

Hard Statistics in Economics: Investigating Economic Drivers in Individual Confidence in the DOW in the US*

A time series analysis of economic trends and their effect on stock market confidence.

Liam Wall

December 3, 2024

Over the pervious decades, an investor being able to forecast the US stock market has been a way to become very successful very fast, however we rarely focus on being able to predict how the stock market can forecast the investor. This paper investigates how well “hard” statistics, in regards to economics, can predict a soft statistic: the investor’s confidence in the stock market for the next year. Using the Federal Reserve Economic Data (FRED) API and a long time survey managed by the Yale International Center for Finance, we employ a vector autoregression model and find that despite a large effort to find a good economic predictors, the best predictor of the investor’s confidence index is last quarter’s investor confidence index. This paper also identifies unexpected positive relationships of some of the variables used in analysis and shows just how difficult consistent predictive modeling is in economics.

1 Introduction

In recent decades, a well-known goal of statisticians in economics has been to develop reliable predictive models for various aspects of the stock market or the U.S. economy. Using consistently available economic indicators such as the unemployment rate or the VIX, these models aim to forecast whether the market will rise, fall, or remain stable. However, less attention has been given to the inverse problem: using hard data from the stock market or the economy to predict how investors will feel.

*Code and data are available at: <https://github.com/Lwall02/US-Economics>

Investors often rely on external data and mathematical models to guide their decisions, substituting raw confidence with quantifiable evidence to validate their instincts. The investor hopes that their own personal biases are avoided in relying on such methods however I argue that this is unavoidable. In this paper, I take the opposite approach: how can the stock market forecast the investor? I aim to use economic and stock market data not to predict market movements directly, but to forecast investor sentiment. By understanding how investors might feel, one can anticipate their likely reactions to market conditions. This information provides a strategic advantage, enabling investors to predict the behavior of others and act accordingly to capitalize on those expectations. Instead of deriving predictions purely from economic data, this paper focuses on deriving predictions based on the projected confidence of investors.

The estimand of this paper is the consumers' confidence in the US stock market one year from now. In other words, and which will be described in detail later, we want to be able to find the best predictor of the percentage of the population that is confident that the US stock market will increase one year from today.

The remainder of this paper is structured as follows. Section 2 discusses the raw data, cleaning process, variables of interest, and offers visual representations of the data through tables and graphs. Section 3 introduces and justifies the vector autoregression model used in the analysis of the data in predicting the relationship between variables as well as the confidence index. Section 4 deals with analyzing the trends and correlations showcased by the model in more detail. Section 5 discusses the real-world implications of the results uncovered in the prior sections, and finally Section 5.5 discusses the limitations and weaknesses of the analysis conducted, as well as the next steps that could be taken to improve the overall reliability of the paper. In the Appendix you can find additional information regarding the model details as well as a discussion of the survey used as our training and test data.

2 Data and Measurement

This paper uses six datasets to begin its analysis. Five datasets are provided through the Federal Reserve Bank of St. Louis, specifically through the Federal Reserve Economic Data (FRED) API, and one dataset is provided by the Yale International Center for Finance. The FRED API provides easy access to all of the Federal Reserve Bank of St. Louis's datasets, which include current and historical financial data for the US and all around the world. In particular in this paper we obtain datasets on the core consumer price index, real GDP growth rate, the VIX valuation, the federal funds effective rate, and the unemployment rate. From Yale's International Center for Finance we obtain a dataset regarding the estimated individual and institutional confidence index in the US stock market.

All data analysis was done through R (R Core Team 2023) with the aid of the following packages: `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), `ggplot2` (Wickham 2016), `knitr` (Xie 2014), `vars` (Pfaff 2008), `arrow` (Richardson et al. 2023), `kableExtra` (Zhu

2021), and `testthat` (Wickham 2011). Moreover, the FRED API was accessed using the package `fredr` (Boysel and Vaughan 2021).

2.1 FRED API

In order to first obtain data from the Federal Reserve Bank of St. Louis, we were able to use the `fredr` package to access the API and easily obtain any FRED datasets. In order to use their API, we first needed to create an API key through the FRED’s website (<https://fred.stlouisfed.org>). After saving and storing the key, we can then use the `fredr` package to obtain the data. We use a code associated with each dataset on the FRED website to quickly download the dataset through R. We can further specify the dates we are looking for as well as the frequency of the data. To have consistent data across all variables, we use quarterly data in the paper. This gives us data for each year on January 1, April 1, July 1, and October 1. We take data from January 1, 1990 to January 1, 2020.

The process of obtaining the data from the FRED API is the exact same for each of the five FRED datasets. Below is a discussion of the value and measurement of the datasets.

2.1.1 Real US GDP Growth Rate and Measurement

The US Real Growth Domestic Product (GDP) growth rate measures how the inflation-adjusted value of goods and services produced in the US economy changes over a specific period, typically from one quarter to the next or year-over-year. It’s one of the most widely used indicators to gauge economic performance. In this paper, we downloaded the US’s Real GDP dataset from FRED and obtained the percent change of each quarter’s real GDP to that of one year ago. Specifically we calculated $(\text{Real GDP in current quarter} - \text{Real GDP in the quarter 1 year ago}) / (\text{Real GDP in the quarter 1 year ago})$.

The US’s Real GDP data we obtain from FRED is in units of “billions of chained 2017 dollars, seasonally adjusted annual rate.” What seasonally adjusted means is that all data in dollars from this dataset is expressed as their purchasing power in 2017, meaning adjusted for inflation up to 2017. The annual rate simply means that the GDP listed for each quarter is the estimated annual GDP. Since we are looking at the growth rate, the data in this paper is in percents and we do not need to focus on the units of Real GDP but rather how it changes each quarter. (Economic Analysis 2024b)

We can see the Real GDP Growth Rate for all years considered in this dataset below Figure 1

Time Series of Quarterly GDP Growth Rate From 1990 to 2024

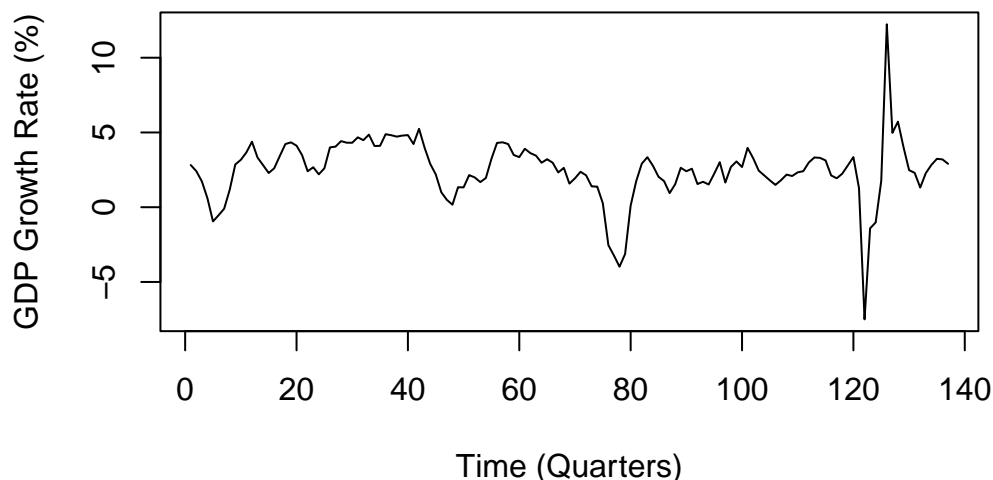


Figure 1: The quarterly US Real GDP Growth Rate for 1990 to 2024 measured in percent change compared to that of one year ago.

2.1.2 Core CPI Data and Measurement

The Core Consumer Price Index (Core CPI) is a measure of inflation that tracks the change in prices for a typical basket of goods and services, excluding food and energy items. Food and energy prices tend to be highly volatile and can fluctuate significantly due to factors like weather conditions, geopolitical events, or supply disruptions (like oil price shocks or crop failures). By excluding these items, Core CPI is able to provide a clearer picture of long-term inflation trends that are not distorted by these short-term price movements. Core CPI is one of the most widely used measures of inflation because it is considered a more reliable indicator of the general inflation trend.

Core CPI is calculated using the same methodology as the overall CPI but with food and energy prices removed. The Bureau of Labor Statistics collects data on prices from a wide range of goods and services purchased by urban consumers and lists the Core CPI as the percent increase of the cost of a typical basket of goods and services from a listed baseline year. In this case 1984 is the baseline year so a Core CPI of 264 at time t implies that the typical basket of goods and services is 164% more expensive at time t than it was in 1984. (Economic Analysis 2024a)

We can see the Core CPI for all years considered in this dataset below [Figure 2](#)

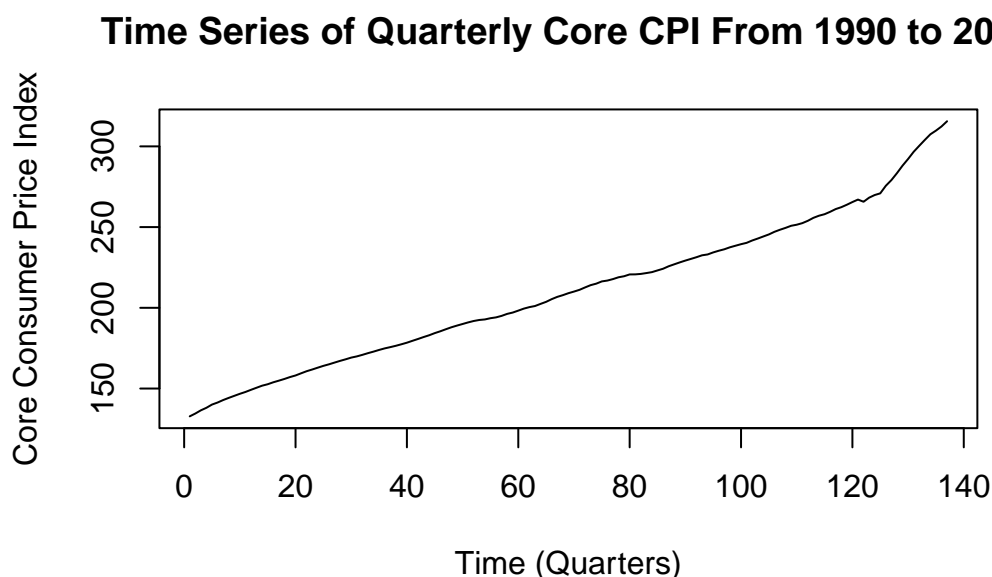


Figure 2: The quarterly Core CPI for 1990 to 2024 measured as an index of change from the baseline year of 1984.

2.1.3 Unemployment Rate and Measurement

The unemployment rate measures the number of unemployed persons in the US as a percentage of the labor force. The labor force label is restricted to people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (like mental facilities or elderly homes), and who are not on active duty in the Armed Forces. (Economic Analysis 2024c)

We can see the Unemployment Rate for the years considered in the dataset in [Figure 3](#).

2.1.4 Federal Funds Effective Rate

The Federal Funds Rate is the interest rate at which depository institutions, such as commercial banks, lend reserves to each other overnight. These reserves are balances held at Federal Reserve Banks. When a bank has surplus reserves, it lends to other banks that need additional liquidity to meet reserve requirements. The rate at which these transactions occur is determined through negotiations between the lending and borrowing banks. The Federal Funds Effective Rate is the weighted average of all such rates across the market, and while it is largely influenced by market forces, the Federal Reserve may manipulate it to adjust the rate to its target. This target is determined by the Federal Open Market Committee (FOMC),

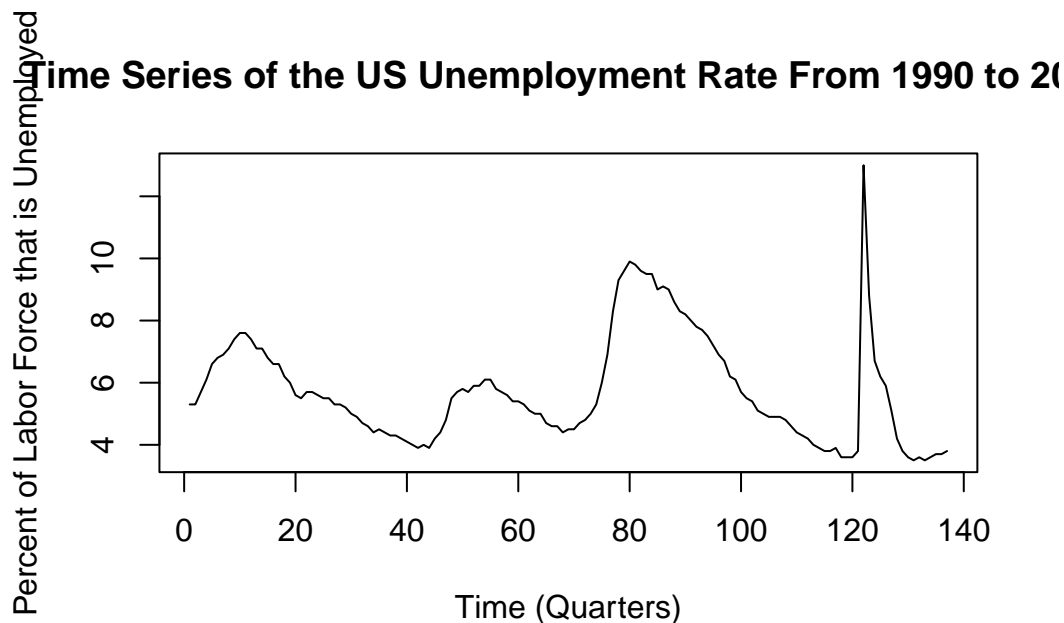


Figure 3: The quarterly Unemployment Rate for 1990 to 2024 measured as a percent of the labor force.

which meets eight times a year to set the federal funds rate target. The rate is a central tool for influencing the broader economy, impacting the cost of borrowing and the level of liquidity in the financial system.

The Federal Reserve manipulates the federal funds rate through buying or selling government bonds to influence the liquidity available in the market. If the economy is overheating, the FOMC may sell government bonds to reduce liquidity and raise the federal funds rate. On the other hand, if the economy is sluggish, the FOMC may buy bonds to increase liquidity and lower the federal funds rate, encouraging borrowing and investment. The federal funds rate is the most important interest rate in the U.S. financial system, as it influences other rates like, mortgages, loans, and savings rates. Changes in the federal funds rate can affect consumer wealth, confidence, and overall economic activity, making it a critical tool for the Federal Reserve in managing the economy's growth. (Governors of the Federal Reserve System (US) 2024)

We can see the Federal Funds Effective Rate for all years considered in this dataset below Figure 4.

Series of the quarterly Federal Funds Effective Rate From 19

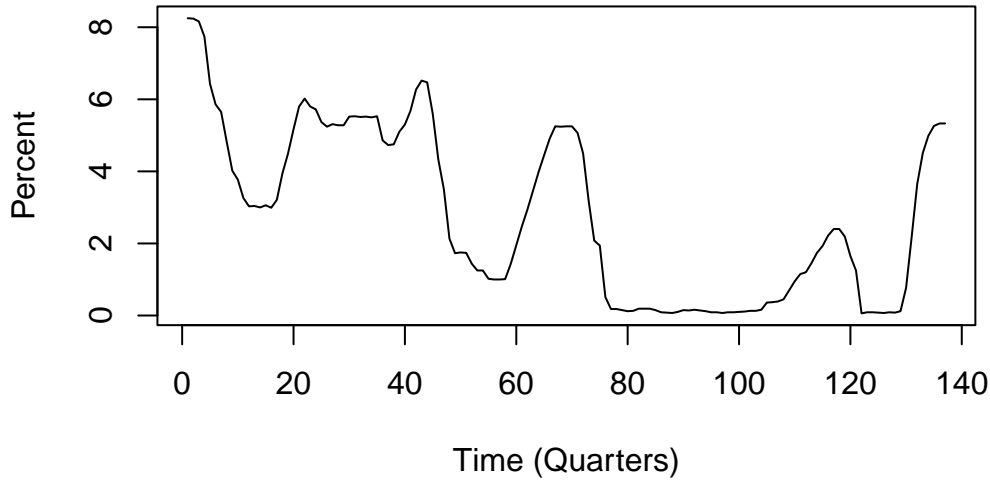


Figure 4: The quarterly Federal Funds Effective Rate from 1990 to 2024.

2.1.5 The VIX Dataset and Measurement

The VIX (Volatility Index) measures the market's expectation of near-term volatility based on the prices of stock index options. Often referred to as the “fear gauge,” the VIX reflects the consensus on the level of volatility expected in the coming 30 days. It is calculated by the Chicago Board Options Exchange (CBOE) using the prices of options on the S&P 500 index. The VIX is expressed as an annualized percentage, which indicates the expected volatility over the next 30 days. Higher VIX values indicate a higher expected volatility, signaling increased uncertainty or fear in the market, while lower values suggest stability or confidence. (Exchange 2024)

We can see the VIX for all years considered in this dataset below Figure 5.

2.2 Yale International Center for Finance

The sixth and last dataset used in this paper comes from the Yale International Center for Finance (ICF). Specifically it comes from survey data that the ICF publishes coming from a longstanding survey of wealthy American investors and American institutions. Both individuals and institutions are included in the sample and their responses are recorded separately.

Time Series of the quarterly VIX From 1990 to 2024

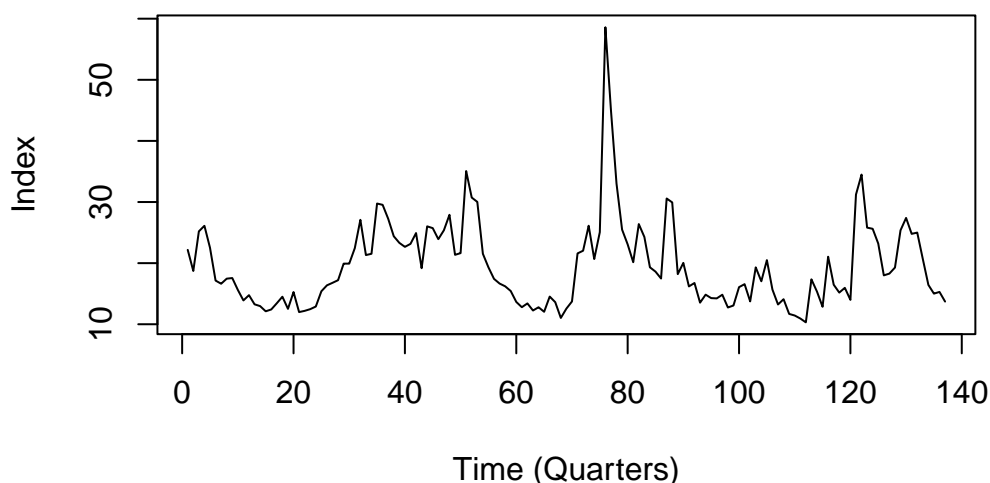


Figure 5: The quarterly VIX from 1990 to 2024.

Starting in July 2001, the beginning date of analysis for this paper, the survey was given monthly and its results are the average of the previous six months of surveys. For example the number for January 2018 is an average of results from surveys between August 2017 and January 2018. We go into more detail about the questions and calculation of the answers in the next section as well as in the appendix. (Finance 2023)

2.2.1 US One-Year Confidence Index

The data for the estimate of the US confidence index in the stock market for the next year comes from the survey data released by the ICF. Specifically, this confidence index is the percent of people who said that the DOW would increase one year from taking the survey. This question was always asked on the survey from 1989 to present. The question is offered to both institutions and wealthy investors, and they consistently display similar findings. More on this survey question and it's calculation can be found in the appendix. (Finance 2023)

Below we can see the one year confidence index for institutions [Figure 6](#)

Time Series of the 1 Year Confidence Index From 2001 to 2020

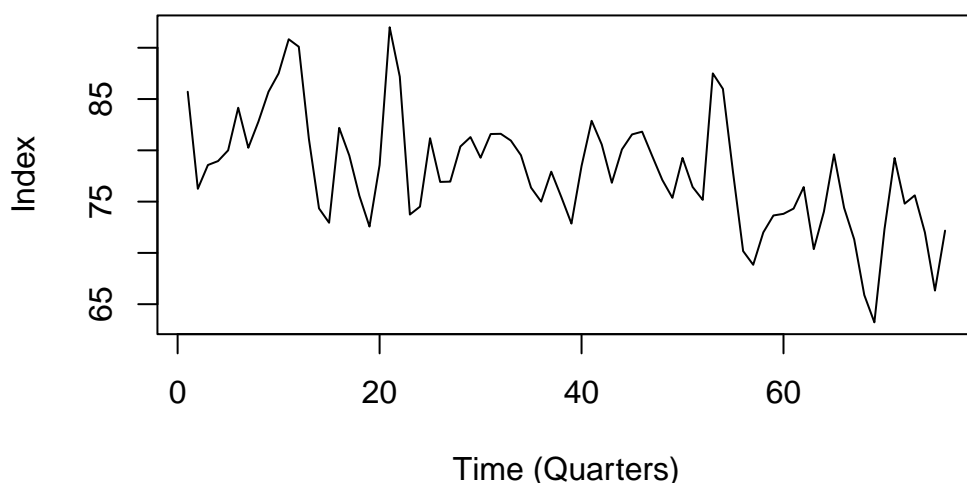


Figure 6: The quarterly 1 year confidence index from 2001 to 2020.

2.3 Cleaned Dataset

The cleaned and concise dataset used in this paper and especially for the model is table with 7 columns and 76 observations. That is 4 observations for each year from January 1, 2001 to January 1, 2020. The reason for starting in 2001 is because prior to 2001, the ICF survey only collected data every six months and we aim to look at quarterly data in this paper. The reason for ending in 2020 is because the COVID-19 pandemic had a very large effect on the US economy and thus on every variable in this analysis in different ways. To try and get a better predictive model, we will only consider these years because the confounding factors of COVID-19 are beyond the scope of this paper.

The 7 columns indicate the date, core CPI, GDP growth rate, federal funds effective rate, the VIX, the unemployment rate, and the one-year confidence index.

3 Model

As the purpose of this paper is to investigate the relationship between various economic factors and the U.S. consumer confidence index in the stock market, we will employ a model that allows us to examine how these factors contribute to predicting investor sentiment over time. Specifically, we aim to understand how variables such as GDP growth, unemployment rates,

the VIX, the federal fund effective rate, and the core CPI interact with the one-year confidence index. By using this information, we will develop a predictive model that estimates the future confidence index based on current and past economic data. The goal is to determine whether these economic indicators are effective predictors of the investor one-year confidence, or if other factors beyond the economic data (such as investor psychology or external market events) play a larger role in shaping this market sentiment.

We justify the use of the vector auto regression (VAR) model used to investigate predictors of market confidence in the Model Justification section. Background details and diagnostics are included in the Appendix.

3.1 Model set-up

The particular model that this paper will utilize is a vector auto regression (VAR) model of the one-year confidence index as a function of the economic variables present in our analysis. In particular, the model is as follows:

$$confidence_t = \beta_0 + \beta_1 gdp_{t-1} + \beta_2 cpi_{t-1} + \beta_3 unemployment_{t-1} \quad (1)$$

$$+ \beta_4 federal - rate_{t-1} + \beta_5 vix_{t-1} + \beta_6 confidence_{t-1} \quad (2)$$

In the above model:

- β_0 is the coefficient for the intercept.
- β_1 is the coefficient for the predicted change in confidence index given a one unit increase in the GDP Growth Rate at time $t - 1$
- β_2 is the coefficient for the predicted change in confidence index given a one unit increase in the Core CPI at time $t - 1$
- β_3 is the coefficient for the predicted change in confidence index given a one unit increase in the Unemployment Rate at time $t - 1$
- β_4 is the coefficient for the predicted change in confidence index given a one unit increase in the Federal Funds Effective Rate at time $t - 1$
- β_5 is the coefficient for the predicted change in confidence index given a one unit increase in the VIX at time $t - 1$
- β_6 is the coefficient for the predicted change in confidence index given a one unit increase in the One-Year Confidence Index at time $t - 1$

Note that this model employs all variables at a lagged time as the predictors. What is most different about this model as opposed to some of the other models we considered is that the predictors are a function of time. Even the one-year confidence index at a lagged time can be a predictor for the current one-year confidence index. Since all of the variables are a function

of time, we are running a model on a multivariate time series. Discussion of choosing the lag as well as making sure the input data is appropriate for such a model is all in the appendix.

Note that we use the function `VAR` from the package `vars` to run this regression.

3.1.1 Model justification

The variables chosen for this model (GDP growth rate, core CPI, unemployment rate, VIX, federal funds effective rate, and previous one-year confidence index values) are all grounded in economic theory and literature regarding their influence on market confidence/sentiment. The one-year confidence index reflects individuals' expectations about the stock market, specifically their belief that the Dow Jones will increase over the next year, making it closely linked to these macroeconomic variables.

We expect a linear relationship between these variables and the confidence index. Specifically, we anticipate that higher GDP growth and a lower unemployment rate will be associated with higher confidence levels, as they signal economic stability and opportunities. Conversely, increases in the VIX and federal funds rate are expected to reduce confidence, as they indicate higher uncertainty or a tightening economic environment.

We chose the VAR model because it is able to model the time dependent interplay of multiple interdependent variables. The variables in this analysis are closely related with one another, as well as being dependent on time, and we wanted a model that could reflect this. Not only do the variables effect each other, but more importantly they effect themselves. For example, as the percentage of people who lose confidence in the market increases it is more than likely that percentage will continue to increase as time continues. A similar argument can be made for each of the other variables. That is why we chose to incorporate a lag into the model, knowing that the values at previous times do have an effect on the value currently for each variable.

4 Results

4.1 Model Coefficients

After running the regression based on the above model, we receive the following coefficient values as showcased in the Appendix. Note that this model summary displays the coefficients for each variable as a function of all the other variables at time $t-1$. What we are interested in is the very last table of coefficients displaying the one-year confidence index vales as a function of all the other variables at time $t-1$.

We can see that in the last table of estimated coefficients, the VAR model estimates the lagged coefficients for each of the variables including the lagged one-year confidence index as a predictor. What we find is rather surprising. Firstly, that increases in unemployment

Table 1: ?(caption)

| | Time | Predicted | Actual |
|---|------|-----------|--------|
| 1 | 1 | 72.99945 | 69.70 |
| 2 | 2 | 73.29827 | 76.92 |
| 3 | 3 | 73.91431 | 71.28 |
| 4 | 4 | 74.42457 | 63.89 |
| 5 | 5 | 75.00227 | 66.30 |
| 6 | 6 | 75.57795 | 70.10 |
| 7 | 7 | 76.17936 | 75.38 |
| 8 | 8 | 76.79510 | 75.93 |
| 9 | RMSE | 5.57948 | NA |

positively correlate with increases in the stock market confidence. All other variables have less of an effect except for the lagged one-year confidence interval. In fact, the lagged one-year confidence interval is the only statistically significant finding from this model with a p-value much less than 0.05. Thus, this model tells us that the previous quarter's confidence index is the biggest predictor of the current confidence index. Followed by the unemployment rate as the second most likely predictor. All other coefficients have relatively high p-values and are thus less significant in prediction of the confidence index.

4.2 Actual vs. Predicted Confidence Indexes

In order to determine predictive performance of this model, we can look at the predicted confidence indexes as given by the VAR model compared to the observed confidence indexes we purposely omitted from the dataset. This can be seen in **?@tbl-predicted**. Here we ask the VAR to predict the next 8 quarters of one-year confidence indexes.

This result shows us that the model does not have the most powerful predictive performance. Note that this model was optimized (discussed in the appendix) and these residual errors being so large indicates that prediction of these confidence indexes is difficult based on the variables we used. However, the model is trying to predict the next two years of confidence indexes which leads into 2020 and 2021. It is very important to note that 2020 and 2021 were in the middle of the COVID-19 pandemic and it is hard to force this model to reflect such a turbulent time in the US economy. Therefore, these somewhat large residuals do not justify this model as insufficient but rather predict the confidence index for the economy without having gone through COVID-19.

5 Discussion

5.1 One-Year Market Confidence Index at Time $t-1$ is the Best Predictor

The most statistically significant result of the model is that the confidence index at time $t - 1$ is the most predictive variable for confidence at time t . This self-reinforcing relationship suggests that market confidence is largely shaped by past sentiment rather than by the economic variables included in the model. This result is both expected and unexpected. On one hand, it is unsurprising that factors beyond traditional economic indicators—such as public sentiment and behavioral patterns—play a significant role in shaping investor outlooks. Investors often rely on social cues and collective behavior, which may carry more weight than metrics like GDP growth, the VIX, or interest rates. On the other hand, it is surprising that past confidence, a largely subjective measure, outweighs variables with direct impacts on economic performance. This raises a key question: why would investor sentiment continue to drive confidence, when forward-looking investors typically aim to act ahead of the crowd? It underscores the complex interplay between psychology and market behavior, where perceptions often matter more than objective data.

5.2 Unemployment Rate and One-Year Confidence Index Exhibit a Positive Relationship

Another unexpected result is the positive relationship between unemployment increases and market confidence. This finding challenges the expectations we laid out earlier in the paper. We would typically associate rising unemployment with declining confidence due to its negative implications for economic stability and personal financial security. The explanation for such a finding can lie in many areas. Perhaps investors consider the context more than model can. For example, investors may act this way because rising unemployment could lead to government stimuli or even interest rate cuts. Not only is this positive relationship surprising alone but also in that it was the second best predictor of market confidence, beating out GDP growth rate and the VIX.

5.3 Economic Variables Are Poor Predictors of Market Confidence

Despite spending time to choose a model with normalized errors and the smallest mean square errors based on predictions, this model still does not perform well and its interpretation lacks good understanding from an investors standpoint. As discussed in the data section, we do not expect this model to perform perfectly given that it is attempting to predict one-year confidence indexes during the peak of COVID-19, but nonetheless we cannot conclude it has great predictive performance. Its performance indicates that these economic factors are not all good predictors of market confidence. It shows that the market is not influenced by these economic factors alone and there is a much more complex and intricate system of factors that

play into how an investor may feel about the future. It would have been nice to find that the most common economic statistics, like GDP growth or the VIX, play a large role in ones confidence in the market, but there is no obvious linear relationship.

5.4 Takeaways

All things considered, there is still a valuable outcome to this paper and it is that “hard” economic statistics are not enough to be able to predict whether an individual feels confident that the market will increase in one year. It shows that our suspicions may have been true that any confidence or lack of confidence in the market come from other places. Perhaps it is the knowledge of smaller sectors of the market or maybe that a sort of market psychology is a better predictor. The only concrete results we can see from this paper is that unemployment has a positive effect on market confidence and that previous market confidence is the biggest predictor of its future. This indicates that the future outlook on the market is more psychological and less responsive to economic fundamentals than we might have expected.

5.5 Weaknesses and next steps

Despite considerable efforts to develop a robust model, its predictive power remains limited. This limitation may stem from several factors. One potential reason is that the model attempts to predict outcomes during a period when COVID-19 was at its peak, which may have introduced unique, unaccounted-for influences on market confidence. Additionally, it is possible that the economic factors included in the model are simply not strongly related to market confidence. Another important consideration is that the model relies solely on quantitative data, overlooking the potential impact of categorical factors that could play a crucial role in predicting the one-year confidence index.

There are several next steps that could enhance this study. First, expanding the model to incorporate additional quantitative economic factors, such as industry-specific data, could provide a more comprehensive understanding of market confidence. Moreover, given our findings that market confidence may be influenced by psychological factors, incorporating elements from behavioral economics, such as investor sentiment and speculation, could improve prediction accuracy. It is also possible that the predictive challenges stem from the model’s inability to capture complex, non-linear relationships. In such a case, leveraging machine learning techniques might help identify and model these intricate patterns. Finally, some of the unexpected findings, such as the positive relationship between rising unemployment and market confidence, merit further exploration to better understand the underlying causes and implications.

Appendix

A Additional data details

A.1 ICF Survey Details

In the US one-year confidence index dataset, the data all comes from a monthly survey given by the ICF to a sample of wealthy investors and institutions in America. It is important to understand that this data from this dataset is not a true reflection of the market confidence but rather an aggregated estimate from active investors. This survey has been conducted since 1989 and the ICF publishes all its results. This paper only uses data from one question. The survey question pertaining to this paper is exactly: How much of a change in percentage terms do you expect in the following (use + before your number to indicate an expected increase, or - to indicate an expected decrease, leave blanks where you do not know). The surveyed will then respond to five categories putting down a percent as an answer. The five categories are: in 1 month, in 3 months, 6 months, in 1 year, and in 10 years. Although the surveyed people will answer all five questions, the one-year confidence index is calculated as the percentage of respondents who give a number strictly greater than zero for “in one year.” (Finance 2023)

Since 1989, the wording of this question has not changed except for the addition of the “in 1 month” and “in 10 years” category. This survey question is deserving of its own section because there are many confounding factors that are to be considered when examining the one-year confidence index. For example we do not know the magnitude of each respondents response. It is possible that every respondent who answers with a positive number could have a number very close to zero. In this way, can we confidently say that these people are all “confident” in the market increasing in the coming year?

Lastly, and most importantly, the documentation for this data and survey details says that the question is presented in the exact format as above. However, in the documentation explaining its calculation, it mentions that this is the percentage of people expecting the DOW to increase in one year’s time. The issue here is that we now assume the respondents answer is based on their confidence in the DOW. We now have to ask is this answer to be reflected across the whole market? It is understood the DOW does reflect the market to some extent however is this question asked properly enough in order to make that assumption? The issue is that the documentation on the survey question does not make it explicitly clear the respondent is speaking of the only the DOW or the answer is meant to be generalized to the entire US stock market/economy. Clearly the statistic is used for such purposes but it begs the question anyways.

B Model details

`?@fig-summary` shows the summary output of the VAR model discussed above. Note that the model specific to this paper is last output of estimated coefficients where the one-year confidence index is a function of all other variables at time $t - 1$.

The first step in creating the VAR model, after deciding on the VAR, was to normalize our data. In other words, we needed to transform our data such that there was a constant mean across the time span of our data. To do this we ‘differenced’ all our data using the `acf.test` from the `vars` package in order that every dataset was stationary with p-value less than 0.05. Once stationary, we can assume that each dataset has a roughly constant mean and we can continue with the VAR model. Note that core CPI and unemployment rate needed to be differenced twice to become stationary. The VIX and GDP growth rate needed to be differenced once. And the federal funds effective rate data did not need to be differenced.

The second step in creating the model was to determine the lag in our model. That is determine the number of lags for each variable we wanted to include. To do this we did employ an `acf.test` from the `vars` package which tests for the optimal lag in each variable. It turns out that the optimal lag among all variables was 9. From there we created two models, one with lag 1 and the other with lag 9. That means the second model included a $t - n$ term on each coefficient for $n = 1, \dots, 9$. We found that they both had a mean square error of approximately 7.33.

Now to choose the ideal model for this paper, we set out to find the simplest model with the lowest mean square error. Since both lag 1 and lag 9 models had similar mean square error we chose the lag 1 model as it is much simpler. To further optimize this model we found that the low variance and very linear behavior of core CPI played almost no role in predicting the outcome of the one-year confidence index. Core CPI consistently had very small coefficient estimates so therefore we excluded it.

In conclusion of the development of this model, we had originally incorporated all 5 of the economic statistics with up to 9 lag components and we ended up excluding the core CPI and keeping only 1 lag component.

VAR Estimation Results:

=====

Endogenous variables: `gdp_diff`, `unemp_diff_2`, `vix_diff`, `eff_rate`, `inst_conf_index`

Deterministic variables: `trend`

Sample size: 73

Log Likelihood: -539.103

Roots of the characteristic polynomial:

0.9782 0.9782 0.4008 0.3347 0.1188

Call:


```
VAR(y = data_matrix_2, type = "trend")
```

```
Estimation results for equation gdp_diff:
```

```
=====
```

```
gdp_diff = gdp_diff.l1 + unemp_diff_2.l1 + vix_diff.l1 + eff_rate.l1 + inst_conf_index.l1 + 1
```

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|-----------|------------|---------|----------|
| gdp_diff.l1 | 0.309947 | 0.119924 | 2.585 | 0.0119 * |
| unemp_diff_2.l1 | -1.031733 | 0.412169 | -2.503 | 0.0148 * |
| vix_diff.l1 | 0.013640 | 0.017118 | 0.797 | 0.4284 |
| eff_rate.l1 | -0.049705 | 0.062793 | -0.792 | 0.4314 |
| inst_conf_index.l1 | 0.001808 | 0.003008 | 0.601 | 0.5499 |
| trend | -0.001083 | 0.004359 | -0.249 | 0.8045 |

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.7683 on 67 degrees of freedom
```

```
Multiple R-Squared: 0.1916, Adjusted R-squared: 0.1192
```

```
F-statistic: 2.647 on 6 and 67 DF,  p-value: 0.02303
```

```
Estimation results for equation unemp_diff_2:
```

```
=====
```

```
unemp_diff_2 = gdp_diff.l1 + unemp_diff_2.l1 + vix_diff.l1 + eff_rate.l1 + inst_conf_index.l1
```

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|------------|------------|---------|------------|
| gdp_diff.l1 | -0.0096834 | 0.0354727 | -0.273 | 0.78571 |
| unemp_diff_2.l1 | -0.3517849 | 0.1219169 | -2.885 | 0.00525 ** |
| vix_diff.l1 | 0.0025648 | 0.0050635 | 0.507 | 0.61416 |
| eff_rate.l1 | 0.0324634 | 0.0185737 | 1.748 | 0.08508 . |
| inst_conf_index.l1 | -0.0016989 | 0.0008898 | -1.909 | 0.06049 . |
| trend | 0.0020240 | 0.0012894 | 1.570 | 0.12118 |

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2272 on 67 degrees of freedom
```

```
Multiple R-Squared: 0.1457, Adjusted R-squared: 0.06916
```

```
F-statistic: 1.904 on 6 and 67 DF,  p-value: 0.09293
```

Estimation results for equation vix_diff:

=====

vix_diff = gdp_diff.l1 + unemp_diff_2.l1 + vix_diff.l1 + eff_rate.l1 + inst_conf_index.l1 + 1

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|----------|
| gdp_diff.l1 | 0.23860 | 0.96113 | 0.248 | 0.805 |
| unemp_diff_2.l1 | 2.60511 | 3.30334 | 0.789 | 0.433 |
| vix_diff.l1 | -0.16160 | 0.13720 | -1.178 | 0.243 |
| eff_rate.l1 | 0.66446 | 0.50325 | 1.320 | 0.191 |
| inst_conf_index.l1 | -0.02461 | 0.02411 | -1.021 | 0.311 |
| trend | 0.02003 | 0.03494 | 0.573 | 0.568 |

Residual standard error: 6.157 on 67 degrees of freedom

Multiple R-Squared: 0.05135, Adjusted R-squared: -0.0336

F-statistic: 0.6044 on 6 and 67 DF, p-value: 0.7258

Estimation results for equation eff_rate:

=====

eff_rate = gdp_diff.l1 + unemp_diff_2.l1 + vix_diff.l1 + eff_rate.l1 + inst_conf_index.l1 + 1

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|-----------|------------|---------|------------|
| gdp_diff.l1 | 0.053409 | 0.058609 | 0.911 | 0.365 |
| unemp_diff_2.l1 | -0.025672 | 0.201435 | -0.127 | 0.899 |
| vix_diff.l1 | -0.013147 | 0.008366 | -1.571 | 0.121 |
| eff_rate.l1 | 0.985654 | 0.030688 | 32.119 | <2e-16 *** |
| inst_conf_index.l1 | -0.001845 | 0.001470 | -1.255 | 0.214 |
| trend | 0.003476 | 0.002130 | 1.632 | 0.107 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3755 on 67 degrees of freedom

Multiple R-Squared: 0.972, Adjusted R-squared: 0.9695

F-statistic: 387.7 on 6 and 67 DF, p-value: < 2.2e-16

Estimation results for equation inst_conf_index:

=====

inst_conf_index = gdp_diff.l1 + unemp_diff_2.l1 + vix_diff.l1 + eff_rate.l1 + inst_conf_index.l1 + 1

| | Estimate | Std. Error | t value | Pr(> t) |
|--|----------|------------|---------|----------|
|--|----------|------------|---------|----------|

| | | | | |
|--------------------|-----------|----------|--------|------------|
| gdp_diff.l1 | 0.094927 | 0.767826 | 0.124 | 0.902 |
| unemp_diff_2.l1 | 3.443139 | 2.638960 | 1.305 | 0.196 |
| vix_diff.l1 | -0.006316 | 0.109603 | -0.058 | 0.954 |
| eff_rate.l1 | 0.061391 | 0.402037 | 0.153 | 0.879 |
| inst_conf_index.l1 | 0.989211 | 0.019259 | 51.363 | <2e-16 *** |
| trend | 0.011841 | 0.027909 | 0.424 | 0.673 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.919 on 67 degrees of freedom
Multiple R-Squared: 0.9964, Adjusted R-squared: 0.996
F-statistic: 3064 on 6 and 67 DF, p-value: < 2.2e-16

Covariance matrix of residuals:

| | gdp_diff | unemp_diff_2 | vix_diff | eff_rate | inst_conf_index |
|-----------------|-----------|--------------|----------|-----------|-----------------|
| gdp_diff | 0.590239 | -0.019506 | -1.8123 | -0.008526 | 0.2276 |
| unemp_diff_2 | -0.019506 | 0.051642 | 0.5463 | 0.005917 | -0.1021 |
| vix_diff | -1.812336 | 0.546302 | 37.9126 | 0.296568 | 0.6050 |
| eff_rate | -0.008526 | 0.005917 | 0.2966 | 0.140975 | 0.0400 |
| inst_conf_index | 0.227616 | -0.102112 | 0.6050 | 0.039996 | 24.1774 |

Correlation matrix of residuals:

| | gdp_diff | unemp_diff_2 | vix_diff | eff_rate | inst_conf_index |
|-----------------|----------|--------------|----------|----------|-----------------|
| gdp_diff | 1.00000 | -0.11172 | -0.38312 | -0.02956 | 0.06025 |
| unemp_diff_2 | -0.11172 | 1.00000 | 0.39043 | 0.06935 | -0.09138 |
| vix_diff | -0.38312 | 0.39043 | 1.00000 | 0.12828 | 0.01998 |
| eff_rate | -0.02956 | 0.06935 | 0.12828 | 1.00000 | 0.02166 |
| inst_conf_index | 0.06025 | -0.09138 | 0.01998 | 0.02166 | 1.00000 |

B.1 Diagnostics

To further verify this model is not biased or has too much noise we look at the ‘normalness’ of its residuals. In Figure 7 we show that the optimized model has normal residuals. We can see that the residuals display some variance however they all center around 0. There is no evidence of bias in any of the variables.

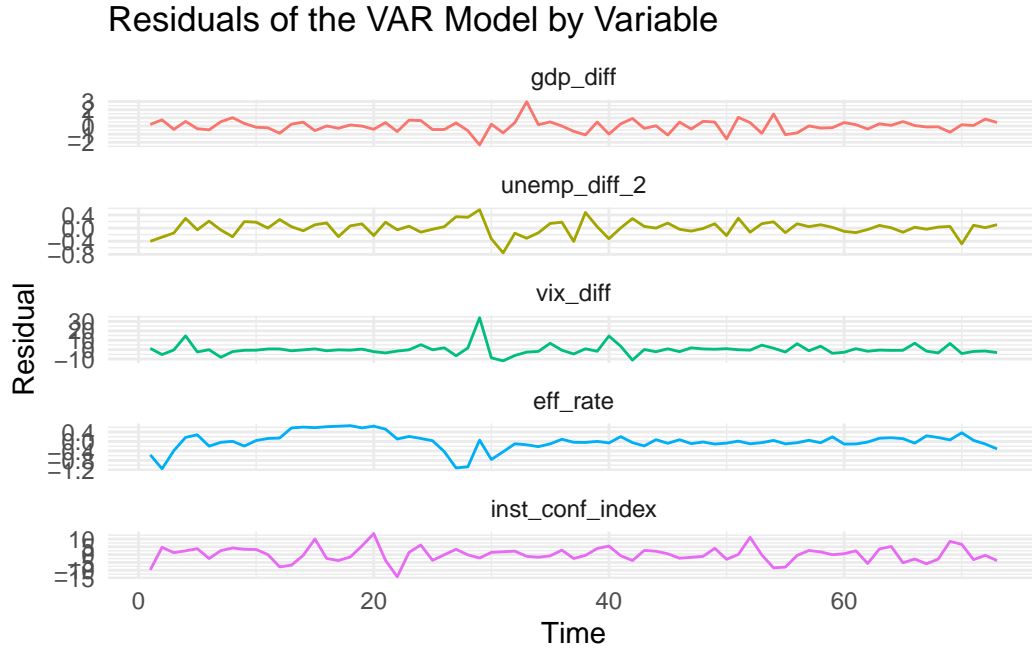


Figure 7: This facet plot shows the residuals for the VAR model as time increases. We can see these residuals have some variance while consistently oscillating around 0.

References

- Boysel, Sam, and Davis Vaughan. 2021. *Fredr: An r Client for the 'FRED' API*. <https://CRAN.R-project.org/package=fredr>.
- Economic Analysis, U. S. Bureau of. 2024a. *Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in u.s. City Average [CPILFESL]*. <https://fred.stlouisfed.org/series/CPILFESL>.
- . 2024b. *Real Gross Domestic Product [GDPC1]*. <https://fred.stlouisfed.org/series/GDPC1>.
- . 2024c. *Unemployment Rate [UNRATE]*. <https://fred.stlouisfed.org/series/UNRATE>.
- Exchange, Chicago Board Options. 2024. *CBOE Volatility Index: VIX [VIXCLS]*. <https://fred.stlouisfed.org/series/VIXCLS>.
- Finance, Yale International Center for. 2023. *United States Stock Market Confidence Indices*. <https://som.yale.edu/centers/international-center-for-finance/data/stock-market-confidence-indices/united-states>.
- Governors of the Federal Reserve System (US), Board of. 2024. *Federal Funds Effective Rate [FEDFUNDS]*. <https://fred.stlouisfed.org/series/FEDFUNDS>.
- Pfaff, Bernhard. 2008. *VAR, SVAR and SVEC Models: Implementation Within r Package Vars*. *Journal of Statistical Software* 27(4). <https://www.jstatsoft.org/v27/i04/>.

- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Wickham, Hadley. 2011. “Testthat: Get Started with Testing.” *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.
- . 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.