Grande École - Finance Major

Research Paper

Academic Year 2023 - 2024

Market Timing Using Macroeconomic Data

Aggregating Macroeconomic Variables to Predict Market returns

Under the supervision of Professor ZVADIADZE

Anthony GIRARD
Sarah KOUAME

This is a **Public Report**

Acknowledgements

This paper marks the successful conclusion of a challenging but highly rewarding academic endeavour. We are extremely grateful to all the individuals and organisations that provided invaluable support and guidance throughout this process.

Firstly, we would like to express our sincerest gratitude to our supervisor, Professor Irina Zviadiadze for her guidance, insightful feedback and support.

We extend our thanks to the faculty and staff of the MIF department for providing a stimulating and enriching academic environment that fostered our academic and professional development.

A special acknowledgement goes to our families for their support and encouragement.

Finally, we would like to express our gratitude to the various institutions and databases that provided the data crucial to this study.

Contents

List of Tables

List of Figures

Abstract

This thesis investigates the predictive power of various macroeconomic indicators on the excess returns of the S&P 500 index using an advanced machine learning framework. Leveraging a dataset comprising 125 macroeconomic variables from the FRED-MD database, the study employs Principal Component Analysis (PCA) and Diffusion Index methods to reduce dimensionality and aggregate macroeconomic information. The research reveals that a set of principal components effectively captures a significant economic signal, which exhibit a robust relationship with S&P 500 excess returns. The signal demonstrates strong predictive power during various economic conditions, except for a notable period between 1990 and 1995. Furthermore, a comparative analysis with the consumption-based signal by Atanasov et al. highlights the superior performance of aggregating various macroeconomic indicators versus a single variable in terms of predictive strength, and robustness across market conditions. This comprehensive approach to forecasting can aid investors in making informed decisions by providing insights into the long-term impacts of macroeconomic factors on market returns.

Keywords: macroeconomic indicators, cyclical component; excess stock returns; forecasting; principal component analysis, diffusion index.

* * *

Introduction

Macroeconomic variables are essential for understanding and predicting stock market returns due to their influence on the overall economic environment in which companies operate. These variables, such as GDP growth, inflation rates, and employment figures, affect consumer and business confidence, which in turn influence spending, investment, and financial risk-taking. The relationship between macroeconomic variables and stock market returns has been a long-standing topic of investigation in financial research. Numerous studies have explored the predictive power of various indicators, such as inflation, money supply, unemployment, and interest rates. While some studies have found evidence of significant predictability for specific variables, others have not, leading to concerns about data mining and the need for a more comprehensive understanding of return predictability.

This paper aims at re-examining the predictability of US stock returns using a comprehensive approach. We leverage a large dataset of 125 macroeconomic variables from the FRED-MD monthly database, encompassing various aspects of the economy such as output production, income, labour market, housing, and consumption. This broad scope allows us to capture a more complete picture of the economic landscape and potentially identify stronger predictive signals. To address the challenges of handling numerous features and potential overfitting, we employ and compare two different aggregation methods: Principal Component Analysis (PCA) and diffusion indexes. PCA efficiently reduces dimensionality by transforming correlated variables into uncorrelated principal components, capturing the essential statistical features of the data. Diffusion indexes, on the other hand, provide an alternative approach by measuring the proportion of indicators moving in a specific direction within pre-defined clusters.

Our study encompasses both in-sample and out-of-sample analysis to assess the robustness of the constructed signal. We employ a XGBoost regression model for in-sample analysis from January 1959 to December 2012, then extend the analysis to out-of-sample data until 2023 to further validate our findings. The research framework we have developped strives to achieve an unbiased forecasting process by using point-in-time data analysis and rolling predictions. Our results are analyzed through a rigorous econometric approach, using the Ordinary Least Squares regression (OLS) method enhanced by the Newey-West variance estimator to evaluate our signal statistical significativity and its predictive power. Furthermore, we delve into the signal's characteristics by examining its predictive power across different market conditions, its reliability over time, its performance for various time horizons, and its effectiveness in predicting industry portfolio returns.

Our findings reveal several key insights. First, we identify 15 principal components effectively capturing significant aspects of the overall economy such as production indexes, labour market indicators and price indexes, ensuring that the essential economic signals are well-represented in our analysis. Out of these principal components, we find that the 9 first components under a 20 years rolling window provide the strongest setting to forecasting excess market returns. The signal exhibits a strong relationship with S&P 500 returns, demonstrating the potential for economic factors to predict future market movements. The signal exhibits reliable predictability during the in-sample period, with a notable exception between 1990 and 1995. These results hold even when extending the analysis to the out-of-sample period. The performance of the signal is optimized for a 2-year time horizon, reflecting the long-term influence of economic factors on the market. However, the signal also show strong performance for higher frequencies such as 3 months and 1 year horizons, making it relevant for investment purposes. Furthermore, the analysis identified significant differences in predictability across industries, with sectors like High-Tech and Non-Durables showing stronger signal performance. Finally, using diffusion indexes built on clusters of indicators as an alternative aggregation method, we are able to establish statistically significant relationship with excess returns and build a competing signal aggregating macroeconomic data. While the diffusion index signal demonstrates strong overall performance, it is less robust than the PCA signal when the prediction horizon is shorter.

The first section of the paper provides a review of the literature. Section 2 details our methodology to construct our signals. Section 3 contains the signal construction details and analysis. Section 4 studies the robustness of the signal on the out-of-sample period from 2012 to 2023. Section 5 performs a complementary analysis on different time horizons and portfolios to study the effectiveness of the signal as an investment indicator. Section 6 compares the merits of building a signal aggregating several macroeconomic indicators to a well-performing single indicator based signal built by Atanasov et al. [2]. Section 7 expands our research to another dimensionality reduction methodology based on diffusion indexes. Finally, Section 8 establishes the performance attribution to the predictors built in this paper.

1 Literature Review

Macroeconomic variables significantly influence firms' expected cash flows and the rate at which these cash flows are discounted. For instance, higher GDP growth typically boosts consumer spending and business investment, benefiting firms. Conversely, a decline in GDP growth can reduce demand for goods and services, negatively impacting corporate profits. Although the hypothesis that macroeconomic developments affect equity returns has intuitive appeal, obtaining robust evidence of stock return predictability has proven challenging. Flannery et al. (2002) [10] argue that while macroeconomic influences are theoretically sound, empirical evidence remains limited. For example, Bossaerts and Hillion (1999) [3] confirm in-sample return predictability but fail to demonstrate out-of-sample predictability for international stock markets. The impact of real macroeconomic variables on aggregate equity returns has been difficult to establish, possibly due to their nonlinear and time-varying effects (Flannery and Protopapadakis, 2002 [10]).

Despite these challenges, considerable empirical studies have made progress in investigating the predictability of stock returns using macroeconomic variables, especially over longer time horizons. Guru-Gharan et al. (2009) [13] found that the explanatory power of selected macroeconomic variables on U.S. stock market returns increases significantly when the time-frame changes from monthly to yearly. Since the pioneering work of Chen, Roll, and Ross (1986) [6] and in line with Merton's (1973) [19] theoretical contributions, researchers have explored the possibility that state variables such as inflation and economic growth may be sources of systematic investment risk and can explain the cross-sectional dispersion of stock returns. Empirical evidence indicates that stock market performance is highly correlated with economic fundamentals, making a model based on this relationship essential for predicting future trends (Morck et al., 2000 [20]; Rapach et al., 2005 [22]; Ahn et al., 2019 [1]).

Recent literature has demonstrated that deviations of factor prices from values implied by macroeconomic conditions predict both in-sample and out-of-sample factor returns, resulting in significant economic gains for mean-variance investors (C. Favero, A. Melone et al., 2020 [9]). Engle et al. (2013) [7] found that macroeconomic factors such as industrial production growth, interest rates, inflation, and unemployment often determine stock market movements. Analysis of 12 industrialized countries by Rapach et al. (2005) [22] found that interest rates are the most effective macroeconomic predictors of stock returns. Liu and Kemp (2019) [15] investigated the predictive accuracy of three macroeconomic variables in forecasting excess returns of the U.S. oil and gas industry stock index, finding that macroeconomic variables offer valuable insights into forecasting future stock market performance, particularly during bear market conditions.

The literature also proposes various methodological approaches regarding factor timing and stock return. Factor models, such as those by Fama and French (1993) [8], often abstract from the predictability of factors and focus on their ability to generate cross-sectional dispersion in asset risk premiums. Haddad, Kozak, and Santosh (2020) [14] propose a new statistical approach to predict anomaly portfolios by predicting their principal components (PCs) using their book-to-market ratios. However, using many factors to predict stock returns can complicate model accuracy and reliability. To address this, researchers employ methods like the least absolute shrinkage and selection operator (LASSO) or the elastic net (ENET) to maintain model stability with numerous variables. Ma, Lu et al. [17] found that a new set of economic attention indices (MAI) developed by Fisher et al. (2021) could predict stock returns better than traditional methods, even during unusual events like the COVID-19 pandemic. Rapach and Zhou (2021) [21] further advanced this field by using machine-learning techniques to estimate a sparse principal components (PCs) regression for 120 monthly macroeconomic variables from the FRED-MD database. Each sparse PC is a sparse linear combination of the underlying macroeconomic variables, allowing for their economic interpretation.

In our analysis, we use the FRED-MD database to construct a stable predictive model for the S&P 500 index. Based on 125 macroeconomic variables, we investigate whether these FRED-MD time series can better predict stock market returns compared to a single variable such as aggregate consumption, suggested by Atanasov et al. (n.d.) [2]. We extract the cyclical components of the macroeconomic time series likely containing information for predicting excess market returns. In addition, we build our model using two dimensionality reduction methods: PCA and Cluster Diffusion index, with the aim to compare the impact of both methods in the prediction of stock return. This research contributes to the broad empirical studies on stock market prediction and aims to prove predictability with a more comprehensive model encompassing a large set of accessible macroeconomic variables. Practically, this study can forecast the return signal of the S&P 500 index, aiding investors in their decision-making processes in the U.S. stock market.

2 Methodology

2.1 Exploratory Data Analysis

For our analysis, we collected data through the FRED database which stands "for Federal Reserve Economic Data". The FRED-MD is a large macroeconomic dataset publicly accessible containing 135 monthly U.S. indicators suitable for empirical analysis requiring economic data.

To explore the dataset, we selected nine common macroeconomic indicators resulting in a sub-sample of nine time series for the exploratory data analysis (EDA). The selected indicators are: Industrial Production (INDPRO), Initial Claims for Unemployment Insurance (CLAIMSx), Building Permits (PERMIT), Inventory-to-Sales Ratio (ISRATIOx), Consumer Sentiment Index (UMCSENTx), Real M2 Money Stock (M2REAL), Moody's Baa Corporate Bond Minus FEDFUNDS (BAAFFM), Personal Consumption Expenditures: Chain Index (PCEPI), and the S&P 500 P/E Ratio.

Through EDA, we carefully examined the subset of 9 time series, providing crucial insights into the dynamics and inherent properties of our data as well as hihghlighting common problematics with handling economic time series :

Trend and Stationarity: We observed that some series exhibit significant trends and lack stationarity. Such characteristics can undermine the predictive power of models that assume data stationarity, necessitating detrending or differencing to stabilize the mean.

Variability: High variability in some series can impact the stability and reliability of predictions. This calls for normalization techniques or transformation methods such as log-scaling to manage extreme values and make the series more homoscedastic.

Distribution and Skewness: Our analysis reveals that the series often do not follow a normal distribution and exhibit considerable skewness. This deviation from normality suggests that standard assumptions of many statistical tests and models might be violated, requiring non-parametric methods or data transformation to correct skewness.

Noise Levels: The presence of significant noise in the data series complicates modeling by obscuring underlying patterns. Smoothing techniques, such as moving averages or exponential smoothing, might be necessary to clarify the data's signal.

These preliminary observations are critical for the subsequent feature engineering phase. They guide our strategy for preprocessing and transforming data to enhance model accuracy and reliability. By addressing these characteristics proactively, we aim to maximize the informational content of features, crafting a robust analytical framework that leverages advanced statistical techniques and machine learning models to generate reliable forecasts. This approach not only strengthens the validity of our results but also enhances our ability to generalize findings to the out-of-sample period, thus ensuring the practical applicability of our research in real-world scenarios.

2.2 Cyclical Component Extraction

To effectively use macroeconomic variables in predicting stock market returns, it is important to not only examine their raw values but also extract their cyclical components. This approach focuses on deviations from long-term trends, capturing more volatile fluctuations that can provide valuable information about future economic conditions. By isolating the cyclical component, analysts can better identify the timing of economic expansions and contractions, which are closely linked to market performance. Furthermore, the frequency of the cyclical component that can be extracted from economic time series matches more closely that of the excess return time series, likely providing stronger signals for forecasting returns.

The extraction of cyclical components helps mitigate the misleading effects of long-term trends, such as secular increases in productivity or demographic shifts, which might not be directly relevant for predicting short-term market movements. The focus on the cyclical nature of economic indicators aligns with the broader financial concept that financial market prices are heavily influenced by changes in expectations about the future rather than long-term trend movements.

To avoid any forward-looking bias, it is necessary to use techniques that extract cyclical components solely based on past data points. One method used by Atanasov et al. (n.d. [2]) involves using the regression error of a regression on remote lagged values. The authors conduct the following regression on the consumption time series C_t :

$$c_t = b_0 + b_1 c_{t-k} + b_2 c_{t-k-1} + b_3 c_{t-k-2} + b_4 c_{t-k-3} + \omega_t$$

Where $c_t = log(C_t), k =$ lag in months. The regression error ω_t is the measure of cyclical consumption cc_t:

$$cc_t = \omega_t = c_t - (b_0 + b_1 c_{t-k} + b_2 c_{t_k-1} + b_3 c_{t-k-2} + b_4 c_{t-k-3})$$

We will use this method and compare it to a simpler and faster alternative that does not add layers of complexity.

The second method consists in detrending the time series by subtracting a moving average from the time series, this moving average representing the trend component. Denoting (X_t) our time series, we can extract the cyclical component cc_t as :

$$\forall t, \quad cc_t = X_t - \frac{1}{L} \sum_{l=1}^{L} X_{t-l}$$

Where L is the size of the moving average window.

Whenever possible, that is when the series does not initially contain negative or null values, we apply this method to the log-transformation of the time series X_t, $x_t = log\,(X_t)$:

$$\forall t, cc_t = x_t - \frac{1}{L} \sum_{l=1}^{L} x_{t-l}$$

We favour log-transformed variables because they tend to linearize time series and smooth variations naturally. Both methods need to be calibrated well by testing different values for k and L. We privilege a 1-year window ($L = 12$) to extract a higher frequency cyclical component in order to have a better match with the frequency of the excess returns variations, while not including too much noise contained in even higher frequency variations.

2.3 Principal Components Analysis

Handling more than a hundred features in predictive models poses significant challenges, primarily due to the risk of overfitting. Additionally, macroeconomic variables tend to be highly correlated, making it hard to determine which variable really contributes to improving the prediction. To mitigate these issues, this study employs Principal Component Analysis (PCA) on the cyclical components of our time series data. PCA is instrumental in dimensionality reduction, symplifying the modeling process without the loss of substantial information. It achieves this by transforming highly correlated variables into a new set of linearly uncorrelated variables termed principal components, which encapsulate distinct statistical features of the data.

Principal Component Analysis is a sophisticated statistical technique that simplifies high-dimensional data while preserving essential trends and patterns. This is achieved through the following methodological steps:

Standardization: Prior to analysis, it is imperative to standardize each feature by centering around the mean and scaling to unit variance. This standardization is crucial as PCA's sensitivity to the variances of initial variables can significantly influence the results.

$$\mathcal{C} = \left(\frac{cc_i - \mu_i}{\sigma_i} \right)_{i=1,..,N}$$

Covariance Matrix Computation: PCA involves the computation of the covariance matrix of the data, which provides insights into how variables co-vary from their mean. We note this covariance matrix

$$\Sigma = \frac{1}{T-1} \mathcal{C}^T \mathcal{C}$$

Eigenvalue Decomposition: The covariance matrix is subjected to eigenvalue decomposition to extract its eigenvectors and eigenvalues. This decomposition reveals the directions of maximum variance in the data (the principal components) and the relative significance of these directions.

$$\Sigma = V \Lambda V^T, V \in O_N\left(R\right)$$

Component Selection: The eigenvectors are ordered by their corresponding eigenvalues in descending order. Selection of principal components is typically based on those that account for the majority of the variance observed in the original dataset

$$\Lambda = Diag\left(\lambda_1, \ldots, \lambda_N\right), \quad \lambda_1 \geq \ldots \geq \lambda_N$$

Dimensionality Reduction: Data projection onto the selected principal components transforms the original high-dimensional data into a new space with reduced dimensions.

$$T \;=\; \left(PC_1, \ldots, PC_K\right) = \mathcal{C} \; \underbrace{\left(V_1, \ldots, V_K\right)}_{\text{First } k \text{ eigenvectors}}$$

This transformation ensures that the first principal component captures the highest possible variance, and each subsequent component, in turn, has the highest variance possible under the constraint that it is orthogonal to the preceding ones. This process effectively reduces the dimensionality of the data by transforming it onto a new coordinate system defined by the principal components. Thus, we are able to engineer relevant features suitable for forecasting models.

2.4 The target variable

Our target variable corresponds to log excess market returns for the S&P500 index over a certain horizon of h months. It is computed as follows :

$$y_h = R_h - r_F$$

$$R_h = (R_{h,t})_{t=1,...,T} = \left(log \left(\frac{S_{t+h}}{S_t} \right) \right)_{t=1,...,T} = (log\,(S_{t+h}) - log\,(S_t))_{t=1,...,T}$$

Where S_t is the index value of the S&P500 at time t, R_h is the log return of the S&P500 over a h-month horizon and r_F is the “risk free” rate which is proxied by the 3-months treasury bill secondary market rate.

Firstly, the use of excess returns, defined as the return on the S&P500 over and above the risk-free rate (here proxied by the 3-months Treasury bill rate), is crucial for isolating the true performance of the market from that which can be earned without bearing any risk. This adjustment allows for a more accurate assessment of the risk-adjusted returns, making it particularly relevant in economic scenarios where risk-free rates may fluctuate due to policy changes or other economic events.

Secondly, the log transformation of returns, as opposed to simple returns, offers several benefits. Log returns are time-additive, a property that simplifies the aggregation of returns over time, making them particularly useful for analysis across different horizons. Moreover, log transformations tend to normalize the returns, reducing the skewness and kurtosis typically observed in financial time series data. This transformation, therefore, helps meet the normality assumptions required in many statistical techniques used for inference and prediction in financial econometrics.

Furthermore, extending the return calculation over a longer horizon h (several months, in this case) helps in smoothing out short-term volatility and provides a clearer view of the underlying trends. Longer horizons filter out short-term noise and are particularly effective in capturing the effects of macroeconomic variables, which typically influence markets over medium to long-term. Macroeconomic variables such as GDP growth rates, inflation, and employment statistics tend to evolve over several months or quarters, and thus, using a longer horizon aligns the target variable more closely with the periodicity of predictive macroeconomic indicators.

Therefore, using log excess returns calculated over an extended horizon as a target variable in modeling the S&P500 not only enhances the robustness of the statistical analysis but also ensures that the insights derived are more reflective of the underlying economic fundamentals, thereby improving the ability to leverage macroeconomic variables for prediction. However, reducing the frequency of excess returns variations by extending the horizon can reduce the practical use for investment purposes, by reducing the number of independent bets a portfolio manager can make. This problem will be further studied later in this paper.

2.5 The prediciton model

Before diving into the prediction model we used to build the signal, it is important to dwell on how we ensure the absence of forward looking biases in the way the model is trained and makes predictions. We implement a rolling window model that works as follows :

- Select a training window $[W_0 : W_L]$ and a test window $[W_{L+1} : W_{L+J}]$ and the subset of data that will be used to train and test the model $(X_t)_{t\in\ [W_0:W_L]}$, $(y_{h,t})_{t\in\ [W_0:W_L]}$, and test the model $(X_t)_{t\in\ [W_{L+1}:W_{L+J}]}$.
- Apply the transformations to the train and test window data in order to obtain the features used in the model, which correspond to principal components of the cyclical components :

$$(X_t)_{t\in[W_0:W_L]} \rightarrow (cc_t)_{t\in[W_0:W_L]} \rightarrow (T_t)_{t\in[W_0:W_L]}$$

$$(X_t)_{t\in[W_{L+1}:W_{L+J}]} \rightarrow (cc_t)_{t\in[W_{L+1}:W_{L+J}]} \rightarrow (T_t)_{t\in[W_{L+1}:W_{L+J}]}$$

- Train the model on the training subset to build a predictive function $\hat{F}$ such that :

$$(y_{h,t})_{t\in\ [W_0:W_L]} = \hat{F}\left((T_t)_{t\in[W_0:W_L]}\right) + \epsilon$$

- Make the predictions on the test subset :

$$(\hat{y}_{h,t})_{t\in[W_{L+1}:W_{L+J}]} = \hat{F}\left((T_t)_{t\in[W_{L+1}:W_{L+J}]}\right)$$

- Move the time window by advancing J time periods, repeat the process to make new predictions

$$(\hat{y}_{h,t})_{t\in[W_{L+1+J}:W_{L+2J}]}$$

- Repeat until the whole period has been covered.

By iterating over the whole dataset, we are able to make predictions past an initial training window while avoiding any forward-looking biases. The model thus only use past data to make computations and forecasts. Calibration for computations such as PCA or regressions are executed on a rolling fashion, thus extending the validation sample to better assess the ability of the model to perform on unseen data.

Employing a rolling window model for forecasting economic variables offers significant advantages over traditional static train/test split methods, particularly in the dynamic context of economic environments.

The rolling window approach better accommodates structural breaks and evolving economic conditions that can significantly affect model performance. Economic variables are inherently influenced by changes in policy, market conditions, and global economic events. A static train/test methodology might train a model on historical data that no longer reflects current economic realities, leading to poor out-of-sample performance. In contrast, a rolling window model continuously updates both the training and testing datasets, allowing the model to adapt and respond to new information as it becomes available. This dynamic updating is crucial for maintaining the relevance and accuracy of the model predictions.

Secondly, this method enhances the robustness of the forecasting model by repeatedly testing it across different time periods. This iterative testing not only improves the generalizability of the model across various market phases but also mitigates the risk of overfitting to a specific period's idiosyncratic features. In economic forecasting, where market regimes can shift unpredictably, the ability of rolling models to integrate and learn from new data continuously provides a clear edge.

Moreover, the rolling window approach naturally aligns with the flow of economic data, which is typically released at regular intervals (monthly, quarterly, etc.). This alignment allows the model to incrementally refine its predictions, integrating the latest economic indicators and thus consistently refining the forecast accuracy. In essence, the rolling window method provides a flexible, adaptive modelling framework that is particularly well-suited to the fluid nature of economic environments, ensuring that the predictions remain pertinent and are based on the most current data. This methodological approach is especially valuable in economics, where the landscape can change rapidly, and historical patterns may not always be reliable indicators of future behaviours.

Eventually, we can modify the size of the rolling window depending on whether we want the model to integrate more datapoints and generalize over several economic scenarios or be more tailored to the economic conditions of the "last 20 years" for instance.

When it comes to the predictive model, we have chosen to work with the XGBoost (Extreme Gradient Boosting) regressor which is a highly sophisticated machine learning model that has proven to deliver strong results in many use cases. This model is particularly well-suited for regression tasks where the underlying data and relationships between variables can be complex and non-linear, thus adressing the observation in Flannery et al. (2002) [10]. The choice of XGBoost over simpler models such as Ordinary Least Squares (OLS) regression is motivated by several of its features and capabilities, particularly relevant when dealing with economic data such as principal components (PCs) that exhibit non-stationarity and strong autocorrelation.

2.5.1 XGBoost Regression

XGBoost is an implementation of gradient boosting machines designed to be highly efficient, flexible, and portable. It operates by constructing an ensemble of weak decision tree learners in a sequential manner. Each subsequent tree attempts to correct the errors made by the previous ones in the ensemble. The model is built in the following way:

1. Objective Function

   The objective function that XGBoost attempts to optimize consists of a loss function and a regularization term Ω, the latter being used to penalize the complexity of the model:

$$Obj\left(\Theta\right) = \sum_{i} L\left(y_i, \widehat{y_i}\right) + \sum_{k} \Omega\left(f_k\right)$$

Where Θ denotes the parameters of the model, y_i the observed values, $\hat{y}_i$ the predictions, and f_k the individual trees.

2. Boosting

   Boosting involves adding new models that solve for net errors left by the previous models. In XGBoost, this is achieved using a gradient descent algorithm on the loss function, specifically tailored for tree-based learners. Each new tree is fitted on the negative gradients of the loss function, effectively reducing the residuals of previous trees.

3. Regularization

   XGBoost incorporates both L1 and L2 regularization, which helps prevent overfitting—a significant advantage over standard boosting methods. This regularization is critical when models are trained on data with noise, ensuring that the model does not learn the noise.

2.5.2 Advantages of using XGBoost

While OLS regressions - widely used for forecasting with economic variables - offer simplicity and interpretability, XGBoost excels in predictive accuracy, especially when dealing with nonlinear relationships and high-dimensional data like principal components.

We found several advantages of using XGBoost over OLS to make predictions using economic data. Firstly, XGBoost can handle non-stationary data, which is common in economic datasets, especially in derived components like PCs from Principal Component Analysis (PCA). OLS relies on strict assumptions of stationary predictors and independently and identically distributed errors (homoscedasticity), while XGBoost is robust to violations of these assumptions, making it more adaptable to real-world data complexities. Then, in cases where PCs exhibit strong autocorrelation, OLS may encounter mis-specification issues and biased estimates, whereas XGBoost, with its tree-based learners, can capture complex patterns such as lag effects and nonlinear relationships, prioritizing predictive accuracy over inference. Also, XGBoost's flexibility and predictive power enable it to model intricate nonlinear relationships and handle various data irregularities including missing values and outliers, which are prevalent in economic datasets. Lastly, XGBoost does not impose any specific distribution requirements on the residuals, unlike OLS, which is advantageous in economic forecasting scenarios where error terms may not adhere to normality due to external shocks or anomalies. Overall, XGBoost offers a robust and powerful framework for regression with PCs in economic analysis.

2.6 Signal Construction

2.6.1 Parameters

Using the described methodology, we build our signal in a rolling fashion with XGBoost Regressor. Two parameters are of main interest for our model : the rolling window size and the number of principal components used to make predictions.

The Size of the Rolling Window

This parameter is relevant because it is related to the memory of our model : How far in the past should the model look back to learn patterns and make predictions for the future ? Economic environments evolve significantly over time due to changes in policy, technology, market structures, and global interactions. Patterns and relationships that held in the distant past may no longer be applicable. For example, the economic environment of the 1970s and 1980s was characterized by issues like high inflation and the oil crisis, which have been quite different from the challenges and economic landscape of the last 20 years. Of course, these issues now echo particularly well since 2021.

Furthermore, economies undergo structural changes over time due to various factors including technological advancements, regulatory changes, and shifts in consumer behavior. These changes can render long-term historical data less informative for future predictions. For instance, the digital revolution and globalization have transformed economic dynamics in ways that data from 50 to 100 years ago could not anticipate. A shorter rolling window also has the advantage of homogeneity. Shorter time windows tend to include data that is more homogeneous in terms of economic policy, consumer behavior, and market conditions. This homogeneity can simplify the modeling process and increase the accuracy of predictions, as the underlying economic conditions are more uniform.

The Number of Principal Components

This parameter relates to balancing simplicity and capturing sufficient variability to make accurate predictions. While later principal components (PCs) account for smaller variations and may appear as noise, they can sometimes contain important information about the data's structure that the first few components miss. This is especially valuable when economic phenomena of interest are influenced by factors not captured by the dominant trends and cycles. However, using more PCs can lead to overfitting issues, which can degrade the model's performance on unseen data and potentially introduce noise into the model.

Using fewer PCs simplifies the model, making it easier to interpret and understand. The first few PCs often have a clear economic interpretation, such as reflecting major economic trends or cycles, which can be directly linked to underlying economic theories. These major economic signals can be sufficient to predict the impact of the overall economy on the stock market.

Therefore, while additional PCs might provide more nuanced insights, the risk of overfitting and complexity must be carefully managed to maintain model robustness and predictive power.

We then construct several signals to determine the best parameters that will strike the balance between accurate predictions and overfitting issues.

2.6.2 Measure of performance & significance

We note $\hat{y}$ the signal thus constructed, ignoring the horizon of predictions h for clarity :

$$\hat{y} = \hat{F}\left(P\left(X\right)\right)$$

Where $P\left(.\right)$ is the pre-processing function that transforms raw data into cyclical components and then principal components. Our goal is to evaluate the information and predictive value of this signal for market timing.

A widely used metric for assessing the strength of the signal is the information coefficient (IC), typically measured using Spearman's rank correlation. In financial analytics, particularly in evaluating market-timing strategies, the IC plays a crucial role. It quantifies the strength and direction of the predictive relationship between a forecast and actual market returns. Utilizing Spearman's rank correlation to compute the IC offers several advantages, making it a pertinent measure for market-timing evaluation.

The IC, defined as the correlation between predicted and actual asset returns, indicates how well a model's predictions align with subsequent outcomes. A higher IC suggests more consistent alignment between predictions and actual returns, indicating a stronger predictive signal. Unlike Pearson's correlation, which assumes linearity and normal data distribution, Spearman's correlation accommodates non-linear relationships and non-normally distributed data. This flexibility is crucial in financial markets where data distributions often exhibit skewness or contain outliers due to market anomalies or high volatility. By evaluating the rank-ordering of data rather than actual values, Spearman's correlation assesses monotonic relationships, whether linear or not, aligning well with the objectives of market-timing strategies.

The Spearman's rank correlation measure is :

$$r = 1 - \frac{6\sum d_t^2}{T\left(T^2-1\right)}, \quad d_t = Rank\left(y_t\right) - Rank\left(\widehat{y_t}\right)$$

Where d_t is the difference between the two ranks of each observation.

To test the null hypothesis H_0 that there is no correlation versus the alternative of a significant correlation between our signal and market returns, we employ the correlation test statistic and its associated p-value. Denoting R_α the rejection region associated with a test at level α for H_0 :

$$t = \frac{r}{\sqrt{\left(1-r^2\right)\left(T-2\right)}}, \text{ p-value} = \inf\{\alpha : t \in R_\alpha\}$$

In addition, we adopt an econometric approach to evaluate our signal and evaluate its predictive power. One robust method is to perform an Ordinary Least Squares regression (OLS) enhanced by the Newey-West variance estimator :

$$y_t = \beta \widehat{y_t} + \epsilon_t$$

In this method, the slope coefficient should ideally approximate 1, indicating a strong and direct relationship between the signal and market returns. Additionally, this coefficient should be statistically significant, meaning a p-value lower than 0.05 (significance at the 5% level). Financial time series are often characterized by autocorrelation and heteroskedasticity— violating classic OLS assumptions and leading to biased standard errors and unreliable hypothesis tests. The Newey-West estimator addresses these issues by providing consistent standard errors, even in the presence of autocorrelation and heteroskedasticity, making it particularly valuable in time series analysis. This procedure constructs an estimate of the covariance matrix of the model parameters that considers autocorrelation up to a fixed lag and potential unequal variances across the data, enhancing reliability.

Furthermore, to gauge performance, we test whether the signal demonstrates predictive ability in both favorable and adverse market conditions. Past research suggests that economic indicators may forecast returns during bad times but not in good times. To test this, we estimate a linear two-state predictive regression by including an indicator variable, as suggested by Boyd et al. (2005) [4] :

$$y_t = \beta_{bad} \chi_{bad} \widehat{y_t} + \beta_{good} \left(1 - \chi_{bad}\right) \widehat{y_t} + \epsilon_t$$

Where χ_{bad} take a value of 1 during recessions and 0 otherwise. Bad times are defined using NBER's recession dates. Similarly, the coefficients β_{bad},β_{good} should ideally approach 1 (and be positive) with a p-value below 0.05.

3 Analysis

3.1 Data Analysis

3.1.1 Data description

For our analysis, we used the FRED-MD (Federal Reserve Economic Data - Monthly Database) dataset from January 1959 to December 2023, with an in-sample period from 1959 to 2012 for both the training period (1959-1967) and validation period (1967-2012). This database offers a comprehensive historical perspective spanning 65 years of economic data observed on a monthly basis and provides a holistic view of various aspects of the US economy.

The data analysis acknowledges several key characteristics of the FRED-MD dataset that shape our analytical approach. Firstly, a significant proportion of the data series exhibit non-normality, which is a common feature in economic data. Then, the presence of trends in some or all of the data series reflects the long-term growth and evolution of the US economy. Therefore, we have employed detrending techniques to facilitate accurate analysis. Finally, the prevalence of non-stationarity, demonstrated by the failure of approximately 78% of the data to pass the

Augmented Dickey–Fuller Stationarity test at the 10% level, highlights the need to transform data using methods such as differencing or cyclical component extraction to prevent spurious regressions. A sample of the time series used for our analysis can be visualised in Figure 1.

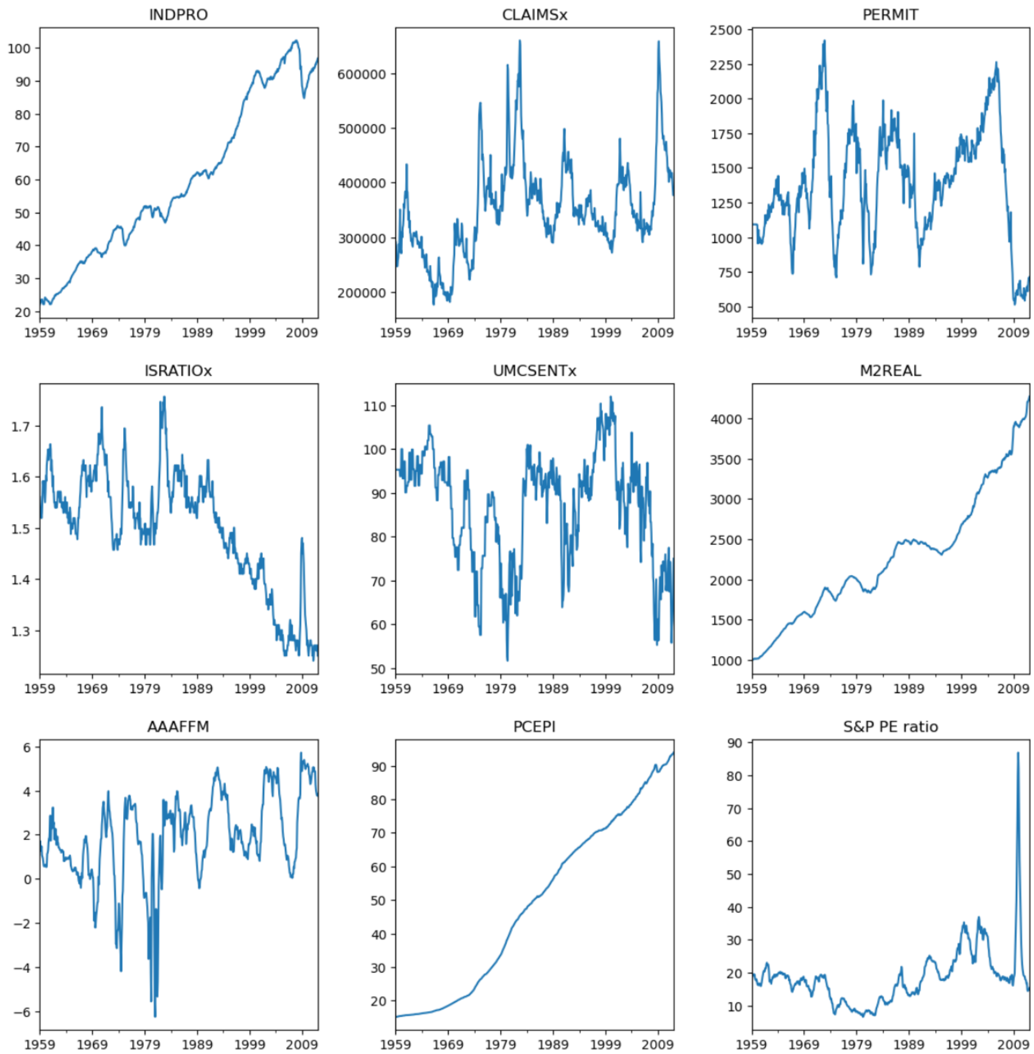

Figure 1: Visual of some macroeconomic time series indicators

The analysis of the correlation matrix reveals high correlation across our set of indicators. One can visualize this observation in the correlation heatmap provided in the Appendix on Figure 19. This is a classic observation for macroeconomic indicators for one country. High correlation across predictors can reduce the efficacy of several models and motivates the implementation of aggregation techniques to reduce the dimensionality of our dataset and provide independent predictors. We thus use a Principal Component Analysis (PCA) to address colinearity issues, ensuring that our data analysis is both comprehensive and robust in examining macroeconomic trends and relationships within the FRED-MD dataset.

3.1.2 Treatment : cyclical components and feature engineering

In our dataset, we employ several steps to enhance the quality and characteristics of our time series data. Firstly, we utilise a one-year moving average detrending method, which allows us to extract the cyclical component from our data. This method enables us to isolate the cyclical patterns from the overall trend, thereby facilitating a clearer analysis of cyclical fluctuations observed in various economic indicators. Additionally, we standardise the data, thus ensuring consistency and comparability across different variables. This, in turn, allows us to provide interpretations of our results that are more meaningful and reliable, thereby enhancing the robustness of the analysis.

Moreover, we employ a winsorizing process at the 95% level, which addresses any potential outliers in our dataset to enhance generalization to unseen data. This process involves capping extreme values, thus mitigating the impact of any such outliers on our analysis. It also serves to ensure the robustness of our results, given that the outliers are removed. Furthermore, we apply an exponential smoothing filter to improve the signal-to-noise ratio in our dataset, improving the forecasting potential of our data.

In summary, these treatments serve to refine the properties of our time series data by reducing skewness, noise, and extreme variance. As a result of these interventions, we observe a notable increase in the level of stationarity across the dataset, with the proportion of non-stationary series decreasing from 78% to 22%. This improvement in stationarity signifies greater stability and reliability of our data, laying a stronger foundation for our subsequent analyses and modelling endeavours. The resulting cyclical components can be observed in Figure 2.

3.2 Principal Components

Across the research sample spanning from 1959 to 2012, 15 principal components can explain roughly 90% of the cyclical components variance which is our target level of captured variance (c.f. Appendix Figure 20). Beyond this level, it is likely that we are capturing more subtle variations or noise that is less likely to drive the market.

3.2.1 Macroeconomic Interpretations

Upon analysing the factor loadings (Appendix, Figure 21) produced by the Principal Component Analysis (PCA) on our dataset, we are able to interpret economically the relevant principal components for our analysis. Although there is no clear grounds on how to interpret the factor loading and the components, this interpretation gives more economic context to the variables we employ for our prediction.

We identified PC1 as an indicator for overall price levels in the economy. The inclusion of various price indices, including PPI for crude materials, intermediate materials, finished consumer goods, and CPI, suggests that PC1 captures inflation across different stages of production and consumption while PC4, PC6, PC12, and PC14 are linked to banking credit and loan indicators.

Table 1 summarizes our interpretation of the principal components by observing the factor loadings produced by the principal component analysis.

Interpreting the later principal components becomes increasingly challenging due to the dispersion of loadings across many variables. However, the interpretation of the initial principal

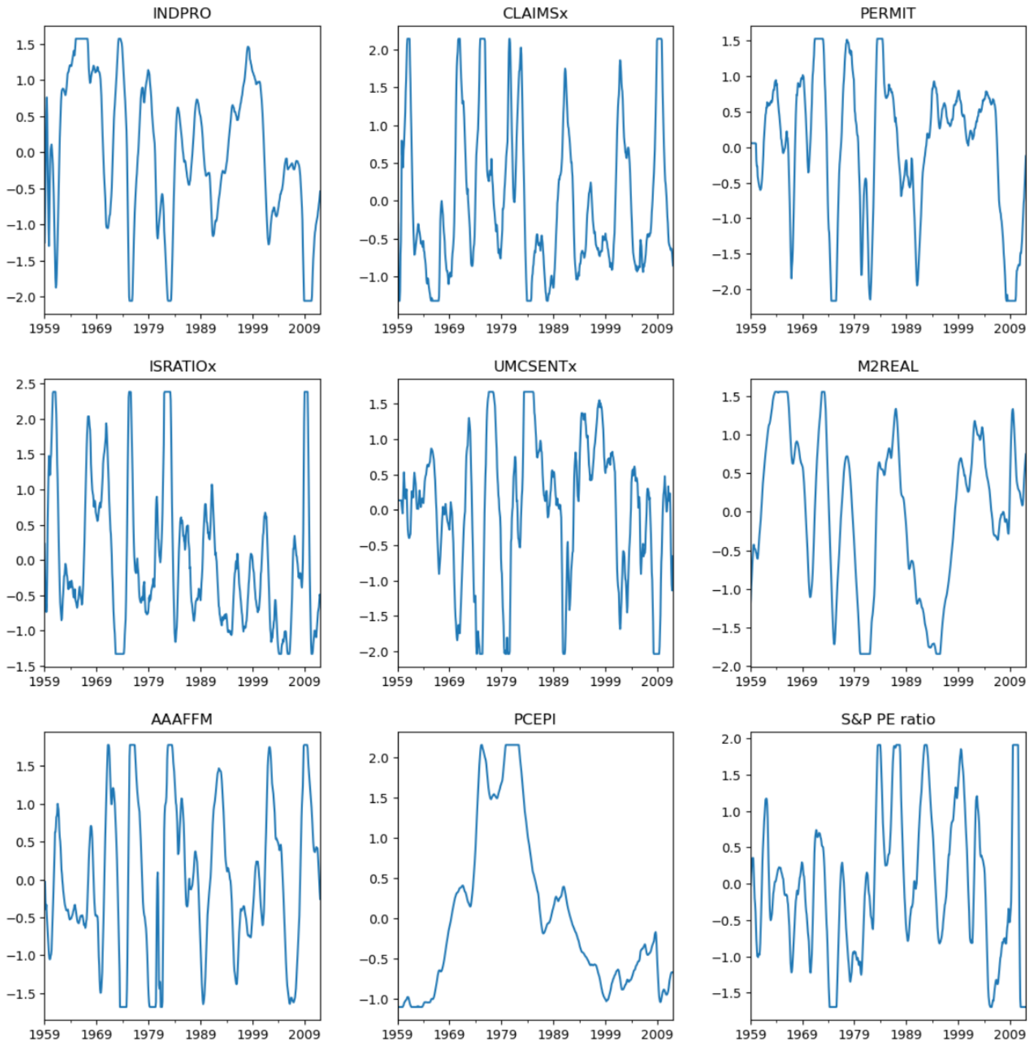

Figure 2: Resulting cyclical component after pre-processing technique

components is relatively straightforward and reassuring. This indicates that our primary principal components effectively capture significant aspects of the overall economy, ensuring that the essential economic signals are well-represented in our analysis.

3.2.2 Time series & distribution analysis of the principal components

Our analysis of the time series properties and distributional characteristics of the principal components revealed that the majority of them are stationary, with only two out of 15 exhibiting non-stationarity at the 10% level. This indicates that the principal components capture underlying economic factors with relatively stable statistics over time, making them suitable for time series analysis techniques. However, it's important to note that all principal components exhibit strong levels of autocorrelation and partial autocorrelation, suggesting persistent underlying patterns. This persistence can lead to issues in OLS models, potentially biasing parameter estimates and undermining the validity of statistical inference.

Furthermore, our analysis reveals that the principal components are not normally distributed, exhibiting strong levels of kurtosis. This departure from normality suggests that the distributions of the principal components are characterised by heavy tails and peakedness,

Table 1: Economic Interpretation of PCs

PC id	Interpretation	Key Indicators
0	Aggregate Supply	Production indexes, Labor market indicators (e.g., Unemployment Rate)
1	Price Component	Price indexes (PPI, CPI)
2	Housing Market	Housing-related indicators (permits, interest rates)
3	Business Activity and Employment	Inventory-to-Sales Ratio, Orders (capital and consumer goods), Employment and Hours Worked indicators
4	Consumer Credit	Consumer loans, Non revolving credit
5	Energy Prices and Stock Market	Energy price indicators (Fuel Price, Oil Price), S&P PE Ratio
6	Consumer Credit	Consumer loan indicators (Consumer Loans, Non-revolving Credit)
7	Foreign Exchange (FX)	US FX indices (Trade-Weighted, exchange rates)
8	Interest Rates	Interest rate spreads (CP-FFR, etc.)
9	Money Stock	Securities in Bank Credit
10	Money Supply and Stock Market	Money supply indicators, S&P PE Ratio
11	International Linkages with Financial Health	Canada FX rate, S&P Dividend Yield, S&P Industrials Index, Total Reserves, Non-borrowed Reserves
12	Banking System and Housing	Securities in Bank Credit, US FX rate, Real Estate Loans
13	Stock Market and Consumer Sentiment	Stock Market indicator, Consumer Sentiment Index
14	Credit and Business Loans	Securities in Bank Credit, Commercial and Industrial Loans, Total Non-revolving Credit

indicating non-Gaussian behaviour. The non-normal distribution of the principal components has implications for the choice of statistical methods and inference procedures, highlighting the need for robust modelling techniques that can accommodate non-Gaussian data. Figure 3 summarizes the time series analysis for the first principal component. This supports the use of alternative non-linear models such as XGBoost regression.

Interestingly, we observe that the 15th principal component displays lower levels of autocorrelation and demonstrates more normal distributional characteristics with lower kurtosis and skewness. This pattern indicates that the more distant principal components may be capturing more noise, hence the convergence towards a Gaussian distribution.

3.3 Signal Analysis

The above analysis enables us to perform the preprocessing phase of the model. We are able to extract cyclical components and aggregate them into principal components, providing the

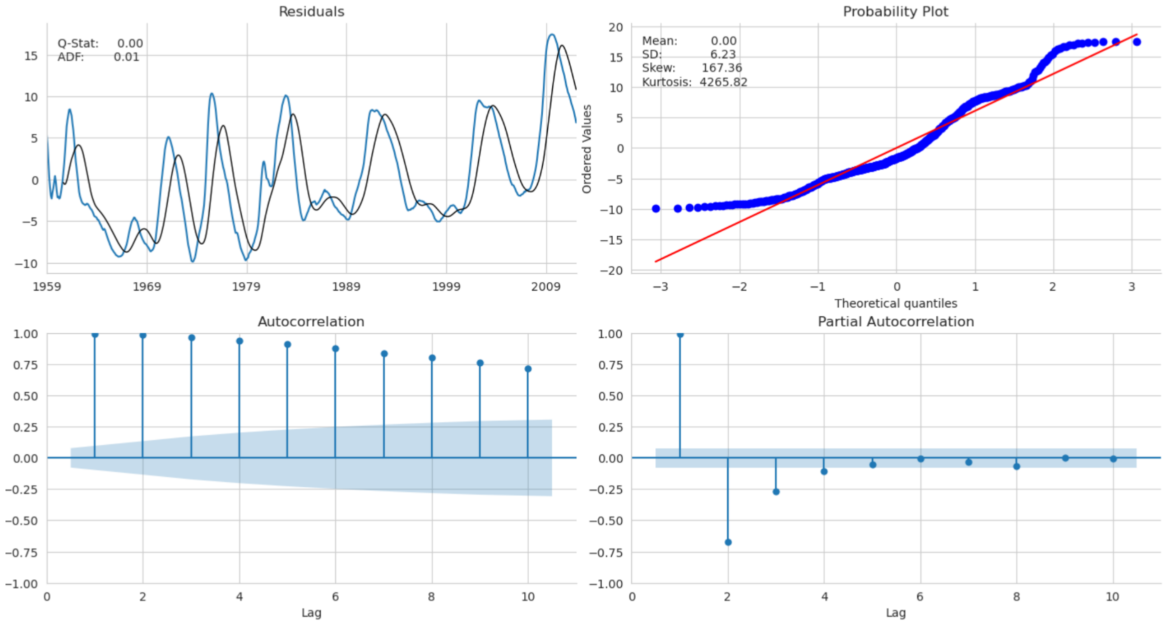

Figure 3: Correlogram & time series analysis for this first PC

predictors for our model. We can now run the model and adjust the parameters to build our macroeconomic signal.

3.3.1 Parameters selection

We need to select the two main parameters discussed earlier for this model, namely the size of the rolling window (which influence the memory of the model) and the number of components used to make predictions.

Table 2 shows performance measures of signals built for different levels of the parameters rolling widow size (W) and number of principal components (PC), for a one year horizon excess return target. The analysis of the result highlights the significance of meticulously selecting parameters in predictive modelling. While larger rolling window sizes may provide more comprehensive data coverage, they may also introduce biases and reduce the model's predictive power. Similarly, while increasing the number of PCs can enhance predictive performance up to a certain point, an excessive number of PCs may lead to overfitting and diminish the model's ability to generalise to new data. Consequently, a balanced approach to parameter selection is essential for optimizing the predictive accuracy and robustness of the model.

The analysis of the table reveals that the best performing combination is achieved through a window of 240 months (20 years) and 9 principal components. This is the combination of parameters we will thus use for the out of sample and robustness analysis.

Figure 4 is the representative graph of our macroeconomic signal juxtaposed with the 1-year future excess return over the in-sample validation period, spanning from 1967 to 2012, thus eliminating the training period. It is notable that both curves exhibit in general similar patterns, often mirroring each other's movements and occasionally overlapping. This alignment suggests a relationship between the identified macroeconomic signal and subsequent market performance, as evidenced by their synchronised fluctuations. The shading, which denotes

Table 2: Statistical Data for Different Window Sizes and Principal Components, h=12 (1 year)

	W = 120			**W = 240**			**W = 360**		
PCs	**IC**	$\boldsymbol{R^2}$	$\boldsymbol{\beta}$	**IC**	$\boldsymbol{R^2}$	$\boldsymbol{\beta}$	**IC**	$\boldsymbol{R^2}$	$\boldsymbol{\beta}$
1 PC	0.18***	0.02	0.16*	0.12***	0.01	0.17	−0.00	0.00	0.06
2 PCs	0.33***	0.08***	0.35***	0.17	0.02	0.21***	0.21***	0.04***	0.29***
4 PCs	0.28***	0.09***	0.36***	0.34***	0.11***	0.44***	0.30***	0.09***	0.44***
6 PCs	0.25***	0.06**	0.31***	0.26***	0.06***	0.33***	0.20***	0.03**	0.21**
9 PCs	0.24***	0.08***	0.43***	**0.32*** **	**0.11*** **	**0.49*** **	0.16***	0.03	0.27*
12 PCs	0.25***	0.09***	0.44***	0.25***	0.09***	0.49***	0.14***	0.03	0.29*
15 PCs	0.20***	0.06***	0.35***	0.24***	0.10**	0.49*	0.09**	0.04**	0.38**
20 PCs	0.24***	0.08***	0.40***	0.24***	0.09***	0.51***	0.17***	0.07**	0.52**

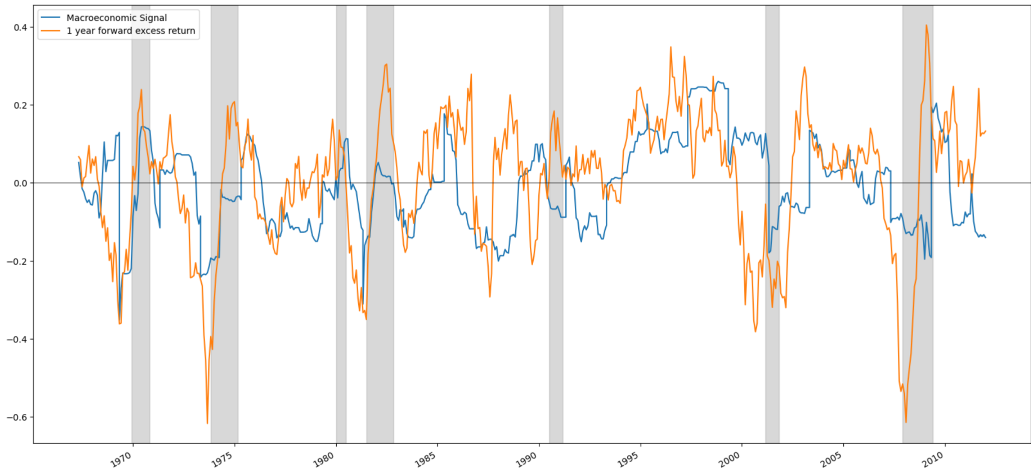

Figure 4: Macroeconomic Signal and 1-year Excess Return, Validation Period (1967-2012)

NBER-dated recessions, provides crucial contextualisation, highlighting periods of economic downturns and their potential impact on the observed dynamics.

3.3.2 Signal performance during good and bad times

To evaluate the predictive power of our signal during different market conditions, we have run the two-state predictive regression. We found that the signal has a statistically significant predictive power in good times and bad times. The performance of our signal in good and bad times is summarized in Table 3.

Table 3: Signal analysis during good and bad times

$\mathbf{R^2 = 12.3\%}$	**Coefficient**	**p-value**
$(1 - \chi_{\text{bad}})y_t$	$\beta_{\text{good}} = 0.432$	0.000***
$\chi_{\text{bad}}y_t$	$\beta_{\text{bad}} = 0.8379$	0.028**

Previous studies have highlighted the significant influence of market sentiment and volatility on the effectiveness of predictive models. In particular, signals are often observed to exhibit stronger predictive power during periods of market stress (Liu and Kemp, 2019 [15]). Our signal seems to work well in both conditions, which plays in favor of its robustness. This can be explained by the wide range of variables taken into account into building it while previous research often focus on a single indicator.

3.3.3 Signal stability over the prediction period

To assess the reliability of our signal's predictions across different periods, we analysed the variation in information coefficients (IC) over the prediction timeline. In fact, even the most robust signals will not be 100% reliable due to the unpredictable nature of markets. However, it is crucial for informed decision-making to be able to discern when the signal is underperforming.

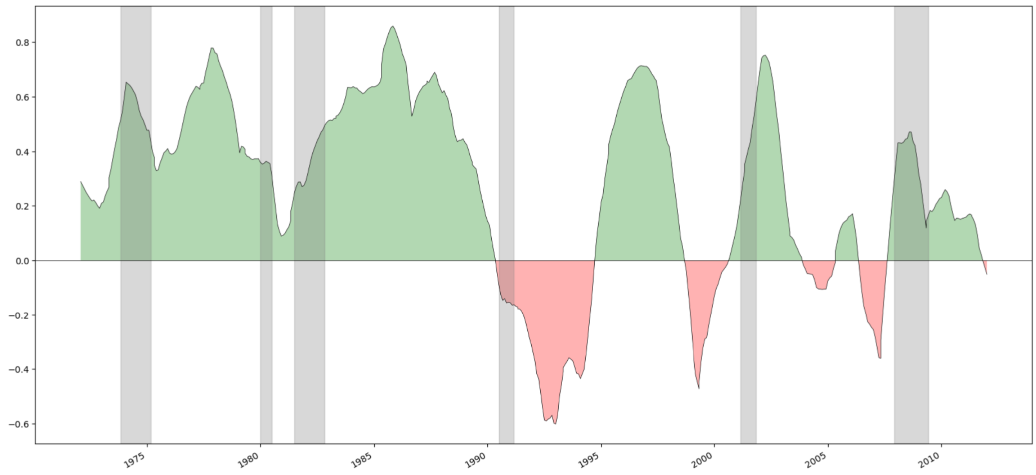

Figure 5: Macroeconomic Signal IC variations

Figure 5 displays the five-year moving average of the signal's IC during the predictive period. As illustrated, we can assume that the signal exhibits notable reliability before 1990, characterised by a consistent performance. However, a downturn is evident between 1990 and 1995, marked by a five-year period of negative performance. Subsequent to this, intermittent negative periods emerged in 2000, 2005, and 2009, indicating periods of weaker signal reliability

4 Robustness checks

A crucial aspect of any predictive model is its ability to perform well on unseen data. To address this, we compute the prediction using the same signal construction process by including the out of sample period of 2012-2024, which was entirely excluded from the model calibration process. This ensures the model is not simply memorizing patterns in the training data but can identify genuine relationships between economic factors (PCs) and stock market returns.

In the robustness check, we first conduct the same tests of predictive power and statistical significance as we did before. The out-of-sample analysis yield impressive results ($IC =$ 0.290***, Correlation= 0.350, $R^2 =$ 0.381***, $\beta =$ 1.154***) that are statistically significant and even exceed the in-sample performance. While the relatively short out-of-sample period might contribute to this stronger performance, it nonetheless demonstrates that the model's ability to predict returns generalizes well to unseen data. This creates confidence in the model's effectiveness for future predictions.

Figure 6 shows that the additional out-of-sample signal (delimitated by the black horizontal dotted line) still follows the same pattern as the 1-year forecast return but with more discrepancies.

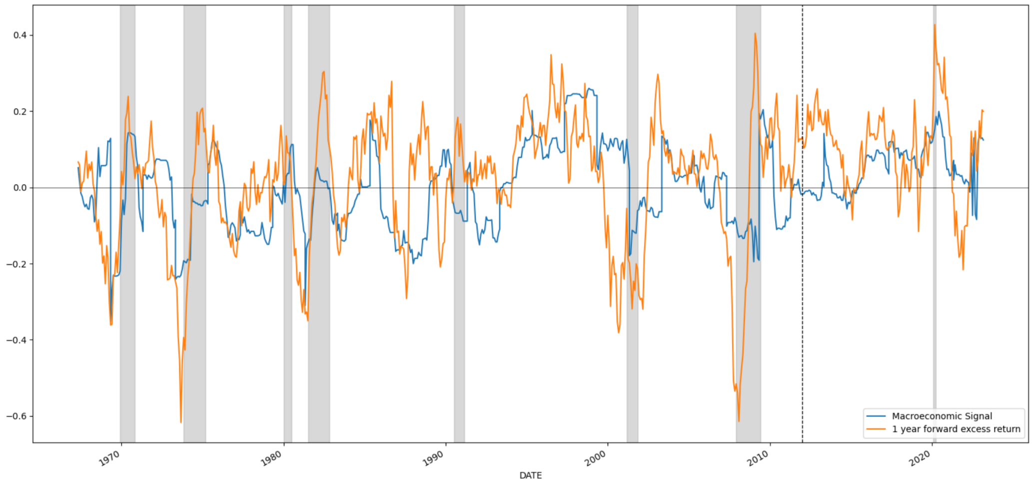

Figure 6: Macroeconomic Signal and 1-year Excess Return

In addition, we check the predictability of the signal in both good and bad states, although it makes less sense given the small period of recession that occurred only during the covid pandemic. Again, the signal performs well in both conditions and is significant. Results are summarized in Tabe 4.

Table 4: Signal analysis during good and bad times, out-of-sample

$\mathbf{R^2 = 40.9\%}$	**Coefficient**	**p-value**
$(1-\chi_{\text{bad}})y_t$	$\beta_{\text{good}} = 1.06$	0.00^{***}
$\chi_{\text{bad}}y_t$	$\beta_{\text{bad}} = 2.17$	0.00^{***}

Furthermore, we compute the variations in IC using a five-year moving average. We can observe in Figure 7 that the signal has been particularly strong in the out-of-sample period.

5 Complementary Analysis

5.1 Signal Performance for Several Time Horizons

To further assess the model's robustness, we investigated its effectiveness for predicting returns across different time horizons: 3 months, 6 months, 1 year, and 2 years. A 1-year horizon emerged as a promising balance, offering sufficient time for the signal to capture slower economic fluctuations while remaining valuable for investment decisions. Table 5 summarizes the results

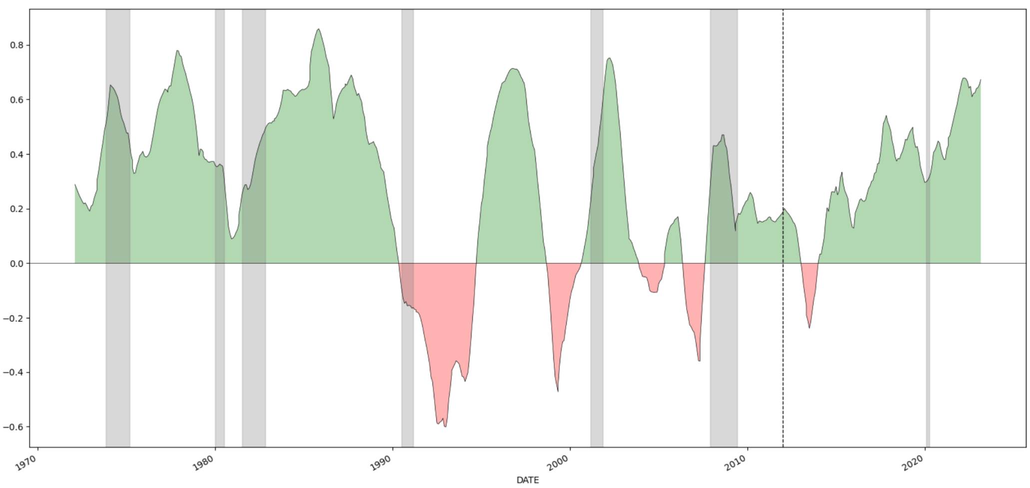

Figure 7: Macroeconomic Signal IC variation, 1967-2023

we obtain for several time horizons for the same parameters configuration already selected (9 PCs, 240 months window). Its analysis revealed interesting insights. While the signal performed well overall, it exhibited some nuances across timeframes. For shorter horizons (3 months), the ability to predict good market periods was weaker, although the signal remained strong at predicting bad times. This is a common challenge in economic signal design for market timing. Conversely, the 2-year horizon yielded excellent results, unsurprising considering the long-term influence of economic factors on the market. However, a 2-year timeframe might be less practical for investment strategies, reducing the number of independent bets a portfolio manager can make. The analysis highlights a key trade-off: shorter time horizons offer more frequent opportunities to leverage the signal but might be noisier for identifying positive market trends. In fact, this trade-off between the signal's strength and the number of bets that can be made using this signal is illustrated by Treynor & Black's (1973) Information ratio [23] :

$$IR = IC \times \sqrt{BR}$$

Where the breadth BR represents the number of independent bet a portfolio manager can make.

5.2 Signal Performance across industry portfolios

In this section, we explored the applicability of the signal beyond overall market returns by investigating its effectiveness in timing industry performance. This analysis aimed to determine if the signal could be used to manage a portfolio of industries based on their forecasted performance. This would reinforce the signal's practical use for investment management by increasing the number of independent investment decisions a portfolio manager could make based on the signal. Using an investment horizon h in months, and 10 industry portfolios, we can make $BR = 10 \times \frac{12}{h}$ independent bets per year.

To conduct this investigation, we import 10 industry portfolio returns from Ken French's online library [11] and evaluate the signal's predictive power for industry-specific log-excess

Table 5: Statistical Analysis for Various Horizons and Samples

	Validation Set: 1968-2012			
Horizon	**h=3**	**h=6**	**h=12**	**h=24**
IC	0.256***	0.294***	0.322***	0.306***
R^2	4.7%	7.1%	11.3%	8.1%
β	0.360***	0.414***	0.494***	0.352**
β_{bad}	0.259***	0.303***	0.838**	0.337***
β_{good}	0.769**	1.06***	0.432***	0.494***
	Out of Sample: 2012-2024			
Horizon	**h=3**	**h=6**	**h=12**	**h=24**
IC	0.209**	0.132	0.290***	0.35 * **
R^2	5.3%	9.7%	38.1%	51.9%
β	0.385*	0.618**	1.154***	0.962**
β_{bad}	1.942***	2.32***	2.17***	0.925***
β_{good}	0.306	0.507*	1.06***	1.427***

returns across different time horizons.

Table 6: Performance Across Various Industry Sectors for Different Time Horizons

	NoDur	**Durbl**	**Manuf**	**Energy**	**HiTec**	**Telcm**	**Shops**	**Hlth**	**Utils**	**Other**
				k=3 (3-months horizon)						
IC	0.29***	0.26***	0.32***	0.11***	0.34***	0.20***	0.29***	0.31***	0.24***	0.34***
R^2	10.6%	8.9%	10.1%	1.6%	9.8%	7.4%	8.3%	12.7%	9.9%	13.8%
β	0.46***	0.55***	0.45***	0.22**	0.44***	0.38***	0.42***	0.45***	0.41***	0.57***
				k=12 (1-year horizon)						
IC	0.26***	0.22***	0.29***	0.16***	0.37***	0.18***	0.26***	0.27***	0.28***	0.18***
R^2	19.7%	11.9%	13.5%	6.6%	16.8%	8.2%	18.7%	18.1%	13.3%	9.9%
β	0.60***	0.64***	0.52***	0.39***	0.68***	0.39***	0.66***	0.56***	0.53***	0.46***
				k=24 (2-years horizon)						
IC	0.43***	0.15***	0.32***	0.21***	0.39***	0.29***	0.31***	0.37***	0.32***	0.32***
R^2	47%	9.6%	29.4%	16.1%	15.6%	15.3%	33.6%	40.1%	24.6%	21.8%
β	0.91***	0.40***	0.73***	0.48***	0.44***	0.72***	0.81***	0.64***	0.56***	0.56***

Our results from Table 6 illustrate well the concept of "Fundamental Law of Active Management" by Grinold & Kahn [12], which suggests that increasing the number of independent bets (industries) can lead to better portfolio performance if the best bets have positive information coefficients (IC). As the investment horizon shortens, market returns become noisier from an economic predictability standpoint, leading to lower IC values.

The results revealed a trade-off between predictive ability and the number of potential bets. By increasing the horizon from 3 months to 2 years, the average IC across industries more than doubled (0.09 to 0.21). However, this also reduced the available bets by a factor of 4, potentially impacting the information ratio (a measure of risk-adjusted return).

Furthermore, the analysis identified significant differences in predictability across industries. Certain sectors, like HiTech and Non-Durables, exhibited much stronger signal performance compared to others. This disparity can be attributed to the inherent characteristics of these industries. Defensive industries like Non-Durables tend to have performances highly sensitive to economic fluctuations, as they tend to be favored by investors during bad times, making them more susceptible to economic signal predictions. Similarly, growth-oriented industries like HiTech often rely on future cash flows, which are heavily discounted by the prevailing interest rates. Since interest rates are influenced by economic factors, these industries become more responsive to economic signals.

The variation of information coefficients for the 1 year horizon signal relative to all 10 industries can be observed in Figure 8. Overall, the different signals evolve in a consistent pattern. Table 7 illustrates the trade-off between breadth and information for the different time horizons. While the 3 months horizon has the lowest cross-sectional correlation, it could prove the most helpful in practice for portfolio managers as it has the highest information ratio.

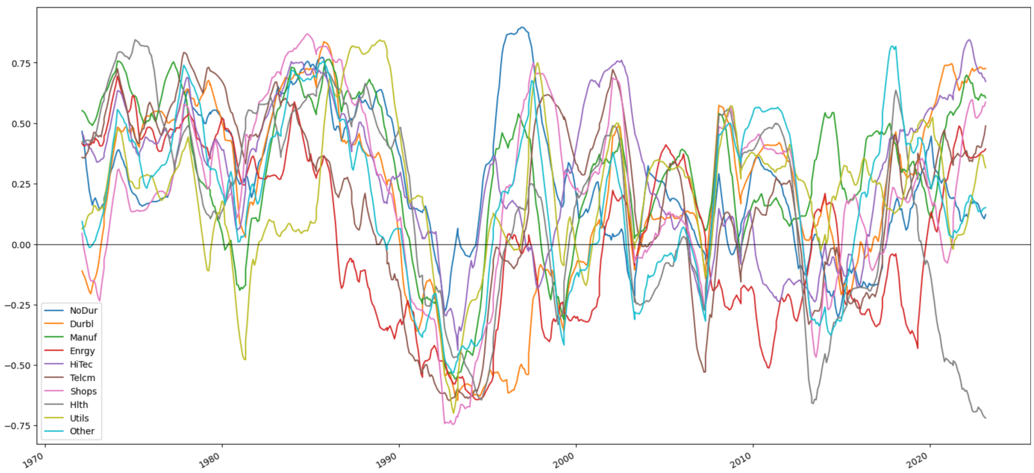

Figure 8: Macroeconomic Signal IC variation for 10 industries portfolios, 1967-2023

Table 7: Information Ratio, XS Correlation and Breadth for different time horizons

	XS Correlation	**Breadth**	**Information Ratio**
k=3 (3-months horizon)	0.090	40	0.567
k=6 (6-months horizon)	0.098	20	0.412
k=12 (1-year horizon)	0.170	10	0.432
k=24 (2-years horizon)	0.212	5	0.468

6 Performance relative to a single variable signal

6.1 The cyclical consumption signal

Atanasov et al. (n.d.) [2] introduced a consumption-based variable - the cyclical consumption - to predict stock returns. They demonstrate the predictive power of cyclical consumption for the same excess return we have used for our study. To build their signal, the authors extract the cyclical component of the aggregate seasonally adjusted consumption expenditures on nondurables and services from the National Income and Product Accounts (NIPA) built by the Bureau of Economic Analysis. In their paper, they prove that cyclical consumption's performance is not confined to bad times and do better than many popular forecasting variables.

Our macroeconomic signal is comparable to the cyclical consumption signal in the sense that it extracts the cyclical component of macroeconomic variables (like consumption) before aggregating them using a principal component analysis. However, our signal and study present significant departures from the authors' methodology. First, we consider a different cyclical component extraction methodology as we eliminate the trend component by substracting the moving average to the log-transformed variables. Secondly, we aggregate several macroeconomic indicators while the authors focus on a single variable. Additionally, we define our signal as the point-in-time predictions of the stock market excess-returns resulting from our rolling model, while the authors directly employ the cyclical component of consumption as their signal. The authors' approach has the advantage of being very simple and transparent, but at the expense of lower comparability with excess returns, as can be observed in Figure 9.

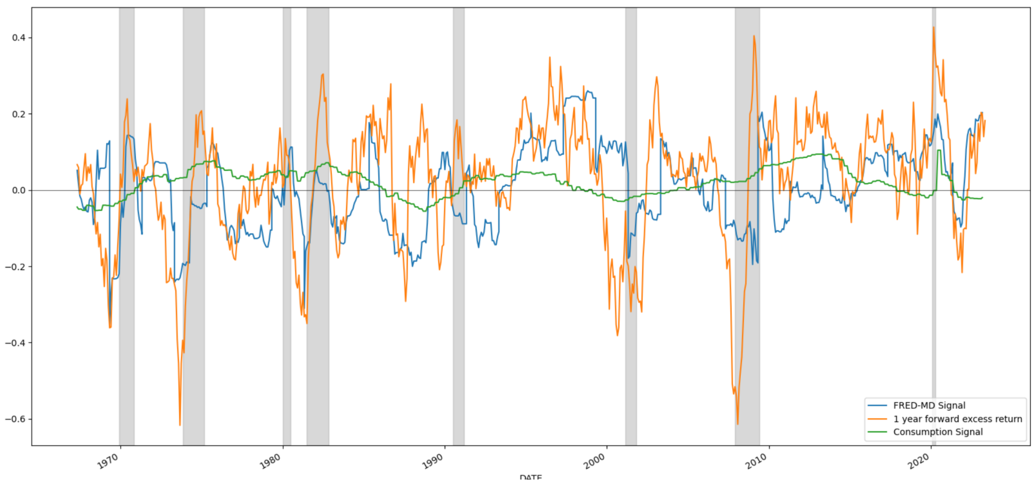

Figure 9: Macroeconomic signal vs Consumption signal vs Excess Return, 1967-2023

It is noteworthy that the cyclical consumption signal is slower than the macroeconomic signal (FRED-MD signal) as it moves cyclically over long periods of times, while excess returns vary over shorter periods. The amplitude of variations also do not match these of excess returns. Overall, the cyclical consumption signal built by the authors are less comparable to excess returns.

We can observe in Figure 10 the Information Coefficient (IC) variations from the negative value of the cyclical consumption variable that we have replicated from the authors' paper. The signal performs overall very well, exhibiting opposite variations as the market excess returns, which verifies the authors observation that "future expected stock returns are high (low) when aggregate consumption falls (rises) relative to its trend". Interestingly, the consumption signal also exhibits the same weakness as our macroeconomic signal between 1990 and 1995.

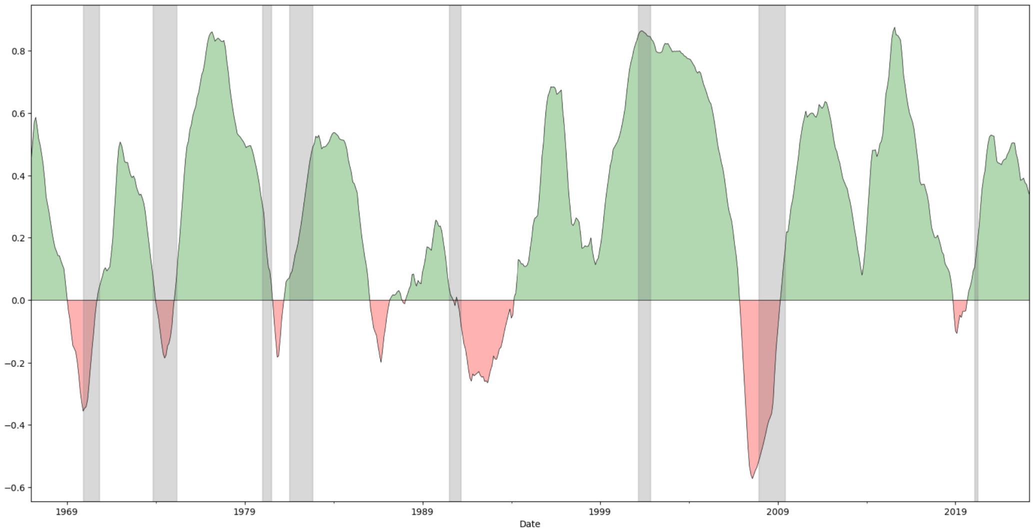

Figure 10: Opposite of cyclical consumption's IC relative to excess market returns, 1967-2023

6.2 Comparison of performance

To compare the relative performance of our signal (FRED-MD signal) versus the cyclical consumption signal, we run the same regression and correlation analysis that we used so far. The results are summarized in Table 8. We used quarterly data since the consumption original data is quarterly released.

Table 8: Comparison of Statistical Measures for Consumption and FRED-MD Signals

	Consumption Signal	**FRED-MD Signal**
IC	0.224***	0.409***
R^2	5.5%	17.9%
β	-1.10***	0.66***
β_{good}	-1.11***	0.60***
β_{bad}	-1.29	1.00***

We can observe that the macroeconomic aggregate signal we have built performs better on all counts. This is not surprising given that we aggregate more than a hundred macroeconomic variables while the cyclical consumption signal is based only on consumption. In the comparable period (1967 - 2023) we note that the consumption signal is not statistically significant in bad times, while the macroeconomic signal is. Nonetheless, the performance of the consumption signal is strong given it is based on a single variable.

6.3 Additivity

When comparing signals it is interesting to study how they perform together. This refers to the concept of additivity whereby several signals can be combined to produce a stronger signal. Signals have high chances to be additive when they cover different datasets and are built using different methodologies, thus producing less correlated outputs. The Pearson and Spearman correlation between both signals is respectively 2.1% and 6.5% which is relatively low and bodes well for additivity properties. To test the addtivity of the cyclical consumption with our macroeconomic signal, we can build the mean signal or include both signals as predictors in a multi-dimensional regression.

6.3.1 Multi-dimensional regression

The multi-dimensional regression including both the FRED-MD and Consumption signal provide the results summarized in Table 9.

Table 9: Additivity Analysis for the multi-dimensional model

Model & Signal	**Coefficient**	**p-value**
Simple Model (R-square = 23.7%)		
Consumption Signal	$\beta_1 = -1.14$	0.007***
FRED-MD Signal	$\beta_2 = 0.66$	0.000***
Two States Model (R-square = 25.5%)		
Consumption Signal Good Times	$\beta_4 = -1.04$	0.015**
Consumption Signal Bad Times	$\beta_5 = -1.83$	0.007***
FRED-MD Good Times	$\beta_6 = 0.58$	0.000***
FRED-MD Bad Times	$\beta_7 = 1.14$	0.006***

These results provide evidence that there is additivity between both signals as the presence of both signals in the regression significantly increases the power of the model. We also observe that signals are statistically significant in all states (good and bad times) which supports the relevance of both signals.

6.3.2 Average signal

Another way to test the additivity of both signals is to consider a mean signal, thus aggregating the information contained in both signals in a simple manner. Note that this method is similar to the multi-dimensional regression model except it is more restrictive in the way the model sets the coefficient since there remain only one dimension. We obtain the results summarized in Table 10.

This mean signal also demonstrates the additivity between both signals and present a slightly stronger performance than the previous method. Overall, this additivity between both signals is very auspicious as it enables to build a stronger signal and aggregate different sources of

Table 10: Additivity Analysis for the aggregate signal

	Coefficient	p-value
Simple Model (R-square = 24.2%)		
Mean Signal	$\beta_1 = 0.723$	0.000^{***}
Two States Model (R-square = 25.8%)		
Mean Signal Good Times	$\beta_4 = 0.65$	0.000^{***}
Mean Signal Bad Times	$\beta_5 = 1.18$	0.002^{***}

information together. It also has the advantage of employing a different methodology for building a consumption signal which can be more tailored to the specificities of consumption while employing a more general method to all indicators contained in the FRED-MD database.

7 An alternative aggregation method : Diffusion Indexes

Diffusion indexes are commonly used by many economic institutions to aggregate and forecast economic data through leading indicators. For instance, The Conference Board in the United States employs diffusion indexes extensively in its composite indexes, such as the Leading Economic Index (LEI), which aggregates various economic indicators to forecast economic activity. Similarly, other organizations, such as purchasing managers' associations in various countries, use diffusion indexes to gauge economic conditions across manufacturing and service sectors [5].

Diffusion indexes serve as an alternative to more complex methods like Principal Component Analysis (PCA) when it comes to aggregating information from a pre-formed cluster of indicators. Unlike PCA, which reduces data dimensionality by transforming correlated variables into a smaller number of uncorrelated variables, diffusion indexes typically measure the proportion of indicators that are moving in a pattern consistent with the overall index. This method can be more intuitive and simpler to interpret, particularly when the indicators are already grouped logically or thematically. Analysts often rely on these diffusion indexes to provide a measure of the breadth of the change in a composite index.

However, there are advantages and drawbacks to using diffusion indexes. One major advantage is their simplicity and ease of interpretation, as they provide a straightforward percentage of indicators showing improvement. This can be particularly useful for policymakers and analysts looking for quick insights into economic trends. On the downside, diffusion indexes may oversimplify complex dynamics because they do not account for the magnitude of changes in the underlying data, only the direction. Additionally, the choice and weighting of indicators in a diffusion index can significantly affect its outputs, potentially introducing bias or overemphasizing certain sectors or variables. This contrasts with PCA, which inherently considers the variance and correlation structure of the data to determine the principal components.

7.1 Choosing the index components

We use the eight groups of macroeconomic indicators derived from the framework established by Ludvigson and Ng (2009) [16]. In their paper, the authors sought to improve the predictability of bond risk premia using macroeconomic factors derived from the same dataset. Using prior information, they organized 131 monthly macroeconomic time series into eight coherent groups, aiming at simplying the complexity inherent in such a vast dataset while retaining the essential economic signals. The initiative came from the observation that principal component analysis provides factors that are difficult to interpret. Thus, they organized the dataset into 8 clusters before building latent factor for each macroeconomic cluster using a factor augmented regression (FAR) technique. The grouping helps isolate different facets of the macroeconomic environment, ensuring that significant but distinct economic activities are adequately represented in the analysis. The resulting clusters refer to (1) Output and Income, (2) Labor Market, (3) Housing Market, (4) Consumption and Orders, (5) Prices and Inflation, (6) Money and Credit, (7) Interest Rates and Spread, (8) Stock Market (see Appendix in Table 15 to Table

Ludvigson and Ng successfully demonstrated that these eight groups could be used to construct factors with significant predictive power for bond risk premia. By leveraging the economic relevance of these clusters, the authors were able to enhance the interpretability and robustness of their factor models. The factors derived from these groups showed strong correlations with economic activity and provided valuable insights into the dynamics of financial markets.

We find this analysis useful for our research as it incorporates macroeconomic insights into our quantitative analysis. The grouping is based on established economic theory and empirical findings, providing a solid foundation for the analysis. This enhances the credibility and relevance of the results, making them more interpretable and actionable for investors and policymakers.

7.2 Computing the Diffusion Indexes

We depart significantly from Ludvigson and Ng as well as follow-up research on the FRED-MD 8 groups such as Michael W. McCracken and Serena Ng [18] to build our factor diffusion indexes. The authors choose to smooth-out high frequency variations and emphasize low-frequency trends from the underlying components by constructing diffusion indexes based on partial sums : for cluster k, $DI_k = \frac{1}{N_k}\sum_{i=1}^{N_k} F_{i,t}, F_{it} = \sum_{j=1}^{t} f_{ij}$, where f_{ij} is the value of the i-th component at time j of cluster k, after the component is made stationary. This process effectively smoothes out high-frequency noise present in the original variables and highlights the low-frequency trends by accumulating the effects of the common factors over time. We find this method to be less effective for our methodology for several reasons. First, the authors apply the partial sum to series made stationary using a differentiation method whereas we have chosen to extract the cyclical component of the macroeconomic indicators. While the method effectively reduces the noise of differentiated time series, there is no use for cyclical components which are far less noisy. Secondly, we do not wish to focus on lower frequency variations with cyclical components as they are designed to capture the relevant variation frequency of economic indicators in regard to market fluctuation. Instead, we chose a simpler framework.

We computed the Diffusion Indexes with a methodology similar to the Conference Board's method to track the growth of the 10 constituents of its Leading Economic Indicator (LEI). First, we employ a classification system to categorize the underlying economic indicators within each cluster. Each indicator in a cluster is assessed for its monthly change. If the change exceeds 0.05%, a value of 1 is assigned, indicating a positive movement. A change less than 0.05% receives a value of 0.5, signifying a neutral change and indicators with a decrease exceeding 0.05% are assigned a value of 0, representing a negative movement. Then, these values are averaged across all component indicators for each economic clusters.

For all clusters $K_k = (x_1, \ldots, x_{N_k})$, we compute its diffusion index (DI_k) :

$$DI_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \left[1 \times I\left(\frac{\Delta x_i}{x_i} > 0.0005 \right) + 0.5 \times I\left(\left| \frac{\Delta x_i}{x_i} \right| < 0.0005 \right) \right]$$

Diffusion indexes can be susceptible to noise. To mitigate this, we apply an additional smoothing filter before utilizing the DI in our predictions. Fortunately, diffusion indexes are naturally scaled between 0 and 100, eliminating the need for further standardization or clipping.

Note that we build this signal on the same basis as the rolling PCA methodology as we work with the cyclical components of our economic indicators.

7.3 Diffusion Index and Macroeconomics clusters

We use the macroeconomic clusters discussed earlier to conduct a first investigation on which component of the economy are the most relevant to predicting market returns. To do so, we build diffusion indexes for each clusters which are a good approximation of the fluctuation dynamics of the variables contained in each cluster (supposing they are positively correlated).

Figure 11 provides insights into the factors influencing stock market performance. Among our selected clusters, Housing, Money & Credit, Labor Market and Interest rates & FX provide diffusion indexes that exhibit a statistically significant linear relationship at the 10% level in regards to excess returns. This can help identify which clusters contain leading variables which are likely to help forecast excess returns. This prior analysis does not guarantee better performance for these clusters however as it merely identify simple linear relationship across the whole sample spanning from 1959 to 2012. The model we employ will likely identify more complex non-linear relationships that can exist between diffusion indexes and excess returns. This is all the more important that diffusion indexes are not built using a linear transformation.

Diffusion indexes have slightly better properties than our principal components. In fact, they are all stationary at the 1% level and have a closer resemblance to a normal distribution. However, diffusion indexes still exhibit characteristics like high autocorrelation and partial correlation , similar to the PCA-derived components. This can be observed in Figure 12 containing the time series analysis of the Labor Market Diffusion Index.

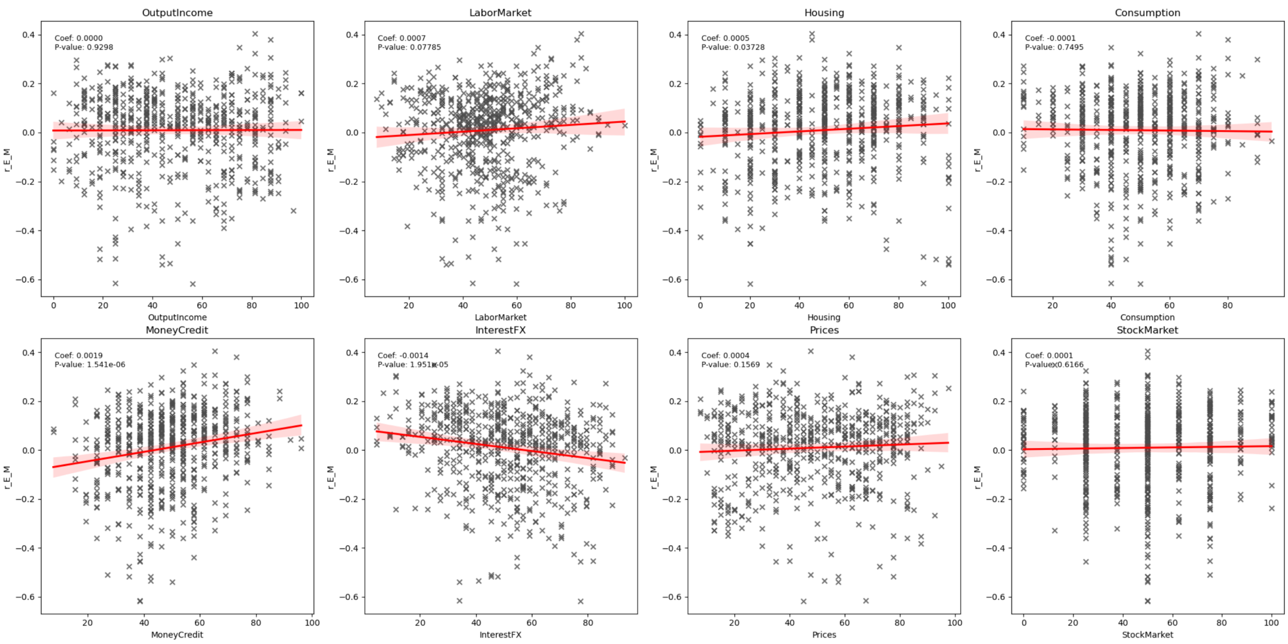

Figure 11: Regression analysis of excess return target versus diffusion indexes

7.4 Signal Construction

We can directly use our diffusion index as predictors with the XGBoost rolling model. Compared to the analysis with the principal components, which involved parameter selection for both the number of components and the rolling window size, diffusion indexes only require optimization of a single parameter, namely the length of the rolling window.

It's important to note that since principal components and diffusion indexes capture economic information differently, the optimal rolling window size for each approach might differ as well. To determine the optimal window size for the diffusion index signal, we evaluated various options using the validation sample spanning from 1967 to 2012. The results of this analysis are summarized in Table 11.

Table 11: Signal Performance over different window sizes

W = 120	**W = 240**	**W = 360**	**W = 480**
IC = 0.193***	IC = 0.292***	IC = 0.384***	IC = 0.319***
R^2 = 3.7%	R^2 = 7.5%	R^2 = 11.0%	R^2 = 7.7%
β = 0.259**	β = 0.419***	β = 0.501***	β = 0.434***

Our analysis revealed that the optimal rolling window size for the diffusion index signal is 10 years longer compared to the window identified for the PCA-based signal. This difference suggests that diffusion indexes might require a longer window to capture the underlying economic dynamics relevant for predicting stock market returns.

In addition, we test the predictive power of the signal in good and bad times, results are summarized in Table 12. Encouragingly, the diffusion index signal exhibited strong predictive power in both good and bad market periods. It demonstrated high statistical significance and

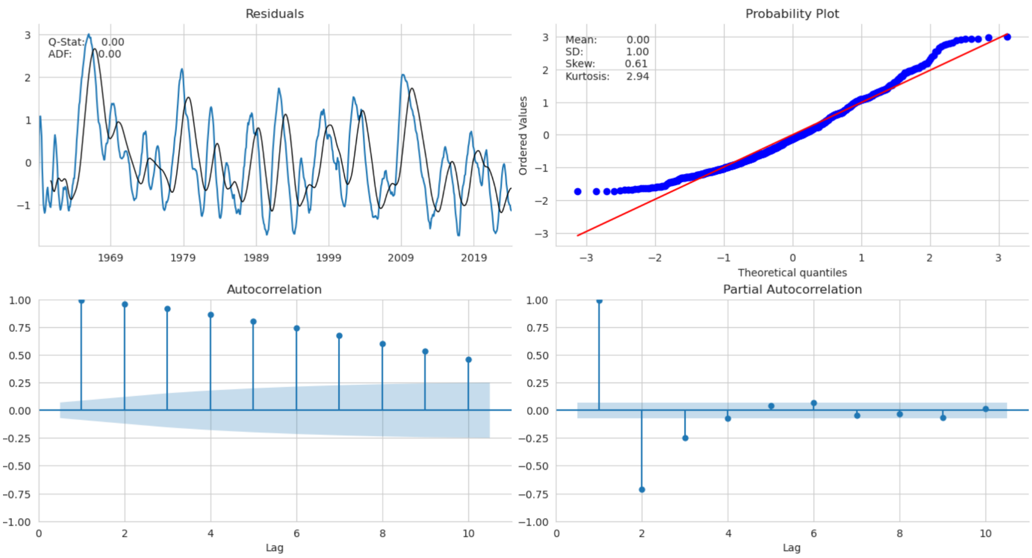

Figure 12: Labor Market Diffusion Index

a relatively high level of predictive ability across market conditions. This suggests the signal's effectiveness in identifying potential opportunities and risks, regardless of the overall market direction.

Table 12: Signal analysis during good and bad times, out-of-sample

$\mathbf{R^2 = 15.2\%}$	**Coefficient**	**p-value**
$(1-\chi_{\text{bad}})y_t$	$\beta_{\text{good}} = 0.384$	0.00^{***}
$\chi_{\text{bad}}y_t$	$\beta_{\text{bad}} = 1.316$	0.00^{***}

7.5 Signal Robustness

We evaluate the robustness of the economic signal by performing the out-of-sample test. The diffusion index signal demonstrates a strong out-of-sample performance that are statistically significant with IC$= 0.414^{***}$, Correlation$= 0.325$, $R^2 = 0.442^{***}$, $\beta = 1.114^{***}$.

By comparing the out-of-sample performance of both signal construction methods (PC and diffusion indexes), we found that the diffusion index approach exhibits a slightly higher IC and R-squared. This suggests that diffusion indexes might be more effective in capturing the economic relationships relevant for predicting future stock market returns.

While the diffusion index signal demonstrates stronger out-of-sample performance, a closer look reveals some interesting nuances. The signal built upon diffusion indexes can be observed in Figure 13. Compared to the PCA signal, the diffusion index signal exhibits slightly higher volatility. However, it compensates by performing better during the 1990-1995 period, which was a weak spot for the PCA signal. This suggests the diffusion index might be more adaptable to different market conditions. Despite its overall strength, the diffusion index signal does appear somewhat noisier on average.

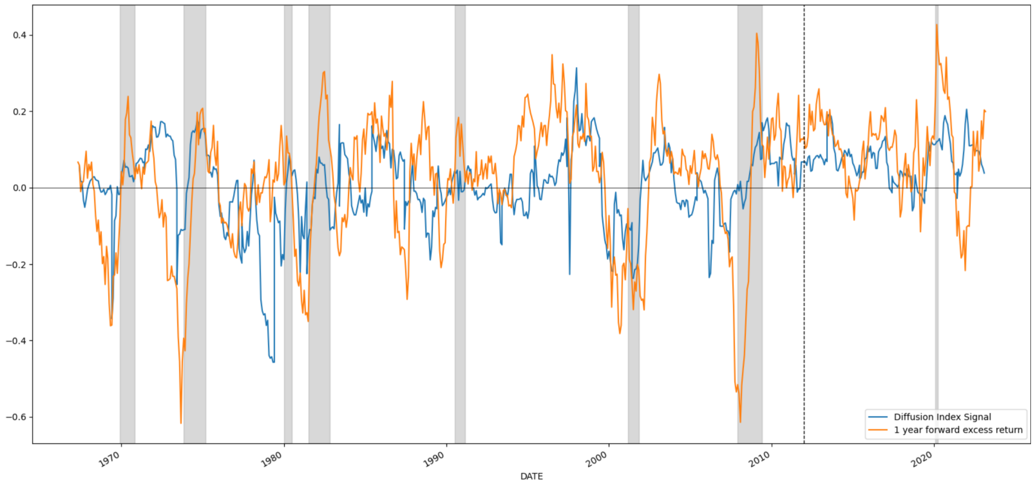

Figure 13: Macroeconomic DI signal vs 1-year horizon excess return

In addition, similar to the PCA signal, the diffusion index signal maintains its robustness across good and bad market periods, as shown in Table 13. This reinforces the signal's effectiveness in identifying potential opportunities and risks regardless of the prevailing market trend.

Table 13: Diffusion Indexes Signal analysis during good and bad times, out-of-sample

$\mathbf{R^2 = 48.8\%}$	**Coefficient**	**p-value**
$(1 - \chi_{\text{bad}})y_t$	$\beta_{\text{good}} = 1.041$	0.00^{***}
$\chi_{\text{bad}}y_t$	$\beta_{\text{bad}} = 2.900$	0.00^{***}

Furthermore, an analysis of the IC curve on Figure 14 provides further evidence of the signal's robustness. The curve exhibits a highly favorable pattern with minimal regions of weakness. This indicates that the signal has consistent predictive power across a broad spectrum of market conditions, solidifying its overall robustness.

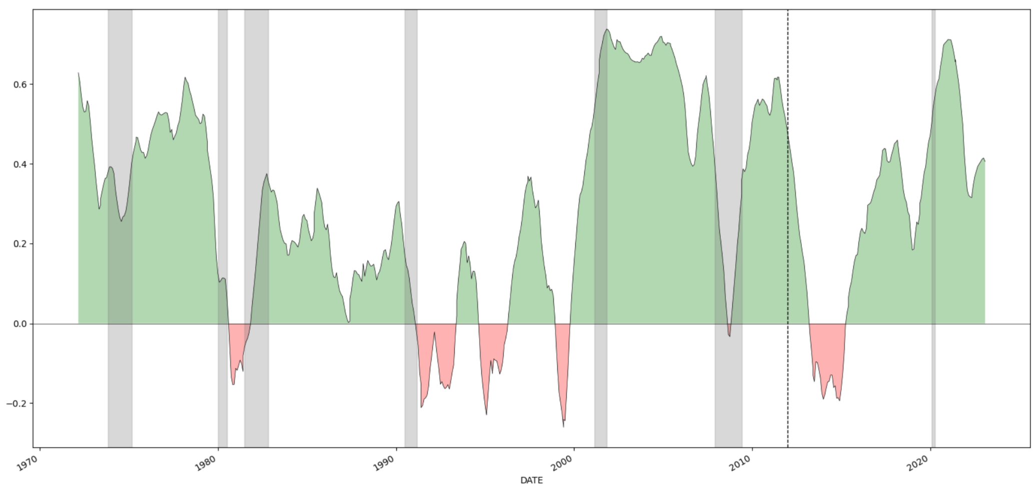

Figure 14: IC curve of the Diffusion Index, full sample

While the diffusion index signal demonstrates strong overall performance, it exhibits a key difference compared to the principal component signal: its sensitivity to the time horizon for return predictions.

Table 14: Different time horizons results for the diffusion index signal, full sample 1967-2023

	h=3	**h=6**	**h=12**	**h=24**
IC	0.073	0.141	0.415***	0.292***
R^2	0.7%	2.2%	14.5%	8.5%
β	0.163	0.246**	0.591***	0.498***
β_{bad}	0.092	0.402	1.389***	0.885***
β_{good}	0.181**	0.227***	0.483***	0.424***

As shown in Table 14, the diffusion index signal appears less robust than the PCA signal when the prediction horizon is adjusted. This suggests that diffusion indexes might be better suited for longer timeframes, such as one year or more. This result makes sense from an economic point of view because diffusion indexes are less influenced by strong variations as the principal components. Indeed, variations are only taken into account by the proportion of indicators moving into one direction, whereas principal components also take into account the magnitude of the variation.

For shorter time horizons, market returns can be more susceptible to short-term fluctuations and noise. Diffusion indexes, with their focus on directional changes without considering magnitude, might be less effective at filtering out this noise. Principal components, by incorporating both direction and magnitude, might be better equipped to handle such short-term variations.

8 Performance attribution

Now that we have built macroeconomic signals to forecast stock market excess returns, one final note is to take a look at the most informative variables. To do so, we can analyse the feature performance of our set of predictors.

8.1 Feature Importance Computation

Our models leverage the XGBoost architecture, which relies on decision trees for making predictions. Decision trees make predictions by iteratively splitting the data depending on the value of a specific feature at each node. To measure the importance of a feature, we use a combination of two common metrics :

Gain : Measures the relative contribution of the corresponding feature to the model by calculating the improvement in accuracy brought by a feature to the branches it is on. For a regression tree, the gain from a particular split is the difference in MSE before and after the split. Given a node S that splits into two nodes S_1 and S_2, the gain G from the split is :

$$G = MSE(S) - \left(\frac{n_1}{n}MSE(S_1) + \frac{n_2}{n}MSE(S_2)\right)$$

where n, n_1, n_2 are the numbers of samples in S, S_1 and S_2, MSE is the common mean square error loss measure.

Frequency or Weight : Reflects the number of times a feature is used to make splits across all trees, regardless of the depth or the purity gain. To prevent overfitting issues, the depth of trees used for prediction is capped, making frequency a more relevant measure.

$$\text{Frequency} = \frac{\text{Number of splits using feature}}{\text{Total splits all features}}$$

We combine these measure to have a broader picture of the performance of our features. To do so, we start by normalizing the scores obtained from Gain and Frequency scores. We use an average of our normalized scores :

$$\text{Combined Importance} = \frac{1}{2}\text{Normalized Gain} + \frac{1}{2}\text{Normalized Frequency}$$

We compute this metric for each features and at each iteration. Since we use a rolling model, the model updates itself at each new iteration, thus we are able to compute the evolution of features importance. This helps in analyzing the importance of our predictors depending on the period and the performance of predictions, which is illustrated by the rolling IC curves.

8.2 Feature performance

We compute the combined performance across the whole period (1967-2023) by averaging the results of each iterations. The total performance for both the Principal Component XGBoost Regression and Diffusion Index XGBoost Regression can be observed on Figure 15 and Figure 16.

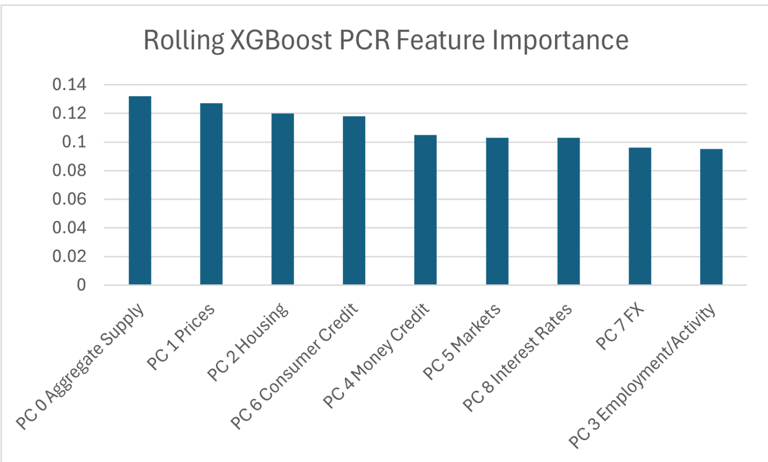

Figure 15: Feature Importance, XGBoost Principal Component Regression, 1967-2023

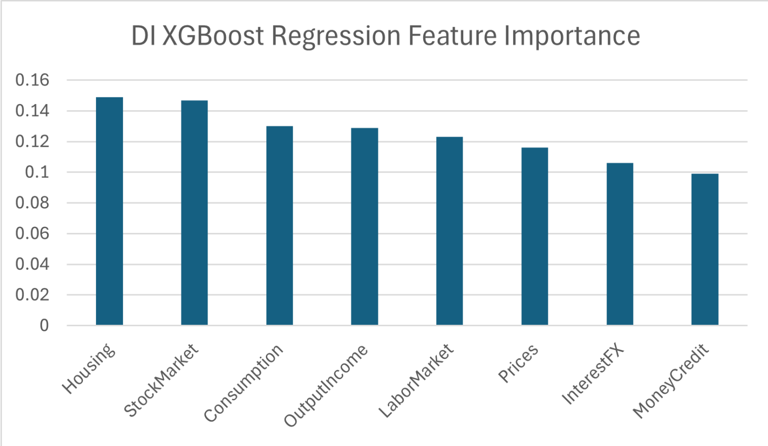

Figure 16: Feature Importance, XGBoost Diffusion Index Regression, 1967-2023

Overall, all the predictors used, whether principal components or diffusion indexes have strong feature importance in average. This suggests that most components contribute meaningfully to the prediction process and highlights the advantage of models like decision trees, which can effectively utilize a multitude of macroeconomic features. Housing and Aggregate Supply are the best performing categories for the model used, but the difference with least performing categories is not significant. This observation suggests that our models can leverage the full range of provided predictors for forecasting, potentially explaining their superior performance compared to models relying on single variables.

8.3 Evolution of performance

The evolution of combined importance of the predictors used for our model can be observed on Figure 17 & Figure 18 in the heatmaps. First, we can observe the advantage of using a rolling model because the importance of the predictors is evolving across the sample period, which is enabled by the 20 and 30 years rolling window. This would not have been possible using a static model. Secondly, we can observe that the best and least performing features are very different for different periods of time. During the late 80s and 1990, PC 6, which is linked to consumer credit, exhibits high performance in a context of shifting monetary policy (changes in interest rates) and increased consumer spending. During the great financial crisis, the PC 2 related to the Housing Market is the most important one. The housing market was both a catalyst and a barometer of the economic downturn, with the collapse of housing prices leading to widespread mortgage defaults and the subsequent failure of financial institutions heavily invested in mortgage-backed securities Around 2018, the Prices component featured prominently. Remarkably, the first principal component, capturing the aggregate supply, has consistent performance across the whole sample, which can be expected for the main principal component resulting from a PCA. For the Diffusion Index regression, we observe similar mechanics, with Housing and Output & Income being key features after 2020. This makes sense in the post-pandemic context of that period, tainted with increasing demand for houses and disrupted production.

This rapid analysis reveals the benefit of using rolling windows models that can take into account shifting environments and multivariate models, able to combine the information contained into multiple predictors.

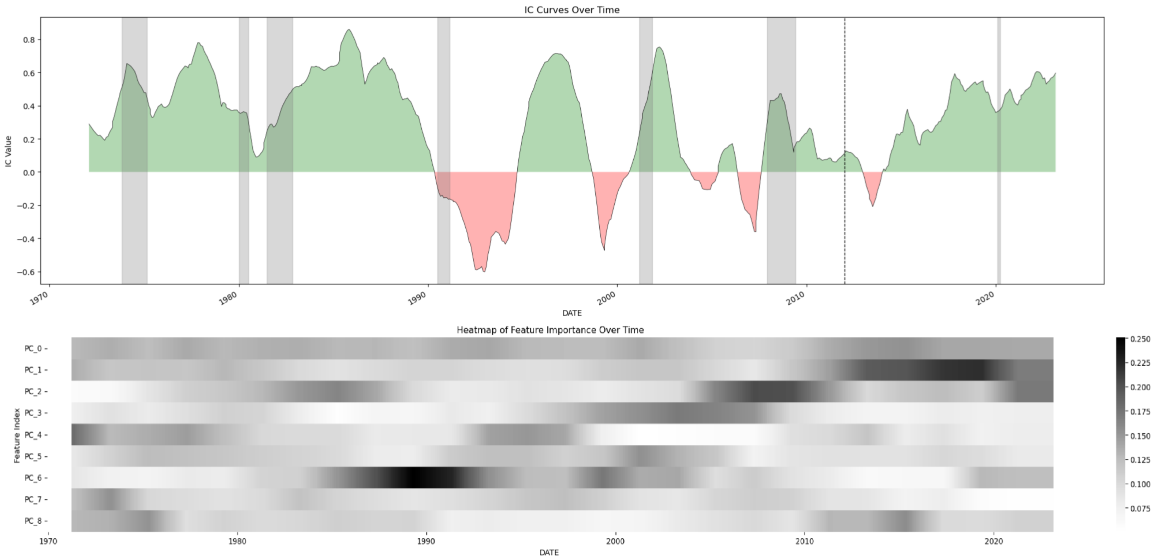

Figure 17: Feature Importance Evolution, XGBoost Principal Component Regression, 1967-2023

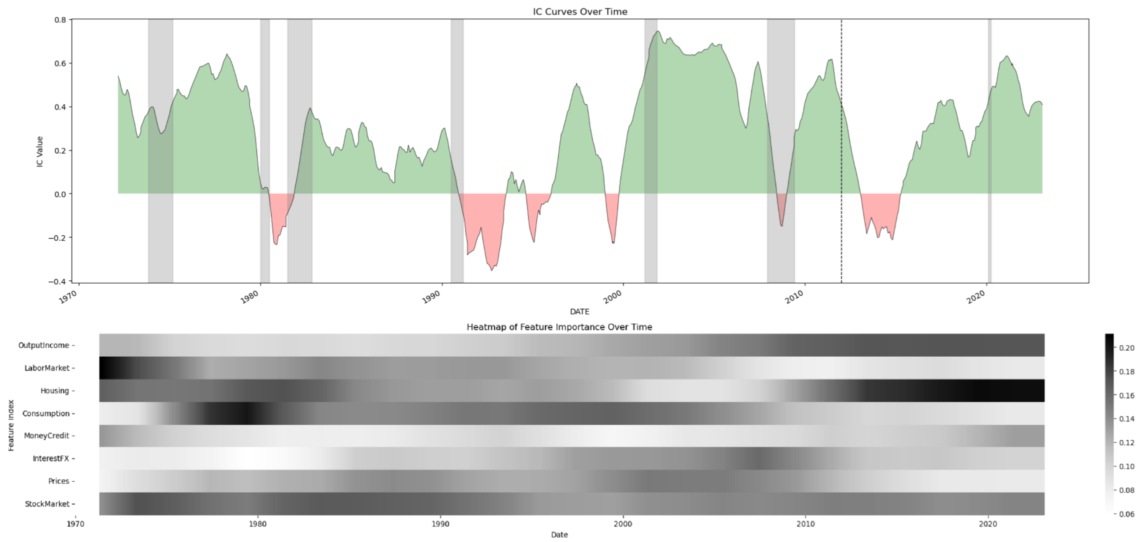

Figure 18: Feature Importance Evolution, XGBoost Diffusion Index Regression, 1967-2023

Conclusion

This paper has examined the complex relationship between macroeconomic factors and US stock market returns. We employed a comprehensive approach, utilizing a large dataset of 125 macroeconomic variables and two distinct aggregation methods: Principal Component Analysis (PCA) and diffusion indexes. Our macroeconomic signal, constructed with both PCA and diffusion indexes aggregation method, exhibits a statistically significant relationship with

S&P 500 returns, demonstrating the potential of economic factors for predicting future market movements. The signal's predictive power remains strong in both good and bad times. While the signal exhibits overall reliability, a notable exception exists between 1990 and 1995. These findings hold true even when extending the analysis to out-of-sample data on which the signal performs well. Furthermore, the signal's performance is optimized for a 2-year time horizon, reflecting the long-term influence of economic factors on the market. In addition, there are significant differences in predictability across industries, with sectors like High-Tech and Non-Durables showing stronger signal performance.

These findings provide valuable insights for investors and researchers seeking to leverage economic factors for informed investment decisions. While further research is needed to fully understand the complex dynamics at play, this study offers a significant step forward in utilizing macroeconomic data for factor timing and stock market prediction.

References

[1] K. Ahn, D. Lee, S. Sohn, and B. Yang. Stock market uncertainty and economic fundamentals: an entropy-based approach. *Quantitative Finance*, 19(7):1151–1163, 2019.

[2] V. Atanasov, S. V. Møller, and R. Priestley. Consumption fluctuations and expected returns. URL `https://ssrn.com/abstract=3098057`.

[3] P. Bossaerts and P. Hillion. Implementing statistical criteria to select return forecasting models: What do we learn? *The Review of Financial Studies*, 12(2):405–428, 1999. doi: 10.1093/rfs/12.2.405.

[4] JOHN H. BOYD, JIAN HU, and RAVI JAGANNATHAN. The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *The Journal of Finance*, 60(2):649–672, 2005. doi: https://doi.org/10.1111/j.1540-6261.2005.00742.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2005.00742.x`.

[5] CFA Institute. *CFA Program Curriculum 2023 Level I Volume 1: Quantitative Methods and Economics.* CFA Institute, Charlottesville, VA, 2023.

[6] N.-F. Chen, R. Roll, and S. A. Ross. Economic forces and the stock market. *Journal of Business*, pages 383–403, 1986.

[7] R. F. Engle, E. Ghysels, and B. Sohn. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3):776–797, 2013.

[8] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. doi: 10.1016/0304-405X(93)90023-5. URL `https://www.jstor.org/stable/26654694`.

[9] C. A. Favero, A. Melone, and A. Tamoni. Factor models with drifting prices. 2020.

[10] M. J. Flannery and A. A. Protopapadakis. Macroeconomic factors do influence aggregate stock returns. *The Review of Financial Studies*, 15(3):751–782, 2002. doi: 10.1093/rfs/15.3.751.

[11] K. R. French. Data library. detail for 10 industry portfolios, 2024. URL https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data$_{L}ibrary/det_{1}0_{i}nd_{p}ort.html$.

[12] Richard C. Grinold. The fundamental low of active management. *The Journal of Portfolio Management*, 15(3):30–37, 1989.

[13] K. K. Guru-Gharan, M. Rahman, and S. Parayitam. Influences of selected macroeconomic variables on u.s. stock market returns and their predictability over varying time horizons. *Academy of Accounting and Financial Studies Journal*, 13(1), 2009.

[14] Valentin Haddad, Serhiy Kozak, and Shrihari Santosh. Factor Timing. *The Review of Financial Studies*, 33(5):1980–2018, 2020. ISSN 0893-9454. doi: 10.1093/rfs/hhaa017. URL https://doi.org/10.1093/rfs/hhaa017.

[15] J. Liu and A. Kemp. Forecasting the sign of us oil and gas industry stock index excess returns employing macroeconomic variables. *Energy Economics*, 81:672–686, 2019.

[16] Sydney C. Ludvigson and Serena Ng. Macro Factors in Bond Risk Premia. *The Review of Financial Studies*, 22(12):5027–5067, 2009. ISSN 0893-9454. doi: 10.1093/rfs/hhp081. URL https://doi.org/10.1093/rfs/hhp081.

[17] F. Ma, X. Lu, J. Liu, and D. Huang. Macroeconomic attention and stock market return predictability. Working Paper.

[18] M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. 2014. URL http://research.stlouisfed.org/econ/mccracken/.

[19] R. C. Merton. An intertemporal capital asset pricing model. *Econometrica*, pages 867–887, 1973.

[20] R. Morck, B. Yeung, and W. Yu. The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics*, 58(1-2):215–260, 2000.

[21] D. Rapach and G. Zhou. Sparse macro factors, 2021. Working Paper.

[22] D. E. Rapach, M. E. Wohar, and J. Rangvid. Macro variables and international stock return predictability. *International Journal of Forecasting*, 21(1):137–166, 2005.

[23] Jack L. Treynor and Fischer Black. How to use security analysis to improve portfolio selection. *The Journal of Business*, 46(1):66–86, 1973. ISSN 00219398, 15375374. URL http://www.jstor.org/stable/2351280.

Appendix

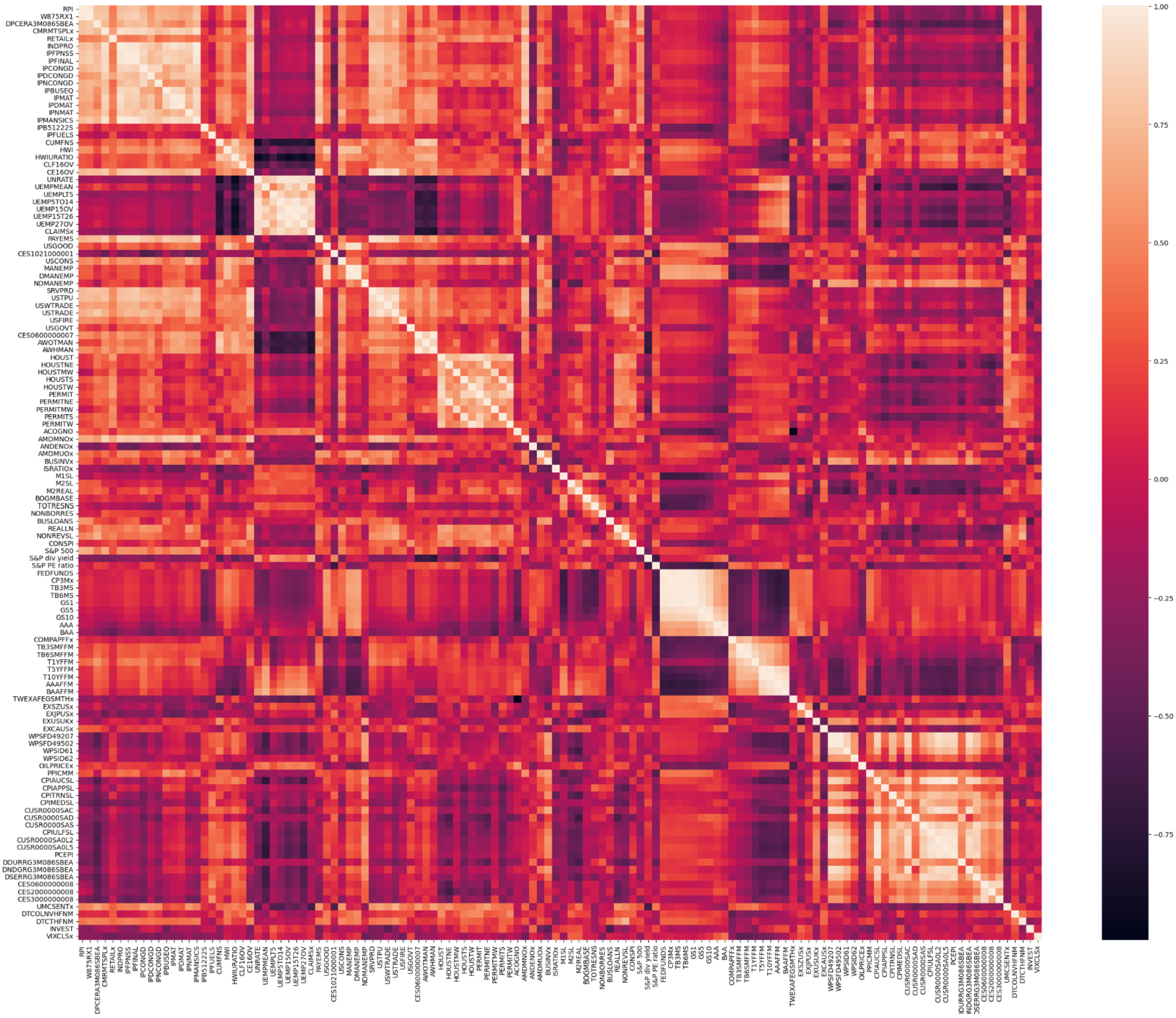

Figure 19: Correlation matrix heatmap between cyclical components

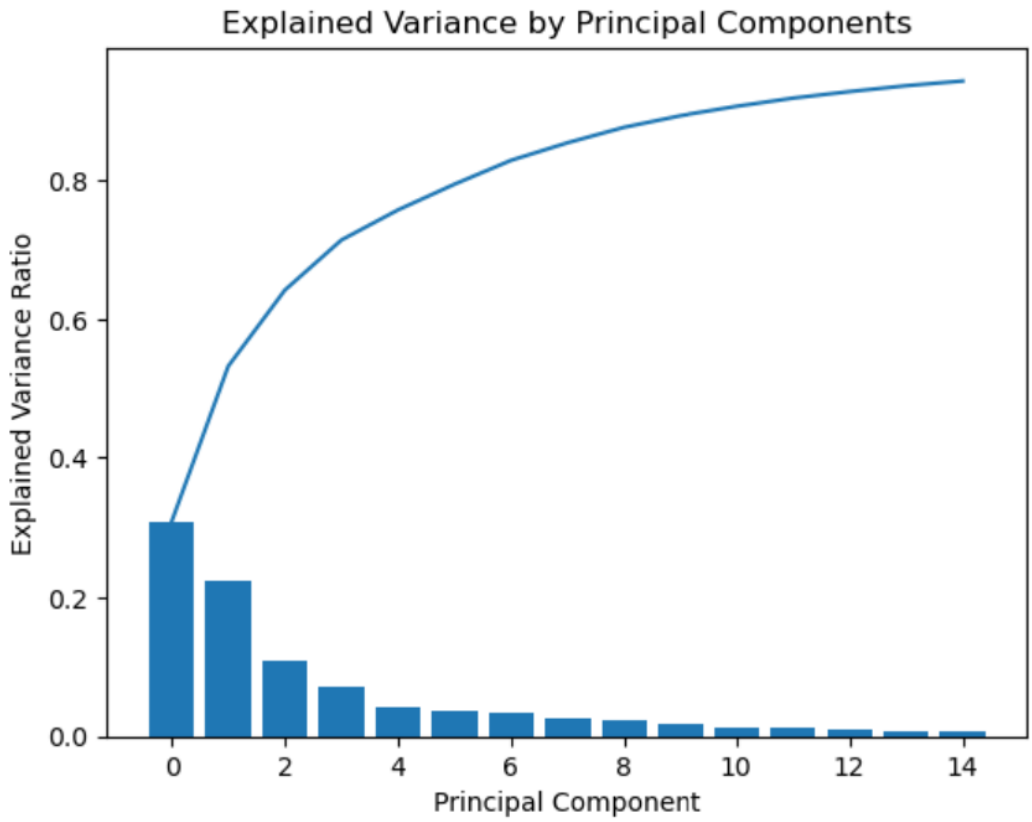

Figure 20: Principal components analysis on cyclical components, Explained Variance

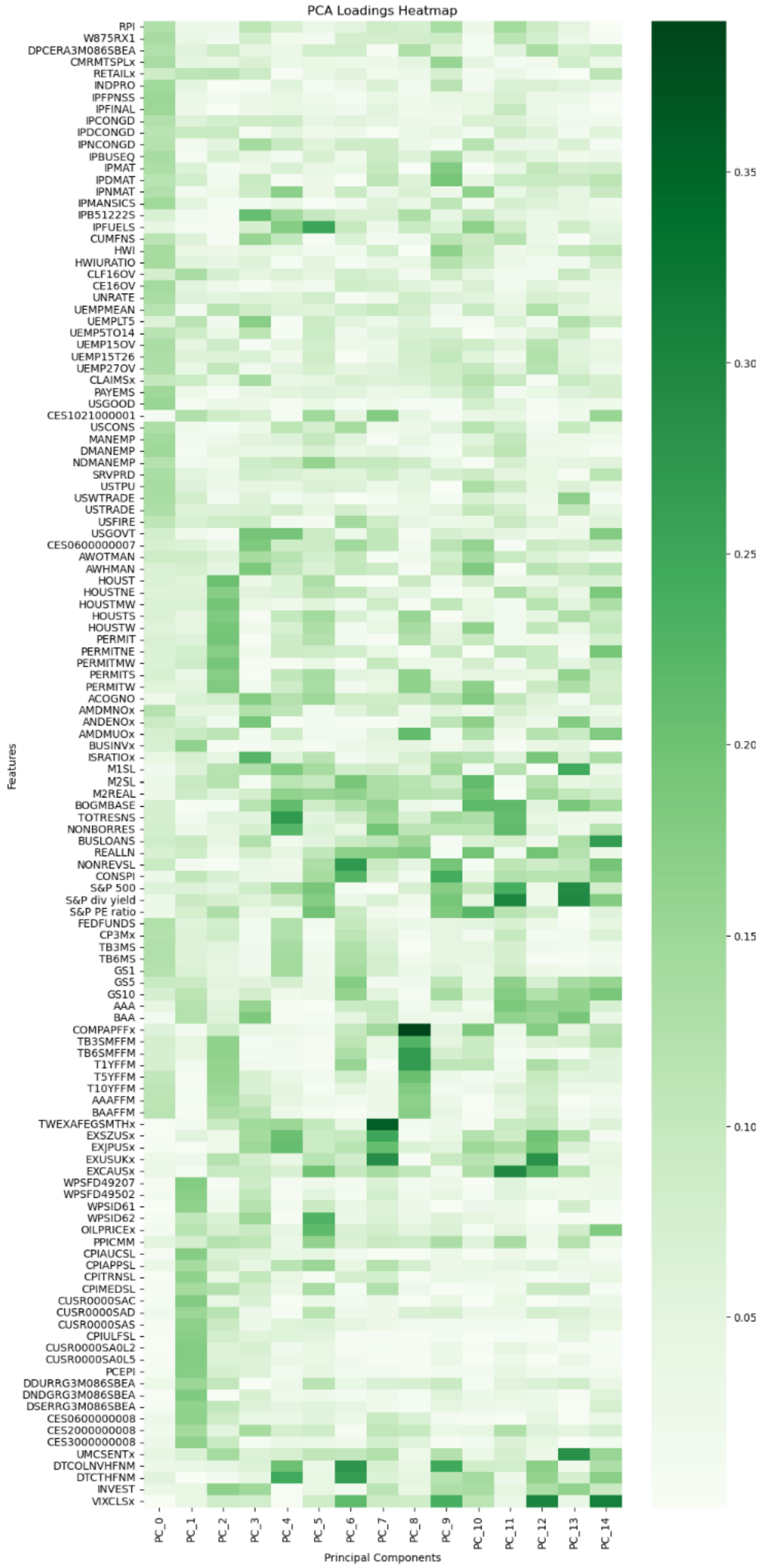

Figure 21: Principal components factor loadings

Macroeconomic clusters

The column `tcode` denotes the following data transformation for a series x:

1. no transformation;
2. Δx_t;
3. $\Delta^2 x_t$;
4. $\log(x_t)$;
5. $\Delta \log(x_t)$;
6. $\Delta^2 \log(x_t)$.

The `fred` column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column `gsi`.

Table 15: FRED macroeconomic cluster 1, Output & Income

Group 1	id	tcode	fred	description	gsi	gsi:description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	RPI ex. Transfers	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing	M_116461013	IP: mfg
14	17*	5	IPB51222S	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	19	1	NAPMPI	ISM Manufacturing: Production	M_110157212	NAPM prodn
17	20*	2	CAPUTLB00004S	Capacity Utilization: Manufacturing	M_116461602	Cap util

Table 16: FRED macroeconomic cluster 2, Labor Market

Group 2	id	tcode	fred	description	gsi	gsi:description
1	21*	2	HWI	Help-Wanted Index for US		Help wanted indx
2	22*	2	HWIURATIO	Help Wanted to Unemployed ratio	M_110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed <5 Weeks	M_110156527	U < 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed 5-14 Weeks	M_110156523	U 5-14 wks
9	29	5	UEMP15OV	Civilians Unemployed >15 Weeks	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed 15-26 Weeks	M_110156525	U 15-26 wks
11	31	5	UEMP27OV	Civilians Unemployed >27 Weeks	M_110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: TT&U	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Hours: Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Overtime Hours: Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Hours: Manufacturing	M_14386098	Avg hrs: mfg
29	49	1	NAPMEI	ISM Manufacturing: Employment	M_110157206	NAPM empl
30	128	6	CES0600000008	Ave. Hourly Earnings: Goods	M_123109182	AHE: goods
31	129	6	CES2000000008	Ave. Hourly Earnings: Construction	M_123109341	AHE: const
32	130	6	CES3000000008	Ave. Hourly Earnings: Manufacturing	M_123109552	AHE: mfg

Table 17: FRED macroeconomic cluster 3, Housing Market

Group 3	id	tcode	fred	description	gsi	gsi:description
1	50	4	HOUST	Starts: Total	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Starts: Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Starts: Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Starts: South	M_110155543	Starts: South
5	54	4	HOUSTW	Starts: West	M_110155544	Starts: West
6	55	4	PERMIT	Permits	M_110155532	BP: total
7	56	4	PERMITNE	Permits: Northeast	M_110155531	BP: NE
8	57	4	PERMITMW	Permits: Midwest	M_110155530	BP: MW
9	58	4	PERMITS	Permits: South	M_110155533	BP: South
10	59	4	PERMITW	Permits: West	M_110155534	BP: West

Table 18: FRED macroeconomic cluster 4, Consumption & Orders

Group 4	id	tcode	fred	description	gsi	gsi:description
1	3	5	DPCERA3M086SBEA	Real PCE	M_123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real M&T Sales	M_110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
4	60	1	NAPM	ISM: PMI Composite Index	M_110157208	PMI
5	61	1	NAPMNOI	ISM: New Orders Index	M_110157210	NAPM new ordrs
6	62	1	NAPMSDI	ISM: Supplier Deliveries Index	M_110157205	NAPM vendor del
7	63	1	NAPMII	ISM: Inventories Index	M_110157211	NAPM Invent
8	64	5	ACOGNO	Orders: Consumer Goods	M_14385863	Orders: cons gds
9	65*	5	AMDMNOx	Orders: Durable Goods	M_14386110	Orders: dble gds
10	66*	5	ANDENOx	Orders: Nondefense Capital Goods	M_178554409	Orders: cap gds
11	67*	5	AMDMUOx	Unfilled Orders: Durable Goods	M_14385946	Unf orders: dble
12	68*	5	BUSINVx	Total Business Inventories	M_15192014	M&T invent
13	69*	2	ISRATIOx	Inventories to Sales Ratio	M_15191529	M&T invent/sales
14	131*	2	UMCSENTx	Consumer Sentiment Index	hhsntn	Consumer expect

Table 19: FRED macroeconomic cluster 5, Prices & Inflation

Group 5	id	tcode	fred	description	gsi	gsi:description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	6	M3SL	MABMM301USM189S in FRED, M3 for the United States	M_110155013	Currency
4	73	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
5	74	6	AMBSL	St. Louis Adjusted Monetary Base	M_110154995	MB
6	75	6	TOTRESNS	Total Reserves	M_110155011	Reserves tot
7	76	6	NONBORRES	Nonborrowed Reserves	M_110155009	Reserves nonbor
8	77	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
9	78	1	REALLN	Real Estate Loans	BUSLOANS	DC&I loans
10	79	6	NONREVSL	Total Nonrevolving Credit	M_110154564	Cons credit
11	80*	2	CONSPI	Credit to PI ratio	M_110154569	Inst cred/PI
12	132	6	MZMSL	MZM Money Stock	N.A.	N.A.
13	133	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans	N.A.	N.A.
14	134	6	DTCTHFNM	Total Consumer Loans and Leases	N.A.	N.A.
15	135	6	INVEST	Securities in Bank Credit	N.A.	N.A.

Table 20: FRED macroeconomic cluster 6, Money & Credit

Group 6	id	tcode	fred	description	gsi	gsi:description
1	85	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	86*	2	CP3M	3-Month AA Comm. Paper Rate	CPF3M	Comm paper
3	87	2	TB3MS	3-Month T-bill	M_110155165	3 mo T-bill
4	88	2	TB6MS	6-Month T-bill	M_110155166	6 mo T-bill
5	89	2	GS1	1-Year T-bond	M_110155168	1 yr T-bond
6	90	2	GS5	5-Year T-bond	M_110155174	5 yr T-bond
7	91	2	GS10	10-Year T-bond	M_110155169	10 yr T-bond
8	92	2	AAA	Aaa Corporate Bond Yield		Aaa bond
9	93	2	BAA	Baa Corporate Bond Yield		Baa bond
10	94*	1	COMPAPFF	CP - FFR spread		CP-FF spread
11	95	1	TB3SMFFM	3 Mo. - FFR spread		3 mo-FF spread
12	96	1	TB6SMFFM	6 Mo. - FFR spread		6 mo-FF spread
13	97	1	T1YFFM	1 yr. - FFR spread		1 yr-FF spread
14	98	1	T5YFFM	5 yr. - FFR spread		5 yr-FF spread
15	99	1	T10YFFM	10 yr. - FFR spread		10 yr-FF spread
16	100	1	AAAFFM	Aaa - FFR spread		Aaa-FF spread
17	101	1	BAAFFM	Baa - FFR spread		Baa-FF spread
18	102	5	TWEXMMTH	Trade Weighted U.S. FX Rate		Ex rate: avg
19	103	5	EXSZUS	Switzerland / U.S. FX Rate	M_110154768	Ex rate: Switz
20	104	5	EXJPUS	Japan / U.S. FX Rate	M_110154755	Ex rate: Japan
21	105	5	EXUSUK	U.S. / U.K. FX Rate	M_110154772	Ex rate: UK
22	106	5	EXCAUS	Canada / U.S. FX Rate	M_110154744	EX rate: Canada

Table 21: FRED macroeconomic cluster 7, Interest Rates & FX

Group 7	id	tcode	fred	description	gsi	gsi:description
1	107	6	PPIFGS	PPI: Finished Goods	M_110157517	PPI: fin gds
2	108	6	PPIFCG	PPI: Finished Consumer Goods	M_110157508	PPI: cons gds
3	109	6	PPIITM	PPI: Intermediate Materials	M_110157527	PPI: int materials
4	110	6	PPICRM	PPI: Crude Materials	M_110157500	PPI: crude materials
5	111*	6	oilprice	Crude Oil Prices: WTI	M_110157273	Spot market price
6	112	6	PPICMM	PPI: Commodities	M_110157335	PPI: nonferrous
7	113	1	NAPMPRI	ISM Manufacturing: Prices	M_110157204	NAPM com price
8	114	6	CPIAUCSL	CPI: All Items	M_110157323	CPI-U: all
9	115	6	CPIAPPSL	CPI: Apparel	M_110157299	CPI-U: apparel
10	116	6	CPITRNSL	CPI: Transportation	M_110157302	CPI-U: transp
11	117	6	CPIMEDSL	CPI: Medical Care	M_110157304	CPI-U: medical
12	118	6	CUSR0000SAC	CPI: Commodities	M_110157314	CPI-U: comm.
13	119	6	CUUR0000SAD	CPI: Durables	M_110157315	CPI-U: dbles
14	120	6	CUSR0000SAS	CPI: Services	M_110157325	CPI-U: services
15	121	6	CPIULFSL	CPI: All Items Less Food	M_110157328	CPI-U: ex food
16	122	6	CUUR0000SA0L2	CPI: All items less shelter	M_110157329	CPI-U: ex shelter
17	123	6	CUSR0000SA0L5	CPI: All items less medical care	M_110157330	CPI-U: ex med
18	124	6	PCEPI	PCE: Chain-type Price Index	gmdc	PCE defl
19	125	6	DDURRG3M086SBEA	PCE: Durable goods	gmdcd	PCE defl: dlbes
20	126	6	DNDGRG3M086SBEA	PCE: Nondurable goods	gmdcn	PCE defl: nondble
21	127	6	DSERRG3M086SBEA	PCE: Services	gmdcs	PCE defl: service

Table 22: FRED macroeconomic cluster 8, Stock Market

Group 8	id	tcode	fred	description	gsi	gsi:description
1	81*	5	S&P 500	S&P: Composite	M_110155044	S&P 500
2	82*	5	S&P: indust	S&P: Industrials	M_110155047	S&P: indust
3	83*	2	S&P div yield	S&P: Dividend Yield		S&P div yield
4	84*	5	S&P PE ratio	S&P: Price-Earnings Ratio		S&P PE ratio