A Resilient Macroeconomic Proxy for Consumer Confidence Resistant to Economic Uncertainty

Victoria Puck-Karam
The Pennsylvania State University
College of Engineering

Michael Licata
The Pennsylvania State University
College of Engineering

Kaamran Raahemifar
The Pennsylvania State University
College of IST

***Abstract*—The Composite Macroeconomic Consumer Index (CMCI) is a novel consumer sentiment index constructed using macroeconomic data from 14 European nations and the U.S. (1995 Q1 to 2022 Q4), acting as a substitute for the Consumer Confidence Index (CCI). The CMCI builds on on Malovaná et al.'s CMCI index. Unlike survey-based indexes like the CCI, the novel CMCI index avoids costly and biased survey data collection, relying solely on macroeconomic indicators constructed using a factor model and kaman filter. By encompassing data from the U.S. and the pandemic era, the CMCI addresses the susceptibility of the CCI to economic volatility. Its responsiveness to economic shifts is evaluated through a GARCH volatility model, while accuracy is assessed via Stein thinning and MCMC analysis.**

***Index Terms*—Consumer Sentiment, Volatility, Macroeconomic Indicators, etc.**

I. Introduction

In the aftermath of the COVID-19 pandemic, the global economy has experienced unprecedented shifts. The pandemic gave rise to supply chain disruptions, hindered economic growth, and introduced economic challenges on a global scale. These multifaceted issues have contributed to heightened volatility in the stock market and unpredictable fluctuations in business dynamics. Despite the effectiveness of traditional economic tools such as GDP Now models and indicators like the Consumer Confidence Index (CCI) in the past, their predictive capabilities have been challenged by the unique confluence of challenges brought about by the pandemic. Further, the previous iteration of the CMCI index was found to have a limited capacity to predict major events [3] that lead to sharp swings in spending, income, and wealth (Coibion et al. 2020). This paper aims to optimize the CMCI index to be more resilient to these extremes by including more data

Notably, Fuster et al. (2010) propose that households tend to harbor inaccurate beliefs about the state of the economy, displaying excessive optimism during prosperous periods and unwarranted pessimism during economic downturns [11]. Compounding this, the CCI relies exclusively on consumer surveys that pose questions such as whether consumers expect their economic situation to deteriorate, improve or remain unchanged [13]. In conjunction with the insights of Fuster et al. (2010), this reliance on surveys may result in less accurate predictions, particularly during times of heightened economic uncertainty. The notion of trends becoming temporarily self-fulfilling, as suggested by Shiller (2000), and the delayed impact of changes in consumer sentiment on the economy [12], coupled with negativity bias [2], underscores the necessity for a more robust and responsive method to assess the state of the economy.

One significant challenge arises from declining survey-response rates, which have undermined the reliability of economic data [16]. Economists have raised concerns about the skewed measures of inflation and job market conditions due to these declines [15]. This situation suggests that government and central bank officials may have been operating on flawed data, potentially leading to misguided policy decisions. Moreover, the diminishing faith in survey data reflects broader societal trends, where economic inequality and belief in conspiracy theories can distort perceptions of economic health [14]. Notably, despite positive macroeconomic indicators like low unemployment rates and steady economic growth, there's a pervasive sense of dissatisfaction with the economy among Americans. This discontent is fueled by various factors, including heightened political polarization, biased news consumption, and mischaracterizations of economic trends. Surveys reveal that a significant portion of the population incorrectly perceives the economy as being in recession, highlighting a disconnect between economic reality and public perception [14].

This discrepancy in economic perception, underscores the need for a more nuanced approach to gauge consumer sentiment. The reliance on traditional consumer surveys, as highlighted by Fuster et al. (2010), may not capture the complexities of consumer sentiment, particularly during times of heightened uncertainty. The proposed Composite Macroeconomic Confidence Index (CMCI) aims to address these limitations by incorporating a diverse set of macroeconomic indicators from multiple high-earning countries, providing a more comprehensive assessment of consumer confidence and expectations.

Addressing the limitations of the CCI and single-value indicators, this paper proposes the development of a composite index serving as a proxy for consumer sentiment. This CMCI index incorporates eleven distinct macroeconomic indicators collectively influencing consumer confidence and expectations. The construction of the CMCI index utilizes data from 14 high-earning countries in the OECD, including the United States, spanning the years 1995 to 2022, with a specific focus on the years impacted by the COVID-19 pandemic.

Through the utilization of the proposed CMCI index, this study aims to assess the relationship between the CCI and CMCI index, examining their divergences and similarities. This analysis will be particularly pertinent when contextualized against significant events related to the COVID-19 pandemic and their abilities to remain reliable during times of economic volatility. Understanding how these indices respond to economic volatility is crucial for policymakers and analysts to make informed decisions in an increasingly complex and uncertain economic landscape.

II. Literature Review

This Literature Review section discusses three research papers that have made a significant mark in the field of economics by developing methods to predict economic conditions based on consumer confidence or consumer sentiment. The CMCI is based on the index developed in the third of these three papers. For each of these papers the problem identified, the approach and methods used to solve the identified problem, and the results as they pertain to the construction of consumer confidence indices and their volatility within their home economy will be discussed.

[1] Consumer Sentiment and its Power in Predicting Economic Success

This paper aims to determine how consumer sentiment can be used to predict different indicators of economic conditions such as recession, business performance, and coincident index. Siemes et. al. investigates how consumer sentiment affects business performance and economic activity, and how states can be clustered by industrial composition and business cycles in terms of coincident index, where clustering geographically has not previously been done in this area of research [1]. It is important to note that this clustering was done to bridge the gap between the differences in economies throughout the United States and how these factors play into the economic conditions. Overall, this experiment poses four main research questions, with the overarching theme of the ability to predict economic conditions using three main indicators. To test this hypothesis, the data was structured using the sliding window technique for time series, to evaluate the accuracy of these indicators as economic predictors.

The data was assembled using a data pipeline that collected historical information, by state, on the three factors. Then, it was split into geographical regions, which were then categorized as either a part of the test or training data set: rain set (76%) composed of Northern, Midwestern, and Southern states, and the test set (24%) composed of Western states. The sliding window technique was used to create a time series of this historical economic data that was then used to train various models including Elastic Net, Support-Vector Machine, Random Forest, PART, k-NN, and others. The authors chose to test a variety of machine learning models to evaluate more thoroughly the effectiveness of this type of data. They found that random forest consistently performed the best, achieving accuracy rates of up to 86% using the testing set against a baseline model.

Despite the success of the random forest model, this paper concluded the opposite of what was hypothesized: the economic recession has the power to predict sentiment. Fundamentally, this adds little to no value to the area of research that is economic modeling, as consumer sentiment has often been used as a tool to gauge and predict what turns the economy will take, not vice versa.

[2] Dynamics and asymmetries between consumer sentiment and consumption in pre and during-COVID-19 time: Evidence from the US

Abosedra et. al. aims to establish a relationship between consumption expenditures and consumer sentiment, in particular following and during the impact of the COVID-19 pandemic. They note that consumer sentiment is a widely agreed upon tool of economic measurement but aim to uncover empirical evidence regarding whether COVID-19 has had any impact on its effectiveness and how sensitive it is as a measure to extreme economic conditions. Further, they point out that Fuhrer's study (1993) indicated that consumer sentiment mirrored economic factors as opposed to driving them, playing a more passive than active role.

To fill this gap, Abosedra et. al. perform vector autoregression specifications to evaluate the dynamics between consumer spending and sentiment which allows for an empirical perspective that has ceased to exist till this point. They also aim to answer the question of asymmetric behavior between consumer expenditures and sentiment. This particular question is answered using a GARCH model, which is a tool used to uncover insights about the volatility and symmetry between variables. Lastly, the paper aims to answer the burning question of whether COVID-19 has made this asymmetry more distinct or has had the opposite effect. Economists and subject matter experts have yet to fully understand the long and short-term impacts that COVID-19 had on the economy, and how it impacted consumers.

Overall, this paper found that behavior became more asymmetric, and volatility became higher because of COVID-19. Further, Abosedra et. al. emphasize that typical modeling and relationships known between consumer sentiment and spending become less reliable and less accurate as economic conditions become unknown or unpredictable. They suggest that the investigated relationships become somewhat useless due to the volatility that occurs during times of economic unrest, particularly the unrest caused by a pandemic. Moreover, during the examination of asymmetric behavior in consumption expenditures, it was discovered that the coefficient for asymmetry was negative and exhibited statistical significance in the subperiod before the onset of the COVID-19 pandemic. Conversely, the coefficient for changes in consumer sentiment was deemed statistically insignificant.

[3] What Does Really Drive Consumer Confidence?

Like the research presented above, Malovaná et. al. aim to measure consumer confidence and use it to predict the behavior of consumers within the economy. While the prior researchers lean heavily on machine learning to perform this task, Malovaná et. al. aims to build an index that gauges consumer confidence. Their goal is to have the index mimic the behavior of the consumer confidence index (CCI), the most popular index for measuring consumer behavior. The main issue with indexes like the CCI, and with the other two papers that have been discussed, is that they are based heavily on consumer data. This type of data is often unreliable, as consumers can lie in responses, not respond at all, or the sampling of consumers can contain biases [3]. The index proposed by Malovaná et. al. called the HOME index, is based entirely on 11 macroeconomic indicators collected from 22 European countries between 2002 and 2019, which avoids the use of unreliable consumer data.

The CMCI index is developed using macroeconomic indicators such as gross disposable income, bank interest rate on loans, and share price index. The macroeconomic data from the 22 European countries is piped through two factor models which both make estimations of the desired factors, F, using a Kalman filter [3]. This initially helps decrease the dimensionality of the data by compressing 11 macroeconomic indicators for 22 countries in 3 factors for each country. A data table is then created which contains, for each country, the variability of the three factors computed from the factor models.

The resulting database is then used to produce a CMCI index for each country in the study which they compare to the CCI. The relationship between these two indexes is explained by two key findings. First, the CMCI index was able to explain a good number of changes in the CCI's value, revealing the direct impact of macroeconomic indicators on consumer confidence [3]. These results were statistically significant "at a 1% level" [3]. Second, the CMCI index best covered the variance seen in the CCI when the CCI was evaluated at time t, and the CMCI index was evaluated at time $(t+2)$ [3]. After developing this index, the researchers perform a case study on their home nation of the Czech Republic. They determine that in the short term, the CMCI index can predict the opening of consumer loans and in the long term the CMCI index can predict the opening of mortgage loans [3].

III. Novel Ideas

While adhering closely to the methodologies outlined in the original paper [3], this research introduces modifications aimed at enhancing the performance of the factor model.

A. *Temporal and Geographical Expansion*

The newly proposed CMCI index incorporates an expanded data set covering the years 1995 to 2023, in contrast to the previous time series of 2000 to 2018. Additionally, data from the United States will be integrated, along with information from 14 other countries. This extension of the time series by 10 years contributes to the overall refinement of the factor model's accuracy.

The augmented time series encompasses macroeconomic data spanning the periods before, during, and after the COVID-19 pandemic. This inclusion allows for a detailed examination of how the turbulent economic conditions associated with COVID-19 influenced the effectiveness of the Consumer Confidence Index (CCI). Beyond the incorporation of a more extensive time frame and data from a volatile period, the methodology for collecting data on mortgage and interest rates has been updated. In the previous study [3], there were numerous missing values for Norway, Poland, Austria, Sweden, and Hungary. However, the novel index sources data directly from the central banks of these countries, ensuring a more comprehensive data set, particularly considering the significance of interest rates and mortgage rates in assessing consumer sentiment.

B. *USA Case Study*

The investigative techniques employed by Malovaná et al. will be utilized to perform a case study focused on the United States. Adapting Malovaná et al.'s methods to the U.S. context aims to address a research gap in the examination of consumer sentiment indexes. Previous studies have not explored the efficacy of a macroeconomic-based consumer index in an economy of the magnitude and influence of the United States. Additionally, the diverse industries and economic structures of other nations examined by Malovaná et al. differ significantly, making a direct application without U.S. data less reliable.

Within the framework of the case study, the innovative index for the United States will be compared against the Consumer Confidence Index, similar to the approach in the research paper [3]. The comparison will extend to include the Consumer Sentiment Index from the University of Michigan and the VIX Index of Market Volatility. Utilizing a correlation matrix and plotting their values over the entire data set time series, as well as up to 2020 to mitigate the impact of COVID-19 volatility, aims to assess how closely the new index aligns with industry standards. This novel approach extends the research conducted in paper [3] by examining the index's performance both in the presence and absence of a disruptive event like COVID-19.

Furthermore, the analysis will encompass correlation matrices and time series graphs considering a lag in the new index. Given that many of the indices used as comparison project further into the future than the CMCI index, this examination will offer additional insights into the practical applications of our index when compared to well-established consumer indices with existing peer reviews.

C. *Principle Component Analysis (PCA)*

By focusing on the principal components, we aim to discern patterns and uncover the primary drivers shaping the predictive power of the CMCI index. Principal component analysis (PCA) will be employed on the original data set utilized in constructing the novel index. Through PCA, the analysis will identify the extent to which each macroeconomic indicator contributes to the variance. This insight will offer valuable information about which indicator carries the most significance for the CMCI index, thus identifying the variable that the index is best suited to predict.

D. Monte-Carlo Posterior Evaluation (MCMC)

After identifying the feature with the greatest contribution to cumulative variance through Principal Component Analysis (PCA), a Markov Chain Monte Carlo (MCMC) model will be developed. This model will rely on a linear regression framework, utilizing the novel index to predict the value of the feature with the greatest contribution to cumulative variance. Subsequently, the model will generate a posterior distribution for the beta values associated with the new novel index. This process offers an estimation of the index's utility in effectively predicting the value of the feature with the greatest contribution to cumulative variance.

E. Stein Thinning

The Markov Chain Monte Carlo (MCMC) output chain will be post-processed using a new method called Stein Thinning, which decreases the computational requirements of the model while maintaining the accuracy of the output chain. This model will be evaluated in the cases where the new index is lagged by $2, 3, 4, 5, 6, 7$, and 8 quarters. By providing this analysis, the predictive power of this new index on an essential macroeconomic indicator, the feature with the greatest contribution to cumulative variance, can be assessed, adding to the literature on building proxy indexes for consumer sentiment within the United States economy.

F. GARCH Volatility Model

The generalized autoregressive conditional heteroskedasticity (GARCH) process, as employed by Abosedra et al., will be utilized to analyze the data comprising the novel CMCI index. Widely acknowledged as the industry standard for estimating volatility in financial markets and the economy, the application of the GARCH process aims to unveil insights into the dynamics between consumer sentiment (represented by the CMCI index) and macroeconomic factors. This approach seeks to identify symmetries between variables, with a particular focus on discerning the variables that contribute the most to the cumulative variance of the dataset.

IV. Methods

A. Dataset Creation

To create the CMCI index, the eleven macroeconomic values were collected from various sources. Some of the data was sourced from the OECD, but many nations did not report their interest rates or mortgage rates to the OECD, so those values were manually added to the database. Once the raw data was collected, a Python script was created, using libraries such as pandas and numpy, to standardize the time series and formatting of the data, to be inputted into the factor model. Further, the data was also put into a pivot table and made wider to create the final CMCI index.

B. Factor Model and Kalman Filter

Once the corresponding data was collected for all the relevant nations, a factor model was created. This factor model is written in MATLAB, utilizing its extensive linear algebra built-in libraries. The factor model is constructed by first checking for correct dimensions of data, with eleven macroeconomic factors across the included years in the time series. Then, the factor model checks for outliers in the data and initializes the corresponding variables.

TABLE I
Macroeconomic Data Sources

Variable	Source
Gross domestic product	OECD [17]
Gross disposable income	OECD [18]
Gross savings	OECD [19]
Employee compensation	OECD [20]
Employment rate	OECD [21]
Consumer loans interest rates	National banks [21] [22] [23] [24] [25]
Mortgage interest rates	ECB & National banks
Residential property prices	BIS [29] & National banks [30]
Share price index	OECD [27]
Effective exchange rate	BIS [26]
Terms of trade	OECD [28]

Next, a lag matrix is constructed with the endogenous variables, by built by stacking lagged versions of the endogenous variables, and then exogenous variables are added in after the stacking. From there, the lag matrix undergoes single value decomposition, decomposing the lag matrix into three matrices, two orthogonal matrices, and one diagonal matrix. Next, Ordinary Least Squares are performed on the lag that creates a matrix of estimated coefficients. This coefficient matrix and the actual values in the lag matrix are compared to calculate residual values which are then in turn used to create a covariance matrix. Lastly, the program outputs the covariance matrix, the coefficient matrices, and the diagonal matrix created by running Single Value Decomposition (SVD).

This entire process follows the estimation of a var model, using OLS and SVD, that is then fed into a Kalman filter. The model's differentiation of exogenic and endogenic variables allows for the most successful and accurate output from the Kalman filter. The Kalman filter is also constructed in MATLAB and takes all the outputs of the initiation creation of matrices from the data. It is successfully able to return smoothed estimates, smoothed covariance, and the log-likelihood, all estimated from the parameters explained above.

C. Comparison of CCI and novel CMCI Index

After the creation of the new HOME index, a comparison between the CMCI index and the CCI was conducted for each country, very similar to the methods in the research paper [3]. The correlation coefficient between the CMCI index and CCI was determined for each country along with its corresponding P-value to determine statistical significance. The correlation coefficients were found for the HMCI and CCI at time *t*, for the HMCI at time (*t+4*) and CCI at time *t* as the CCI is meant to predict 4 quarters ahead, for the HMCI and CCI at time *t* up to 2020, to exclude the variance due to COVID-19, and for the HMCI at time (*t+4*) and CCI up to 2020.

D. CMCI Index Comparison

A robust methodology is employed to conduct a comparative analysis of the novel CMCI index against established benchmarks, including the Volatility Index (VIX), Consumer Sentiment Index (CSI), and Consumer Confidence Index (CCI). This comprehensive approach aims to elucidate the unique insights provided by the CMCI index concerning both consumer sentiments and broader economic conditions. This multifaceted approach to comparing the CMCI index across these industry-standard benchmarks ensures a comprehensive exploration of its utility and relevance. It not only facilitates a nuanced understanding of the intricacies within the housing market but also provides valuable insights into the interplay between housing dynamics, consumer sentiments, and overall economic volatility.

V. United States Case Study

Following correlation analysis of the new CMCI index with the CCI, an industry-standard consumer index, a case study is conducted in the United States in which the predictive power of the new CMCI index is evaluated.

A. PCA Analysis

To start, to further investigate the value of each macroeconomic index included in the creation of the CMCI index, Principal Component Analysis is conducted to uncover the cumulative variance explained by each of the factors. This data is represented using a bi-plot, which is created in R using standard CRAN libraries in the R-Studio interface. Overall, the methods used provide valuable information about the initial raw dataset as well as the outputted CMCI index, and how much cumulative variance each of the variables contributed to the index, which allows for a better understanding of what features are valuable and where the index can be improved. In this analysis, it is found that the majority of the variance is explained by the GDP macroeconomic indicator, which reveals that the new CMCI index could best be used to predict future values of the GDP.

B. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Process

For a focused exploration of the USA CMCI Index, we employ GARCH modeling to scrutinize the volatility dynamics inherent in its historical data. By selecting the USA as our focal point, we aim to capture the nuanced volatility patterns specific to the American consumer confidence index.

In the methodological approach to analyzing the CMCI index, a critical step involves the application of a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, a widely used tool in financial time series analysis for modeling volatility. The implementation utilizes the arch library in Python, a robust framework designed for estimating and forecasting volatility models. The GARCH model is initialized with parameters that define the order of the autoregressive conditional heteroskedasticity (ARCH) and GARCH components, essential elements for capturing the dynamic nature of volatility over time. The model specifications are as follows:

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t | \Psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Following the model initialization, the fit method is employed to optimize the model's parameters based on the observed time series data, using an optimization algorithm.

Once the model is fitted, a comprehensive summary is generated, offering insights into the estimated parameters, their significance, and the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). This systematic approach allows for a detailed exploration of the volatility structure within the CMCI index, providing valuable information for risk assessment and decision-making within financial markets.

C. Markov Chain Monte Carlo Analysis

To determine the predictive power of the new CMCI index on the United States GDP, a simple linear regression model is used. This model was chosen as GDP and the new CMCI index are two quantitative standardized variables and a model like linear regression can assist in determining the relationship between them. The equation for the model is as follows where $\boldsymbol{y}$ represents the GDP and $\boldsymbol{x}$ represents the new CMCI index.

$$\boldsymbol{y} = \alpha + \beta * \boldsymbol{x} + \sigma$$

To understand the predictive power of the new CMCI index within a linear regression model, an MCMC chain is built to predict the posterior distribution of the Beta value assigned to the CMCI index. To build this chain, the Stan statistical modeling tool is utilized. To determine the posterior distribution of the Beta values, a model needs to be created that contains a conditional likelihood function of the GDP given the Beta value for the CMCI index and a guess of the prior distribution of the Beta coefficient.

The likelihood function of the GDP given the Beta coefficient is simply a normal distribution where the μ value is $\alpha + \beta$ * The CMCI index and the variance are explained by the sigma value from the linear regression equation above. This is represented in the equation below:

$$\boldsymbol{y} \sim N(\alpha + \beta * \boldsymbol{x})$$

The prior distribution that is used is a t-distribution, which is a distribution widely used in Bayesian statistics and specifically when conducting linear regression analysis. The choice of this prior comes from research done by Gelman et al and Ghosh et al [5], [6]. Gelman et al revealed that the t-distribution outperforms Laplace and Gaussian priors, which were the previous standards for linear regression. In their research, they use only one degree of freedom for the t-distribution, which effectively makes it equivalent to the

Cauchy distribution [5]. Ghosh et al then found empirically that the tails of these distributions were far too large, and determined that to obtain a more accurate posterior, a prior with smaller tails was required. For this reason, they hypothesized and found that a t-distribution with degrees of freedom greater than one, μ equal to zero and σ equal to 0.5 was preferred for linear regression analysis [6]. There is no recommendation of what value specifically should be used for the degrees of freedom, but it is generally considered good practice to choose $3 <= \mu <= 7$. For this reason the following prior is used for the beta coefficient:

$$\beta \sim StudentT(\nu = 3, \mu = 0, \sigma = 0.5)$$

D. Stein-Thinning

To fully explore the posterior distribution of the β coefficient, Markov chains of over ten thousand values are required. Working with data of this size to understand the behavior of this distribution is computationally taxing. Due to this a method known as Stein-Thinning is performed to reduce the size of the outputted chain, allowing for easier analysis of the distribution itself. Stein-Thinning creates a small sample of points from the outputted chain while maintaining all information about the exploration done within the chain [7], [8]. This is done using the kernel Stein Discrepancy, which can be used to guarantee convergence of the new chain to the actual prior while simultaneously selecting the best subsample from the original chain of over ten thousand points [8].

By using the thinned chain, the mean and variance of the prior distribution of the β coefficient can be computed. Along with this, the Shapiro Test can be performed in R, which provides information regarding the possibility that the β value follows a normal distribution. This all provides a better understanding of the true value of β.

VI. Results

A. Data Analysis

B. The CMCI Index

In the analysis of the CMCI Index for the year 2022, we observed distinctive variations among the selected countries. The CMCI Index reflects the diverse economic and socio-political factors influencing each nation as well as the differing economic structures and governments that exist across the 11 nations.

C. CCI vs CMCI Correlation

D. Principle Component Analysis (PCA)

Following the initial results and the determination of the new CMCI index's covariance with the Consumer Confidence Index (CCI), we delve into a comprehensive Principal Component Analysis (PCA). This multivariate statistical technique aims to uncover the latent structures within our dataset, revealing the key variables that contribute significantly to the cumulative variance observed in the macroeconomic factors comprising the CMCI index.

TABLE II
CMCI Index in 2022

Country	CMCI Index (2022)
AUS	-1.657232539
AUT	-1.676532889
BEL	1.185018794
DEU	-1.106387182
DNK	-1.442298969
ESP	0.198100961
GBR	-0.153156436
IRL	0.174687179
NLD	1.266854981
POL	-2.352274697
PRT	-2.823344193
SVN	-1.823292923
SWE	1.137822774
USA	0.629886252

TABLE III
CMCI Index in 2022

Country	Correlation	p-value
AUS	0.2826	0.0445
AUT	0.1308	0.3235
BEL	-0.3290	0.0659
DEU	-0.1090	0.4156
DNK	0.6291	0.0000
ESP	0.7305	0.0000
GBR	0.2271	0.0836
IRL	0.4350	0.0009
NLD	0.1730	0.1901
POL	0.4449	0.0010
PRT	-0.1386	0.2952
SVN	0.0380	0.7752
SWE	-0.2950	0.0807
USA	0.395	0.0000

The first step in analyzing the newly created CMCI index involved conducting Principal Component Analysis (PCA) to explore the cumulative variance explained by each macroeconomic factor. Notably, the analysis revealed that a substantial portion, about 25.21% of the variance is explained by the Gross Domestic Product (GDP) macroeconomic indicator. This finding underscores the significance of GDP in predicting future values using the CMCI index. The second largest contributor to the variance of the data set is household disposable income, contributing 12.49% of the variance. The variance breakdown for each of the components can be seen in IV.

The scree plot, a graphical representation of the eigenvalues of the principal components, provides a visual guide to the diminishing returns of adding successive components. Notably, the scree plot in our analysis shows a distinctive leveling off after the first 10 of the 11 components. This suggests that the majority of the variance in our data is effectively captured by this subset of principal components, highlighting the intrinsic structure of the macroeconomic variables and their collective impact on the CMCI index.

To further interpret the results, we turn our attention to the bi-plot generated from the PCA. The bi-plot visually represents the relationship between variables and principal components, allowing us to identify patterns, correlations, and the relative importance of each variable in shaping the variability of

Component	Variance %
PC1: Quarterly GDP	25.36%
PC2: Household Disposable Income	12.59%
PC3: Gross Household Savings	12.02%
PC4: Employee Compensation	10.98%
PC5: Employment Rate	9.10%
PC6: Share Price Index	8.09%
PC7: Effective Exchange Rate	7.48%
PC8: Terms of Trade	5.56%
PC9: House Price Index	4.87%
PC10: Interest Rates of Mortgages	2.33%
PC11: Interest Rates on Consumer Loans	1.59%

TABLE IV
VARIANCE EXPLAINED BY EACH COMPONENT IN THE CMCI INDEX

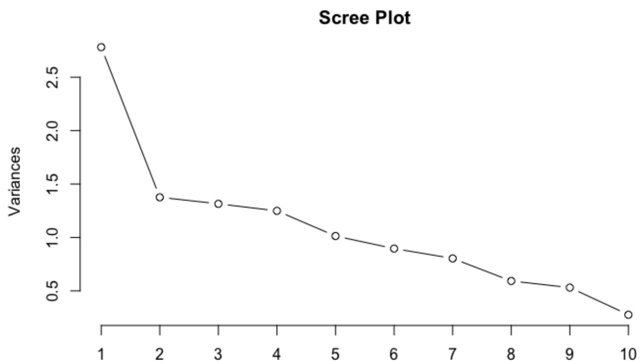

Fig. 1. Scree plot of the eleven macroeconomic features

the dataset. Of particular interest is the contribution of each macroeconomic indicator to the principal components, with a focus on understanding the dominant factors driving the predictive power of the CMCI index.

E. GARCH Volatility Analysis

The GARCH model, implemented through the arch library in Python, allows us to model conditional heteroskedasticity, providing a detailed depiction of how volatility changes over time. The results of the GARCH analysis for the USA CMCI Index will shed light on the presence of volatility clustering, leverage effects, and the persistence of past volatility. This analysis goes beyond conventional measures of risk and return, offering a dynamic perspective on the underlying volatility.

TABLE V
GARCH MODEL RESULTS FOR USA CMCI INDEX

Parameter	Coefficient	Standard Error
Mean Model		
Constant Mean (μ)	-0.1476	0.01209
Volatility Model		
Constant Variance (ω)	1.9712e-03	2.159e-03
ARCH Coefficient ($\alpha[1]$)	0.8758	0.229
GARCH Coefficient ($\beta[1]$)	0.1242	0.04481
Model Fit		
Log-Likelihood	-9.09875	
AIC	26.1975	
BIC	32.0604	

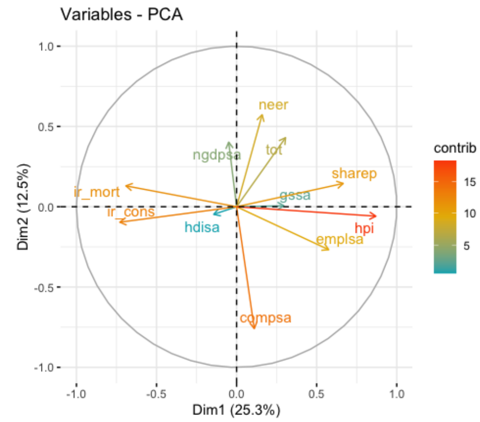

Fig. 2. Biplot of the eleven macroeconomic features

The GARCH model implemented on the CMCI index, specifically on the United States from Q1 1997 to Q4 2022 indicates that past volatility shocks significantly impact current volatility, suggesting some level of persistence. The negative constant mean coefficient (μ) of -0.1476 suggests a tendency for the CMCI index to experience below-average returns during the analyzed period. In terms of the volatility model, the positive ARCH coefficient ($\alpha[1]$) of 0.8758 implies a significant impact of past volatility shocks on the current volatility level, indicating volatility clustering. Additionally, the GARCH coefficient ($\beta[1]$) of 0.1242 signifies the persistence of past volatility's impact on future volatility, showcasing a degree of memory in the volatility process. The low log-likelihood and associated AIC and BIC values indicate a relatively good fit of the GARCH model to the data.

F. Markov Chain Monte Carlo and Stein-Thinning

The Markov Chain Monte Carlo model, implemented in R using the RStan library powered by Stein thinning enables parameter estimation and exploration of the posterior distribution. The GDP lead variable accounts for the different quarterly delays that were implemented in the CMCI index time series to account for the predictive nature of consumer confidence indices. The results are in the form of the β values and their corresponding distribution. The β values corresponding μ (posterior mean) and their posterior variance values can be found in table VI.

The β μ values represent the posterior means of the regression coefficients associated with the GDP lead variable. In simpler terms, they indicate the average estimated impact of a one-unit change in the GDP lead on the expected value of the CMCI index. For instance, a β μ of 0.0208 for a GDP Lead of 2 suggests that, on average, a two-quarter lead in GDP is associated with an increase of approximately 0.0208 units in the CMCI index.

The β Variance values represent the posterior variances of the corresponding regression coefficients. These values quantify the uncertainty or variability in the estimated impact of the GDP lead on the CMCI index. Higher variance values indicate greater uncertainty in the estimate.

GDP Lead	β μ	β Variance
2	0.02081689	0.04522726
3	0.05061047	0.05133639
4	0.06168754	0.04276023
5	0.03781971	0.04524802
6	0.03659416	0.0550438
7	0.01811133	0.02335502
8	0.01354306	0.02292844

TABLE VI
GDP LEAD MCMC AND STEIN THINNING RESULTS

The distribution of the found β values can be found in the distribution chart below.

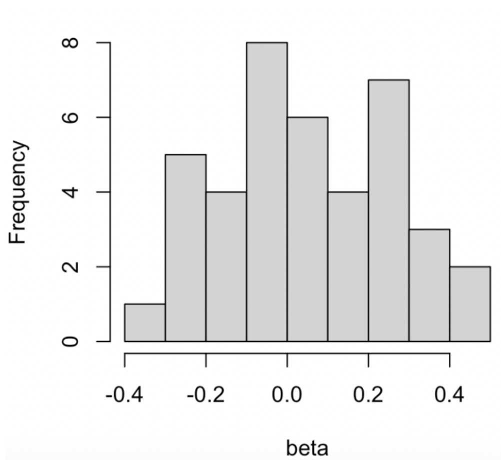

Fig. 3. Distribution of beta (β) values

Next, a Shapiro-Wilk test is conducted to determine whether a sample comes from normally distributed data. The Shapiro-Wilk test produces two values: a W statistic, a value used to make this determination for the posterior distribution of the regression coefficient and a p statistic which tests for normality on the posterior distribution. The results of the Shapiro-Wilk test can be found in table VII.

GDP Lead	β Shapiro W-Value	β Shapiro P-Value
2	0.9484374	0.06700807
3	0.9498701	0.07505408
4	0.9565656	0.1276759
5	0.9431516	0.04420681
6	0.9609504	0.1804908
7	0.9534789	0.09993471
8	0.9585037	0.1488487

TABLE VII
GDP LEAD MCMC AND STEIN THINNING RESULTS

A higher W statistic suggests a distribution closer to normal, while a lower p-value indicates evidence against the null hypothesis of normality. For instance, at a GDP lead of 5, the Shapiro-Wilk test yields a W statistic of 0.9432 and a p-value of 0.0442. The lower p-value signals a departure from normality in the posterior distribution for this specific scenario. These results contribute valuable insights into the assumptions underlying the model, allowing for a more comprehensive interpretation of the MCMC and Stein thinning outcomes and bolstering the reliability of the regression coefficient estimates.

G. CMCI Index Comparisons

To comprehensively assess the potential applications and significance of the innovative CMCI index, a thorough analysis involves the comparative evaluation of index values across various industry-standard benchmarks reflecting consumer sentiment and economic volatility. This comparative study specifically includes key metrics such as the Volatility Index (VIX), Consumer Sentiment Index (CSI), and Consumer Confidence Index (CCI).

1) **The Volatility Index (VIX)** The VIX, often referred to as the "fear index," measures market volatility and investor sentiment. By juxtaposing the CMCI index against the VIX, we gain insights into how changes in consumer sentiment align with broader market volatility trends. This comparison helps discern the relative stability or turbulence in the housing market concerning overall economic conditions.

Metric	**Value**
Correlation Coefficient	-0.0134
P-value	0.8923
Statistical Significance	No

TABLE VIII
CORRELATION ANALYSIS: CMCI INDEX VS. VIX

In the correlation analysis between the VIX and the CMCI index, the obtained metrics are presented in Table VIII. The results indicate that the composite CMCI index, designed, does not exhibit a statistically significant correlation with the Volatility Index (VIX) over the studied quarterly periods. In particular, the p-value of 0.8923 is not indicative of a statistically significant correlation, and the visualization confirms that there no statistically significant correlation between the VIX and the CMCI index.
This suggests that the movements in the CMCI index, reflecting changes in 11 macroeconomic factors, are not strongly tied to the overall market volatility as measured by the VIX. The absence of a significant correlation implies that the CMCI index may capture different aspects of economic conditions compared to the VIX, highlighting its potential unique role as a leading economic indicator rather than a direct indicator of market volatility.

2) **The Consumer Sentiment Index (CSI)** The CSI is a pivotal indicator of consumer confidence and perception regarding current and future economic conditions. Assessing the CMCI index in conjunction with the CSI provides a nuanced understanding of how consumer sentiments specific to the housing sector correlate with broader economic outlooks.
In the correlation analysis between the CSI and the CMCI index, the obtained metrics are presented in

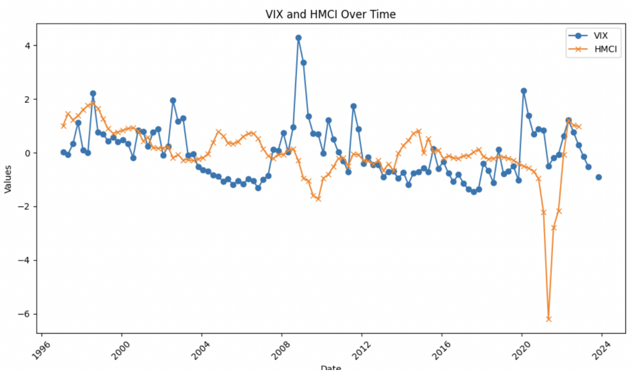

Fig. 4. VIX versus CMCI Index

Table IX. The correlation coefficient between the two indices is calculated to be 0.65, indicating a moderately positive correlation. The associated p-value is extremely small (3.29×10^{-12}), leading to the conclusion that the correlation is statistically significant. Therefore, there is compelling evidence to suggest a meaningful relationship between the CSI and the CMCI index.

Metric	**Value**
Correlation Coefficient	0.65
P-value	3.29×10^{-12}
Statistical Significance	Yes

TABLE IX
CORRELATION ANALYSIS: CSI VS. CMCI INDEX

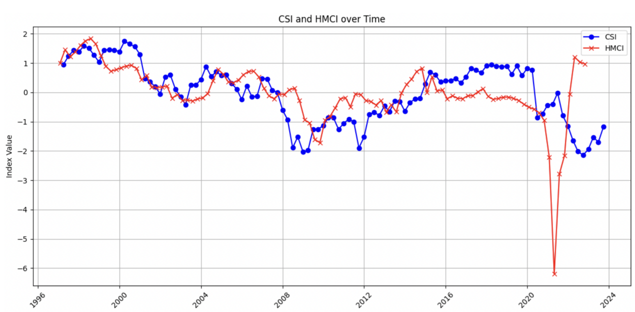

Fig. 5. CSI versus CMCI Index

3) **The Consumer Confidence Index (CCI)** Finally, the index that the CMCI index aims to offer an alternative to, the CCI. Akin to the CSI, the CCI gauges consumer confidence but encompasses a broader spectrum, including views on current economic situations and expectations for the future.
Firstly, the correlation coefficients were found for the HMCI and CCI at different t values. The resulting correlation coefficients and information about their significance at time t can be found in table X.
And the same comparison with the CMCI index at time $t + 4$ can be found in table XI.
In the correlation analysis between the CCI and the CMCI index, the results are presented in Table XII. The

TABLE X
COVARIANCE BETWEEN CMCI INDEX AND CCI AT TIME t

Country	**CMCI (t) vs. CCI (t)**
AUS	0.282**
AUT	0.130
BEL	-0.329*
DEU	-0.108
DNK	0.629***
ESP	0.730***
GBR	0.227*
IRL	0.434***
NLD	0.173
POL	0.444***
PRT	-0.138
SVN	0.037
SWE	-0.294*
USA	0.389***

***, **, and * indicate P-values less than 1%, 5%, and 10% respectively.

TABLE XI
COVARIANCE BETWEEN CMCI INDEX AND CCI AT TIME $t + 4$ PRE-2020

Country	**CMCI (t+4) vs. CCI (t) pre-2020**
AUS	0.304*
AUT	-0.412***
BEL	0.215
DEU	-0.654***
DNK	0.237
ESP	0.477***
GBR	-0.421***
IRL	0.262
NLD	0.174
POL	0.113
PRT	-0.809***
SVN	-0.402***
SWE	0.374
USA	0.682***

***, **, and * indicate P-values less than 1%, 5%, and 10% respectively.

correlation coefficient between the CCI and the CMCI index is calculated to be 0.65, indicating a moderately positive correlation. The associated p-value is remarkably small (2.63×10^{-12}), providing strong evidence of statistical significance. Hence, the analysis suggests a meaningful and statistically significant relationship between the CCI and the CMCI index.

Metric	**Value**
Correlation Coefficient	0.65
P-value	2.63×10^{-12}
Statistical Significance	Yes

TABLE XII
CORRELATION ANALYSIS: CCI VS. CMCI INDEX

Overall, the results of the comparison between CCI at time t and CMCI at time (t+4) confirmed that the HOME index initially created by Malovaná, S et. al could be applied to the United States successfully, as visualized by figure 6.

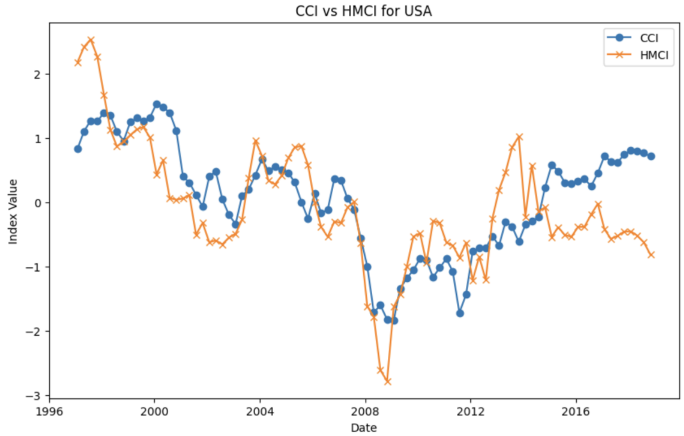

Fig. 6. CCI versus CMCI Index

VII. Conclusions

The study conducted a thorough examination of the CMCI index, a novel economic indicator crafted from eleven macroeconomic parameters. Utilizing data from diverse sources, including the OECD, national banks, the ECB, and the BIS, the index underwent a rigorous analytical process involving factor models, single value decomposition, Ordinary Least Squares, and a Kalman filter. The proposed CMCI index is comprised of publicly available macroeconomic data, allowing for the CMCI index to be replaceable and applicable to other nations.

Results indicated a significant positive correlation between the CMCI index and the Consumer Confidence Index (CCI), suggesting its efficacy in predicting consumer sentiments up to four quarters in advance. Comparative analyses with established benchmarks, such as the Volatility Index (VIX) and Consumer Sentiment Index (CSI), revealed distinct patterns. While the CMCI index displayed no significant correlation with VIX, signaling its independence from overall market volatility, a noteworthy positive correlation with CSI underscored its specialization in consumer sentiment.

A case study focused on the United States reinforced these findings, showcasing the CMCI index's meaningful correlation with CCI and employing additional analyses, including Principal Component Analysis, GARCH modeling, and Markov Chain Monte Carlo. In conclusion, the study affirmed the CMCI index's potential as a valuable tool for predicting economic trends and volatility, particularly during times of economic uncertainty and without the need for direct surveying of consumers. The CMCI index hedges against the consumers' tendencies to over-correct their feelings about the economy in times of extreme good or bad, which in turn allows for a more reliable and accurate estimate of the state of the economy.

References

[1] Siemes, I. (2019). Consumer Sentiment and its Power in Predicting Economic Success (Doctoral dissertation, Tilburg University).
[2] Abosedra, S., Laopodis, N. T., and Fakih, A. (2021). Dynamics and asymmetries between consumer sentiment and consumption in pre- and during-COVID-19 time: Evidence from the US. Journal of economic asymmetries, 24, e00227.
[3] Malovaná, S., Hodula, M., and Frait, J. (2021). What Does Really Drive Consumer Confidence?. Social indicators research,155(3),885–913.
[4] Fuhrer, J. C. (1993). What role does consumer sentiment play in the US macroeconomy?. New England Economic Review, (Jan), 32-44.
[5] Andrew Gelman. Aleks Jakulin. Maria Grazia Pittau. Yu-Sung Su. ”A weakly informative default prior distribution for logistic and other regression models.” Ann. Appl. Stat. 2 (4) 1360 - 1383, December 2008. https://doi.org/10.1214/08-AOAS191
[6] Ghosh, J., Li, Y., & Mitra, R. (2018). On the use of Cauchy prior distributions for Bayesian logistic regression. Chicago
[7] South, L. F., Riabiz, M., Teymur, O., & Oates, C. J. (2022). Postprocessing of MCMC. *Annual Review of Statistics and Its Application*, *9*, 529-555.
[8] Dwivedi, R., & Mackey, L. (2021). Kernel thinning. *arXiv preprint arXiv:2105.05842*. Chicago
[9] Hsu, J., PhD (Director). (2023, September 8). Prices, Interest Rates, and Consumer Sentiment. Survey Research Center, University of Michigan.
[10] Curtin, R., PhD (Director). (2021, October 22). Confidence in Financial Institutions. Survey Research Center, University of Michigan.
[11] Fuster, A., Laibson, D., & Mendel, B. (2010, December). Natural Expectations and Macroeconomic Fluctuations. Journal of Economic Perspectives, 24(4), 67-84.
[12] Shiller, R. (2000). Measuring bubble expectations and investor confidence. Journal of Psychology and Financial Markets, 1, 49–60.
[13] Organisation for Economic Co-operation and Development. (n.d.). Statistical insights: The OECD Consumer Barometer.
[14] Harris, B., and Sojourner, A. (2024, January 5). Why are Americans so displeased with the economy? Brookings Institution.
[15] Ip, G. (2023, November 1). The economy is great. Why are Americans in such a rotten mood? The Wall Street Journal.
[16] Zumbrun, J. (2023, March 10). Falling survey-response rates undermine economic data. The Wall Street Journal.

Data Sources

[17] OECD (2023), Gross domestic product (GDP) (indicator). doi: 10.1787/dc2f7aec-en
[18] OECD (2023), Household disposable income (indicator). doi: 10.1787/dd50eddd-en (Accessed on 22 November 2023)
[19] OECD (2023), Household savings (indicator). doi: 10.1787/cfc6f499-en (Accessed on 22 November 2023)
[20] OECD (2023), Employee compensation by activity (indicator). doi: 10.1787/7af78603-en
[21] OECD (2023), Employment rate (indicator). doi: 10.1787/1de68a9b-en
[22] Board of Governors of the Federal Reserve System (US).(2023) Finance Rate on Personal Loans at Commercial Banks, 24 Month Loan [TERMCBPER24NS]. Retrieved from FRED, Federal Reserve Bank of St. Louis
[23] Oesterreichische Nationalbank. (2023). Deposit and Lending Rates - Outstanding Amounts. Oesterreichische Nationalbank
[24] Sveriges Riksbank. (2023). Policy Rate, Deposit, and Lending Rate. Sveriges Riksbank.
[25] Reserve Bank of Australia. (2023). Indicator Lending Rates. Reserve Bank of Australia.
[26] Bank for International Settlements. (2023). Table I.1: Nominal effective exchange rates, Broad (64 economies) indices Bank for International Settlements.
[27] OECD (2023), Share prices (indicator). doi: 10.1787/6ad82f42-en
[28] OECD (2023), Terms of trade (indicator). doi: 10.1787/7722246c-en
[29] Bank for International Settlements. (2023, October 23). BIS Statistics: Selected Policy and Public Interest Papers. Bank for International Settlements.
[30] Bank for International Settlements, Real Residential Property Prices for United States [QUSR628BIS], retrieved from FRED, Federal Reserve Bank of St. Louis.