EMPIRICAL ASSET PRICING IN LATENT FACTOR MODELS

Master Thesis

Submitted: Jan 28th, 2024

By: Tung-Lin Wu
Peißenberg str.10, 81547 München

Matriculation number: 1824429

Supervisor: Qi Zhang
Reviewer: Prof. Stefan Ruenzi

University of Mannheim
Chair of International Finance

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
BS	Barillas and Shanken Selected 6 Factors Model
CAPM	Capital Asset Pricing Model
FF3	Fama-French 3 Factors Model
FFC4	Fama-French Carhart 4 Factors Model
FF5	Fama-French 5 Factors Model
FFC6	Fama-French Carhart 6 Factors Model
HXZ	Hou et al. Q Factor Model
IPCA	Instrumented Principal Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory
PCA	Principal Component Analysis
RNN	Recurrent Neural Network
SY	Stambaugh and Yuan 4 Factors Model

List of Symbols

Symbols	Definition
α	Loadings of intercept
β	Loadings of systemic risk factors
γ	Intercept of LASSO regression
δ	Loadings of pre-specified systemic risk factors
ε	individual firm residual
θ	Coefficient of macroeconomic factor in LASSO regression
Γ	Matrix of mapping
υ	Intercept of α, β, *g*
$\boldsymbol{\Theta}$	Sum of θ
b (LSTM)	Bias term
c	(*L*-1 X 1) vector of firm characteristics
C (LSTM)	Cell state
$\tilde{C}$ (LSTM)	Cell gate
d	Residual of managed portfolio
f	Systematic risk factor for instruments
f (LSTM)	Activation function
G	Pre-specified systematic risk facto
h (LSTM)	Hidden State
I (LSTM)	Input Gate
J	Amount of macroeconomic features
k	Error term for LASSO regression
L	Amount of firm characteristics and a constant
N	Amount of stocks
O (LSTM)	Output gate

q	Random time index
r	Return of individual stock
t	Time (Month)
W (LSTM)	Weight
x	Return of managed-portfolio
X (LSTM)	Input
t	Instrumented Principal Component Analysis
Y	J x 1 vector of macroeconomic features
Y (LSTM)	Output
z	L x 1 vector of instruments and a constant

1. Introduction

In the past decades, researchers have strived to explain the behaviors of capital markets, especially the differences in average returns among different assets. To understand the underlying composite of financial products, diverse research approaches have been applied.

Beginning with inspiration from the macroeconomics academia, econometrics built up a general model rooted in the Euler equation for explaining the variation of excess return. The assumption of a Stochastic Discounted Factor (SDF) has therefore emerged within this theoretical framework. It represents a stochastic factor component that captures the relationship between risk and expected return over time and across different assets under the non-arbitrage assumption.

Nevertheless, this belief in an existing common risk factor faces several challenges due to the unobservable attribute of the risk and the risk premiums. Hence, two diverse methodologies have been broadly applied by researchers in order to address this issue. The first involves utilizing pre-specified factors that have already been demonstrated to have a statistically significant influence on stock premiums. One of the most popular examples in this category is the Fama-French 3 Factor Model (1993). In contrast, the second approach employs advanced statistical techniques, such as principal components. This method transforms the original set of variables into a smaller set of uncorrelated components that capture the maximum variance. Via this method, researchers could gain valuable insight without prior knowledge regarding the specific risk factors.

However, besides its convenience and a better fit with data, the static property of PCA makes it unable to capture dynamics over time. Furthermore, the limitation of incorporating additional data beyond returns is a significant drawback when evaluating the conditional dependencies of diversified risk. Kelly et al. (2019) recognize these confinements and improve the PCA model by introducing instrumental loading, allowing flexible and dynamic input. Besides introducing an enhanced structure, Kelly et al. (2019) also compare the performance of the IPCA model to other existing models. The result shows that the IPCA model not only outperforms its peers but is also highly efficient.

In the original paper of Kelly et al. (2019), they incorporate 36 firm characteristics in the IPCA model presented to have superior explanatory power among other existing factor models. Furthermore, it is shown that the incorporation of observable factors could not genuinely improve its performance. This stimulates my interest and leads to my research question: *Whether including features besides firm characteristics can further enhance the overall explanatory power of the IPCA model, and how should they be applied?* Furthermore, a more detailed understanding of the composite of systematic risk could be attained after incorporating these new features. Therefore, the idea of taking macroeconomic features into account is generated after reading through various pieces of literature.

The main goal of this paper is to improve the performance of the IPCA model from Kelly et al. (2019) after including a non-linearly transformed selected macroeconomic factor. A further discussion aims to identify the weakness of the linear model and describe how the newly incorporated macroeconomic features influence returns within the IPCA framework.

The following content is organized as follows: Section 2 provides a comprehensive review of related literature. Section 3 illustrates the model and methodology employed in the study. Section 4 outlines the hypothesis test. In Section 5, a detailed description of the data is provided. Section 6 presents the empirical results, and finally, Section 7 presents the conclusion.

2. Literature Review

The further examination of this paper is referred to or inspired by the understanding of several literatures. Based on the content mentioned in the papers, the literature can be summarized into three primary categories.

2.1 Asset Pricing Model

The initial strand within this category believes that factors are observable, originating from the Capital Asset Pricing Model (CAPM) proposed by Sharpe (1964) and Lintner (1965). This model decomposes the risk of an asset into two components: systematic risk and unsystematic risk. Expanding on this framework, Fama and French (1993, 2013) further develop the notion of systematic risk by introducing market risk, size risk (SMB), value risk (HML), profitability factor (RMW), and investment factor (CMA). Carhart (1997) extends Fama and French's exploration by delving into the performance tendencies of securities and introducing the momentum factor. Hou et al. (2012) draw inspiration from Tobin's Q theory and integrate the market and size factors accompanied by two factors derived from investment and return on equity. Stambaugh and Yuan (2017) subsequently refined the process of ranking characteristic to achieve a less noisy measure of stock pricing. Finally, Barillas and Shanken (2018) consolidate the existing comprehension of pricing model. By employing statistical methods and comparing different models, they encapsulate a six-factor model and are deemed as the most comprehensive description of asset pricing dynamics. This strand underscores the development based on prior empirical experience, simultaneously advantageous with enhanced economic intuition.

Diverging from the first strand, the second strand of literature constructs factors through the application of latent factor analysis directly on collected datasets. This approach is grounded in the foundation of Ross's (1976) Arbitrage Pricing Theory, which expands the CAPM into a multifactor model. Within this framework, the expected return on an asset class is linked to its exposure to various systematic risk factors, operating under the non-arbitrage assumption. Chamberlain and Rotshield (1982) as well as Conor and Korajczyk (1985) employ principal components analysis (PCA) on returns. A notable advantage inherent in this approach is it does not necessitate ex-ante empirical knowledge.

In contrast to solely employing return data, an alternative strand of literature integrates asset-specific or firm-specific characteristics. Rosenberg (1974) offers a foundational theoretical analysis of factor exposure as a function of factor characteristics. Ferson and Harvey (1991) introduce dynamic betas, considering asset-specificity and size-grouped functions about macroeconomic variables. Both bodies of literature either assume or depend upon observable features.

The fourth strand of literature attempts to assemble high dimensional series of characteristics by applying statistical techniques that are designed for dimension reduction. Starting with Lewellen (2015), investigates the combined predictive power of 3, 7, and 15 predictors in OLS regression. On the other hand, Light et al. (2016) apply partial least squares to derive a common factor from 26 firm characteristics, while Freyberger et al. (2017) utilize LASSO for factor selection from a set of 36 characteristics. Kozak et al. (2018) further discovered that the principal components which are extracted from 15 anomaly portfolios (originating from Novy-Marx et al. (2014)) exhibit comparable explanatory power to those portfolios with insignificant alphas. However, Kelly et al. (2019) highlight the limitations of the PCA method, emphasizing its static nature and shortcomings with alternative sets of test assets. They propose an improvement through the introduction of Instrumented Principal Component Analysis (IPCA), enabling to capture the transitions in asset loadings as characteristics evolve.

In conclusion, the IPCA model stands out from other methodologies by incorporating firm characteristic and beta models and simultaneously estimating the risk premium. This method provides a highly efficient estimation process and offers clear insight into the contribution of the feature. Additionally, they contribute a robust framework that allows for further extension by incorporating alternative characteristics. This criterion allows the model to select pricing factors based on its incremental variance relative to total variance and subsequently assess whether the composites of loadings account for explaining variations in average returns.

2.2 Macroeconomics Characteristics

This segment of the literature is divided into individuals' responses to macroeconomic information and the examination of the interaction between economic information and assets.

In the first realm, Lucas (1976) posits that individuals rationally adapt their expectations by taking into account all accessible economic information and their understanding of the economic structure. In contrast, Mankiw and Reis (2001) provide empirical evidence supporting the stickiness of information distribution and reaction, while Sims (2003) deviates from the assumption of rational behavior by proposing that individuals often contend with information-processing constraints. Malmendier and Nagel (2016) demonstrate that individuals' expectations of inflation are highly influenced by their life experiences. Subsequently, Nagel and Xu (2019) further extend the discourse by incorporating asset pricing considerations, demonstrating that the fading memory of data contributes to induced volatility in asset prices.

Besides the content mentioned above, the relationship between macroeconomic information and financial markets is also widely studied, with literature investigating various macroeconomic indicators. For instance, Pearce and Roley (1983) observe that the announcement of an unexpected change in money supply is accompanied by the change in stock price. Chen et al. (1986), conducted empirical research demonstrating that the pricing of stocks aligns with the exposures of stock to 8 economic variables. Additionally, Patelis (1997) found that monetary policy indicators impact excess stock returns. Yet, Chan et al. (1998) presented evidence which contradicts these findings. Flannery and Aris (2002) further validate the influence of economic factors on aggregate stock returns and volatility by estimating a GARCH model, where realized daily return and conditional volatility depend on 17 macroeconomic series. Ludvigson and Ng (2005) extended the previous research by encompassing hundreds of macroeconomic series then extracting a small set of principal components. However, Chen et al. (2019) highlight a limitation of Principal Component Analysis (PCA), noting that it is unable to capture the current state of an economic system that relies on dynamics. Hence, they exploit Recurrent Neural Networks containing Long-Short-Term-Memory cells.to deal with large dimensional macroeconomic data and allow partial dependency on long-term memory. It is worth noting that both papers find a significant relationship between excess return and economic indicators.

2.3 Factor selection

This subsection summarizes the additional literature related to factor selection that is applied in this paper.

As mentioned in the previous segment, one of the classic methods is PCA. It reduces the dimensionality by extracting a smaller number of principal components that contain the most relevant information from the original data. Nonetheless, PCA does not mainly serve as feature selection since its main purpose is to create a minor subset of data while retaining as much variance as possible from the original dataset. As opposed to PCA, the least absolute shrinkage and selection operator (LASSO) published by Tibshirani (1996) promotes model regularization and sparsity by including a penalty on the coefficient within the linear regression model. This characteristic makes LASSO a method specifically designed for feature selection and is broadly exploited in numerous fields. For example, AI (2022) selects the cancer genomics feature based on LASSO, while Li et al. (2022) employ it to predict the stock market volatility via selecting factors from high dimensional predictors. Furthermore, LASSO exhibits flexibility to integrate with other models. One of the examples is Yang et al. (2022) combining LASSO and LSTM to predict stock movement. The integration of these two models allows for variable selection and the processing of time-series data, similar to the methodology presented in this paper.

3. Model and Methodology

3.1 Instrumented Principal Component Model (IPCA) model

3.1.1 Fundamental IPCA

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \varepsilon_{i,t+1} \tag{2}$$

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

$$z_{i,t} = \ [\ c_{i,t}\,, 1]'$$

Kelly et al. (2019) introduce the general IPCA model for studying the relationship between excess return of individual stock $r_{i,t+1}$ and common risk factors. IPCA model permits the inclusion of time-variant factor loadings $\alpha_{i,t}$ and $\beta_{i,t}$, which respectively represent the anomaly intercepts and systematic risk factors. These factor loadings can be decomposed into a time series of *L x 1* instrument vector $z_{i,t}$, in conjunction with two static matrices Γ_α and Γ_β and intercept $\upsilon_{\alpha,i,t}$ and $\upsilon_{\beta,i,t}$. These intercepts capture the residual components that are not explained by the specified parameter.

The vector $z_{i,t}$ is derived from the combination of the vector $c'_{i,t}$, which stores the characteristic of individual stocks at time *t,* and a constant. On the other hand, the matrices Γ_α and Γ_β serve as a weighted linear transformation for mapping the large combination of characteristics into a concise risk factor exposure. This mechanism facilitates $\alpha_{i,t}$ and $\beta_{i,t}$ to vary over time driven by the changes in characteristics of each stock. Besides, it enables an extension in the amount and diversity of characteristics that can be considered.

Compared to a normal PCA model, IPCA enhances its capabilities by allowing to capture the evolution of the sensitivity of each stock to the underlying risk factor over time and provides a more intuitive interpretation.

3.1.2 Restricted IPCA (Γ_α = 0) and Unrestricted IPCA ($\Gamma_\alpha \neq$ 0)

The definition provided by Kelly et al. (2019) characterizes an "anomaly" as the compensation of return without bearing associated risk. In the context of IPCA, the objective is to model the expected return of individual firms by incorporating individual-specific characteristics, with the exposure to the systematic risk factor, f_{t+1}, captured through time-varying factor loadings. Therefore, two distinct assumptions are introduced: one allows the possibility of intercept, where $\Gamma_\alpha \neq 0$, while the alternative assumes $\Gamma_\alpha = 0$.

When Γ_α is assumed to be zero, none of the idiosyncratic attributes is mapped into the IPCA model. In this respect, IPCA can be expressed in the stacked form of vectors as follows:

$$r_{t+1} = Z_t \Gamma_\beta f_{t+1} + \varepsilon^*{}_{t+1} \tag{3}$$

where r_{t+1} is an N x 1 vector which stores all individual firm return, Z_t is the *N x L* matrix which stacks all individual firm characteristics, and $\varepsilon *_{t+1}$ is an *Nx1* vector storing individual firm residual.

The objective function for the estimation is to minimize the sum of squared residuals over N assets and T periods:

$$\min_{\Gamma_\beta, f_{t+1}} \sum_{t=1}^{T-1} (r_{t+1} - Z_t \Gamma_\beta f_{t+1})'(r_{t+1} - Z_t \Gamma_\beta f_{t+1}) \tag{4}$$

$$\hat{f}_{t+1} = (\Gamma'_\beta Z_{\prime t} Z_t \Gamma_\beta)^{-1} \Gamma'_\beta Z'_t r_{t+1}, \quad \forall t \tag{5}$$

$$vec(\hat{\Gamma}'_{\beta}) = \left(\sum_{t=1}^{T-1} Z'_t Z_t \otimes \hat{f}_{t+1}\hat{f}'_{t+1}\right)^{-1} \left(\sum_{t=1}^{T-1} [Z'_t \otimes \hat{f}'_{t+1}]' r_{t+1}\right). \quad (6)$$

For restricted IPCA, the objective function is minimized by doing partial derivatives on $\Gamma_{\beta}\ and\ f_{t+1}$ and subsequently solving the resulting first-order conditions to obtain the optimal values for these parameters as shown in equation (5) and (6). These equations resemble to the structure of cross-sectional regression where the return regressed on β_t, and the return regressed on the Kronecker product involving individual characteristics and factors. However, equation (6) involves a complex structure, making it challenging to find a closed-form solution directly. Therefore, Kelly et al. (2019) employ alternative least squares (ALS) to solve this numerical problem.[1]

In contrast, a less constrained formulation allows expected returns to depend on characteristics in a manner not solely explained by exposure to common risk. The stacked form and the objective function are written as follows:

$$r_{t+1} = Z_t\Gamma_{\alpha} + Z_t\Gamma_{\beta}f_{t+1} + \varepsilon^*_{t+1} \quad (7)$$

$$\min_{\Gamma_{\alpha},\Gamma_{\beta},f_{t+1}} \sum_{t=1}^{T-1} (r_{t+1} - Z_t\Gamma_{\beta}f_{t+1} - Z_t\Gamma_{\alpha})'(r_{t+1} - Z_t\Gamma_{\beta}f_{t+1} - Z_t\Gamma_{\alpha}) \quad (8)$$

The model can be rewritten into a cumulative form $r_{t+1} = Z_t\tilde{\Gamma}\tilde{f}_{t+1} + \varepsilon^*_{t+1}$ where $\Gamma = [\,\Gamma_{\alpha}, \Gamma_{\beta}]$,and $\tilde{f}_{t+1} = [\,1, f_{t+1}]$. The first-order condition of Γ is similar to the restricted model shown in equation (6) except replacing f_{t+1} with $\tilde{f}_{t+1}$, while the estimation of f_{t+1} changes into:

$$\hat{f}_{t+1} = (\Gamma'_{\beta}Z'_tZ_t\Gamma_{\beta})^{-1}\Gamma_{\beta}'Z'_t(r_{t+1} - Z_t\Gamma_{\alpha}), \quad \forall t \quad (9)$$

[1] See, Kelly et al. (2019) for more mathematical details.

Details regarding the selection between the two models will be discussed in the following sections, accompanied with empirical results.

3.1.3 Managed Portfolio Interpretation of IPCA

Kelly et al. (2019) propose a managed portfolio interpretation for IPCA by introducing a managed portfolio interpretation to derive the initial value for Alternating Least Squares (ALS) estimation. This idea is rooted in the complex structure of IPCA. Beginning with the restricted IPCA, the original objective function (4) is modified by substituting the first-order condition (5) into it, resulting in a concentrated function for Γ_β:

$$\max_{\Gamma_\beta} tr\left(\sum_{t=1}^{T-1}(\Gamma'_\beta Z'_t Z_t \Gamma_\beta)^{-1}\Gamma'_\beta Z_t r_{t+1} r'_{t+1} Z_t \Gamma_\beta)\right) \tag{10}$$

In contrast to a PCA model, where *β* remains static, IPCA introduces $\Gamma'_\beta Z'_t Z_t \Gamma_\beta$ term that varies over time. This dynamic attribute brings the challenge finding a single eigenvector solution for Γ_β. Nonetheless, the similarity between static and dynamic problems brings inspiration for the solution. Similar to PCA, instead of applying singular value decomposition (SVD) directly on return matrix r_{t+1}, Kelly et al. (2019) demonstrate that the objective function of IPCA can be solved approximately by employing SVD on a combined term where characteristics interact with the return.

$$x_{t+1} = \frac{Z'_t r_{t+1}}{N_{t+1}} \tag{11}$$

As illustrated in equation (11), x_{t+1} denotes a 1 *x L* vector of the return of *L* managed portfolios at month t+1.On the right-hand side, Z_t is a *N x L* matrix containing *L* characteristics of *N* firms, r_{t+1} is the return of *N* firms, and N_{t+1} presents the non-missing data for each *t* +1 month. One of the highlights of this equation is that the instruments Z_t are transformed into $[-0.5, 0.5]$ range.

This transformation aligns the managed portfolio with a long-short portfolio, resembling a traditional anomaly portfolio.

In order to approximately solve the objective function, Kelly et. al (2019) replace part of the denominators $Z_t'Z_t$ to a related constant. For instance, the average of $Z_t'Z_t$, given that $Z_t'Z_t$ is not too volatile. Since the constant is stable, the solution of Γ_β can be approximately set as the first *K* eigenvectors of the variance-covariance matrix $X'X = \sum_{t=1}^{T} x_t x'_t$. Simultaneously, the estimation of f_{t+1} involves capturing the first *K* principal components which explains the most variability of the managed-portfolio panel.

In summary, the characteristics of managed-portfolio interpretation offer several advantages. Firstly, it simplifies the treatment of unbalanced panel data by focusing solely on non-missing values each month. This facilitates a more straightforward handling of data gaps. Secondly, this approach can be seen as a time-series cross-section regression. When *K* < *L*, the estimation involves a constrained regression, resulting in a reduction in rank while still capable of explaining a significant portion of variability. Lastly, the characteristics-managed portfolio is equivalent to using individual returns as the test asset, providing the advantage of lower dimensionality, and averaging out a portion of the idiosyncratic risk.

3.1.4 Extended form of IPCA

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \delta_{i,t} g_{t+1} + \varepsilon_{i,t+1} \tag{12}$$

$$\alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \upsilon_{\alpha,i,t}, \quad \beta_{i,t} = z'_{i,t}\Gamma_\beta + \upsilon_{\beta,i,t}, \quad \delta_{i,t} = z'_{i,t}\Gamma_\delta + \upsilon_{\delta,i,t} \tag{13}$$

The IPCA framework exhibits flexibility, enabling extensions within its general structure. Besides the common risk factor, f_{t+1} , derived from the characteristics of individual firms, the extended form further incorporates the pre-specified systematic risk factor $g_{i,t+1}$. Compared to f_{t+1}, which is estimated from the optimization process of the objective function, $g_{i,t+1}$ represents a *M* ×1 vector of observed common risk factor that directly interacts with the excess return. Hence, it should have been demonstrated to have explanatory power on excess return in previous empirical studies.

There are only minor differences between the extended form and the fundamental model, mainly from the introduction of $g_{i,t+1}$ and dynamic factor loadings $\delta_{i,t}$. Similar to Γ_α and Γ_β, Γ_δ is a *L* x *M* static matrix that determines how the individual firm characteristics influence the factor loadings. The estimation process is also identical to the fundamental model, and the objective function is now extended as:

$$\min_{\Gamma_\alpha,\Gamma_\beta,\Gamma_\delta,f_{t+1}} \sum_{t=1}^{T-1} (r_{t+1} - Z_t\Gamma_\beta f_{t+1} - Z_t\Gamma_\alpha - Z_t\Gamma_\delta)'(r_{t+1} - Z_t\Gamma_\beta f_{t+1} - Z_t\Gamma_\alpha - Z_t\Gamma_\delta) \tag{14}$$

The model can be expressed in a reformulated manner $r_{t+1} = Z_t\tilde{\Gamma}\tilde{f}_{t+1} + \varepsilon_{t+1}$ where $\Gamma = [\,\Gamma_\alpha, \Gamma_\beta, \Gamma_\delta]$ and $\tilde{f}_{t+1} = [\,1, f_{t+1}, g_{i,t+1}]$. The first-order condition of Γ is similar to the restricted model shown in equation (6) except replacing f_{t+1} with $\tilde{f}_{t+1}$, while the estimation of f_{t+1} changes into:

$$\hat{f}_{t+1} = (\Gamma'_\beta Z'_t Z_t \Gamma_\beta)^{-1}\Gamma'_\beta Z'_t(r_{t+1} - Z_t\Gamma_\alpha - Z_t\Gamma_\delta), \quad \forall t \tag{15}$$

The extended model serves two main purposes within this paper. First, it evaluates the additional explanatory power offered by previously observed factors, such as Fama-French factors, on the fundamental IPCA model, in order to facilitate a straightforward comparison between different pre-specified factors, the fundamental model imposes Γ_α = 0 restriction[2].

[2] Γ_α can be easily added ack to the framework. However, the necessity of the inclusion of alpha will be further discussed

Secondly, the extended model allows the incorporation of general risk factors, such as macroeconomic characteristics. Unlike the firm-specific characteristics, economic factors are not confined to individual firms but could be inclusive components for asset pricing considerations. Furthermore, these factors are regarded as control variables of the overall economic state within this extended framework.

3.2 Preparation for Macroeconomic Characteristics

The primary objective of this paper is to examine whether the compensation for risk exposure to macroeconomic characteristics can account for the unexplained variation in individual firm characteristics. I therefore exploit the attribute of the extended form of IPCA to incorporate observable factors. However, a large set of pure macroeconomic characteristics can introduce noise, redundancy, and potentially multicollinearity into the data. Hence, the combination of the two models is exploited to leverage their strengths. The subsequent sub-section will outline the methodologies, covering factor selection and the approach for extracting the short-term and long-term dynamics of the economic state.

3.2.1 Least Absolute Shrinkage and Selection Operator (LASSO)

$$\tilde{r}_{t+1} = \gamma_t + \sum_{j=1}^{J} \theta_j Y_{j,t} + k_{t+1} \tag{16}$$

While inputting all the information collected for explaining stock return, there is a high probability that exists multicollinearity between different features and includes irrelevant characteristics. LASSO introduces a penalty term and allows sparsity in the estimation of coefficients. In other words, when a feature exhibits multicollinearity or is deemed irrelevant, LASSO effectively shrinks the corresponding coefficient to zero. As shown in equation (16), where $\tilde{r}_{t+1}$ represents the medium of $r_{i,t+1}$ for different *i*, $Y_{j,t}$ denotes the information of different characteristics *j*, θ_j represents the coefficient , k_{t+1} is the error term and γ_t *is the intercept*. The medium of return $r_{i,t+1}$ is employed as the dependent variable since it is less sensitive to outliers and represents the center of cross-sectional data distribution in every time *t.* In general, Equation (16) is similar to OLS, with the main difference existing in the objective function, where LASSO includes a penalty term $\lambda \mid \mathbf{\Theta} \mid$. The mathematical expression is shown in equation (17),and the characteristics is written into a J x1 vector form, $Y_{t=}[\, Y_{1,t}, \, Y_{2,t} \ldots \, Y_{j,t}]$.

$$\min_{\theta} \sum_{t=1}^{T} (\tilde{r}_{t+1} - \gamma_t + \theta Y_t)(\tilde{r}_{t+1} - \gamma_t + \theta Y_t)' + \lambda \mid \mathbf{\Theta} \mid \tag{17}$$

$$\mathbf{\Theta} = \sum_{j=1}^{J} \theta_j \tag{18}$$

The penalty term contains two components. The first is the absolute sum of the coefficient θ_j , encouraging a lower sum of term. The second component λ serves as a tuning parameter by setting the level of penalization. A higher value of λ results in stronger penalization, influencing the sparsity of the estimated coefficients. This mechanism considers only the characteristic that contributes significant relationship, and the irrelevant features would be forced to shrink to 0. That is, one of the significant attributes of LASSO is its ability to automatically select features during the estimation of coefficients.

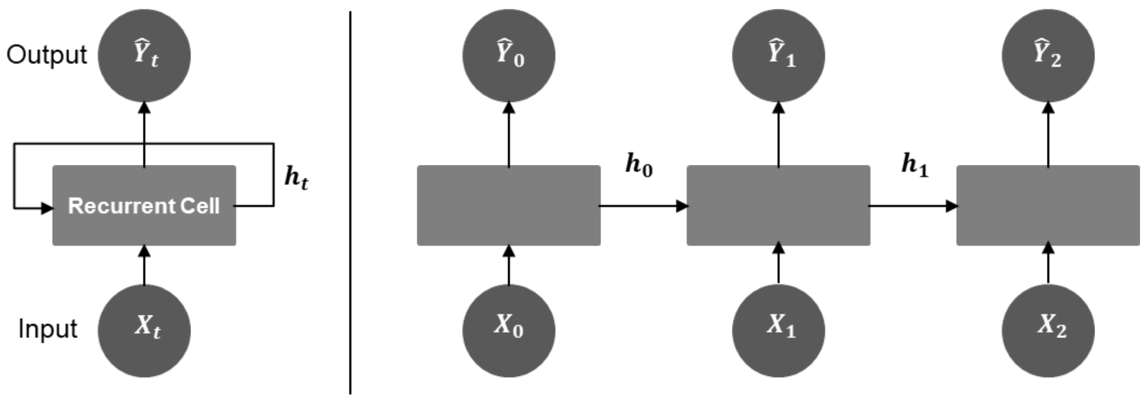
3.2.2 Recurrent Neural Network and Long-Short Term Memory Cell

All Mankiw and Reis’s (2001), Malmendier and Nagel's (2016), and Nagel and Xu's (2019) theories highlight the stickiness of reaction to economic information and the significance of long-term memory are crucial in shaping expectations, whether related to macroeconomic characteristics or asset prices. All the evidence supports the notion that incorporating both long-term and short-term memory is important for a comprehensive understanding of the interaction between economic patterns and asset return.

As suggested by Chen et al. (2021), the attribute of Recurrent Neural Network (RNN) aligns with the requirement for incorporating both long and short-term memory in modeling. The architecture of a vanilla RNN is summarized in Figure 1.

Equation (19) provides a clear expression of each hidden state at time *t.* For each time *t,* the hidden state, which summarizes the information that the network went through, is not only influenced by the current input X_t but also partially by the previously hidden state h_{t-1}. b_h is the bias term that captures the pattern which is not dependent on the input. This recurrent mechanism maintains and captures the dependencies and patterns across the sequential inputs. The determination of how much X_t and h_{t-1} input into the estimation of the current hidden state depends on the weight matrices W_{hx} and W_{hh} which are learned via the model training process.

Figure 1: Recurrent Neural Network (RNN)



This figure presents a vanilla Recurrent neural network and the unfolded RNN. X_t denotes the input, h_t is the hidden state, and $\hat{Y}_t$ represents the output at time t.

The outer symbol f represents the activation function for handling complex and non-linear relationships. These components collectively define a recurrent cell, and an RNN neural network is constructed by connecting these recurrent cells over user-defined time intervals, also known as time steps.

$$h_t = f(W_{hx}X_t + W_{hh}h_{t-1} + b_h) \tag{19}$$

However, an underlying concern exists in the vanilla RNN.The "vanishing gradient" problem occurs during the training process while calculating the gradient and backpropagating to layers. Consequently, the weight will vanish or become negligible, causing the omission of past information in the estimation of the current hidden state. One of the solutions is to use a Long Short-Term Memory (LSTM) cell instead of a basic recurrent cell. The structure of an LSTM cell is as follows:

Figure 2: Long Short-Term Cell

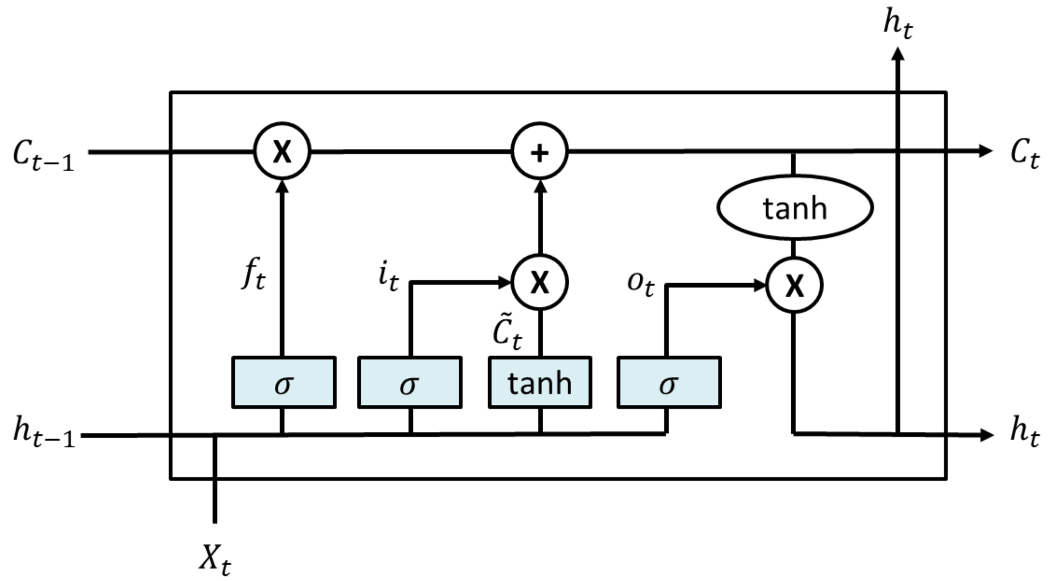


Figure 2 shows the architecture of a LSTM cell. It is constructed with 3 Gates (Forget Gate f_t, Input Gate i_t, Candidate Cell Gate $\tilde{C}_t$, and Output Gate o_t) and 2 States (Hidden State h_t and Cell State C_t).

In contrast to a vanilla recurrent cell, LSTM solves the gradient vanishing issue by employing a gate mechanism and a cell state. These expansions adjust the information that flows into the estimation. In detail, besides the hidden state h_t, the introduction of cell state C_t is a specific storage for retaining long -term data from the defined time step. At the initial phase, input data and the previous hidden state are added into the LSTM cell. The forget gate f_t and input gate i_t would further decide how much portion of the previous cell state C_{t-1} and temporal data $\tilde{C}_t$ should be updated into cell state respectively by applying sigmoid function σ as the gate mechanism. The output gate ultimately decides the level of the current cell state's exposure. Subsequently, it interacts with the activation function *tanh*, then updates the current hidden state. In conclusion, this subsection introduced an RNN neural network augmenting with an LSTM cell which is capable of capturing the long-term and short-term non-linear dynamics present within economics time series.

The examination in this paper makes use of both the LASSO and LASSO-LSTM results. A performance comparison and discussion about the approach choices are further conducted.

4. Asset Pricing Test and Performance Evaluation

4.1 Asset Pricing Test

In this section, four hypothesis tests recommended by Kelly et al. (2019) are introduced. The first test assesses the zero-alpha condition, evaluating the necessity of alpha's existence. The second and third tests, which are similar, aim to determine whether the inclusion of observable factors or macroeconomic features significantly enhances the explanatory power of the IPCA model. For macroeconomic factors, there's an additional test to assess the predictive capability while controlling other significant economic features. The fourth test focuses on the additional significance of individual firm characteristics while controlling for all other instruments.

4.1.1 Testing $\boldsymbol{\Gamma_\alpha}$ = 0

Revisiting in equation (2), $\beta_{i,t}$ captures the sensitivity of return to systematic risk and provides a premium f_{t+1} according to the exposure, while $\alpha_{i,t}$ captures the anomalies unique to each instrument. Therefore, the levels of estimation $\hat{\alpha}_{i,t} = z'_{i,t}\hat{\Gamma}_\alpha$ and $\hat{\beta}_{i,t} = z'_{i,t}\hat{\Gamma}_\beta$ offer insight into how characteristics interface with return and the magnitude of influence exerted by a common risk factor on asset prices.

By setting $\hat{\alpha}_{i,t} = z'_{i,t}\hat{\Gamma}_{\alpha}$ to zero allows for the exclusive examination of the common risk component. If the existence of the idiosyncratic component is necessary, then the restricted model would have a relatively poor fit compared to the unrestricted model. Hence, a test is introduced with the null hypothesis

$$H_0 : \ \Gamma_{\alpha} = 0_{L \times 1}$$

against the alternative hypothesis

$$H_1 : \ \Gamma_{\alpha} \neq 0_{L \times 1}.$$

While the rejection of the null hypothesis doesn't equal to total elimination of the intercept that captures the unsystematic risk. Instead, the restricted model is just a constrained alternative that focuses only on the anomaly expected return on systematic risk to all firms. This logic is advantageous as it allows for a nuanced examination of alpha and clarifies which characteristic is responsible for the rejection.

Additionally, a Wald-type test statistics is applied for calculating the distance between restricted and unrestricted models,

$$W_{\alpha} = \hat{\Gamma}'_{\alpha}\hat{\Gamma}_{\alpha}.$$

The test is processed with residual wild bootstrap. At the initial state, the parameters $\hat{\Gamma}_{\alpha}$, $\hat{\Gamma}_{\beta}$, and $\hat{f}_t$ from the unrestricted model are estimated. Subsequently, the null hypothesis is imposed for every iteration *b*. Since it is equivalent and more efficient according to equation (10), the portfolio residuals are taken into account for resampling instead of stock-level residuals. The utilization of residuals bootstrap stems from the complexity of its attribute. By facilitating the distribution of residuals from model estimation, this approach provides a solution for capturing uncertainty and brings the simulation closer to realism.

$$x_{t+1} = Z'_t r_{t+1} = (Z'_t Z_t)\Gamma_\alpha - (Z'_t Z_t)\Gamma_\beta f_{t+1} - d_{t+1},\ d_{t+1,} \\ = Z'_t \varepsilon^*_{t+1} \tag{20}$$

$$\tilde{x}^b{}_t = (Z'_t Z_t)\hat{\Gamma}_\beta \hat{f}_{t+1} + \tilde{d}^b_t,\ \ \hat{d}^b_t = q^b_1 \tilde{d}_{q^b_2} \tag{21}$$

Starting with equation (20), the definition of managed portfolio with vector form, the estimated residual $\{\tilde{d}_t\}_{t=1}^{T}$ is extracted after the estimation and is input into equation (21) imposing null hypothesis for the next *b*=1,…,1000 iterations. In every iteration, a random time index, which is represented by q^b_2 , would be randomly and uniformly drawn. Additionally, each draw is multiplied by a Student *t* random variable, q^b_1, which has unit variance and five degrees of freedom. This allows different levels of impact for each draw and provides a robust simulation. In the last stage of each *b*, the unrestricted model is re-estimated accompanied by the test statistic $\widetilde{W}^b_\alpha = \hat{\Gamma}'^b_\alpha \hat{\Gamma}'^b_\alpha$. Finally, the *p*-value is calculated as the proportion of $\widetilde{W}^b_\alpha$ that exceeds the W_α from actual data.

4.1.2 Testing extended IPCA Model against fundamental IPCA Model

The extended form of IPCA, introduced in section 3.1.4, has the capability to incorporate observable factors and analyze the risk premium associated with those factors. A test is formulated to assess whether the inclusion of each observable factor yields a statistically significant enhancement in explanatory power compared to the fundamental IPCA model. This test is conducted for both the pre-specified factors (e.g. FF factors) and macroeconomic factors in this paper. The null hypothesis is as follows:

$$H_0 :\ \ \Gamma_\delta = 0_{L\times M}$$

against the alternative hypothesis

$$H_1 : \ \Gamma_\delta \neq 0_{L \times \mathrm{M}}.$$

Similar to the previous test, the Wald-type test statistics for estimated parameter $\hat{\Gamma}_\delta$ is constructed as

$$W_\delta \ = \ \widehat{\Gamma'}_\delta \hat{\Gamma}_\delta.$$

It implies the distance between the extended model and the fundamental model which excludes the observable factors. The test is processed with residual wild bootstrap. Starts from the estimation of the residual from the extended model imposing an alternative hypothesis with managed portfolio form. Subsequently, these estimated residuals are randomly drawn to construct resampled managed portfolios under the imposed null hypothesis. Then, from each bootstrap resample, Γ_δ is re-estimated for constructing W_δ. Finally, the *p*-value is calculated as the proportion of $\widetilde{W}_\delta^b$ that exceeds the W_δ from actual data.

In line with the paper's objective, this significance test can be viewed as a method for selecting factors among macroeconomic features. It aims to identify the singular or grouped factors that genuinely contribute to improvements to the fundamental IPCA model. After applying the significance test for the first time, two scenarios may emerge. In the first scenario, if only one macroeconomic factor demonstrates statistical significance, the test procedure should be halted. In the second scenario, if multiple factors exhibit statistical significance, the identified significant factors will be emphasized, and the test will be restarted from the estimation stage with the joint input of these factors and check whether they bring economically significant improvement, respectively. The procedure iterates continuously, repeating the cycle until all composite or single factors exhibit statistical significance. This approach allows us to delve into the interaction between different macroeconomic factors and avoid the appearance of multicollinearity.

4.1.3 Testing Instrument Significance

This test assesses the significance of individual firm characteristics while controlling other features. By focusing on the restricted model ($\Gamma_\alpha = 0_{L \times 1}$), the significance of the incremental power provided by each firm characteristic could be determined from the results of the test. "It could also be applied to the extended form IPCA model with the same objective, controlling for the impact of macroeconomic factors. The null hypothesis is constructed as below:

$$H_0: \ \Gamma_\beta = \left[\tau_{\beta,1}, \dots, \tau_{\beta,l-1}, 0_{K\times1}, \dots, \tau_{\beta,L}\right]$$

against the alternative hypothesis

$$H_1: \ \Gamma_\beta = \left[\tau_{\beta,1}, \dots, \tau_{\beta,L}\right],$$

where $\tau_{\beta,l}$ represents the *K* x 1 vector factor loadings for *K* factors from characteristic *l.* Hence, the target characteristic is replaced by a *K x 1* zero vector in order to exclude the impact of it. The Wald-type test statistics for estimated parameter $\hat{\Gamma}_\delta$ is therefore constructed as

$$W_{\beta,l} = \hat{c}'_{\beta,l}\hat{c}_{\beta,l}.$$

Like the earlier test, commence by estimating parameters while imposing the alternative hypothesis using a managed portfolio form. Then, extract the residuals for residuals wild bootstrap and resample the managed portfolio under the imposed null hypothesis. Subsequently, re-estimate the alternative model from each bootstrap resample to construct $W_{\beta,l}$. Calculate the p-value by determining the proportion of $\widetilde{W}^b_{\beta,l}$ values surpassing the observed $W_{\beta,l}$.

4.2 Evaluation of Performance

When assessing the performance of restricted, unrestricted, and extended forms of IPCA, two kinds of evaluation methods are taken into consideration. The first measure is the total R^2, written in the form of individual returns[3]:

$$\text{Total } R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1} + \hat{\Gamma}_\delta g_{t+1}))^2}{\sum_{i,t} r_{i,t+1}^2} \tag{21}$$

It represents the fraction of the variation of realized return that accounts for both the time-variant risk premium and conditional loadings, aggregated over all assets and all time. The equation can also be rephrased into a characteristic managed-portfolio form by simply replacing $r_{i,t+1}$ with x_{t+1}.

The other evaluation is predictive R^2, which can also be calculated in individual return or managed portfolio form. It is defined as in equation (21), where $\hat{f}_{t+1}$ and g_{t+1} are substituted by the time-series mean of $\hat{f}_{t+1}$ and $\bar{g}$ respectively, they can be seemed as the average risk premium.

$$\text{Predictive } R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda} + \hat{\Gamma}_\delta \bar{g}))^2}{\sum_{i,t} r_{i,t+1}^2} \tag{21}$$

It represents the fraction of the variation of realized returns explained by the expected risk premium. The distinction between total R^2 and predictive R^2 lies in the use of realized or expected rewards of systematic risk, with the latter incorporating a static interpretation of the risk premium. This approach allows the integration of predictive information into return forecasts solely through the instrumented loadings. Hence, the ideal number of principal components *K* should be determined based on the effectiveness of predictive R^2.

5. Data

[3] $\hat{\Gamma}_\alpha$, $\hat{\Gamma}_\delta g_{t+1}$ can be omitted or included depends on the specific model being employed.

This section provides the details of the source of the raw data and the methodology for data preparation before the final input. According to their content, data could be divided into three categories. The generalized road map of data processing is presented in figure 3 for a clear overview of the procedure.

Figure 3. Expected Road Map for Data Processing

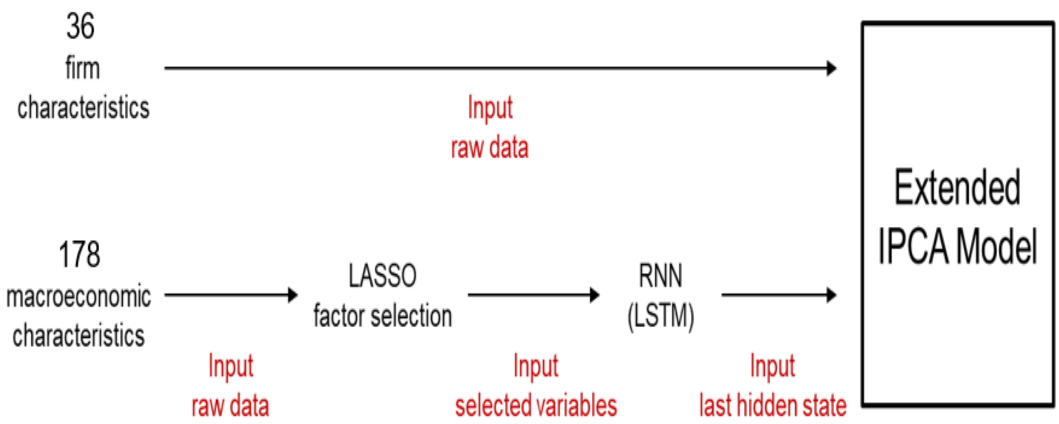


Figure 3 presents the procedure of data processing for firm characteristics and macroeconomics characteristics.

5.1 Stock Return and Firm Characteristics

The stock returns and firm characteristics used in this study are sourced from Kelly et al. (2019)[4], which, in turn, originated from Freyberger et al. (2017). The dataset consists of the return of 12,813 firms includes the following 36 characteristics for each firm: market beta (beta), assets-to-market (a2me), total assets (assets), sales-to-assets (ato), book-to-market (beme), cash-to-short-term-investment (c), capital turnover (cto), capital intensity (d2a), ratio of change in PP&E to change in total assets (dpi2a), earnings-to-price (e2p), fixed costs-to-sales (fc2y), cash flow-to-book (free_cf), idiosyncratic volatility comparing to FF3 model (idio_vol), investment (investment), leverage (lev), market capitalization (mktcap), lagged turnover (turn), net operating assets (noa), operating accruals (oa), operating leverage (ol), price-to-cost margin

[4] Firm characteristics and observable factors data are available on https://sethpruitt.net.

(pcm), profit margin (pm), gross profitability (prof), Tobin's Q (q), closeness to 52-week price (w52h), return on net operating assets (rna), return on assets (roa), return on equity (roe), momentum (mom), intermediate momentum (intmom), short-term reversal (strev), long-term reversal (ltrev), sales-to-price (s2p), SG&A-to-sales (sga2s), bid-ask spread (bidask), and unexplained volume (suv). For further information and summary statistics, you may refer to Freyberger et al. (2017).

In order to mitigate the influence of outliers, the characteristics undergo a cross-sectional standardization. This involves ranking each feature at each period dividing the rank by the total amount of non-missing data then subtracting 0.5 Consequently, this mapping confines the characteristics within the range of $[-0.5, +0.5]$. This standardized approach not only enhances the robustness of the analysis but also provides an additional advantage by facilitating the construction of a long-short portfolio when forming the managed portfolio from stock-level features. The vector of IPCA instrument $z_{i,t}$ = $[1, c_{i,t}]$, including a constant and the characteristic of stock *I* at time *t* represented as $c_{i,t}$.

5.2 Macroeconomics Characteristics

The macroeconomics data are mainly collected from the FRED-MD database, which provides 127 monthly macroeconomic features from January 1959 to January 2015. Since most of the series are not stationary, the transformation suggested by McCracken and Ng (2016) is taken into account for improving the accuracy in future estimation[5].

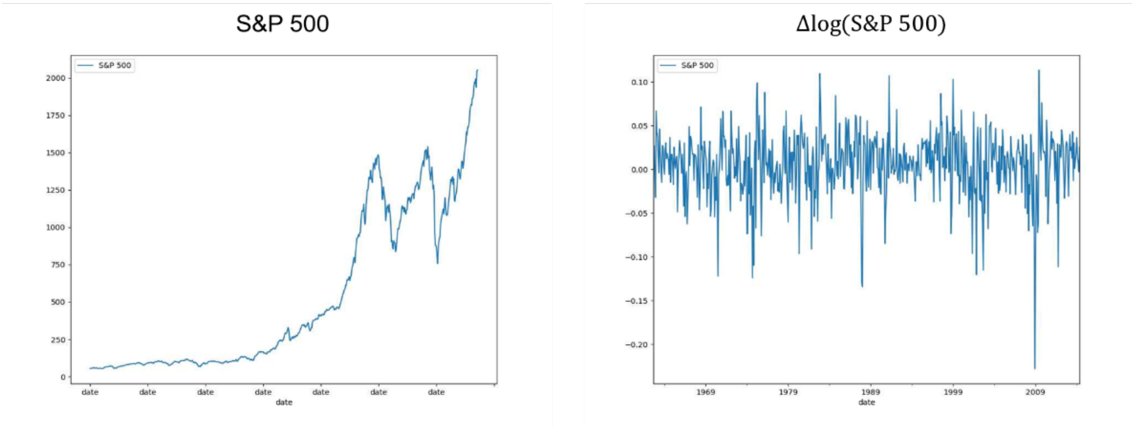
Alongside the 127 features, 15 macroeconomics predictors, which were not included in the FRED-MD database provided by Welch and Goyal (2007)[6]. For details and summary statistics, please see Welch and Goyal (2007).

[5] For a detailed understanding of the transformation process, please refer to McCracken and Ng (2016). Each series is assigned with a code as part of this transformation, which are (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; ; 7) $\Delta(x_t/x_{t-1}$−1.0).

[6] Data is collected from Amit Goyal's website: https://sites.google.com/view/agoyal145.

Additionally, the medians of each firm characteristic for capturing the common trends within the broader economy are included, as recommended by Chen et al. (2021). To summarize, a comprehensive analysis incorporating 178 macroeconomic factors is conducted using LASSO regression to determine their impact on the medium of stock returns. Consequently, 17 variables are selected by the regression.

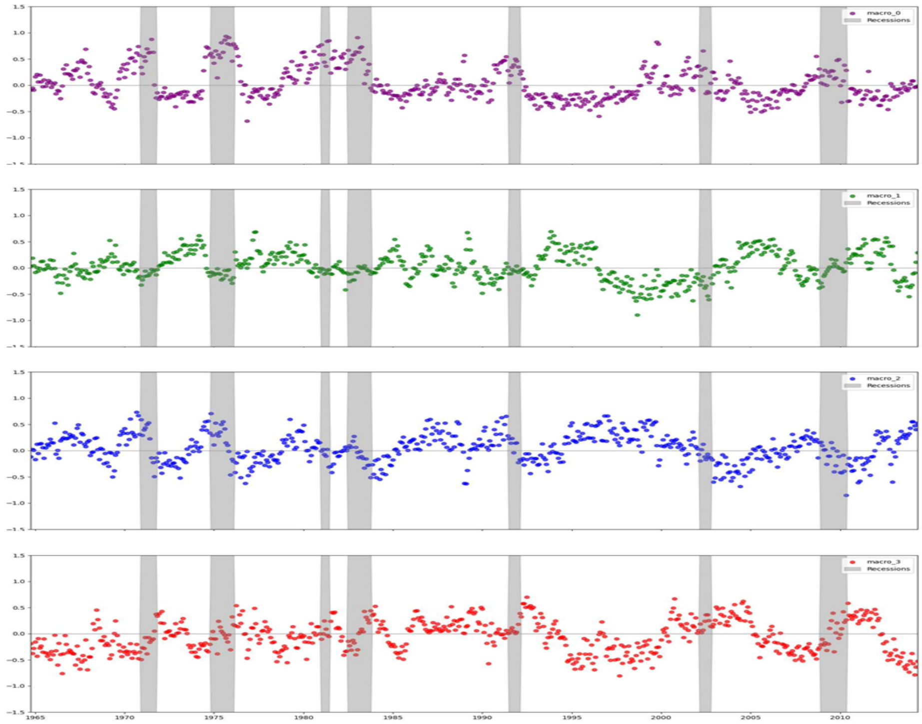
Figure 4. An Example of Macroeconomic Series



This figure presents an example of the application of transformation from McCracken and Ng (2016).

Besides directly inputting the 17 variables into the IPCA model, they are also fed into the Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) to capture the temporal dynamics inherent in the data The next step is to define the time horizon of the sequence, also known as timesteps, which determines the number of observations the model considers in each iteration. Utilizing diverse timesteps, such as 3, 6, and 12 months, determined in this research to capture a variety of temporal information. The rationale behind this approach is to extract dependencies at different intervals, specifically on a quarterly, half-yearly, and yearly basis. Finally, the hidden state from the last time step is extracted, exploiting its capability to encode patterns within the sequential information from previous time steps. This process could also be considered as a dimension reduction and non-linear transformation of macroeconomic data. Since each timestep consists of 4 hidden states, a total of 12 hidden states are prepared for incorporation into the IPCA model[7]. Figure 5 depicts the extracted hidden state for timestep 12, while additional information can be found in Figure A.1. The figures are reported accompanying the recession mark collected from the NBER website[8].

Figure 5 Macroeconomic Hidden State from Last Time Step (Time Horizon for Prediction =12)



This figure presents the four macroeconomic last hidden states extracted from RNN with LSTM cell. The faded area represents the recession periods provided by NBER.

As displayed in Figure 5, the first and third series tend to reach their peaks predominantly during recessionary periods, whereas the second and fourth series exhibit the opposite trend. The observation indicates that these four hidden state series capture distinct aspects of risk respectively.

[7] Information regarding the selected macroeconomics features and importance, see Appendix B.
[8] Business Cycle Expansions & Contractions are downloaded from https://fred.stlouisfed.org.

5.3 Observable Factors

These factors are responsible for a thorough exploration of the interactions between the IPCA model and various observable factors. The dataset utilized in this analysis encompasses various well-known financial factors, including the Fama-French Three-Factor Model (1993), the Carhart momentum factor (1997), the Fama-French Five-Factor Model (1997), the Hou et al. (2015) market factor, size factor, investment factor, and momentum, the Stambaugh and Yuan (2017) market factor, size factor, and two mispricing factors, along with the Barillas and Shanken (2018) selected 6 factors models, including market factor, size factor, momentum factor, value factor, investment factor, profitability factor. These factors are also downloaded from the database created by Seth Pruitt.

6. Empirical Result

6.1 Performance of Fundamental IPCA Model

At the initial step, a comparative analysis is conducted between the restricted ($\Gamma_\alpha = 0$) IPCA model and the unrestricted IPCA model ($\Gamma_\alpha \neq 0$) to assess their respective performance. In addition, a Wald-type test statistic is applied to evaluate the relative impact of allowing Γ_α to be non-zero. The result is presented in Table 1, where *K* implies the number of principal components. In general, the unrestricted model implies a higher performance than the restricted model in both total R^2 and predictive R^2.

For the case where K=1, the restricted model accounts for approximately 14.8% of the total variance, whereas the unrestricted model explains about 15.22%. For predictive R^2, the unrestricted model demonstrates a significantly stronger performance, more than doubling the predictive power compared to the restricted model. This observation indicates that the flexibility of the unrestricted model allows it to capture the variation more effectively and better fits the return data. Additionally, at the managed portfolio level, both models can explain over 90% of the total variance.

Table 1. Fundamental IPCA Model Performance (Restricted and Unrestricted)

		K					
		1	2	3	4	5	6
Panel A: Individual Stocks (r_t)							
Total R^2	$\Gamma\alpha$=0	14.80	16.39	17.42	18.05	18.56	18.88
	$\Gamma\alpha\neq$0	15.22	16.79	17.75	18.37	18.69	18.46
Pred. R^2	$\Gamma\alpha$=0	0.35	0.34	0.41	0.42	0.69	0.68
	$\Gamma\alpha\neq$0	0.76	0.75	0.75	0.74	0.74	0.72
Panel B: Managed Portfolios (x_t)							
Total R^2	$\Gamma\alpha$=0	90.30	95.28	97.09	98.04	98.41	98.76
	$\Gamma\alpha\neq$0	90.84	95.68	97.31	98.24	98.62	98.86
Pred. R^2	$\Gamma\alpha$=0	2.01	2.00	2.10	2.03	2.41	2.39
	$\Gamma\alpha\neq$0	2.61	2.56	2.54	2.51	2.50	2.46
Panel C: Asset Pricing Test							
Wα p-value		0.00	0.00	0.00	0.00	0.024	0.510

Panel A and Panel B report the Total R2 and Pred R2 for both restricted and unrestricted models at the stock level and managed portfolio level. Panel C presents the Wald-test statistics for the null hypothesis test ($\Gamma_\alpha = 0$).

Moving on to the Wald-test statistic, W_α is statistically significant at 0.01 level up to *K*=5, yet fails to reject the null hypothesis $\Gamma_\alpha = 0$. The narrow gap in explanatory power between the restricted and unrestricted models suggests that, at *K*=5, the common risk factors alone can achieve a comparable level of explanation without the need for the intercept term. Furthermore, this observation implies that the impact of idiosyncratic risk becomes relatively negligible. Consistent with the findings in the Wald-test statistics, the restricted model exhibits a notable enhancement in predictive R^2 from 0.42% to 0.69% at *K*=4 and *K*=5. At *K*=6, the restricted model overtakes the unrestricted model in explaining the variance of realized return, as shown in the total R^2.

6.2 Comparison with Existing Models

Table 2 compares the restricted ($\Gamma_\alpha = 0$) IPCA model to the existing factors model and PCA model. The analysis considers different sets of factors, for values *K*=1, 3, 4, 5, 6 . These values sequentially represent the market factor at *K*=1, the market factor, the size factor (SMB), the value factor (HML), forming the Fama-French 3 factors model at *K*=3, FF3 plus the momentum factor (MOM) at *K*=4, The Fama-French 3-factor model, extended with the profitability factor (RMA) and the investment factor (CMA) at *K*=5, and finally the FF5 model plus the momentum factor at *K*=6. In all these models, the intercept is intentionally set to 0. By omitting the intercept term, the analysis isolates and evaluates the impact of systematic risk factors on stock returns across the different modeling approaches.

Panel A restates the result for the restricted fundamental IPCA model. Whereas Panel B presents the performance of the static observable factor model. At first glance, the observable factor model has better performance in total R^2 than the IPCA model. However, the observable factors model requires a higher amount of parameters, which is calculated by the amount of loadings $N_p = N\ firms \times\ K\ factors$. Take *K*=3 for example, the observable factor model requires for 34,356 parameters to explain around 20 % of the total variance. In contrast, IPCA needs only $N_p = L\ characteristics \times$ $K\ factors$ + $T\ time \times\ K\ factors$, which is 95% less of parameters to reach 85% of the performance of FF3 models. Besides the effectiveness, IPCA can better describe the average common risk premium as shown in the higher predictive R^2.

Panel C reports the observable factors using the dynamic factor loadings similar to the IPCA model. The technique is introduced as the extended IPCA model in equation 12 excluding the α and β terms. The results display an elevated accuracy of explanation in common risk compensation indicating by the higher predictive R^2 with lower input of parameters compared to the static model. However, it is worth noting that the performance of the instrumented observable factor model is not comparable to the IPCA model.

Panel D presents the result of the latent factors using static and non-instrumented loadings. The performance in total R^2 predominately overperforms IPCA model but with larger amount of parameters calculated via $N_p = N\ firms \times\ K\ factors + T\ time \times\ K\ factors$. Moreover, another limitation is highlighted by the remarkably low predictive R^2, which even becomes negative after reaching K=3. This emphasizes a notable weakness in the PCA model's ability to accurately predict the common risk premium. At the managed portfolio level, the amount of estimated parameters is the same in the IPCA model. Its performance exceeds all other existing models except IPCA.

Continuing with the analysis of observable factors, Table 3 reports the performances of three additional models introduced in Section 5.3. These models include the four-factor model from Stambaugh and Yuan (2017) denoted as SY, the four-factor model from Hou et al. (2015) denoted as HXZ, and the six-factor model by Barillas and Shanken (2018) denoted as BS. Similar to Table 2, the intercept is excluded to focus on the impact of common risk factors.

Panel A reports the result from the static model. With the amount of 4 factors, HXZ outperforms SY in explaining the total variance but with a lower predictive R^2. When compared to the FFC4 model, the standout observation is the higher predictive R^2 in the SY model, while all other statistical measures from other models indicate a less favorable outcome compared to the FFC4 model. The same outcome can also be seen in Panel B.

Table 2. Restricted Fundamental IPCA Model Compares to Existing Model

		1	3	4	5	6
Panel A. IPCA						
rt	Total R2	14.80	17.42	18.05	18.56	18.88
	Pred R2	0.35	0.41	0.42	0.69	0.68
	Np	636	1908	2544	3180	3816
xt	Total R2	90.30	97.09	98.04	98.41	98.76
	Pred R2	2.01	2.10	2.03	2.41	2.39
	Np	636	1908	2544	3180	3816
Panel B. Observable Factors (no instruments)						
rt	Total R2	13.57	20.44	22.45	23.36	25.18
	Pred R2	0.25	0.30	0.28	0.29	0.26
	Np	11452	34356	45808	57260	68712
xt	Total R2	65.62	85.15	87.48	86.38	88.57
	Pred R2	1.67	2.07	1.98	2.06	1.96
	Np	37	111	148	185	222
Panel C. Observable Factors (instruments)						
rt	Total R2	10.38	14.23	15.27	14.68	15.60
	Pred R2	0.27	0.37	0.33	0.38	0.34
	Np	37	111	148	185	222
xt	Total R2	66.86	87.19	89.49	88.28	90.28
	Pred R2	1.63	2.07	2.19	2.06	1.96
	Np	37	111	148	185	222
Panel D. Principle Components						
rt	Total R2	16.24	22.23	23.98	25.68	27.48
	Pred R2	0.00	0.00	< 0	< 0	< 0
	Np	12051	36153	48204	60255	72306
xt	Total R2	87.64	91.55	93.47	94.07	94.70
	Pred R2	2.01	2.06	2.11	2.14	2.15
	Np	636	1908	2544	3180	3816

This table presents the total R^2, predictive R^2, and the amount of parameters N_p at stock level r_t and managed portfolio level for restricted IPCA model in Panel A, observable factors model, including CAPM, FF3, FFC4, FF5, and FFC6 in K=1, 3, 4, 5, 6 at Panel B, observable factors with instrumented dynamic loadings at Panel C, and PCA model at Panel D.

Moving on to the BS model, it outperforms the FFC6 model in the instrumented case but poorly explains the expected return at the non-instrumented case, with only 0.14% of predictive R^2.

In conclusion, the gathered evidence from Table 2 and Table 3 points out that none of the models with pre-specified factors is competitive as the IPCA model. IPCA directly extracts the pattern and dynamics from the original data, while other rely on predefined factors, which may limit their adaptability to the inherent complexities and evolving nature of financial data. This conclusion restates IPCA's advantage in flexibility and better adaptability.

Although the observable factor model exhibits weaker performance in explaining both the variance or return and risk premium, it raises a possibility that it could serve as a supplementary tool for capturing the variance that remains unexplained by the IPCA model. Therefore, Table 4 shows the improvement in the inclusion of observable factors in the IPCA model. Panel A and Panel B present Total R^2 and Predictive R^2 for the inclusion of 0,1,4,6 observable factors. These factors represent the market factor, FFC4, and FFC6, offering a comprehensive examination of how different observable components contribute to the overall performance of the model.

Overall, the participation of observable factors improves the explanatory power of the IPCA model. Specifically, for *K*=1, total R^2 improves from 14.8% to 17.5% and predictive R^2 improves from 0.35% to 0.5% with the inclusion of FFC6 factors. On the other hand, at *K*=5, total R^2 improves from 18.56% to 18.91%, and predictive R^2 experiences a decline from 0.69% to 0.67% with the inclusion of FFC6 factors. This indicates the gap between improvement narrows and, in some instances, even results negative effect on Predictive R^2.

Table 3. Performance of Other Observable Factor Models

		SY	HXZ	BS
K		4	4	6
Panel A. Without Instrument				
rt	Total R^2	19.7	20.5	23.7
	Pred R^2	0.37	0.18	0.14
xt	Total R^2	19.7	20.5	23.7
	Pred R^2	0.37	0.18	0.14
Panel B. With Instrument				
rt	Total R^2	14.4	14.7	15.64
	Pred R^2	0.41	0.32	0.38
xt	Total R^2	86.47	87.5	90.3
	Pred R^2	2.11	1.81	2.00

This table presents the performance of other observable factor models including and excluding instruments respectively. SY represents the four-factor model from Stambaugh and Yuan (2017) , HXZ is the four-factor model from Hou et al. (2015), and BS denotes the six-factors model from Barillas and Shanken (2018).

The hypothesis test in Panel C consistently reflects this trend. Before K=5, the momentum factor demonstrates statistical significance at the 1% level. However, after reaching K=5, none of the factors retains statistical significance. This evidence supports the notion that incorporating additional factors beyond *K*=5 is redundant, aligning with the results presented in Table 1 where the difference between the restricted and unrestricted models beyond *K*=5 is minor.

The redundancy that occurs in observable factors could be explained by the superiority of the PCA method since the principal components have already captured the maximum variance within the data. Additionally, there is no room for further incorporation of extra observables because the loadings of the latent factor model are generated based on the firm characteristics. This leads to an overlap in the additional observable factors since they are also generated based on firm characteristics.

Table 4. IPCA Model including Observable Factors

	K					
Observable Factors	1	2	3	4	5	6
	Panel A. Total R^2					
0	14.80	16.39	17.42	18.05	18.56	18.88
1	15.79	16.85	17.53	18.09	18.59	18.90
4	17.31	17.85	18.26	18.59	18.82	19.06
6	17.54	18.02	18.37	18.68	18.91	19.11
	Panel B. Predictive R^2					
0	0.35	0.34	0.41	0.42	0.69	0.68
1	0.35	0.4	0.41	0.5	0.69	0.68
4	0.45	0.66	0.67	0.66	0.71	0.69
6	0.50	0.66	0.66	0.66	0.67	0.69
	Panel C. Individual Significance Test p-value					
MKT-RF						
SMB		*	*	*	*	*
HML	*	*				
RMW	*					
CMA	*			*		*
MOM	**	**	**	**	*	*

This table displays the results of incorporating observable factors into the IPCA model. Panel A and B provide insights into the performance. The left-hand side of the table indicates the number of observable factors included in the IPCA model, where Observable factor=1 represents the market factor, 4 corresponds to FFC4, and 6 denotes FF6 factors. Panel C reports the p-values from the Wald test. Significance levels are denoted by ** and *, indicating that an observable factor significantly enhances the model at the 1% and 5% levels, respectively.

6.3 Performance of IPCA Model Including Macroeconomics Factors

The previous evidence in Sections 6.1 and 6.2 indicates that IPCA can perfectly capture the relationship between firm characteristics and returns when compared to other models. Nevertheless, it is worthwhile to conduct further assessments of the IPCA model incorporating factors built from different perspectives beyond firm characteristics. Therefore, macroeconomic factors are taken into account in this subsection according to my research motivation.

The initial evaluation involves inputting the 17 macroeconomic features directly into IPCA to assess its incremental impact and conduct significance testing through Wald-type statistics, as illustrated in Table 5. In general, integrating these new factors enhances the performance of IPCA in terms of both total and predictive R^2. However, the p-values derived from the Wald-type test do not exhibit significance at the 1% level across various values of K and each of the macroeconomic factors selected by LASSO.

The insignificance of the factor is also rooted in the reason for the redundancy of observable factors shown in the previous subsection. Since the principal components have already explained most of the variance, the incorporation of extra factors is irrelevant. However, when the latent factors are excluded from the linear model, both standalone instrumented observable factors and macroeconomic factors have promising explanatory power. The related presentation can be seen in Appendix C.

Although the fundamental IPCA model shows strong explanatory and prediction power. The defect rooted in the linear model draws limitations in explaining the non-linear relationship between factors and returns. Therefore, instead of directly inputting the macroeconomic features selected by LASSO. The series serve as the input of RNN with LSTM cells for a non-linear transformation and finally, reduces the dimension of macroeconomic data.

As mentioned in Section 5.2, three different time frames—3 months, 6 months, and 12 months—are employed in the Recurrent Neural Network (RNN) to capture dependencies of varying time lengths. Macro factors 1 to 4 are derived from a time frame of 3, Macro factors 5 to 8 are generated based on a time frame of 6, and Macro factors 9 to 12 are produced from a time frame of 12. This results in an aggregated set of 12 macroeconomic factors.

Comparing the basic IPCA model to the extended version, the incorporation of. the LASSO-RNN macroeconomic factor generates improvement in total R^2 and predictive R^2 before *K*=5. For *K*=1, total R^2 improves by 0.48% at the stock level after incorporating the 12 macroeconomic factors and predictive R^2 improves by 0.24%. However, the predictive R^2 has a slight decline by 0.0001% at *K*=5. In contrast to Table 4, in each amount of principal components, one or more macroeconomic factors demonstrate economic significance at the 1% level across various principal component quantities except *K*=1.

Table 5. Comparison between Fundamental IPCA and Extended IPCA incorporating macroeconomic factors selected by LASSO

		K					
		1	2	3	4	5	6
Panel A. Fundamental IPCA							
r_t	Total R^2	14.80	16.39	17.42	18.05	18.56	18.88
	Pred R^2	0.35	0.34	0.41	0.42	0.69	0.68
x_t	Total R^2	90.30	95.28	97.09	98.04	98.41	98.76
	Pred R^2	2.01	2.00	2.10	2.03	2.41	2.39
Panel B. IPCA with macroeconomic factors from LASSO							
r_t	Total R^2	16.49	17.33	18.11	18.55	18.85	19.08
	Pred R^2	0.75	0.75	0.74	0.73	0.73	0.72
x_t	Total R^2	93.74	95.84	97.50	98.38	98.71	98.95
	Pred R^2	2.57	2.60	2.53	2.50	2.49	2.46
Panel C. Individual Significance Test p-value							
macro 1							
macro 2							
macro 3							
macro 4							
macro 5							
macro 6	*						
macro 7					*	*	*
macro 8							
macro 9	*	*					
macro 10							
macro 11							
macro 12							
macro 13							
macro 14	*	*					
macro 15	*						
macro 16					*		
macro 17	*	*					

This table presents the performance of both the Basic IPCA and IPCA models incorporating the 17 macroeconomic features selected by LASSO. Panel C provides the p-values from the Wald-test statistics. Significance levels are denoted by ** and *, indicating that a macroeconomic factor significantly enhances the model at the 1% and 5% levels, respectively.

Table 6. Comparison between Fundamental IPCA and Extended IPCA

		K					
		1	2	3	4	5	6
		Panel A. Fundamental IPCA					
r_t	Total R^2	14.80	16.39	17.42	18.05	18.56	18.88
	Pred R^2	0.35	0.34	0.41	0.42	0.69	0.68
x_t	Total R^2	90.30	95.28	97.09	98.04	98.41	98.76
	Pred R^2	2.01	2.00	2.10	2.03	2.41	2.39
		Panel B. Extended IPCA with macroeconomic factors					
r_t	Total R^2	15.28	16.80	17.73	18.31	18.68	18.98
	Pred R^2	0.59	0.59	0.62	0.62	0.68	0.67
x_t	Total R^2	91.11	95.75	97.29	98.21	98.51	98.84
	Pred R^2	2.28	2.29	2.33	2.34	2.38	2.36
		Panel C. Individual Significance Test p-value					
macro 1							
macro 2							
macro 3							
macro 4							
macro 5		*	(**)	**	**	(**)	**
macro 6							
macro 7							
macro 8		*	*		*		(**)
macro 9		*	*	*	*		
macro 10		*					
macro 11							
macro 12		*	*	(**)	(**)		

This table presents the performance of both the Basic IPCA and Extended IPCA models. Panel C provides the p-values from the Wald-test statistics. Each set of 4 characteristics captures dynamics from different time frames (time steps=3, 6, 9). In total, 12 macroeconomic factors are being considered. Significance levels are denoted by ** and *, indicating that a macroeconomic factor significantly enhances the model at the 1% and 5% levels, respectively. The final selection of the factor is illustrated with extra ( ), which remains economically significant after the last iteration of the hypothesis test.

The mixed evidence from Table 4 and Table 6 suggests that when *K*=1, IPCA roughly explains the total variance in returns. Hence, in Table 4, momentum serves as a valuable complement, effectively compensating for the limitations of IPCA at the lower amount of principal components. This is because momentum emerges as a capable standalone factor in explaining the variance. On the other hand, LASSO-RNN macroeconomic factors alone are not good features for explaining the returns. Yet, macroeconomic factors provide value by taking care of aspects of the relationship that the fundamental IPCA model overlooks, given its sole consideration of linear relationship.

In order to select the most effective macroeconomic factor, I applied the iteration for null-hypothesis test introduced in Section 4.1.2. As shown in Panel C of Table 5, only a single factor remains significant for *K* values ranging from 2 to 5. Furthermore, a comparison between the performance of selected and eliminated macroeconomic factors is provided in Appendix D.

To conclude this section, it appears that the direct input of macroeconomic features from LASSO alone does not significantly enhance the explanatory power of the IPCA model. In contrast, the LASSO-RNN macroeconomic factors are capable of capturing the underlying dynamics from economic series, which is not obvious in the pure dataset and hence not detected by the IPCA model since it served as a linear model. Therefore, including the RNN with LSTM cell for macroeconomic data slightly expands the linear IPCA model to have a touch in a non-linear relationship between return and macroeconomic features, which is also not possible to capture latent factors.

6.3.1 Is it necessary to apply LASSO before RNN-LSTM?

The results from the previous subsection indicate that the factors generated from LASSO-RNN provide incremental explanatory power in the fundamental IPCA model. This observation raises up another crucial question: *Is it necessary to apply LASSO before inputting the features into RNN?* Table 7 reveals the answer to this question. In brief, the dynamics extracted from the last hidden state still appear to be an economically significant factor for enhancing the IPCA model. However, take *K*=6 for instance, both total R^2 and predictive R^2 reveal worse performance compared to the LASSO-RNN factors. The inclusion of LASSO selection is therefore concluded to be crucial since it helps to exclude irrelevant information underlying the macroeconomic dataset which is regardless of stock returns. That is, the information series are refined to be the most crucial elements for explaining return variance after passing two filters, constructed by LASSO and RNN.

Table 7. Performance of Extended IPCA incorporating RNN hidden state inputting all macroeconomic features

		K					
		1	2	3	4	5	6
Panel A. IPCA with all macroeconomic series RNN factor							
r_t	Total R^2	15.11	16.63	17.58	18.19	18.66	18.96
	Pred R^2	0.39	0.38	0.44	0.45	0.67	0.66
x_t	Total R^2	90.93	95.55	97.16	98.12	98.48	98.76
	Pred R^2	2.03	2.03	2.11	2.13	2.38	2.35
Panel B. Individual Significance Test p-value							
macro 1							
macro 2							
macro 3							
macro 4							
macro 5		*	**	**	**	**	**
macro 6							
macro 7							
macro 8		*	*		*		*
macro 9		*	*	**			
macro 10							
macro 11							
macro 12		*	*	*	**	**	

This table presents the performance of Extended IPCA incorporating RNN hidden state inputting all macroeconomic features. Significance levels are denoted by ** and *, indicating that a macroeconomic factor significantly enhances the model at the 1% and 5% levels, respectively.

6.3.2 Extended IPCA excluding insignificant macroeconomic factors

After the exclusion of insignificant macroeconomic factors, ta further examination of IPCA's performance is presented in Table 8.The figures incorporate more decimals to facilitate a nuanced comparison, particularly for the model with higher values of *K*.

At *K*=2, The Total R^2 and Predictive R^2 of the Extended IPCA model outperform the fundamental IPCA (without macroeconomic factors) model.

This trend persists until *K*=6, with an increase in total R^2 increases and consistent in predictive R^2[9].

Table 8. Performance of Extended IPCA Model with Selected Macroeconomics Factors

K		2	3	4	5	6
Panel A. Extended IPCA with selected macroeconomic factors						
r_t	Total R^2	16.422	17.548	18.173	18.571	18.887
	Pred R^2	0.347	0.5846	0.5883	0.6878	0.6835
x_t	Total R^2	95.349	97.172	98.106	98.412	98.771
	Pred R^2	2.015	2.285	2.317	2.411	2.393
Panel B. Fundamental IPCA						
r_t	Total R^2	16.393	17.415	18.050	18.561	18.875
	Pred R^2	0.3391	0.4137	0.4189	0.6871	0.6835
x_t	Total R^2	95.281	97.087	98.035	98.405	98.765
	Pred R^2	2.000	2.102	2.213	2.411	2.390

This table compares between Extended IPCA model with selected factors and Fundamental IPCA. K=1 can be ignored since it has no significant economic factors.

The findings suggest that incorporating the selected macroeconomic factors alongside firm characteristics enhances the IPCA model's capability to explain variance in realized returns. Simultaneously, the model maintains or even improves the accuracy of the model-implied conditional expected return. This result affirms the assumption that the performance of the IPCA model can be enhanced by considering factors beyond firm characteristics, supporting the notion that alternative perspectives contribute to its improvement.

6.3.3 Interpreting IPCA Factors

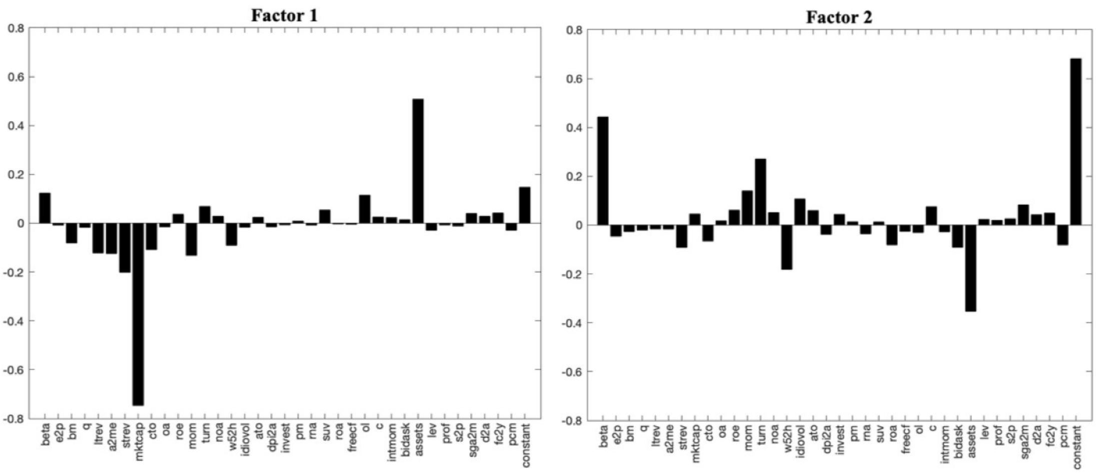
Figure 6 presents the mapping of characteristics for both principal components and a selected macroeconomic factor for *K*=5 model. It

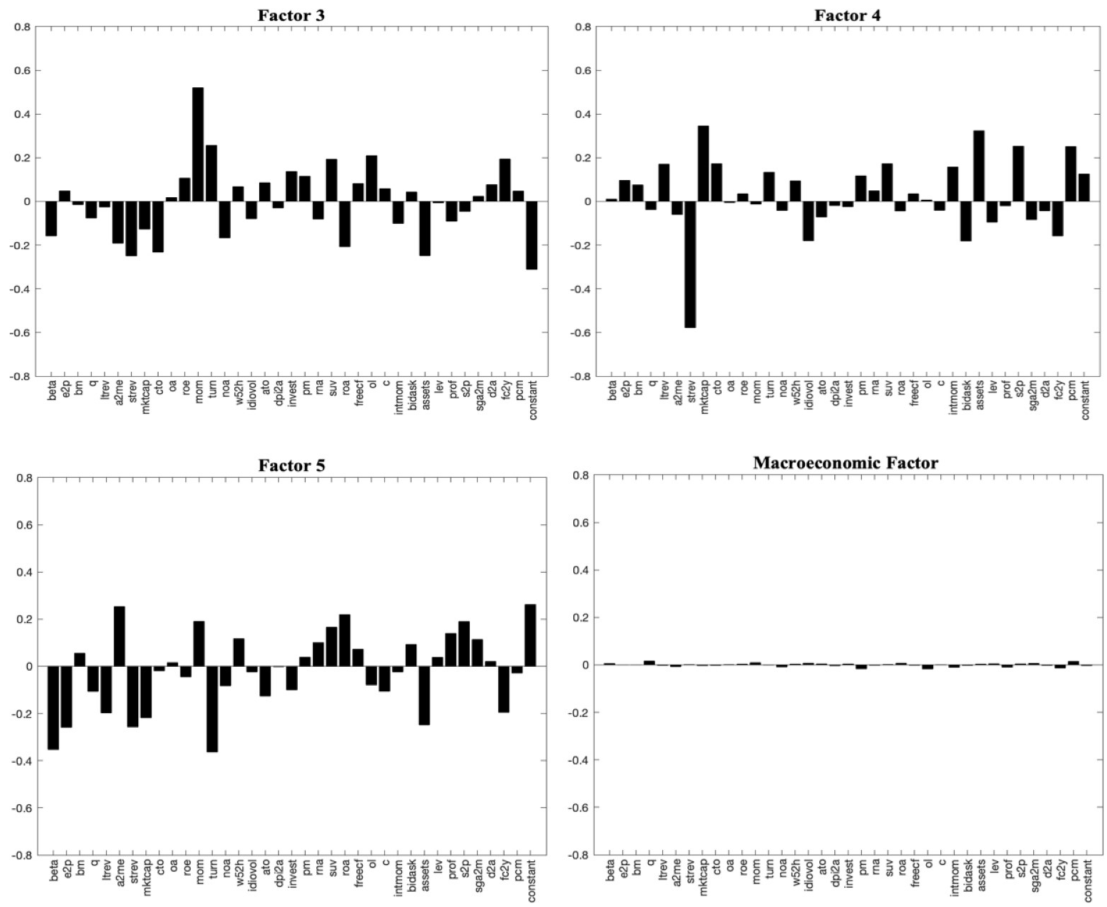
[9] When examined with more decimals, the extended IPCA model shows higher Predictive R^2 compared to the fundamental IPCA, both for stock level and portfolio level.

illustrates the significance of how firm characteristics influence the dynamic factor loadings. The larger the absolute number, the more pronounced the impact.

The interpretation of each principal component exhibits similarities to the Fama-French factors. In Factor 1, asset value (assets) and market capitalization (mktcap) exhibit predominant influences with distinct directions. Specifically, higher asset values correspond to higher expected returns, while smaller market capitalizations result in higher expected returns. This observation aligns with the characteristics of the value factor (HML) and size factor (SMB). However, in the mapping of firm characteristics, the book-to-market ratio shows a relatively minor impact. This can be explained by the fact that Factor 1 simultaneously captures the size factor and distinct elements of the value factor separately, given the shared presence of market capitalization in both factors.

Figure 6. Γ_β Coefficient Estimation





This figure plots the mapping of the firm characteristics of each principal component and macroeconomic factor, with each column representing both a firm characteristic and a constant.

In Factor 2 and 3, market risk (beta) and 12-month momentum (mom) play the pivotal role, therefore it can be defined as the market factor and momentum in the FFC4 model, respectively. Factor 4 is dominated by one-month stock reversal (strev), which is the lagged one-month return of individual stock. The graph illustrates that when the return of the previous month is lower, the expected return of the current month tends to be higher. This implication is aligned with **t**he Short-Term Reversal Factor from the FF5 model. Lastly, Factor 5 is the mixture of all characteristics.

Regarding the macroeconomic factor, the mapping weights of characteristics are minor or nearly zero. This result suggests that the systematic risk depicted by this factor, derived from the last hidden state, is largely unaffected by the variation in firm features. Therefore, it provides an independent source of systematic risk. Furthermore, since the firm characteristics have a similar impact on the loading of the factor, it suggests the diversification of the risk profile across the portfolio.

The findings from sections 6.3.1 and 6.3.2 suggest that while the Fama-French model is on the right path to explain returns, employing the IPCA model provides an alternative approach for a more refined and nuanced understanding of the relationship.

6.4 Out-of-Sample Performance

6.4.1 Out-of-Sample Fits

Table 9. Comparison of Out-of-Sample Performance of =Extended and Fundamental IPCA Model

		K					
		1	2	3	4	5	6
		Panel A. Extended IPCA with macroeconomic factors					
r_t	Total R^2	13.87	15.29	16.33	16.97	17.46	17.77
	Pred R^2	0.34	0.35	0.57	0.63	0.60	0.60
x_t	Total R^2	89.47	94.39	96.09	97.11	97.96	98.51
	Pred R^2	2.21	2.18	2.41	2.51	2.42	2.41
		Panel B. Fundamental IPCA					
r_t	Total R^2	13.87	15.34	16.31	16.93	17.49	17.80
	Pred R^2	0.34	0.33	0.55	0.61	0.60	0.60
x_t	Total R^2	89.47	94.75	96.44	97.38	98.18	98.62
	Pred R^2	2.21	2.15	2.42	2.44	2.42	2.42

This table reports the out-of-sample performance for both extended IPCA with selected macroeconomic factors and fundamental IPCA.

For out-of-sample measures, the backward estimation approach from Kelly et al. (2019) is applied. For every month $t \geq 120$, all data through t is employed to estimate the backward-looking $\hat{\Gamma}_{\beta,t}$. Afterward, $\hat{\Gamma}_{\beta,t}$ is utilized to generate the latent factor $\hat{f}_{t+1,t}$ and managed portfolio for time t+1.

The out-of-sample total R^2 is the fraction of r_{t+1} explained by $Z_t\hat{\Gamma}_{\beta,t}\hat{f}_{t+1,t}$, and x_{t+1} explained by $Z_t\hat{\Gamma}_{\beta,t}\hat{f}_{t+1,t}$. Similarly, the out-of-sample predictive R^2 is defined by substituting $\hat{f}_{t+1,t}$ with its mean through t. Table 9 presents the result of the out-of-sample fit of Extended IPCA with the selected macroeconomic factor and Fundamental IPCA. At K=1, both models have the same performance since none of the factors is statistically significant. From K=2 to 4, the extended IPCA improves the performance in both total R^2 and predictive R^2. For instance, the extended model improves the total R^2 by 0.04% and predictive R^2 by 0.00002% at the stock level at K=4.

However, from *K*=5 onward, extended IPCA exhibits slightly diminished performance in explaining the total variance of realized return. At *K*=5, the disparity is 0.03% at the stock level and 0.22% at the portfolio level.The gap narrowed to 0.03% and 0.11% with 6 principal components. Nevertheless, the result is acceptable since the predictive R^2 maintains at the similar level.

6.4.2 Unconditional Out-of-Sample Sharpe Ratio

The IPCA model, constructed upon conditional instrumented firm features and characterized by its dynamic nature, exhibits flexibility in capturing time-variant risk exposure. However, a practical challenge arises from inefficiencies in updating data in a timely manner. This limitation poses difficulties in keeping the model current with the latest information. Consequently, this section introduces the unconditional Sharpe ratio as a strategic measure. The aim is to assess whether the past estimations of unconditional loadings retain their predictive capability for future asset pricing.

The construction of the tangency portfolio returns utilizes the mean and covariance of estimated factors through *t*. Subsequently, the out-of-sample approach is applied to track the post-formation of t+1 return.
Table 10 reports the out-of-sample unconditional univariate and tangency Sharpe ratio for extended IPCA, fundamental IPCA, and observable Factors

(FFC6). model[10]. The univariate Sharpe ratio is the risk-adjusted performance in specific for each factor *K*, and the tangency Sharpe ratio is calculated with 1 to 6 factors model. In general, the Sharpe ratio of the two IPCA models outperforms the observable factor model since both models depict the risk and the risk premium more accurately. Notably, the tangency Sharpe ratio exhibits a strong outperformance, exceeding that of the observable factor model approximately threefold.

Table 10. The Out-of-Sample Portfolio Unconditional Sharpe Ratio

	K					
	1	2	3	4	5	6
Panel A. Extended IPCA						
Univariate	0.62	0.06	1.27	1.49	0.93	0.58
Tangency	0.62	0.77	3.15	3.48	3.93	4.29
Panel B. Fundamental IPCA						
Univariate	0.62	0.04	1.67	1.33	0.97	0.54
Tangency	0.62	0.62	2.49	3.09	3.89	4.05
Panel C. Observable Factors						
Univariate	0.46	0.32	0.40	0.46	0.61	0.51
Tangency	0.46	0.51	0.78	1.00	1.29	1.37

This table reports the univariate and tangency Sharpe Ratio for Extend IPCA with selected Macroeconomic Factors, Fundamental IPCA, and Observable Factors from FFC6.

In the comparison between the two IPCA models, the extended IPCA model outperforms the fundamental IPCA model in several aspects. Specifically, the tangency Sharpe ratio for the extended IPCA model surpasses that of the fundamental IPCA by 0.2. Furthermore, the extended IPCA model exhibits higher univariate factors through K. It could be attributed to the better diversification of risk among factors after adding the macroeconomic factor and lead to a higher risk-return profile.

[10] The target of tangency portfolio volatility is 1% per month.

6.5 Large Versus Small Stocks

The dataset is then divided into two subsamples: the first comprises the 1,000 "large" stocks, defined by their highest market capitalization each month, while the second encompasses the remaining stocks, classified as "small". Subsequently, the estimation and analysis of performance are conducted to discern the distinct behaviors of firms of large and small sizes. Notably, the parameters are estimated using unified samples, while the R^2 is computed based on the subsamples.

Table 11. In-Sample and Out-of-Sample Performance of Fundamental IPCA and Extended IPCA with Selected Macroeconomic Factor

			K				
			2	3	4	5	6
			Panel A. Small Stocks with Fundamental IPCA				
In-Sample	r_t	Total R^2	15.8	17.0	17.5	18.1	18.3
		Pred R^2	0.69	0.76	0.75	1.10	1.10
Out-of-Sample	r_t	Total R^2	15.7	16.9	17.5	17.9	18.2
		Pred R^2	0.80	1.02	1.08	1.07	1.09
			Panel B. Small Stocks with Extended IPCA				
In-Sample	r_t	Total R^2	14.2	15.1	15.7	16.1	16.4
		Pred R^2	0.36	0.63	0.64	0.75	0.74
Out-of-Sample	r_t	Total R^2	12.7	13.6	14.2	14.7	14.9
		Pred R^2	0.24	0.59	0.55	0.50	0.50
			Panel C. Large Stocks with Fundamental IPCA				
In-Sample	r_t	Total R^2	27.1	29.0	30.0	30.5	31.2
		Pred R^2	0.31	0.40	0.43	0.56	0.53
Out-of-Sample	r_t	Total R^2	25.9	27.3	28.1	29.0	29.7
		Pred R^2	0.32	0.46	0.52	0.41	0.39
			Panel B. Large Stocks with Extended IPCA				
In-Sample	r_t	Total R^2	29.57	31.21	32.12	32.88	33.5
		Pred R^2	0.36	0.55	0.56	0.49	0.61
Out-of-Sample	r_t	Total R^2	28.06	29.75	30.47	31.34	31.98
		Pred R^2	0.40	0.48	0.59	0.60	0.53

This table presents the in-sample and out-of-sample performance of the Fundamental and Extended restricted IPCA model with large and small stocks.

Table 11 reports the results for Fundamental IPCA and Extended IPCA with small stocks and large stocks. Table 9 reports the results for Fundamental IPCA and Extended IPCA with small stocks and large stocks. The findings indicate that the IPCA model solely based on firm characteristics demonstrates better predictive power with small stocks and provides a more comprehensive depiction of the total variance of large stocks. However, both subsets exhibit high performance, extending to out-of-sample fits as well.

Adding the selected macroeconomic factor leads to differing effects across the two subsamples compared to fundamental IPCA, the inclusion of the new factor deteriorates the explanation of small stocks in both total R^2 and predictive R^2. In contrast, there is an improvement in both explanation and predictive power for large stocks. Another notable observation is the narrowing gap in the predictive power of large and small stocks. These observations can be concluded that small stocks, which are normally less liquid and volatile, have volatility driven by firm-specific characteristics. On the other hand, large stocks, with higher liquidity, have a higher impact from general macroeconomic information.

Table 12. Out-of-Sample Sharpe Ratio of Fundamental IPCA and Extended IPCA with Selected Macroeconomic Factor

K	2	3	4	5	6
Panel A. Fundamental IPCA					
Small	0.58	1.10	2.82	4.15	4.19
Large	0.58	1.10	1.41	2.03	2.61
Panel B. Extended IPCA					
Small	0.74	3.70	3.49	4.08	4.26
Large	0.70	2.12	2.71	2.06	2.83

Panel A reports the out-of-sample Sharpe ratio of Fundamental IPCA while Panel B reports the out-of-sample Sharpe Ratio for Extended IPCA Model for small and large stocks.

Table 12 presents the out-of-sample Sharpe ratios for both the fundamental IPCA model and the Extended IPCA model across small and large stocks. Generally, smaller stocks yield higher returns per unit of risk, and the incorporation of macroeconomic factors further enhances the return-to-risk ratio. Even after segmenting the sample, IPCA continues to offer a robust explanation.

6.6 Performance Comparison After Time Split

In this section, I assess the robustness of the IPCA model by employing different time frames: pre-1990, and post-1990, as well as two distinct random splits of the original time horizon for the IPCA model with *K*=5. The results are shown in Table 13.

Table 13. Extended IPCA Model Performance after Time-Split (*K*=5)

	Pre-1990	Post-1990	Random Split 1	Random Split 2
Total R^2	18.78	18.67	16.55	18.13
Predictive R^2	0.80	0.67	0.73	0.73

This table presents the result of the performance of the extended IPCA model after splitting the time interval of the original data. 4 time frames are applied, pre-1990, post-1990, Random Time Split 1, and Random Time Split 2.

The results confirm the robustness of IPCA, demonstrating its applicability across different estimation terms. For instance, considering Random Split 1 and Post-1990, which exhibit the lowest explanatory power in the total variance of realized return and expected return, respectively, the former still achieves 89% of the total R^2 while the latter reaches 97% of the predictive R^2 when using the full sample. This insight suggests reliability as a capable tool for the assessment of stock return.

6.7 Asset Pricing Test for Individual Instruments

Table 14. Individual Firm Characteristic Significance (*K*=5)

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

This table presents the result of the hypothesis test introduced in Section 4.1.3 for individual firm characteristic. The extended model with selected macroeconomic factor at *K*=5 is employed. The Significance levels are denoted by ** and *, indicating that an observable factor significantly enhances the model at the 1% and 5% levels, respectively.

This section dives into whether the individual firm feature brings incremental explanatory power while controlling all other features. The examination utilizes the hypothesis test introduced in Section 4.1.3, and the results are presented in Table 14.

Among the 36 firm characteristics, only 10 of them are economically significant at 1% level and 5 features are significant at 5% level. The statistically significant features are as follows: market beta, long-term reversal, short-term reversal, market capitalization, momentum, turnover, price relative to its 52-week high,

unexplained volume, and total assets. This observation is aligned with the result in Figure 6 while these features contributed the most weight during the mapping process. In comparison to the fundamental model[11], idiosyncratic volatility was no longer significant at the 1% level after the introduction of macroeconomic features. This observation suggests that the macroeconomic

[11] See Appendix E.

feature may partly account for idiosyncratic volatility, potentially leading to multicollinearity and diminishing explanatory power.

Table 15. Model Fits of Extended IPCA Model Excluding Insignificant Firm Features

K			2	3	4	5	6
			Panel A. Stock Level				
r_t	$\Gamma\alpha$=0	Total R^2	16.11	17.20	17.83	18.19	18.41
		Pred R^2	0.34	0.54	0.55	0.61	0.62
	$\Gamma\alpha\neq$0	Total R^2	16.44	17.36	17.98	18.23	18.43
		Pred R^2	0.66	0.66	0.65	0.65	0.65
			Panel B. Portfolio Level				
x_t	$\Gamma\alpha$=0	Total R^2	97.70	98.85	99.52	99.73	99.86
		Pred R^2	2.15	2.32	2.32	2.36	2.37
	$\Gamma\alpha\neq$0	Total R^2	97.93	98.89	99.62	99.77	99.87
		Pred R^2	2.43	2.43	2.42	2.41	2.41

This table reports the performance of the Extended IPCA model at K=5 after excluding all the insignificant firm characteristics. Both stock level and portfolio level are presented, as well as the model including ($\Gamma_\alpha = 0$) and excluding intercept ($\Gamma_\alpha \neq 0$). *K*=1 is excluded from the table since it does not incorporate any macroeconomic factor.

Given that these significant variables explain a substantial portion of systematic risk, it is worthwhile to explore whether they provide a robust explanation for total variance and expected return when excluding the insignificant instruments from the analysis. The result is reported in Table 15. At *K*=2, the combination of significant instruments and the macroeconomic factor can explain approximately 98% of the total R^2 and also 98% of predictive R^2 shown in Table 6, which includes all the firm characteristics While comparing to fundamental IPCA, the combination can explain 98% of total R^2 and 88% of predictive R^2 presented in Table 1. This evidence reinforces the statement that the explanatory power of the IPCA model heavily relies on only a few characteristics. It also demonstrates that the model can achieve strong performance using only a select set of relevant features.

7. Conclusion

This paper re-evaluates, extends, and enhances the IPCA model introduced by Kelly et al. (2019) by incorporating macroeconomic features. The results can be summarized into the following three main points.

First, the fundamental IPCA appears to be a well-performed and robust linear model. The PCA methodology extracts the most relevant information with the aim of maximizing the variance captured from the return data. This attribute leads to the impossibility of incorporating extra factor that explains the linear relationship between return data and the specified common risk. Therefore, although the raw macroeconomic features are revealed to have stand-alone explanatory power of the variance of the data, they still fail to pass the hypothesis test.

Second, since the strength of the linear IPCA model is observed, an alternative methodology to incorporate the factor could only be applied via focusing on the weakness of the linear model. Therefore, LASSO-LSTM is introduced in this paper to extract the hidden state, which is applicable for explaining the non-linear relationship between macroeconomic features and stock returns. Via this approach, the Sharpe ratio of the tangency portfolio improves. Besides, the conditional loadings in the IPCA model reveal to have a better risk diversification after adding this macroeconomic factor.

Last but not least, the empirical results open a path to a non-linear pricing model. It is believed that the methodology provided in this paper can be further refined and improved. In addition, it would be interesting to conduct further research on a more sophisticated both linear and non-linear relationship between firm characteristics and returns, or to incorporate features from distinct aspects (e.g. option characteristics). It is also worth noting to be aware of the complexity of the model and economic intuition.

In conclusion, the IPCA model is shown to be a good performing model in the linear modeling regime. The research question also answers that macroeconomic characteristics do affect stock returns, but the linear relationship has been explained by the principal components. The ultimate solution to improve the IPCA model with firm characteristics is to discover the unobserved, i.e. non-linear, relationship between returns and economic information.

8. References

Amit Goyal, Amit Goval, Goyal, A., Ivo Welch, & Welch, I. (2004). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *National Bureau of Economic Research*.

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221.

Barillas, F., & Shanken, J. (2018). Comparing Asset Pricing Models. *National Bureau of Economic Research*.

Barr Rosenberg, & Rosenberg, B. (1974). Extra-Market Components of Covariance in Security Returns. *Journal of Financial and Quantitative Analysis*, *9*(2), 263–274.

Chamberlain, G., & Rothschild, M. (1982). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *National Bureau of Economic Research*.

Chan, L. K. C., J. Karaceski, Karaceski, J., Karceski, J., & Lakonishok, J. (1997). The Risk and Return from Factors. *Journal of Financial and Quantitative Analysis*.

Chen Ai & Chen Ai. (2022). A Method for Cancer Genomics Feature Selection Based on LASSO-RFE. *Iranian Journal of Science and Technology Transaction A-Science*.

Daniel, K., & Titman, S. (1996). Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. *Journal of Finance*.

Eugene F. Fama, Fama, E. F., Kenneth R. French, & French, K. R. (2013). *A Five-Factor Model for the Size, Value, and Profitability Patterns in Stock Returns*.

Fama, E. F., Kenneth R. French, & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Ferson, W. E., & Harvey, C. R. (1991). THE VARIATION OF ECONOMIC RISK PREMIUMS. *Journal of Political Economy*, *99*(2), 385–415.

Flannery, M. J., & Protopapadakis, A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *Review of Financial Studies*, *15*(3), 751–782.

Freyberger, J., Neuhierl, A., Michael Weber, & Andrew KarolyiEditor. (2017). Dissecting Characteristics Nonparametrically. *National Bureau of Economic Research*.

Gregory Connor, Connor, G., & Korajczyk, R. A. (1985). Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis. *Journal of Financial Economics*.

Jonathan Lewellen, & Lewellen, J. (2015). *The Cross-section of Expected Stock Returns*. *4*(1), 1–44.

Jotjensen, N. F., Chen, N., Richard Roll, Roll, R., Stephen A. Ross, & Ross, S. A. (1986). Economic Forces and the Stock Market. *The Journal of Business*, *59*(3), 383–403.

Junwen Yang, Junwen Yang, Yunmin Wang, Yunmin Wang, Xiang Li, & Xiang Li. (2022). Prediction of stock price direction using the LASSO-LSTM model combines technical indicators and financial sentiment analysis. *PeerJ*, *8*, e1148–e1148.

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.

Kewei Hou, Hou, K., Xue Chen, Xue, C., & Zhang, L. (2012). Digesting Anomalies: An Investment Approach. *Review of Financial Studies*.

Light, N., Denys Maslov, Maslov, D., & Rytchkov, O. (2016). Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach. *Review of Financial Studies*.

Lintner, J. (1965). THE VALUATION OF RISK ASSETS AND THE SELECTION OF RISKY INVESTMENTS IN STOCK PORTFOLIOS AND CAPITAL BUDGETS. *The Review of Economics and Statistics*, *47*(1), 13–37.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, *1*(1), 19–46.

Ludvigson, S. C., & Ng, S. (2005). The Empirical Risk-Return Relation: A Factor Analysis Approach. *Journal of Financial Economics*.

Luyang Chen, Chen. L., Markus Pelger, M., Jason Zhu, & Zhu J. (2019). Deep Learning in Asset Pricing

Malmendier, U., & Nagel, S. (2016). Learning from Inflation Experiences. *Quarterly Journal of Economics*, *131*(1), 53–87.

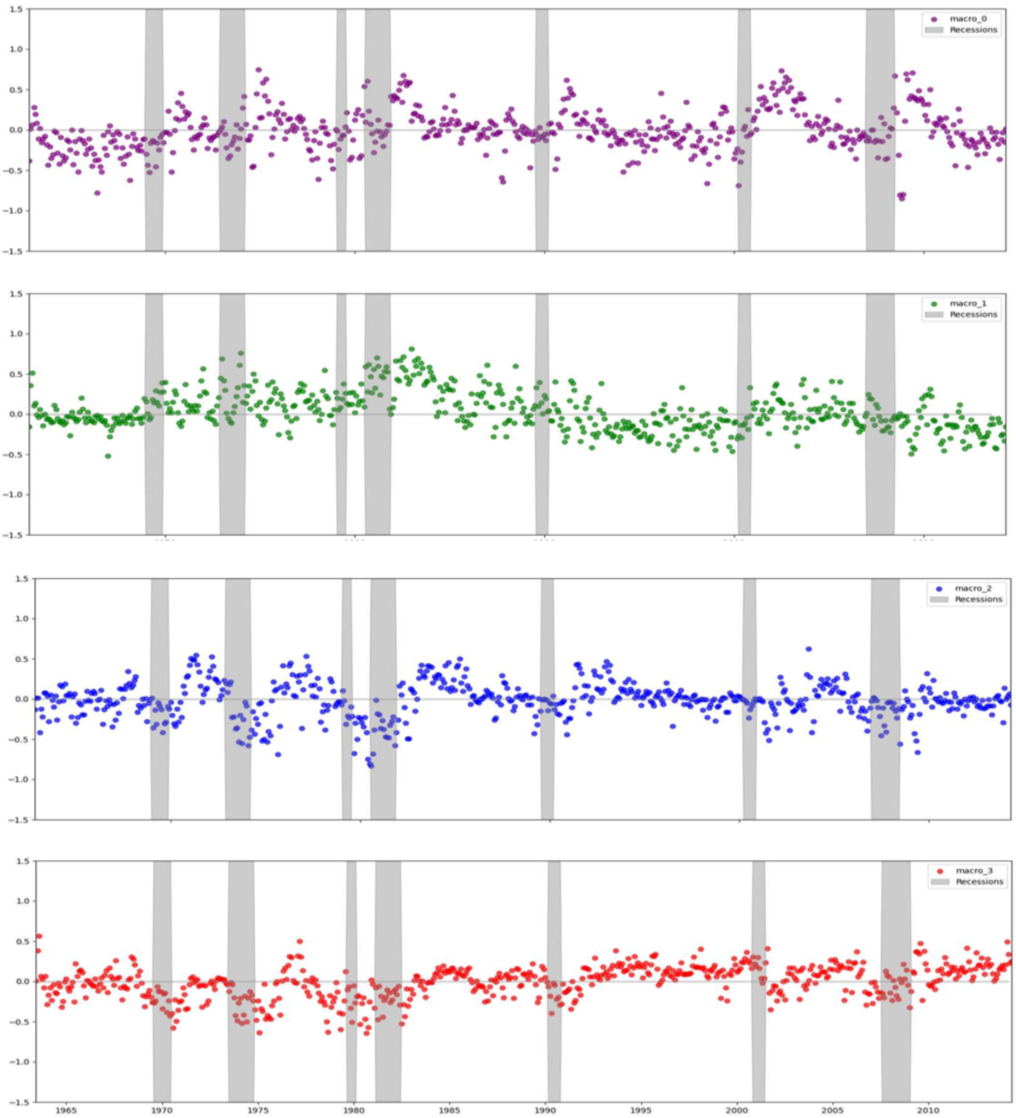
Mark M. Carhart, & Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, *52*(1), 57–82.
McCracken, M. W., & Ng, S. (2015). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.
N. Mankiw & R. Reis. (2001). *Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*.
Patelis, A. D. (1997). Stock Return Predictability and The Role of Monetary Policy. *Journal of Finance*, *52*(5), 1951–1972.
Pearce, D. K., V. Vance Roley, Roley, V. V., & V. Vance Roley. (1983). The Reaction of Stock Prices to Unanticipated Changes in Money: A Note. *Journal of Finance*, *38*(4), 1323–1333.
Robert F. Stambaugh, Robert F. Stambaugh, Yuan Yu, & Yu Yuan. (2017). Mispricing Factors. *Social Science Research Network*.
Robert Novy-Marx, Novy-Marx, R., & Velikov, M. (2014). A Taxonomy of Anomalies and their Trading Costs. *National Bureau of Economic Research*.
Robert Tibshirani, & Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological*, *58*(1), 267–288.
S. Nagel & Zhengyang Xu. (2019). *NBER WORKING PAPER SERIES ASSET PRICING WITH FADING MEMORY*.
Serhiy Kozak, Kozak, S., Nagel, S., & Santosh, S. (2017). Interpreting Factor Models. *Journal of Finance*.
Serhiy Kozak, Kozak, S., Nagel, S., & Santosh, S. (2018). Shrinking the Cross Section. *Journal of Financial Economics*.
Sharpe, W. F. (1964). CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*. *Journal of Finance*, *19*(3), 425–442.
Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.
Stephen A. Ross, & Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360.
Tarun Chordia, Chordia, T., Goyal, A., & Shanken, J. (2013). *Cross-Sectional Asset Pricing with Individual Stocks: Betas versus Characteristics **.

Xiafei Li, Xiafei Li, Chao Liang, Liang, C., Feng Ma, & Feng Ma. (2022). Forecasting stock market volatility with a large number of predictors: New evidence from the MS-MIDAS-LASSO model. *Annals of Operations Research*.

Appendix

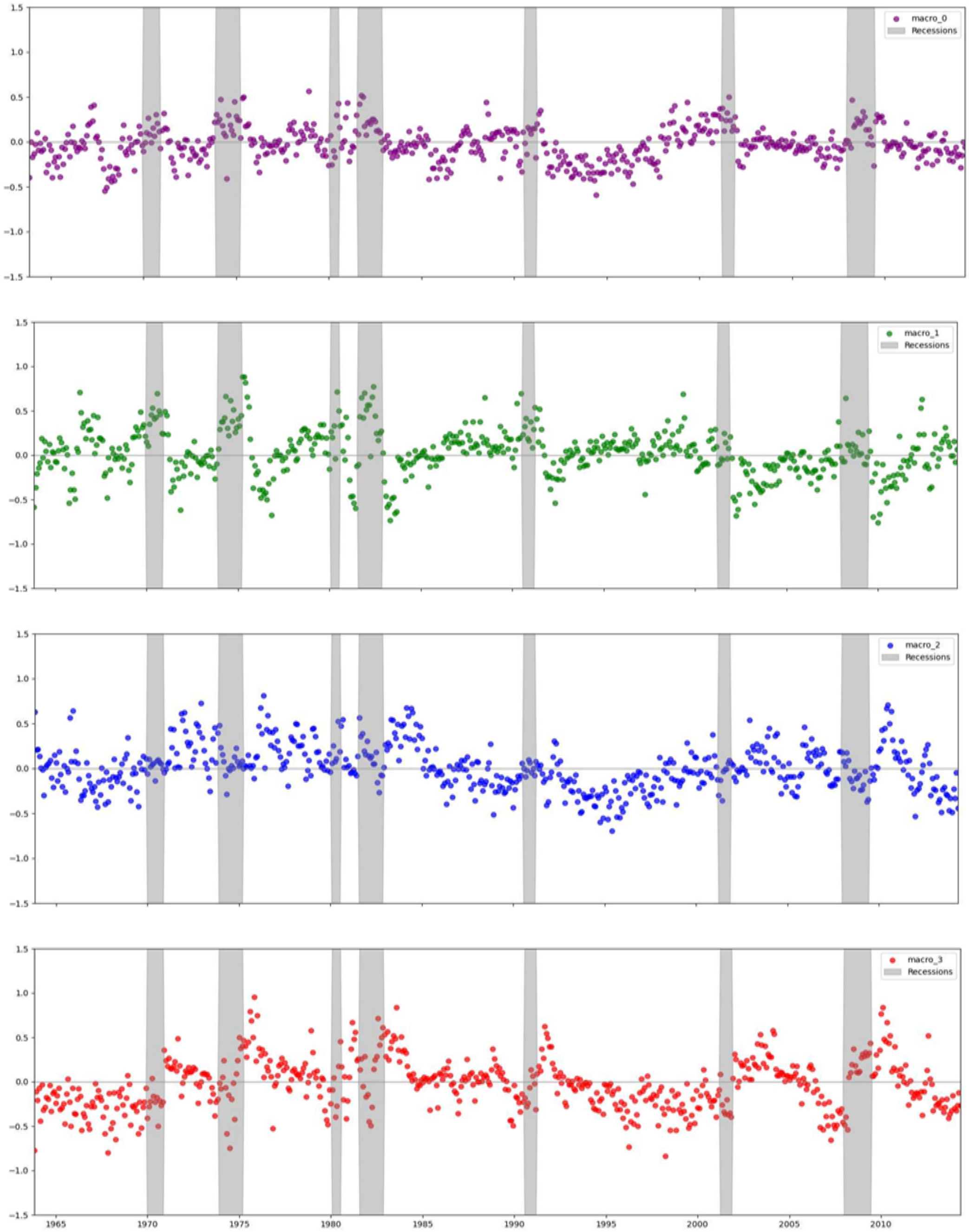
Appendix A. Additional Information from LSTM-RNN

Figure A.1. Macroeconomic Hidden State from Last Time Step (Time Horizon for Prediction=3)



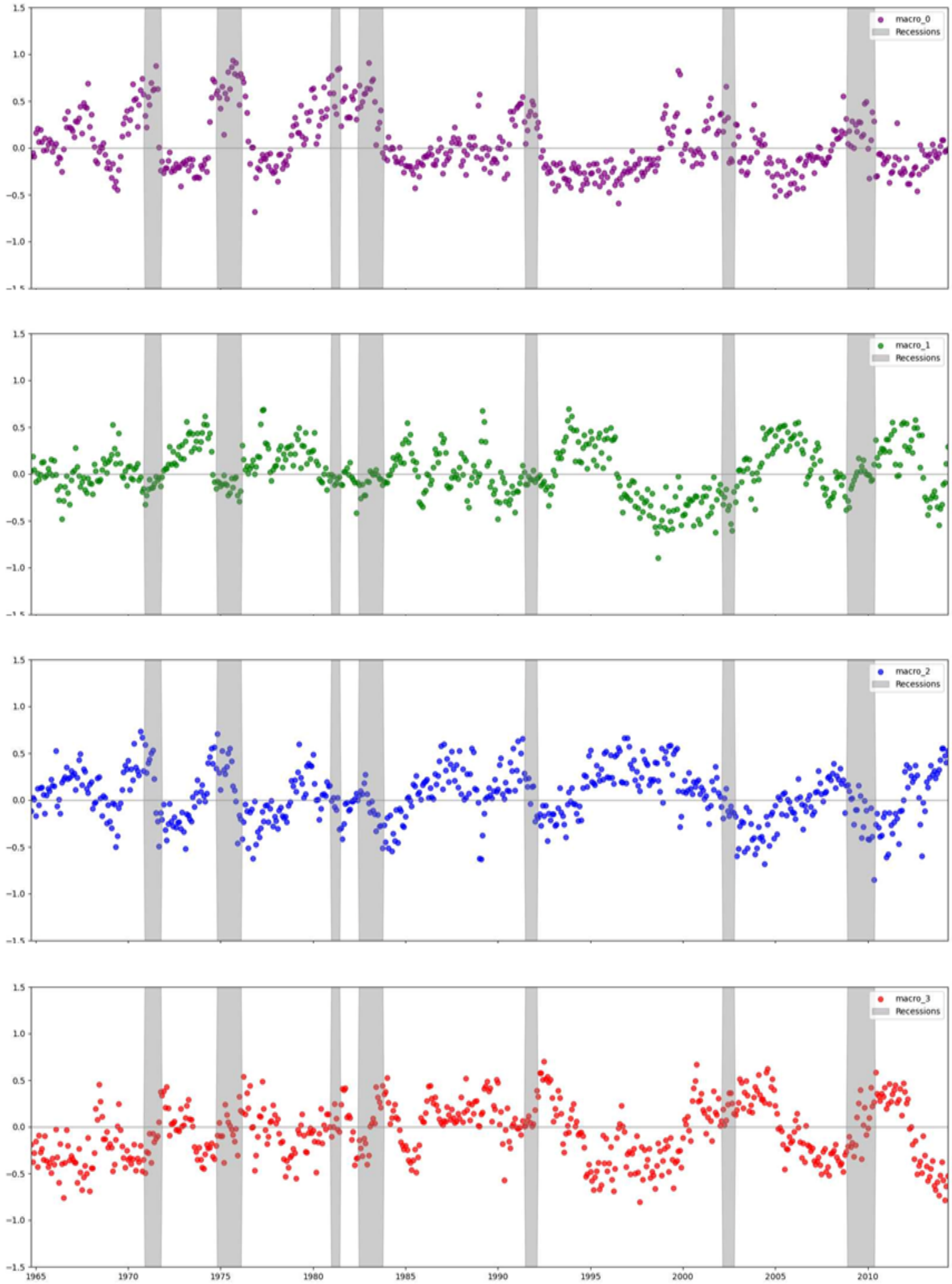
This graph presents the series from the last hidden state from LSTM-RNN with time horizon for prediction = 3

Figure A.2. Macroeconomic Hidden State from Last Time Step (Time Horizon for Prediction=6)



This graph presents the series from the last hidden state from LSTM-RNN with with time horizon for prediction = 6

Figure A.3. Macroeconomic Hidden State from last Time Step (Time Horizon for Prediction=12)



This graph presents the series from the last hidden state from LSTM-RNN with time horizon for prediction =12

Appendix B. Selected Macroeconomic Features and Relative Importance

In total 17 macroeconomic factors are selected as the input of RNN-LSTM model. Table B reports the representative code and name of these factors. Additionally, the table presents feature importance through gradient assessment, indicating the impact of input changes on the output.. In essence, a low gradient signifies minimal influence on the output, while a higher gradient indicates a more substantial impact resulting from changes in the corresponding feature.

When the aggregated timesteps=3, Dividend-Price Ratio and Real Manu. And Trade Industries Sales catch high attention from the model. While timesteps=6, Help-Wanted Index for U.S. shows outstanding weight. At time steps=12, Dividend-Price Ratio comes back into the play accompanies with Help-Wanted Index for U.S.

Table B. Relative Importance of Selected Macroeconomic Features

Factor Name	Time Horizon=3	Time Horizon=6	Time Horizon=12
Real Manu. And Trade Industries Sales	9.56	4.24	7.81
Intellectual Property: Consumer Goods	5.25	8.35	4.98
Help-Wanted Index for United States	6.13	9.34	9.45
Avg Weekly Hours: Goods-Producing	7.13	5.22	4.13
Avg Weekly Hours: Manufacturing	6.61	4.39	7.69
3-month Commercial Paper minus FEDFUNDS	5.21	2.57	6.09
Swiss/U.S. Foreign Exchange Rate	4.22	6.12	5.23
Total Consumer Loans and Leases Outstanding	3.46	6.98	4.3
Median Short-Term Reversal	6.23	6.67	5.23
Median Fixed Costs-to-Sales	7.55	7.24	4.3
Median Operating Accruals	3.77	8.55	5.97
Median Closeness to Relative High	2.64	6.82	5.22
Median Sales-to-Price	4.72	5.94	5.88
Median Unexplained Volume	6.21	3.73	6.34
Dividend-Price Ratio	11.09	4.7	9.85
The Term Spread	3.95	5.49	7.12
Stock Variance	6.28	3.65	2.24

The table represents the relative importance of the selected macroeconomic features decided by LASSO. The figures are in %.

Appendix C. Comparison between Instrumented Macroeconomic and Observable Factors

Table C. Performance of Instrumented Factors

Assets	Statistics	LASSO selected	LASSO-RNN	FFC6
r_t	Total R2	11.15	2.28	15.60
	Pred R2	0.72	0.44	0.34
x_t	Total R2	65.96	9.77	90.28
	Pred R2	2.60	1.40	1.96

This table shows the comparison between LASSO selected macroeconomic features, LASSO + LSTM macroeconomic factors, and FFC6 factors

Appendix D. Comparison between Selected and Eliminated Macroeconomic Factors

Table D. Comparison between selected and eliminated macroeconomic factors

		K		
		3	4	6
	Panel A. Selected factors			
	factor name	macro 12	macro 12	macro 8
r_t	Total R^2	17.548	18.173	18.887
	Pred R^2	0.5846	0.5883	0.6835
x_t	Total R^2	97.172	98.106	98.771
	Pred R^2	2.285	2.317	2.393
	Panel B. Eliminated factors			
	factor name	macro 5	macro 5	macro 5
r_t	Total R^2	17.434	18.063	18.884
	Pred R^2	0.4145	0.4191	0.6839
x_t	Total R^2	97.106	98.048	98.769
	Pred R^2	2.105	2.124	2.393

This table provides the comparison of significant factors that are selected and eliminated by the asset pricing test for K=3,4,6.

Table D offers an insight in the performance of selected and eliminated factors.

Generally speaking, the selected factors outperform in both performance measurement the eliminated factors while incorporated in the IPCA model except at *K*=6. However, a minor difference of 0.0004% is neglectable, the total R^2 of macro 8 in both stock and portfolio level exceed macro 5.

Appendix E. Individual Firm Characteristic Significance for Fundamental IPCA Model

Table E. Individual Firm Characteristic Significance for Fundamental IPCA model (*K*=5)

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

This table presents the result of the hypothesis test introduced in Section 4.1.3 for individual firm characteristic. The extended model with selected macroeconomic factor at *K*=5 is employed. The Significance levels are denoted by ** and *, indicating that an observable factor significantly enhances the model at the 1% and 5% levels, respectively.

Statutory Declaration

Students shall include a bibliography in all their term papers and project-related papers and submit a signed declaration with the following wording:

> „I hereby declare that the paper presented is my own work and that I have not called upon the help of a third party. In addition, I affirm that neither I nor anybody else has submitted this paper or parts of it to obtain credits elsewhere before. I have clearly marked and acknowledged all quotations or references that have been taken from the works of other. All secondary literature and other sources are marked and listed in the bibliography. The same applies to all charts, diagrams and illustrations as well as to all Internet sources. Moreover, I consent to my paper being electronically stores and sent anonymously in order to be checked for plagiarism. I am aware that the paper cannot be evaluated and may be graded "failed" ("nicht ausreichend") if the declaration is not made."

For term papers and project-related papers in English, the English declaration shall be included as well.

München 28.01.24

Place, Date

Tunglin Wu

Signature

Eidesstattliche Erklärung

Zu Prüfende haben ihren schriftlichen Seminar- und Projektarbeiten ein Verzeichnis der benutzten Hilfsmittel beizufügen und eine eigenhändig unterschriebene Erklärung mit folgendem Wortlaut abzugeben:

> „Hiermit versichere ich, dass diese Arbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurden. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Veröffentlichungen in gedruckter oder elektronischer Form sind gekennzeichnet. Sämtliche Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen und die Prüfungsleistung mit „nicht ausreichend“ bewertet werden kann, wenn die Erklärung nicht erteilt wird."

Bei Seminar- und Projektarbeiten in englischer Sprache ist *zusätzlich* die übersetzte Erklärung in Englisch abzugeben.

München 28.01.24

Ort, Datum

Tung Lin Wu

Unterschrift