is in million dollar.

	Industry											
	BusEq	Chems	Durbl	Enrgy	Hlth	Manuf	Money	NoDur	Other	Shops	Telcm	Total
ret	0.012 (0.21)	0.009 (0.13)	0.007 (0.15)	0.007 (0.17)	0.013 (0.19)	0.009 (0.15)	0.008 (0.13)	0.008 (0.14)	0.009 (0.18)	0.009 (0.17)	0.011 (0.20)	0.009 (0.17)
a2me	1.354 (1.92)	1.665 (1.73)	2.345 (3.32)	2.115 (3.06)	1.137 (4.04)	2.349 (3.01)	9.920 (18.40)	2.355 (3.41)	2.569 (6.08)	2.546 (3.83)	2.298 (4.82)	2.828 (7.07)
assets	1.073 (6.21)	2.129 (7.81)	2.907 (25.72)	3.589 (15.27)	1.484 (7.85)	1.367 (14.05)	10.398 (85.56)	1.266 (5.21)	0.938 (3.94)	1.164 (5.43)	7.864 (27.48)	2.415 (28.54)
ato	3.234 (101.06)	2.422 (4.53)	3.223 (22.39)	1.652 (15.89)	2.583 (42.96)	2.405 (18.89)	0.884 (76.80)	1.906 (73.36)	2.946 (57.61)	4.911 (16.73)	2.676 (65.94)	2.760 (59.66)
beta	1.247 (0.69)	0.991 (0.52)	1.049 (0.57)	0.990 (0.66)	1.001 (0.58)	1.017 (0.56)	0.639 (0.54)	0.868 (0.51)	1.022 (0.65)	0.958 (0.56)	1.104 (0.58)	1.000 (0.62)
bidask	0.041 (0.07)	0.028 (0.05)	0.033 (0.04)	0.059 (0.11)	0.035 (0.06)	0.034 (0.05)	0.029 (0.05)	0.033 (0.05)	0.040 (0.07)	0.038 (0.06)	0.028 (0.07)	0.037 (0.07)
bm	0.699 (0.70)	0.773 (0.68)	0.974 (0.97)	0.869 (0.80)	0.546 (0.69)	0.998 (0.90)	0.967 (1.05)	1.052 (1.04)	0.919 (1.23)	1.007 (1.16)	0.844 (1.54)	0.890 (0.98)
c	0.236 (0.21)	0.101 (0.13)	0.105 (0.13)	0.089 (0.12)	0.229 (0.22)	0.090 (0.11)	0.078 (0.10)	0.093 (0.12)	0.145 (0.17)	0.098 (0.12)	0.126 (0.17)	0.134 (0.16)
cto	1.258 (0.75)	1.345 (0.58)	1.706 (1.14)	0.815 (1.69)	1.176 (0.85)	1.418 (0.63)	0.356 (1.34)	1.738 (1.06)	1.421 (1.39)	2.478 (1.58)	0.770 (0.71)	1.405 (1.26)
d2a	0.049 (0.05)	0.042 (0.02)	0.039 (0.02)	0.077 (0.06)	0.039 (0.03)	0.041 (0.02)	0.009 (0.02)	0.038 (0.02)	0.052 (0.06)	0.042 (0.03)	0.060 (0.05)	0.043 (0.04)
dpi2a	0.070 (0.22)	0.081 (0.21)	0.080 (0.22)	0.200 (0.56)	0.079 (0.18)	0.084 (0.23)	0.023 (0.39)	0.071 (0.20)	0.103 (0.36)	0.110 (0.24)	0.111 (0.28)	0.087 (0.29)
e2p	-0.076 (0.49)	0.036 (0.22)	0.009 (0.44)	-0.043 (0.63)	-0.048 (0.86)	0.030 (0.33)	-0.007 (0.62)	0.022 (0.66)	-0.047 (0.70)	-0.003 (0.64)	-0.138 (0.94)	-0.019 (0.59)
fc2y	0.697 (5.79)	0.378 (1.27)	0.240 (0.37)	0.539 (9.18)	1.673 (25.05)	0.253 (3.73)	0.381 (8.96)	0.280 (0.78)	0.573 (7.30)	0.322 (10.13)	0.762 (8.00)	0.518 (9.23)
freecf	-0.254 (5.59)	0.003 (0.83)	-0.116 (4.29)	-0.268 (1.53)	-0.271 (3.93)	-0.052 (2.23)	0.028 (1.99)	-0.045 (3.55)	-0.199 (2.70)	-0.077 (1.48)	-1.468 (31.59)	-0.150 (5.19)
idiovol	0.038 (0.03)	0.024 (0.02)	0.028 (0.03)	0.032 (0.03)	0.035 (0.03)	0.027 (0.02)	0.024 (0.02)	0.026 (0.02)	0.034 (0.03)	0.031 (0.03)	0.031 (0.03)	0.031 (0.03)
intmom	0.094 (0.60)	0.073 (0.32)	0.075 (0.39)	0.073 (0.43)	0.091 (0.53)	0.077 (0.37)	0.063 (0.32)	0.075 (0.37)	0.077 (0.48)	0.085 (0.53)	0.106 (0.59)	0.081 (0.47)
invest	0.181 (0.84)	0.111 (0.32)	0.125 (0.40)	0.207 (0.84)	0.214 (0.62)	0.120 (0.46)	0.145 (1.09)	0.110 (0.36)	0.180 (0.60)	0.150 (0.46)	0.283 (1.03)	0.157 (0.67)
lev	0.195 (0.21)	0.313 (0.20)	0.305 (0.23)	0.361 (0.23)	0.225 (0.22)	0.323 (0.21)	0.501 (0.24)	0.322 (0.22)	0.343 (0.26)	0.321 (0.23)	0.439 (0.27)	0.315 (0.24)
ltrev	0.392 (1.72)	0.316 (0.79)	0.360 (1.23)	0.344 (1.29)	0.411 (1.44)	0.333 (1.04)	0.307 (0.79)	0.321 (0.93)	0.350 (1.53)	0.369 (1.16)	0.420 (1.46)	0.354 (1.29)

Table 2: Continued from Previous Page

	Industry											Total
	BusEq	Chems	Durbl	Enrgy	Hlth	Manuf	Money	NoDur	Other	Shops	Telcm	
mktcap	2.118	2.823	1.320	3.292	3.287	1.114	1.781	1.678	0.848	1.478	6.215	1.842
	(15.11)	(12.13)	(8.90)	(19.79)	(15.62)	(5.87)	(10.36)	(8.86)	(3.38)	(8.69)	(20.73)	(11.38)
mom	0.177	0.137	0.141	0.133	0.175	0.143	0.123	0.142	0.143	0.155	0.201	0.151
	(0.95)	(0.45)	(0.60)	(0.61)	(0.79)	(0.57)	(0.44)	(0.56)	(0.73)	(0.78)	(0.89)	(0.71)
noa	0.607	0.714	0.719	0.824	0.674	0.734	0.328	0.740	0.707	0.714	0.839	0.669
	(0.62)	(0.29)	(0.31)	(0.58)	(0.47)	(0.33)	(0.71)	(0.31)	(0.47)	(0.35)	(0.69)	(0.50)
oa	-0.708	0.055	-1.037	-0.220	-0.532	0.012	3.965	-0.059	-1.042	0.215	-59.924	-0.885
	(46.78)	(4.85)	(61.49)	(17.80)	(62.15)	(5.26)	(968.49)	(4.21)	(123.43)	(35.63)	(3933.18)	(582.38)
ol	1.041	1.088	1.398	0.582	0.951	1.154	0.263	1.450	1.137	2.075	0.574	1.147
	(0.54)	(0.48)	(0.61)	(0.77)	(0.60)	(0.47)	(0.61)	(0.90)	(1.00)	(1.45)	(0.58)	(0.95)
pcm	0.437	0.361	0.288	0.289	0.468	0.286	0.522	0.324	0.285	0.292	0.356	0.356
	(1.02)	(0.71)	(0.16)	(3.84)	(1.19)	(0.28)	(0.23)	(0.42)	(2.76)	(2.16)	(4.03)	(1.71)
pm	-0.175	0.014	0.049	-0.437	-0.786	0.026	0.118	0.041	-0.303	-0.020	-0.569	-0.135
	(6.08)	(1.41)	(0.33)	(10.21)	(16.19)	(3.47)	(8.29)	(0.88)	(8.02)	(10.10)	(13.51)	(7.95)
prof	1.063	1.306	1.093	0.504	1.078	0.925	0.683	1.243	1.322	1.600	4.148	1.148
	(5.30)	(15.31)	(4.43)	(1.64)	(5.33)	(4.04)	(10.74)	(6.36)	(24.17)	(6.21)	(97.40)	(16.05)
q	2.151	1.673	1.458	1.443	2.723	1.381	1.151	1.423	1.698	1.570	1.681	1.674
	(2.40)	(1.39)	(1.16)	(1.08)	(2.82)	(1.13)	(0.74)	(1.03)	(1.87)	(1.34)	(1.89)	(1.73)
rna	0.582	-0.052	0.203	-0.024	1.458	0.161	-0.027	0.070	0.264	0.190	-0.036	0.295
	(34.96)	(13.59)	(2.03)	(9.94)	(77.28)	(1.37)	(9.79)	(6.87)	(11.67)	(1.88)	(14.94)	(25.97)
roa	-0.007	0.052	0.048	0.016	-0.021	0.046	0.013	0.050	0.018	0.043	-0.018	0.023
	(0.26)	(0.12)	(0.13)	(0.16)	(0.36)	(0.11)	(0.08)	(0.13)	(0.18)	(0.12)	(0.25)	(0.19)
roe	-0.007	0.137	0.129	0.040	-0.082	0.099	0.067	0.092	0.010	0.092	-0.147	0.048
	(2.47)	(1.34)	(1.71)	(0.87)	(2.83)	(0.86)	(1.40)	(0.84)	(3.31)	(0.81)	(4.53)	(1.98)
s2p	1.575	2.085	3.513	1.582	1.155	3.017	1.389	3.838	2.853	5.817	1.217	2.781
	(2.44)	(2.66)	(4.91)	(4.02)	(2.09)	(4.01)	(4.31)	(6.07)	(5.15)	(10.02)	(2.70)	(5.47)
sga2s	0.495	0.462	0.606	0.169	0.388	0.479	0.316	0.702	0.488	1.194	0.363	0.560
	(0.73)	(0.62)	(0.96)	(0.34)	(0.61)	(0.64)	(0.80)	(1.06)	(0.92)	(2.23)	(0.84)	(1.12)
strev	0.016	0.013	0.012	0.012	0.016	0.013	0.011	0.012	0.013	0.013	0.014	0.013
	(0.21)	(0.13)	(0.15)	(0.17)	(0.19)	(0.15)	(0.13)	(0.14)	(0.18)	(0.17)	(0.20)	(0.17)
suv	0.248	0.193	0.212	0.189	0.247	0.193	0.399	0.213	0.260	0.218	0.265	0.241
	(3.24)	(2.66)	(2.59)	(3.13)	(2.74)	(2.36)	(3.96)	(2.62)	(3.57)	(2.90)	(2.72)	(3.04)
turn	0.152	0.068	0.073	0.103	0.128	0.072	0.059	0.062	0.107	0.105	0.135	0.099
	(0.31)	(0.16)	(0.12)	(0.19)	(0.21)	(0.15)	(0.12)	(0.14)	(0.23)	(0.22)	(0.31)	(0.21)
w52h	0.681	0.777	0.740	0.724	0.714	0.757	0.794	0.761	0.720	0.732	0.729	0.735
	(0.23)	(0.19)	(0.20)	(0.22)	(0.22)	(0.19)	(0.20)	(0.20)	(0.21)	(0.21)	(0.24)	(0.21)
n	569.5	83.7	100.0	153.9	228.7	449.2	449.2	224.1	322.0	373.4	53.0	2731.3
	(180.25)	(15.81)	(21.69)	(48.62)	(75.05)	(99.43)	(161.61)	(42.46)	(90.04)	(115.28)	(21.01)	(775.85)

3.2. IPCA PERFORMANCE

First, I estimate the IPCA model in each industry group. To do this, I split the database into 11 distinct databases by industry groups. In each case, I cross-sectionally transform the characteristics. Following Gu et al. (2020, 2021), Kelly et al. (2019) and Ma et al. (2023), I rank-normalize the characteristics for each month, mapping them into a $[-0.5, 0.5]$ interval. This transformation emphasizes ordering characteristics within each period rather than their magnitudes, thereby mitigating the influence of outliers. Instead of focusing on absolute sizes, it prioritizes relative sizes in a cross-sectional view, similar to portfolio sorting. Finally, I only included months with at least 50 observations, except in the Total case, where I included months with at least 100 observations.

3.2.1. In-sample performance

When conducting in-sample estimations, the entire sample is utilized to determine the factors and Γ_{β} . This approach offers a comprehensive view of the IPCA model's overall explanatory performance in each industry group. All this is due to the total R^2 measuring the model's ability to explain risk by quantifying the amount of total variation in realized returns described by the IPCA common factors. Table 3 reports the restricted model's in-sample total R^2 for individual stock returns in each industry group, considering $K = 1, \dots, 6$ factors. The Industry Total shows the case where the IPCA model parameters are estimated separately by industry group, but the model's performance is evaluated for all industry groups taken together.

Each industry group has a higher total R^2 for at least three factors, and when industry groups are taken together (Industry Total), the total R^2 is more than five percentage points higher for $K = 6$ factors than for estimating the parameters together (Total). This indicates that the systematic risk factors better explain the realized riskiness for individual stocks by estimating the model separately in each industry group. It is confirmed that there is a higher level of comovements among individual stock returns within the same industry group.

However, the total R^2 values differ significantly among different industries. For example, the highest total R^2 (Telcm) is more than twice the lowest ones (Other, Shops, Hlth) for $K = 6$. This means that Telcm is more than two times more explainable by systematic risk factors than the Other, Shops, or Hlth industry groups, showing that the IPCA model's ability to describe the common variation in realized returns varies widely across industry groups.

Table 4 shows the unrestricted model's total R^2 s for individual stocks. Comparing these with the total R^2 s for the restricted model, revealing that both increase as the number of latent factors increases. Additionally, the gap between the restricted and unrestricted models becomes smaller. For $K = 1$, the restricted models capture more than 97% of the unrestricted model's explanatory power, while for $K = 6$, this number increases to more than 99% in the Industry Total case.

Table 3: Restrictive model's Total R^2

The table reports the total R^2 statistics for the panel of individual stocks and for the restricted IPCA model ($\Gamma_\alpha = 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	17.5%	20.2%	21.7%	23.0%	24.0%	24.9%	125603
Durbl	21.6%	24.7%	27.0%	29.3%	31.3%	33.0%	54550
Manuf	17.9%	20.1%	21.5%	22.7%	23.5%	24.1%	248143
Enrgy + Utils	21.1%	24.0%	25.7%	27.1%	28.3%	29.5%	77829
Chems	20.6%	24.5%	27.2%	29.7%	31.8%	33.8%	46543
BusEq	19.4%	21.5%	22.4%	23.1%	23.6%	24.0%	246001
Telecm	22.7%	28.7%	33.7%	37.1%	40.4%	43.1%	13995
Other	14.2%	16.0%	17.4%	18.3%	19.1%	19.8%	161487
Shops	14.9%	16.9%	17.9%	18.8%	19.5%	20.2%	184043
Hlth	13.6%	16.0%	17.4%	18.4%	19.4%	20.2%	92928
Money	16.9%	20.1%	22.0%	23.5%	24.7%	25.7%	128699
Industry Total	17.4%	19.8%	21.1%	22.2%	23.1%	23.9%	1379821
Total	14.8%	16.4%	17.4%	18.0%	18.6%	18.9%	1403544

Table 4: Unrestrictive model's Total R^2

The table reports the total R^2 statistics for the panel of individual stocks and for the unrestricted IPCA model ($\Gamma_\alpha \neq 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	18.0%	20.7%	22.2%	23.4%	24.3%	25.2%	125603
Durbl	22.1%	25.2%	27.5%	29.7%	31.5%	33.3%	54550
Manuf	18.4%	20.6%	21.9%	23.1%	23.8%	24.3%	248143
Enrgy + Utils	21.6%	24.4%	26.1%	27.5%	28.6%	29.8%	77829
Chems	21.2%	25.1%	27.7%	30.1%	32.2%	34.1%	46543
BusEq	19.9%	21.8%	22.7%	23.3%	23.8%	24.2%	246001
Telecm	23.1%	29.1%	34.1%	37.4%	40.8%	43.4%	13995
Other	14.7%	16.4%	17.8%	18.6%	19.4%	20.0%	161487
Shops	15.3%	17.3%	18.3%	19.0%	19.7%	20.4%	184043
Hlth	14.1%	16.5%	17.7%	18.7%	19.6%	20.5%	92928
Money	17.4%	20.6%	22.5%	23.9%	25.0%	26.0%	128699
Industry Total	17.9%	20.2%	21.5%	22.5%	23.3%	24.1%	1379821
Total	15.2%	16.8%	17.7%	18.4%	18.7%	19.0%	1403544

Tables 5 and 6 report the in-sample predictive R^2 s for the restricted and unrestricted models, respectively. IPCA aims to maximize the total R^2 , which is the proportion of total return variation that the model can explain. However, a high total R^2 may not accurately predict the conditional expected returns, which account for a tiny portion of the total return. Therefore, a high total R^2 does not always indicate a high predictive R^2 . This is precisely what we can see in the case of some industry groups. For example, Telcm has the highest total R^2 and the lowest predictive R^2 for at least $K = 4$ factors.

Table 5: Restrictive model's Predictive R^2

The table reports the predictive R^2 statistics for the panel of individual stocks and for the restricted IPCA model ($\Gamma_\alpha = 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	0.37%	0.37%	0.42%	0.62%	0.72%	0.74%	125603
Durbl	0.24%	0.23%	0.33%	0.36%	0.53%	0.55%	54550
Manuf	0.42%	0.42%	0.47%	0.50%	0.78%	0.88%	248143
Enrgy + Utils	0.18%	0.20%	0.21%	0.33%	0.51%	0.49%	77829
Chems	0.58%	0.57%	0.57%	0.76%	0.80%	0.77%	46543
BusEq	0.40%	0.49%	0.57%	0.67%	0.73%	0.78%	246001
Telcm	0.35%	0.35%	0.30%	0.21%	0.13%	0.13%	13995
Other	0.26%	0.27%	0.30%	0.53%	0.57%	0.60%	161487
Shops	0.31%	0.31%	0.32%	0.48%	0.52%	0.54%	184043
Hlth	0.50%	0.48%	0.75%	0.78%	0.78%	0.77%	92928
Money	0.44%	0.42%	0.44%	0.43%	0.63%	0.68%	128699
Industry Total	0.37%	0.39%	0.45%	0.56%	0.66%	0.70%	1379821
Total	0.35%	0.34%	0.41%	0.42%	0.69%	0.68%	1403544

When the model's performance is evaluated for all industry groups taken together (Industry Total) the predictive R^2 is nearly the same as in the Total case. Moreover, only five of the 11 industry groups have higher predictive R^2 s for $K = 6$ when the model is estimated separately in each industry group. This indicates that the predictive R^2 s vary across the industry groups similarly to the total R^2 , although the order of the industry groups differs considerably in terms of their values. Overall, the results suggest that when considering only returns belonging to the same industry group, the IPCA nearly as accurately describes the average returns for individual stocks on average, as when estimating the model parameters for all industry groups together. However, in some industry groups like Telcm, its effectiveness in describing the average returns is quite poor.

The unrestricted model shows higher predictive R^2 s for all K s and industry groups. When considering only one factor ($K = 1$), the restricted model can capture almost 44.4% of the unrestricted model's predictability in the Industry Total case. However, by increasing the number of factors to six ($K = 6$), this percentage can increase to more than 87.6%. Similarly to

the total R^2 , this demonstrates that the restricted and unrestricted models tend to converge as the number of factors increases.

Table 6: Unrestrictive model's Predictive R^2

The table reports the predictive R^2 statistics for the panel of individual stocks and for the unrestricted IPCA model ($\Gamma_\alpha \neq 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	0.89%	0.88%	0.87%	0.86%	0.86%	0.86%	125603
Durbl	0.80%	0.78%	0.78%	0.77%	0.76%	0.76%	54550
Manuf	0.94%	0.93%	0.93%	0.92%	0.91%	0.91%	248143
Enrgy + Utils	0.64%	0.63%	0.62%	0.61%	0.62%	0.61%	77829
Chems	1.21%	1.19%	1.18%	1.18%	1.17%	1.16%	46543
BusEq	0.85%	0.84%	0.82%	0.82%	0.82%	0.81%	246001
Telcm	0.77%	0.76%	0.70%	0.60%	0.55%	0.50%	13995
Other	0.70%	0.69%	0.68%	0.68%	0.68%	0.68%	161487
Shops	0.69%	0.68%	0.67%	0.67%	0.66%	0.66%	184043
Hlth	0.98%	0.97%	0.97%	0.96%	0.95%	0.95%	92928
Money	0.93%	0.92%	0.91%	0.89%	0.88%	0.89%	128699
Industry Total	0.83%	0.82%	0.81%	0.80%	0.80%	0.80%	1379821
Total	0.76%	0.75%	0.75%	0.74%	0.74%	0.72%	1403544

In summary, the IPCA model is more effective when estimated separately in describing the common variation in individual returns for each industry group. Additionally, it is nearly as effective in describing the average returns of individual stocks. However, its effectiveness is limited to describing the differences in average returns in less than half of the 11 industry groups. These results highlight the homogeneity of stocks within industries and confirm that individual stock returns behave similarly within industry groups. Moreover, the IPCA is less able to capture differences between individual returns in some industry groups such as Telcm or Enrgy + Utils. Comparing the restricted and unrestricted models reveals that by increasing the number of factors, IPCA can accurately explain how risks influence expected returns through the exposure and interaction of instrumental loadings without relying on irrational alpha on average and in most industry groups.

3.2.2. Out-of-sample performance

IPCA's asset pricing performance has been evaluated using factors and dimension-reduction matrices estimated in-sample. To prevent any look-ahead bias, I computed the IPCA model's parameters in a recursive and backward-looking manner and re-evaluated its performance as described in Section 2.2. Tables 7 and 8 report restricted model out-of-sample total and predictive R^2 s.

Table 7: Restrictive model's Out-of-sample Total R^2

The table reports the out-of-sample total R^2 statistics for the panel of individual stocks, with recursive backward-looking estimation scheme for $t \geq 120$, and for the restricted IPCA model ($\Gamma_\alpha = 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	14.2%	16.6%	17.9%	18.9%	19.8%	20.6%	104580
Durbl	17.8%	20.3%	22.1%	23.7%	25.1%	26.5%	43919
Manuf	16.3%	17.9%	18.8%	19.9%	20.5%	21.7%	213963
Enrgy + Utils	19.9%	22.4%	23.6%	24.7%	25.7%	26.5%	68624
Chems	18.3%	21.4%	23.4%	25.2%	27.1%	28.6%	37861
BusEq	18.4%	20.3%	21.1%	21.6%	22.0%	22.4%	232691
Telecm	18.5%	22.1%	24.1%	25.5%	26.2%	28.7%	6174
Other	12.2%	13.8%	14.9%	15.7%	16.2%	16.8%	146165
Shops	13.2%	15.0%	15.5%	16.2%	16.8%	17.3%	164891
Hlth	12.6%	14.4%	15.3%	15.9%	16.7%	17.4%	84523
Money	15.7%	17.6%	19.7%	20.8%	21.7%	22.4%	119507
Industry Total	15.8%	17.7%	18.7%	19.5%	20.1%	20.8%	1222898
Total	13.9%	15.3%	16.3%	16.9%	17.5%	17.8%	1289590

Table 8: Restrictive model's Out-of-sample Predictive R^2

The table reports the out-of-sample predictive R^2 statistics for the panel of individual stocks, with recursive backward-looking estimation scheme for $t \geq 120$, and for the restricted IPCA model ($\Gamma_\alpha = 0_{L \times 1}$) from $K = 1$ to $K = 6$ for each industry group. The *Industry Total* refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and *Total* refers to the aggregate case where we consider all observations and do not apply any industry breakdown. The *N. Obs* column contains the number of observations used for the IPCA model's estimation.

industry	K						N. Obs
	1	2	3	4	5	6	
NoDur	0.39%	0.33%	0.47%	0.53%	0.54%	0.53%	104580
Durbl	0.20%	0.17%	0.20%	0.25%	0.24%	0.25%	43919
Manuf	0.44%	0.43%	0.52%	0.69%	0.78%	0.78%	213963
Enrgy + Utils	0.09%	0.08%	0.19%	0.17%	0.23%	0.22%	68624
Chems	0.55%	0.51%	0.53%	0.57%	0.70%	0.70%	37861
BusEq	0.33%	0.39%	0.56%	0.57%	0.63%	0.62%	232691
Telecm	0.34%	0.18%	0.16%	0.11%	-0.01%	-0.04%	6174
Other	0.23%	0.23%	0.30%	0.44%	0.43%	0.43%	146165
Shops	0.30%	0.28%	0.38%	0.40%	0.42%	0.41%	164891
Hlth	0.41%	0.39%	0.59%	0.59%	0.64%	0.68%	84523
Money	0.37%	0.30%	0.29%	0.26%	0.34%	0.35%	119507
Industry Total	0.33%	0.33%	0.45%	0.50%	0.54%	0.54%	1222898
Total	0.34%	0.33%	0.55%	0.61%	0.60%	0.60%	1289590

IPCA's out-of-sample performance is similar to its in-sample performance. Overall, it performs well out-of-sample in describing the common variation in expected returns, although only 8 out of 11 industry groups have a higher out-of-sample total R^2 compared to the Total case. Furthermore, the out-of-sample total R^2 is reduced by more than 20% for some industries, such as Telcm or Durbl, which may be due to overfitting, as there are few observations in these industries. The predictive R^2 varies widely across industry groups, similar to the in-sample case. However, evaluating all industry groups together (Industry Total) results in a slightly lower value compared to the Total case. Telcm has the lowest out-of-sample predictive R^2 , which is negative for at least five factors, while the highest one is 0.78% in the Manuf industry group for $K = 5$.

Overall, these results show that the IPCA can still effectively describe the common variation in realized returns, even in an out-of-sample manner. However, in line with the in-sample results, it struggles to capture the differences between average returns in some industry groups, but mainly in the Telcm industry.

I analyze out-of-sample unconditional Sharpe ratios for IPCA factors. The in-sample Sharpe ratios may be unrealistically high due to in-sample overfitting, as highlighted by Kan et al. (2019) and Kozak et al. (2018). Therefore, I focus solely on the analysis of out-of-sample IPCA factor returns. I present the univariate annualized Sharpe ratio to describe the unconditional efficiency of individual factors. Additionally, I report the ex-ante unconditional annualized tangency Sharpe ratio for a set of factors to describe its multivariate efficiency. To calculate out-of-sample factor returns, I use the recursive estimation approach as before. Using individual stock weights $(\hat{\Gamma}'_{\beta,t} Z'_t Z_t \hat{\Gamma}_{\beta,t})^{-1} \hat{\Gamma}'_{\beta,t} Z'_t$ that require no information beyond time t , I calculate IPCA factor returns at $t + 1$, according to Equation 5. I also construct the tangency portfolio for a set of factors return based purely on out-of-sample data. This involves using the estimated mean and covariance matrix of the factors up to time t and tracking the post-formation $t + 1$ returns. To ensure consistency, I adjust the tangency weights for each period by targeting a 1% monthly portfolio volatility based on historical estimates. The Sharpe ratios for factor K and the tangency portfolio are based on factors 1 through K .

Tables 9 and 10 report the out-of-sample unconditional annualized univariate and tangency Sharp ratios for each Industry group. The univariate Sharp ratios of the IPCA factors ranged from -0.11 to 1.29. Adding multiple factors to the tangency portfolio increased the tangency Sharp ratios for most industry groups. Moreover, the tangency Sharp ratios peak at $K = 6$ for all industry groups, ranging between 0.57 and 3.29. Compared with the Total case, the tangency Sharp ratios are smaller for all industry groups, starting from $K = 3$.

A high unconditional Sharpe ratio means that the IPCA can capture a lot of co-movements among assets while effectively aligning their factor loadings with differences in average returns. As a result, the smaller Sharp ratios observed in the industry groups indicate that while the IPCA is better at describing the common variation in expected returns when estimated separately for each industry group, it performs worse overall because it is less accurate in describing differences in average returns.

Table 9: Out-of-sample Annualized Univariate Sharpe Ratios

The table reports out-of-sample annualized Sharpe ratios for individual factors (“univariate”) for each industry group from $K = 1$ to $K = 6$. Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown.

industry	K					
	1	2	3	4	5	6
NoDur	0.65	0.27	0.46	0.97	0.44	0.20
Durbl	0.43	0.08	0.03	-0.24	0.15	0.16
Manuf	0.67	0.34	0.26	1.28	0.67	0.49
Enrgy + Utils	0.39	0.41	0.18	0.40	0.47	0.51
Chems	0.67	0.56	0.31	0.31	0.44	0.51
BusEq	0.55	-0.08	0.73	1.29	1.37	1.05
Telcm	0.50	0.30	-0.05	-0.11	-0.08	0.18
Other	0.56	0.08	-0.16	0.88	1.13	0.45
Shops	0.61	0.21	0.51	0.71	0.93	0.99
Hlth	0.67	0.58	0.23	0.43	0.44	0.32
Money	0.69	0.55	0.69	0.83	0.66	0.60
Total	0.62	0.03	1.66	1.33	0.97	0.55

Table 10: Out-of-sample Annualized Tangency Sharpe Ratios

The table reports out-of-sample annualized Sharpe ratios or the mean-variance efficient portfolio of factors in each model (“tangency”) for each industry group from $K = 1$ to $K = 6$. Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown.

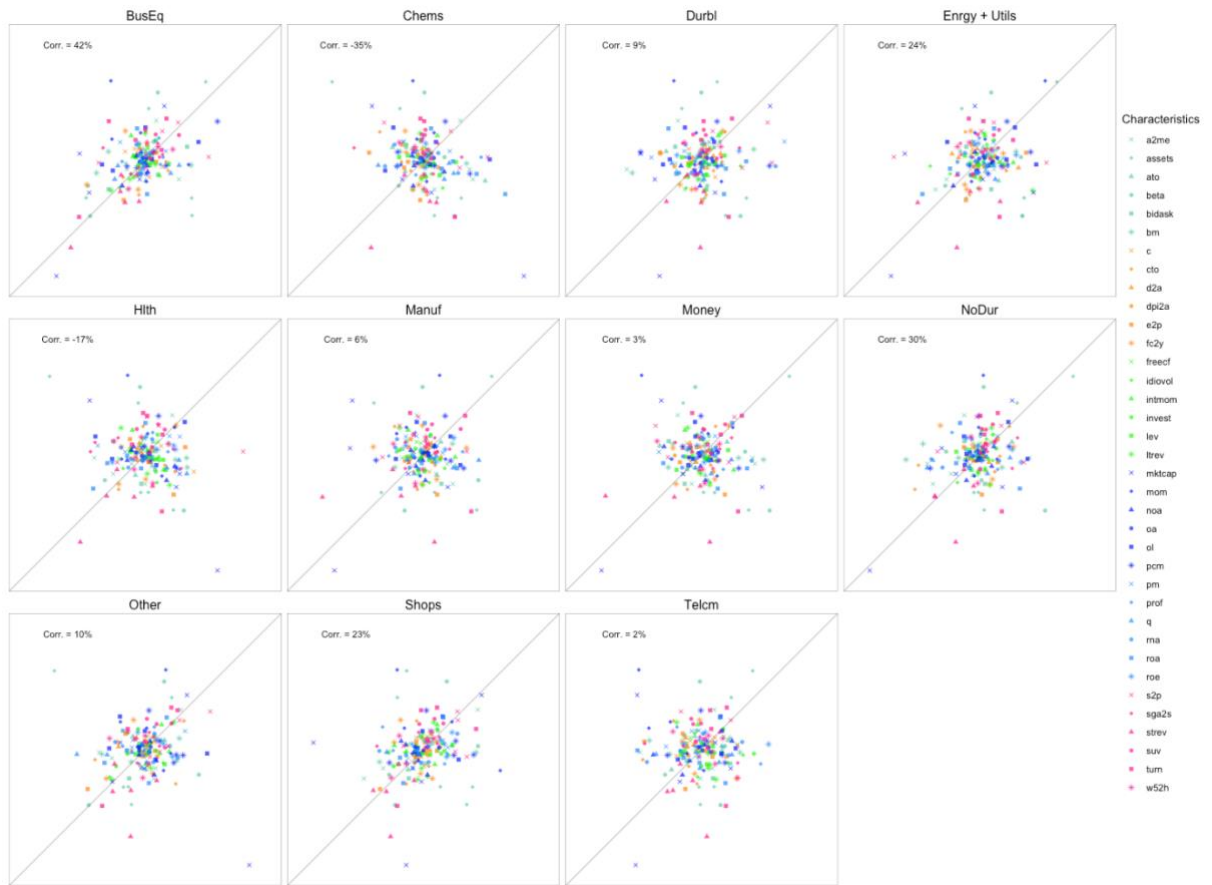
industry	K					
	1	2	3	4	5	6
NoDur	0.65	0.47	1.40	1.93	2.06	2.11
Durbl	0.43	0.39	0.76	0.93	0.93	0.94
Manuf	0.67	0.76	1.10	1.96	2.89	3.29
Enrgy + Utils	0.39	0.40	0.82	1.18	1.80	1.84
Chems	0.67	0.66	0.57	0.72	1.22	1.26
BusEq	0.55	0.89	2.05	2.36	2.86	3.13
Telcm	0.50	0.08	0.29	0.45	0.15	0.57
Other	0.56	0.55	0.89	2.50	2.50	2.57
Shops	0.61	0.49	1.85	2.31	2.38	2.38
Hlth	0.67	0.61	1.64	1.73	1.93	2.15
Money	0.69	0.28	0.25	0.36	0.70	0.76
Total	0.62	0.62	2.48	3.08	3.89	4.04

3.2.3. Parameter stability

Finally, I analyze the stability of the IPCA model between the industry groups. Figure 1 displays the estimated coefficients of Γ_β for the Total case and various industry groups. When the estimated parameters are stable, we would expect them to be clustered close to the 45-degree line. However, for most industry groups, we observe the opposite as the characteristics are scattered equally on the plot. The differences in the parameters are also indicated by the fact that the correlation coefficient ranges between -0.35 and 0.4. Overall, these results show differences in the IPCA parameters among industry groups, confirming industry heterogeneity.

Figure 1: Parameter Stability

The figure plots Γ_β parameter estimates elementwise between industry groups and Total for the five-factor restricted IPCA model. Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown.



3.3. WHICH CHARACTERISTICS MATTER?

In order to determine which characteristics are important in explaining stock returns within each industry group, I used the method outlined in Section 2.3. I started from the five-factor restricted model as Kelly et al. (2019) and determined the importance of each characteristic by setting all elements of Γ_β related to that characteristic to zero and calculating the reduction in total R^2 of individual returns. Next, I normalized the values of each characteristic importance as a percentage of the total characteristic importance, as in the work of Ma et al. (2023). This allowed me to compare the relative importance of each characteristic between the different industry groups. Figure 2 displays the detailed ranking of the importance of firm characteristics for each industry group. The characteristics are ordered based on their rank in the Total case. Additionally, Figure 3 shows the ten most important characteristics of each industry group.

Figure 2: Characteristics Importance

The figure reports the ranking of firm characteristics importance for each industry group for the five-factor restricted IPCA model. Columns correspond to each industry group, and color gradients within each column indicate the most influential (gray) to least influential (white) variables. The Industry Total refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown. Characteristics are ordered based on their ranks of the Total case, with the most influential characteristics on the top and the least influential on the bottom. The variable importance within each industry group is normalized to the sum of one.

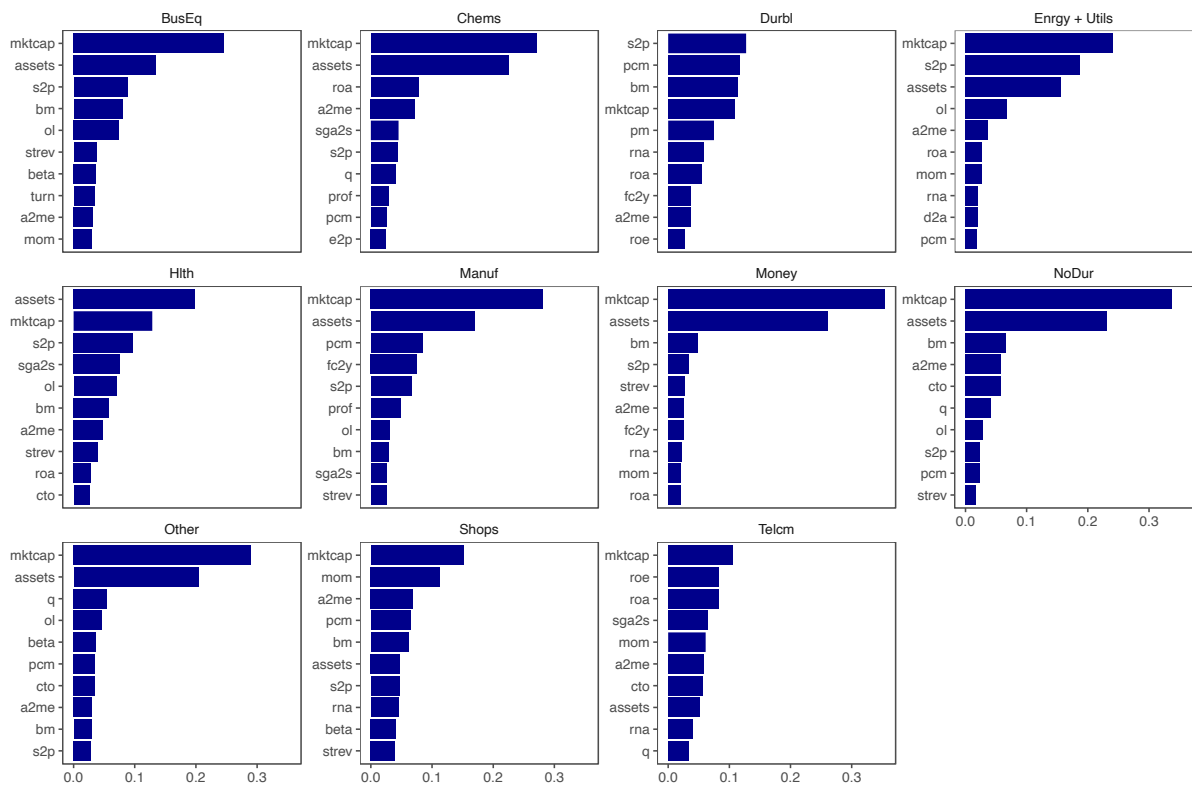
mktcap	0.25	0.27	0.11	0.24	0.13	0.28	0.35	0.34	0.29	0.15	0.1	0.23	0.38
assets	0.13	0.23	0.03	0.16	0.2	0.17	0.26	0.23	0.2	0.05	0.05	0.15	0.22
beta	0.04	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.02	0.07
strev	0.04	0.01	0.01	0.02	0.04	0.03	0.03	0.02	0.02	0.04	0.01	0.02	0.06
mom	0.03	0	0.01	0.03	0.01	0.01	0.02	0	0.01	0.11	0.06	0.03	0.04
turn	0.03	0	0.01	0.01	0.01	0.01	0.01	0	0.02	0.02	0.01	0.01	0.04
a2me	0.03	0.07	0.04	0.04	0.05	0.01	0.03	0.06	0.03	0.07	0.06	0.04	0.02
w52h	0.01	0	0	0	0.01	0	0	0	0.01	0.02	0.03	0.01	0.02
cto	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.06	0.04	0	0.06	0.02	0.02
ol	0.07	0.02	0.03	0.07	0.07	0.03	0	0.03	0.05	0.01	0.02	0.04	0.01
ltrev	0.01	0	0	0.01	0	0	0	0	0.01	0.01	0	0	0.01
fc2y	0.02	0.01	0.04	0.01	0.02	0.08	0.02	0.02	0.02	0.02	0.01	0.02	0.01
roa	0	0.08	0.06	0.03	0.03	0.01	0.02	0.01	0.03	0.01	0.08	0.03	0.01
suv	0.01	0	0	0	0	0	0	0	0.01	0	0	0	0.01
pcm	0.03	0.03	0.12	0.02	0.01	0.08	0.01	0.02	0.04	0.06	0.02	0.04	0.01
s2p	0.09	0.04	0.13	0.19	0.1	0.07	0.03	0.02	0.03	0.05	0.02	0.06	0.01
idiov	0	0	0.01	0.01	0	0	0	0	0.01	0.01	0.01	0.01	0.01
bm	0.08	0.02	0.11	0.01	0.06	0.03	0.05	0.07	0.03	0.06	0.03	0.05	0.01
bidask	0.01	0	0	0	0.01	0	0	0	0	0	0.02	0.01	0
e2p	0	0.02	0.01	0.01	0.01	0	0	0	0.01	0.01	0.02	0.01	0
noa	0.01	0.01	0	0	0.01	0.01	0.01	0	0	0.01	0.02	0.01	0
ato	0	0.01	0.02	0.01	0	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0
sga2s	0.03	0.05	0.02	0.01	0.08	0.03	0.01	0.01	0	0.03	0.06	0.03	0
intmom	0.01	0	0	0.01	0.01	0	0.01	0	0	0.02	0	0.01	0
roe	0	0	0.03	0.01	0.02	0.01	0	0.01	0	0.02	0.08	0.02	0
c	0.01	0	0	0	0.01	0	0	0	0	0.01	0.01	0	0
invest	0	0	0	0	0	0	0	0	0.01	0.02	0.02	0.01	0
pm	0	0	0.07	0.01	0.02	0.01	0.01	0	0.02	0.01	0.03	0.02	0
rna	0	0.02	0.06	0.02	0	0.01	0.02	0.01	0	0.04	0.04	0.02	0
prof	0.01	0.03	0	0	0.02	0.05	0	0.01	0.01	0.03	0.02	0.02	0
d2a	0	0	0.01	0.02	0	0	0	0	0	0	0	0	0
q	0.03	0.04	0.02	0	0.01	0.01	0.01	0.04	0.05	0.03	0.03	0.03	0
lev	0	0.01	0.01	0	0.01	0	0	0	0	0.01	0.01	0.01	0
freecf	0	0	0	0	0.01	0	0	0	0	0.01	0.01	0	0
dpi2a	0	0	0.01	0	0	0.01	0.01	0	0	0	0.01	0	0
oa	0	0	0	0.01	0	0	0.01	0	0	0	0	0	0
	BusEq	Chems	Durbl	Energy + Utilis	Hlth	Manuf	Money	NoDur	Other	Shops	Telcm	Industry Total	Total

Taking all industry groups together (Industry Total), the five most important characteristics are market capitalization (mktcap), total assets (assets), sales-to-price (s2p), book-to-market equity (bm), and assets-to-market equity (a2me). Compared with the Total case, market capitalization, and total assets are the two most important characteristics in both

cases. However, only three of the first five most important characteristics and only five of the first ten most important characteristics are shared between the two cases. We can also see that the most important characteristics vary by industry groups. Although, in nine of 11 eleven industry groups, market capitalization is the most important characteristic. The second most important characteristic is one of the following six different characteristics: total assets (assets), market capitalization (mktcap), momentum (mom), price-to-cost margin (pcm), sales-to-price, or return-on-equity (roe). Figure 4 shows the rank correlation of the characteristic importance between the different industry groups. The rank correlation scores range from 0.3 to 0.8, with an average rank correlation of 0.6 between industry groups and 0.55 between Industry Total and Total.

Figure 3: Top 10 Characteristics by Variable Importance

The figure compares firm characteristic importance in deriving the total R2 for the top ten most influential variables in each industry group for the five-factor restricted IPCA model. The variable importance within each industry group is normalized to the sum of one.



Overall, the results indicate that the characteristics that are most important in explaining individual returns vary across industry groups. If we estimate the model separately in each industry group, we obtain different results. However, market capitalization (mktcap) is among the top five, and total assets (assets) is among the top ten most important characteristics in each industry group. But only these two characteristics are among the 10 most important characteristics of any industry group.

Figure 4: Correlation of Characteristic Importance

The figure reports the rank correlation of characteristics importance between each industry group for the five-factor restricted IPCA model. The Industry Total refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown.

Total	0.73	0.31	0.3	0.59	0.52	0.44	0.45	0.47	0.69	0.49	0.27	0.55	1
Industry Total	0.72	0.81	0.84	0.7	0.84	0.82	0.68	0.89	0.78	0.76	0.75	1	0.55
Telcm	0.33	0.59	0.59	0.41	0.68	0.44	0.44	0.61	0.4	0.52	1	0.75	0.27
Shops	0.68	0.55	0.59	0.51	0.56	0.64	0.65	0.61	0.59	1	0.52	0.76	0.49
Other	0.7	0.62	0.66	0.62	0.61	0.63	0.48	0.8	1	0.59	0.4	0.78	0.69
NoDur	0.63	0.81	0.75	0.57	0.73	0.75	0.58	1	0.8	0.61	0.61	0.89	0.47
Money	0.48	0.6	0.64	0.59	0.52	0.65	1	0.58	0.48	0.65	0.44	0.68	0.45
Manuf	0.67	0.73	0.73	0.56	0.72	1	0.65	0.75	0.63	0.64	0.44	0.82	0.44
Hlth	0.73	0.69	0.61	0.58	1	0.72	0.52	0.73	0.61	0.56	0.68	0.84	0.52
Enrgy + Utils	0.45	0.62	0.7	1	0.58	0.56	0.59	0.57	0.62	0.51	0.41	0.7	0.59
Durbl	0.4	0.69	1	0.7	0.61	0.73	0.64	0.75	0.66	0.59	0.59	0.84	0.3
Chems	0.5	1	0.69	0.62	0.69	0.73	0.6	0.81	0.62	0.55	0.59	0.81	0.31
BusEq	1	0.5	0.4	0.45	0.73	0.67	0.48	0.63	0.7	0.68	0.33	0.72	0.73
	BusEq	Chems	Durbl	Enrgy + Utils	Hlth	Manuf	Money	NoDur	Other	Shops	Telcm	Industry Total	Total

3.3.1. Robustness

So far, I have estimated the five-factor restricted IPCA model for the entire sample in each industry group and determined the importance of each characteristic. However, the importance of these characteristics may vary over time. If we evaluate them for different months, we may obtain different outcomes. In order to get a more robust result, I recursively estimated the five-factor restricted IPCA model for each month where I have observations for at least 120 months, similar to the out-of-sample estimation. Afterward, I determined the relative importance of each characteristic for every month and industry group, using the same approach as previously done. Figure 5 shows the time series average of the importance of each characteristic for all industry groups, and Figure 6 displays the ten most important characteristics by industry group.

Figure 5: Time Series Average Characteristics Importance

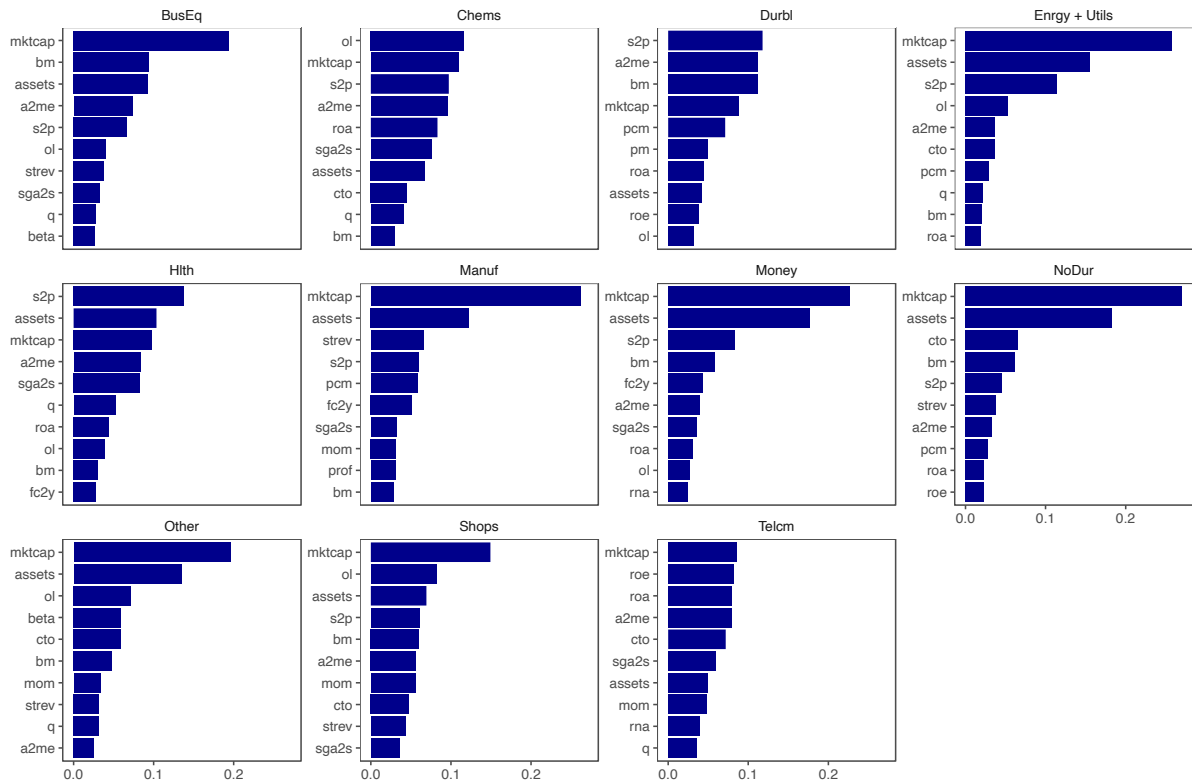
The figure reports the ranking of the time series average of firm characteristics importance for each industry group for the five-factor restricted IPCA model. Columns correspond to each industry group, and color gradients within each column indicate the most influential (gray) to least influential (white) variables. The Industry Total refers to the case where the IPCA model's performance is evaluated for all industry groups taken together, and Total refers to the aggregate case where we consider all observations and do not apply any industry breakdown. Characteristics are ordered based on their ranks of the Total case, with the most influential characteristics on the top and the least influential on the bottom. The variable importance within each industry group is normalized to the sum of one.

mktcap	0.19	0.11	0.09	0.26	0.1	0.26	0.23	0.27	0.2	0.15	0.09	0.19	0.33
assets	0.09	0.07	0.04	0.16	0.1	0.12	0.18	0.18	0.13	0.07	0.05	0.11	0.15
strev	0.04	0.01	0.02	0.02	0.02	0.07	0.02	0.04	0.03	0.04	0.01	0.03	0.09
beta	0.03	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.06	0.02	0.01	0.02	0.07
mom	0.02	0.01	0.02	0.02	0.01	0.03	0.01	0.01	0.03	0.06	0.05	0.02	0.07
turn	0.02	0	0.01	0.01	0	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.03
a2me	0.07	0.1	0.11	0.04	0.08	0.02	0.04	0.03	0.03	0.06	0.08	0.06	0.03
pcm	0.02	0.03	0.07	0.03	0.03	0.06	0.02	0.03	0.02	0.02	0.01	0.03	0.02
sga2s	0.03	0.08	0.03	0.01	0.08	0.03	0.04	0.02	0.02	0.04	0.06	0.04	0.02
fc2y	0.02	0.02	0.03	0.01	0.03	0.05	0.04	0.01	0.01	0.01	0.01	0.02	0.02
ltrev	0.01	0	0.01	0.01	0	0.01	0	0	0.01	0.01	0	0.01	0.02
bidask	0.01	0	0.01	0.01	0.01	0.01	0	0	0.01	0	0.02	0.01	0.02
ol	0.04	0.12	0.03	0.05	0.04	0.02	0.03	0.02	0.07	0.08	0.02	0.05	0.02
cto	0.01	0.05	0.02	0.04	0.02	0.01	0.02	0.07	0.06	0.05	0.07	0.03	0.01
s2p	0.07	0.1	0.12	0.11	0.14	0.06	0.08	0.05	0.02	0.06	0.03	0.07	0.01
w52h	0.01	0	0.01	0	0.01	0.01	0	0	0.01	0.01	0.02	0.01	0.01
bm	0.09	0.03	0.11	0.02	0.03	0.03	0.06	0.06	0.05	0.06	0.02	0.05	0.01
suv	0	0	0	0	0	0.01	0	0	0.01	0	0	0	0.01
roe	0.02	0.02	0.04	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0.08	0.02	0.01
pm	0.01	0.01	0.05	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01
roa	0.02	0.08	0.04	0.02	0.04	0.02	0.03	0.02	0.02	0.03	0.08	0.03	0.01
q	0.03	0.04	0.03	0.02	0.05	0.01	0.02	0.02	0.03	0.03	0.04	0.03	0.01
intmom	0.01	0	0.01	0.01	0.01	0	0.01	0	0.01	0.01	0	0.01	0.01
idiovol	0	0	0	0.01	0.01	0	0	0	0.01	0.01	0.01	0.01	0.01
e2p	0.01	0.02	0.02	0.01	0.01	0	0.01	0.01	0.01	0.01	0.02	0.01	0.01
prof	0.03	0.02	0.01	0.01	0.02	0.03	0.01	0.02	0.02	0.02	0.02	0.02	0
ato	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.03	0.01	0
invest	0.01	0.01	0.01	0.01	0	0.01	0	0	0.01	0.01	0.02	0.01	0
lev	0.01	0.01	0	0.01	0.01	0	0.01	0	0.01	0.01	0.01	0.01	0
rna	0.02	0.02	0.02	0.01	0.02	0.01	0.02	0.02	0.01	0.02	0.04	0.02	0
noa	0	0.01	0.01	0.01	0.01	0.01	0.02	0	0.01	0.02	0.03	0.01	0
d2a	0	0.01	0	0.01	0	0	0.01	0	0.01	0	0.01	0	0
c	0	0	0	0	0.01	0	0	0	0	0.01	0.01	0	0
freecf	0	0	0	0	0.01	0	0	0	0	0	0.01	0	0
dpi2a	0	0	0.01	0	0.01	0.01	0.02	0	0	0	0.01	0.01	0
oa	0	0	0	0	0	0	0.01	0	0	0	0	0	0
	BusEq	Chems	Durbl	Enrgy + Utilis	Hlth	Manuf	Money	NoDur	Other	Shops	Telcm	Industry Total	Total

Overall, my previous findings are robust. Although the order of the characteristics importance has changed in each case, the main conclusions are the same. Taking all industry groups together (Industry Total), the five most important characteristics have not changed, and the comparison with the Total case is also the same. Furthermore, in eight out of the 11 industry groups, the most important characteristic is market capitalization (mktcap), which is in the top five most important characteristics in each case. In addition, the only other characteristics that are in the top ten most important ones in each industry group is also the total assets (assets). Finally, except for these two characteristics, the most important characteristics vary also across industry groups.

Figure 6: Top 10 Characteristics by Time Series Average of Variable Importance

The figure compares the time series average of firm characteristic importance in deriving the total R2 for the top ten most influential variables in each industry group for the five-factor restricted IPCA model. The variable importance within each industry group is normalized to the sum of one.



4. CONCLUSION

This paper applies the IPCA model to determine whether different firm characteristics drive asset prices in different industries. First, I estimated the IPCA model separately for each industry group and investigated its in-sample and out-of-sample performances. The findings indicate that by increasing the number of factors, on average, IPCA can effectively describe risk compensation in expected returns through the exposure of instrumented loadings without relying on an anomaly intercept. Additionally, estimating the model by industry groups proves to be more effective in describing the common variation of individual returns, which confirms the homogeneity of stocks within the same industry. The IPCA model produces lower Sharp ratios compared to the case when the model is estimated without applying any industry breakdown. There are also significant differences between industries in both in-sample and out-of-sample performance. Moreover, the estimated parameters also differ across industry groups, highlighting heterogeneity across them.

Second, the results show that the importance of firm characteristics in explaining the common variation of returns is different for each industry group. However, market capitalization and total assets are one of the most important characteristics in all industry groups. Taking all industries together, the five most important characteristics are market capitalization, total assets, sales-to-price, book-to-market equity, and assets-to-market equity if the model is estimated separately by industry groups. Furthermore, only three out of these characteristics are among the five most important ones when the model is estimated without any industry breakdown, and analyzing the ten most important characteristics, only five are shared between the two cases. Finally, these results are robust over time, as the main findings are unchanged by analyzing the time series average of characteristic importances.

This study suggests two areas for future research. Firstly, while our study has incorporated 36 characteristics to construct the latent factor model, there is a rapid increase in the number of predictive characteristics related to risk premium and mispricing. Thus, incorporating additional characteristics from diverse channels into IPCA could enhance the model's performance and identification of a latent factor structure. Secondly, it would be more informative to determine which characteristics significantly explain stock returns using the assets pricing test proposed by Kelly et al. (2019) and analyze how the performance changes if we re-estimated the model, using only the significant characteristics.

REFERENCES

- Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61(1), 259–299.
<https://doi.org/10.1111/j.1540-6261.2006.00836.x>
- Balakrishnan, K., Bartov, E., & Faurel, L. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics*, 50(1), 20–41.
<https://doi.org/10.1016/j.jacceco.2009.12.002>
- Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. V. (2015). Deflating profitability. *Journal of Financial Economics*, 117(2), 225–248.
<https://doi.org/10.1016/j.jfineco.2015.02.004>
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3–18. [https://doi.org/10.1016/0304-405X\(81\)90018-0](https://doi.org/10.1016/0304-405X(81)90018-0)
- Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks. *Journal of Financial Economics*, 12(1), 129–156.
[https://doi.org/10.1016/0304-405X\(83\)90031-4](https://doi.org/10.1016/0304-405X(83)90031-4)
- Bhandari, L. C. (1988). Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. *The Journal of Finance*, 43(2), 507–528. <https://doi.org/10.1111/j.1540-6261.1988.tb03952.x>
- Bustamante, M. C., & Donangelo, A. (2017). Product Market Competition and Industry Returns. *The Review of Financial Studies*, 30(12), 4216–4266.
<https://doi.org/10.1093/rfs/hhx033>
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), 57–82. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Chamberlain, G. (1983). Funds, Factors, and Diversification in Arbitrage Pricing Models. *Econometrica*, 51(5), 1305. <https://doi.org/10.2307/1912276>
- Chan, L. K. C., Lakonishok, J., & Swaminathan, B. (2007). Industry Classifications and Return Comovement. *Financial Analysts Journal*, 63(6), 56–70.
<https://doi.org/10.2469/faj.v63.n6.4927>
- Chou, P.-H., Ho, P.-H., & Ko, K.-C. (2012). Do industries matter in explaining stock returns and asset-pricing anomalies? *Journal of Banking & Finance*, 36(2), 355–370.
<https://doi.org/10.1016/j.jbankfin.2011.07.016>
- Chung, K. H., & Zhang, H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets*, 17, 94–120.
<https://doi.org/10.1016/j.finmar.2013.02.004>
- Cochrane, J. H. (2011). Presidential Address: Discount Rates. *The Journal of Finance*, 66(4), 1047–1108. <https://doi.org/10.1111/j.1540-6261.2011.01671.x>
- Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT. *Journal of Financial Economics*, 21(2), 255–289. [https://doi.org/10.1016/0304-405X\(88\)90062-1](https://doi.org/10.1016/0304-405X(88)90062-1)

- Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance*, 63(4), 1609–1651. <https://doi.org/10.1111/j.1540-6261.2008.01370.x>
- D’Acunto, F., Liu, R., Pflueger, C., & Weber, M. (2018). Flexible prices and leverage. *Journal of Financial Economics*, 129(1), 46–68. <https://doi.org/10.1016/j.jfineco.2018.03.009>
- Datar, V. T., Y. Naik, N., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2), 203–219. [https://doi.org/10.1016/S1386-4181\(97\)00004-9](https://doi.org/10.1016/S1386-4181(97)00004-9)
- Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997. *The Journal of Finance*, 55(1), 389–406. <https://doi.org/10.1111/0022-1082.00209>
- De Bondt, W. F. M., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, 40(3), 793–805. <https://doi.org/10.1111/j.1540-6261.1985.tb05004.x>
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- Fama, E. F., & French, K. R. (2016). Dissecting Anomalies with a Five-Factor Model. *Review of Financial Studies*, 29(1), 69–103. <https://doi.org/10.1093/rfs/hhv043>
- Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1), 1–25. <https://doi.org/10.1016/j.jfineco.2013.10.005>
- Freyberger, J., Neuhierl, A., & Weber, M. (2017). Dissecting Characteristics Nonparametrically (w23227; p. w23227). National Bureau of Economic Research. <https://doi.org/10.3386/w23227>
- Gandhi, P., & Lustig, H. (2015). Size Anomalies in U.S. Bank Stock Returns. *The Journal of Finance*, 70(2), 733–768. <https://doi.org/10.1111/jofi.12235>
- Garfinkel, J. A. (2009). Measuring Investors’ Opinion Divergence. *Journal of Accounting Research*, 47(5), 1317–1348. <https://doi.org/10.1111/j.1475-679X.2009.00344.x>
- George, T. J., & Hwang, C. (2004). The 52-Week High and Momentum Investing. *The Journal of Finance*, 59(5), 2145–2176. <https://doi.org/10.1111/j.1540-6261.2004.00695.x>
- Gorodnichenko, Y., & Weber, M. (2016). Are Sticky Prices Costly? Evidence from the Stock Market. *American Economic Review*, 106(1), 165–199. <https://doi.org/10.1257/aer.20131513>
- Goyal, A. (2012). Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management*, 26(1), 3–38. <https://doi.org/10.1007/s11408-011-0177-7>
- Green, J., Hand, J. R. M., & Zhang, X. F. (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, 30(12), 4389–4436. <https://doi.org/10.1093/rfs/hhx019>
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5), 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>

- Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, 222(1), 429–450. <https://doi.org/10.1016/j.jeconom.2020.07.009>
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *Review of Financial Studies*, 29(1), 5–68. <https://doi.org/10.1093/rfs/hhv059>
- Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3), 401–439. [https://doi.org/10.1016/0304-405X\(95\)00868-F](https://doi.org/10.1016/0304-405X(95)00868-F)
- Hirshleifer, D., Kewei Hou, Teoh, S. H., & Yinglei Zhang. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38, 297–331. <https://doi.org/10.1016/j.jacceco.2004.10.002>
- Hou, K., Karolyi, G. A., & Kho, B.-C. (2011). What Factors Drive Global Stock Returns? *Review of Financial Studies*, 24(8), 2527–2574. <https://doi.org/10.1093/rfs/hhr013>
- Hou, K., & Robinson, D. T. (2006). Industry Concentration and Average Stock Returns. *The Journal of Finance*, 61(4), 1927–1956. <https://doi.org/10.1111/j.1540-6261.2006.00893.x>
- Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *The Journal of Finance*, 45(3), 881–898. <https://doi.org/10.1111/j.1540-6261.1990.tb05110.x>
- Kan, R., Wang, X., & Zheng, X. (2019). In-Sample and Out-of-Sample Sharpe Ratios of Multi-Factor Asset Pricing Models. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3454628>
- Kelly, B. T., Pruitt, S., & Su, Y. (2017). Instrumented Principal Component Analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2983919>
- Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501–524. <https://doi.org/10.1016/j.jfineco.2019.05.001>
- Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting Factor Models. *The Journal of Finance*, 73(3), 1183–1223. <https://doi.org/10.1111/jofi.12612>
- Lewellen, J. (2015). The Cross-section of Expected Stock Returns. *Critical Finance Review*, 4(1), 1–44. <https://doi.org/10.1561/104.00000024>
- Lyandres, E., Sun, L., & Zhang, L. (2008). The New Issues Puzzle: Testing the Investment-Based Explanation. *Review of Financial Studies*, 21(6), 2825–2855. <https://doi.org/10.1093/rfs/hhm058>
- Ma, T., Leong, W. J., & Jiang, F. (2023). A latent factor model for the Chinese stock market. *International Review of Financial Analysis*, 87, 102555. <https://doi.org/10.1016/j.irfa.2023.102555>
- McLean, R. D., & Pontiff, J. (2016). Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance*, 71(1), 5–32. <https://doi.org/10.1111/jofi.12365>
- Moskowitz, T. J., & Grinblatt, M. (1999). Do Industries Explain Momentum? *The Journal of Finance*, 54(4), 1249–1290. <https://doi.org/10.1111/0022-1082.00146>
- Novy-Marx, R. (2011). Operating Leverage*. *Review of Finance*, 15(1), 103–134. <https://doi.org/10.1093/rof/rfq019>
- Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, 103(3), 429–453. <https://doi.org/10.1016/j.jfineco.2011.05.003>

- Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1), 162–185. <https://doi.org/10.1016/j.jfineco.2011.12.009>
- Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11(3), 9–16. <https://doi.org/10.3905/jpm.1985.409007>
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360. [https://doi.org/10.1016/0022-0531\(76\)90046-6](https://doi.org/10.1016/0022-0531(76)90046-6)
- Sloan, R. G. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *The Accounting Review*, 71(3), 289–315. JSTOR.
- Soliman, M. T. (2008). The Use of DuPont Analysis by Market Participants. *The Accounting Review*, 83(3), 823–853. <https://doi.org/10.2308/accr.2008.83.3.823>
- Windmüller, S. (2022). Firm Characteristics and Global Stock Returns: A Conditional Asset Pricing Model. *The Review of Asset Pricing Studies*, 12(2), 447–499. <https://doi.org/10.1093/rapstu/raab024>