JOURNAL OF
EMERGING INVESTIGATORS

Article

Using economic indicators to create an empirical model of inflation

Jai Kasera[1], Howard Powers[2]
[1] Princeton Day School, Princeton, New Jersey
[2] History Department, Princeton Day School, Princeton, New Jersey

SUMMARY

Inflation affects all aspects of the economy, and it is important for traders, economists, and monetary authorities to understand the behavior of inflation to predict economic growth in the future. In our research, we investigated and calculated the correlation between various economic indicators and inflation over history. We then used the correlation information to develop a rolling linear regression model to predict inflation on an out-of-sample basis. For the measure of inflation, we used the Month over Month Consumer Price Index Seasonally Adjusted (CPI), released by the U.S. Bureau of Labor Statistics (BLS) every month. CPI measures inflation by calculating the change in the price of a basket of goods that consumers pay for. We chose 50 of the most important economic indicators followed by the market, and we hypothesized that by using out-of-sample data, the CPI of the next month could be reasonably predicted by using a regression model of a subset of economic indicators. We concluded that the average gasoline price, U.S. import price index, and 5-year market expected inflation have the most significant correlation with CPI. By using these indicators, we predicted CPI using a linear regression with a mean absolute error of about a tenth of a unit of CPI, where CPI is measured as a percentage. Furthermore, we discuss the possible public policy implications of our study and how inflation may be reduced by focusing on the three economic indicators highly correlated with it.

INTRODUCTION

Inflation, an economic mechanism, refers to a decrease in the purchasing power of money and is often caused by increasing prices of goods and services (1). Inflation is measured through the consumer price index (CPI), which calculates the change in the price of a basket of goods from the previous month (1). For decades, economists have tried to predict inflation—both short-term and long-term—creating various inflation expectation measures to do so. Some examples include groups such as the Survey of Professional Forecasters (SPF), the University of Michigan, and the Cleveland Federal Reserve (2). Each of their inflation models use different metrics, with some using primarily market-based factors while others are more survey-based. Either way, inflation has remained difficult to forecast due to the unpredictability of economic growth and difficulty of understanding what factors even cause inflation.

We hypothesized that by using out-of-sample data, the CPI of the next month could be reasonably predicted by using a regression model of a subset of economic indicators. The economic indicators used were primarily stocks and bonds. To examine our hypothesis, we tried to analyze the connection between different economic indicators and the CPI to understand how inflation functions. We then used the highly correlated indicators to predict CPI in the month ahead. After calculating the correlations between each variable and inflation, we implemented a brute-force method and found that a linear regression consisting of the average gasoline price, average import price, and 5-year market expected inflation (treasury inflation protected securities) could predict CPI one month ahead within 0.107 units of the expected CPI, where CPI is expressed as a percentage.

RESULTS

When we calculated the correlation of each individual economic indicator to CPI, we found a wide range of values from significantly highly correlated variables to variables with almost no correlation **(Figure 1)**.

We then classified these correlations into five categories. We based these cutoffs off the clusters of the correlation numbers, where c is the correlation percentage between each variable and CPI. We classified high correlation as $|c| > 48\%$, medium correlation as $37\% < |c| < 48\%$, low correlation as $18\% < |c| < 37\%$, marginal correlation as $7\% < |c| < 18\%$, and no correlation as $|c| < 7\%$ (3).

High Correlation Economic Variables

- <u>Average Gasoline Price:</u> Gasoline prices have the highest impact on inflation. They act as a tax on consumers and industries, and when gasoline prices increase, they cause transportation prices and factory production costs to increase as well (4). These rising prices subsequently cause inflation to increase, resulting in a significant correlation between gas prices and CPI **(Figure 2)**.
- <u>U.S. Import Price Index:</u> Based on the prices of goods and services traded between the U.S. and foreign countries, the International Price Program creates the import price indexes (5). Over the last few decades, various American companies have set up their supply chain in lower-cost countries. As a result, many goods that were once

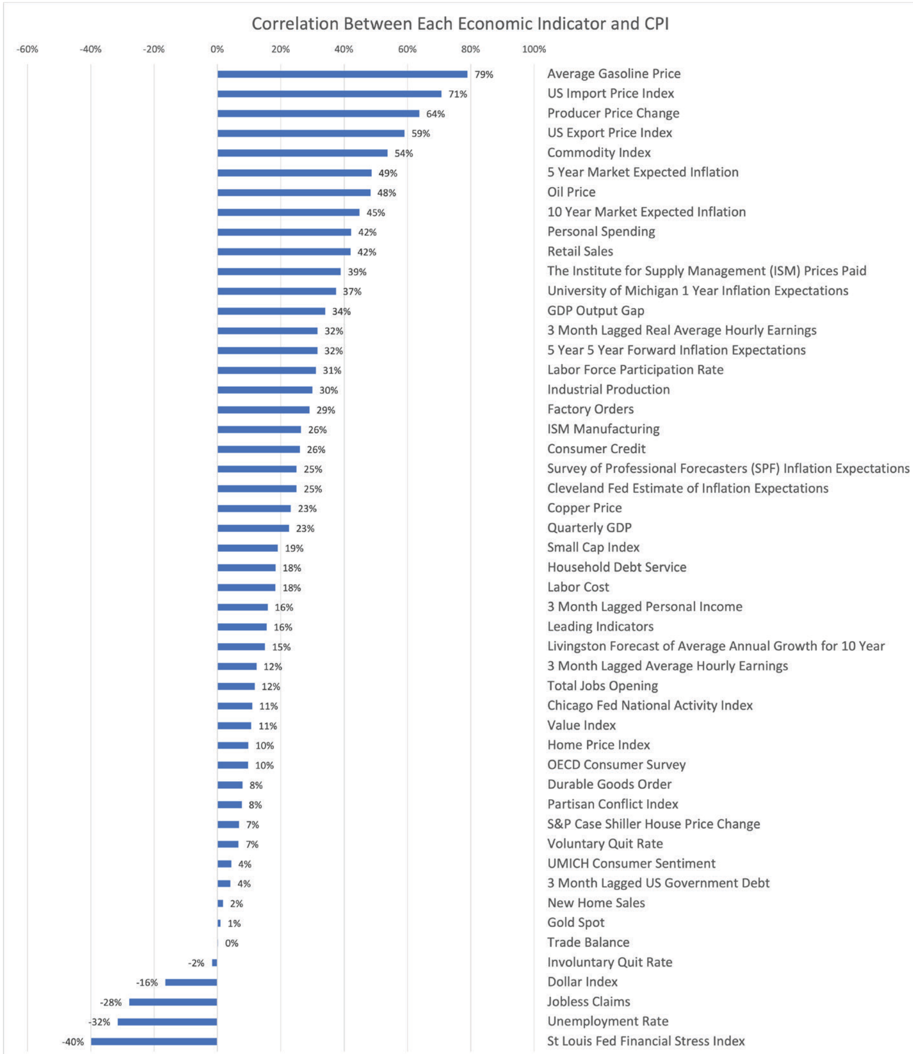

Figure 1: Correlations between economic indicators and CPI. Bar graphs shows correlation value (%) between each economic indicator and CPI. Economic indicators and CPI were converted into same frequency of release. Correlation was then modified to generate a percentage of correlation.

manufactured in the U.S. are now imported, so a rise in import prices will likely have a direct impact on how much consumers pay for goods, increasing inflation **(Figure 3)**.

- Producer Price Change (PPI)
- U.S. Export Price Index
- Commodity Index
- 5 Year Market Expected Inflation: Treasury inflation-protected securities (TIPS) and benchmark treasury bonds measure the real and nominal five-year yield, respectively. The difference between the nominal yield of a five-year benchmark treasury bond and the real yield of a five-year TIPS bond provides a market estimate of inflation for the next five years **(Figure 4)**.
- Oil Price

Medium Correlation Economic Variables

- 10-Year Market Expected Inflation
- Personal Spending
- Retail Sales
- St. Louis Fed Financial Stress Index
- The Institute for Supply Management (ISM) Prices Paid
- University of Michigan 1-Year Inflation Expectations

Low Correlation Economic Variables

- Gross Domestic Product (GDP) Output Gap
- 3-Month Lagged Real Average Hourly Earnings
- Unemployment Rate
- 5-Year, 5-Year Forward Inflation Expectations
- Labor Force Participation Rate
- Industrial Production
- Factory Orders
- Jobless Claims
- ISM Manufacturing
- Consumer Credit
- Survey of Professional Forecasters (SPF) Inflation Expectations

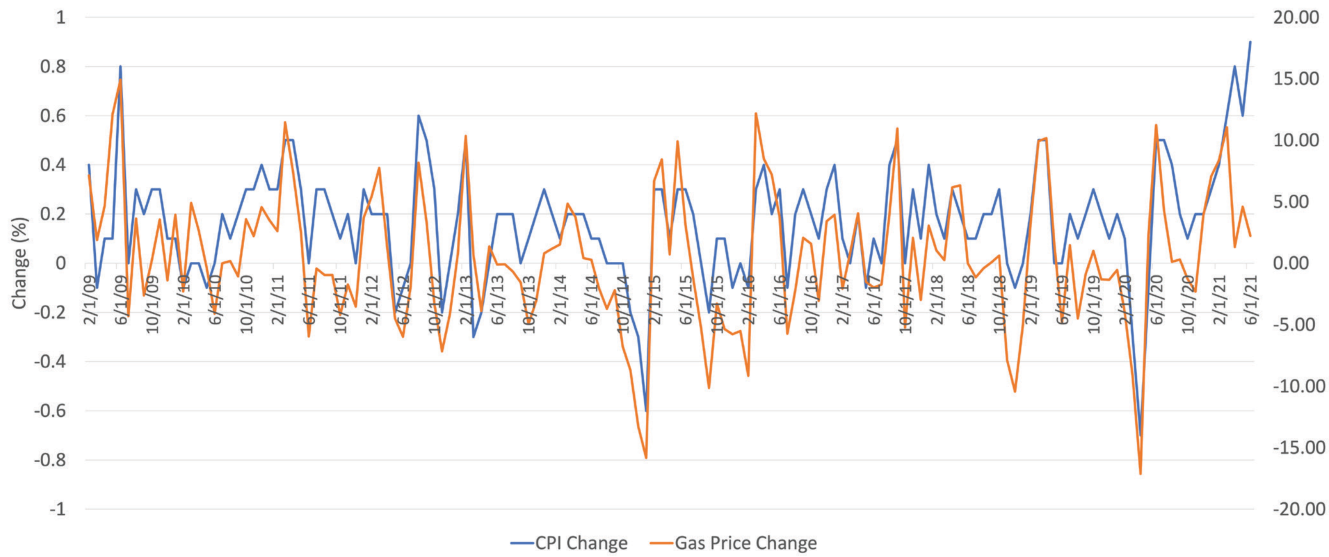

Figure 2: CPI change versus gas price change. Line graph shows the relationship between the CPI percent change (blue) and gas price percent change (orange) from February 2009 to June 2021. Gas price was converted from a daily release to a monthly release to match the monthly release of CPI.

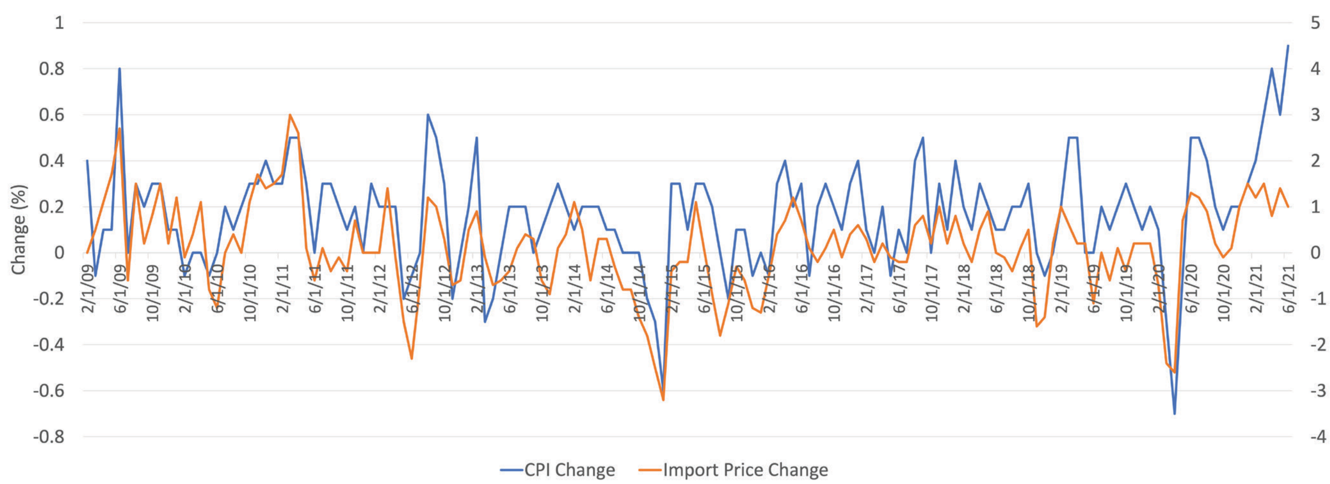

Figure 3: CPI change versus import price change. Line graph shows the relationship between the CPI percent change (blue) and import price percent change (orange) from February 2009 to June 2021.

- Cleveland Federal Reserve Estimate of Inflation Expectations
- Copper Price
- Quarterly GDP
- Small Cap Index
- Household Debt Service
- Labor Cost

Marginal Correlation Economic Variables

- Dollar Index
- 3-Month Lagged Personal Income
- Leading Indicators
- Livingston Forecast of Average Annual Growth for 10 Year
- 3 Month Lagged Average Hourly Earnings

- Total Job Openings
- Chicago Fed National Activity Index
- Value Index
- Home Price Index
- Organization for Economic Co-operation and Development (OECD) Consumer Opinion Survey
- Durable Goods Orders
- Partisan Conflict Index (PCI)

Zero Correlation Economic Variables

- S&P Case Shiller House Price Change
- Voluntary Quit Rate
- UMICH Consumer Sentiment
- 3-Month Lagged U.S. Government Debt
- Involuntary Quit Rate

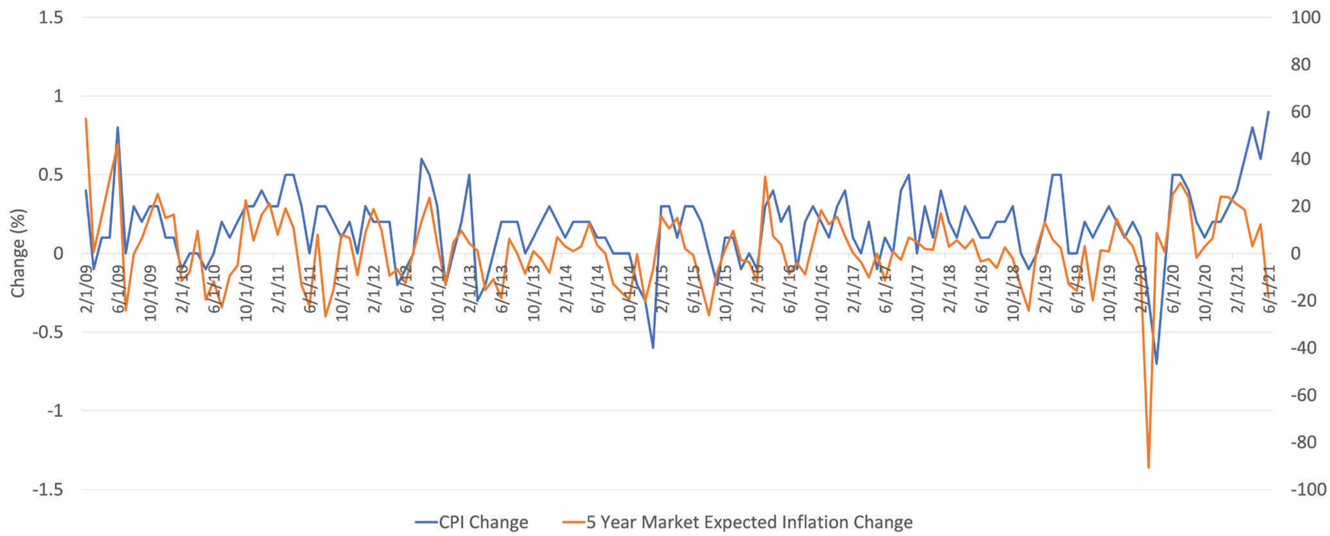

Figure 4: CPI change versus five-year market expected inflation change. Line graph shows the relationship between the CPI percent change (blue) and five-year market expected inflation percent change (orange) from February 2009 to June 2021.

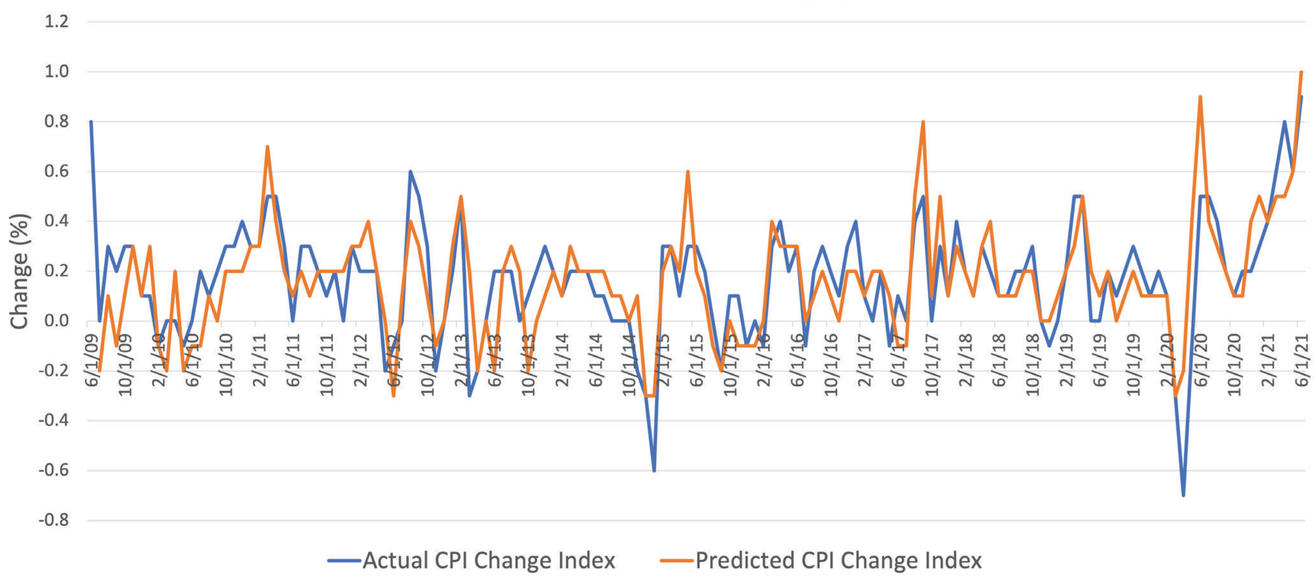

Figure 5: Actual CPI versus predicted CPI change. Line graph shows the relationship between the actual CPI percent change (blue), and the predicted CPI percent change (orange) from June 2009 to June 2021. Predicted CPI was created through a linear regression of average gas prices, average import prices, and 5-year market expected inflation.

- New Home Sales
- Gold Spot
- Trade Balance

The final regression model to predict CPI with the least average absolute error constituted of the average gasoline price, average import price, and 5-year market expected inflation.

When we ran this regression model against the actual CPI, it produced a mean absolute error of 0.107% over the years tested (2009 – 2021) and an average R^2 of 0.76 **(Figure 5)**. The model produced a maximum error of 0.40% and a minimum error of 0.00%. We calculated absolute error by using the following formula: $e = |x_i - x|$, where e = absolute error, x_i = predicted CPI, and x = actual CPI.

In summary, our data revealed that the average gasoline price, average import price, and 5-year market expected inflation have a strong relationship to inflation and can be used to predict it within an average of 0.107% absolute error.

DISCUSSION

This empirical study of the correlation between various economic variables and inflation shows that a combination of gas prices, import prices, and the 5-year market expected inflation can predict inflation, which can help traders forecast the behavior of inflation in the future by tracking as few variables as possible. Additionally, by understanding the highly correlated indicators with inflation, monetary and relevant governmental authorities can attempt to reduce high inflation. We have also shown that economic indicators that have historically been believed to be highly correlated with inflation might not be as correlated as previously thought. Commodities and wage prices, for example, have been used to try to predict inflation. However, their correlation with inflation through this study is not as high as the correlation of a different indicator with inflation, such as gas prices or import prices (3). Furthermore, gold prices have been conventionally considered to be an "inflationary hedge" and used to predict inflation, but this study shows very little correlation between gold and CPI (6).

The variables we discuss in the paper, especially the ones used in our regression, could be helpful for the government to develop public policy to reduce inflation. A general strategy to control inflation is to increase interest rates and slow down the economy (7). However, implementing this solution can crash the economy in the process. Instead, according to our study, if we can reduce gas prices, import prices, and change the public's perception of inflation, inflation might decrease in a way that is less harmful to the economy. Recently, reducing gas prices has been used as a potential mechanism of decreasing inflation (8). To do so, our study suggests that one solution is to ban energy exports for the near term. Seeing as gas is likely highly correlated with import prices, another possible solution is to ban tariffs on imported goods, such as those imposed by the Trump administration, or to provide a direct rebate to consumers purchasing goods with the money earned from imposing these tariffs (9). This would create a discount on the prices of goods in the country, likely lowering inflation. Additionally, following these policy changes and maintaining a vigilant Federal Reserve can lead to changing the public's view of inflation, which affects the 5-year market expected inflation.

One limitation of our model is its inaccuracy in predicting inflation during highly chaotic economic times, such as during the economic crisis in 2008 and the COVID-related market in 2020. These times were characterized by many unpredictable, sharp price changes that can be caused by many factors, some of which can be social or political.

For example, in the 12 months before June 2009—which was during the 2008 financial crisis—CPI was down more than 0.1% and gas prices were down 1.8% on average every month (10). In March 2009, the Federal Reserve tried to combat the recession by introducing quantitative easing to encourage economic activity, in which the Federal Reserve bought billions of dollars of securities aiming to expand the money supply and encourage money lending and investment (11). The Federal Reserve's policy change was followed by the start of the summer driving season in June. As a result, gas prices rose 17.3% and accounted for over 80% of the 0.7% increase in inflation during this month (10). Based on behavior of the 12-month data on which the regression was established, the gap between the actual vs. predicted CPI changes was large **(Figure 5)**.

For March 2020, the regression model made another similar inaccurate CPI prediction. March 2020 was the onset of the COVID-19 pandemic in the U.S., where people were encouraged to isolate themselves, consequently not using as much gasoline or buying goods as the economy came to a standstill (12). This led to the actual inflation heavily decreasing during this time, but the model was not able to foresee these circumstances.

Although an out-of-sample rolling linear regression model was used to predict CPI in this research, we can also utilize machine learning models in the future such as k-nearest neighbors, neural networks, and random forests to try to obtain a more accurate prediction model. Additionally, we can try to predict long-term inflation, such as 6 months or 12 months ahead, instead of short-term inflation. These questions and others are still topics of active research.

Our results provide a model to generate a typically accurate prediction of CPI, which has vast applications. By uncovering the few variables that have a significant correlation to CPI, traders can have a general estimate of the direction CPI is heading in a given month, depending on the performance of indicators like gas prices, import prices, and 5-year inflation market expected inflation. Additionally, governmental agencies can make efforts to fight against inflation by having a better idea of what truly drives it.

MATERIALS AND METHODS

Our study consisted of first deriving the correlations between each individual economic indicator to CPI, and then creating a regression model to predict CPI one month ahead.

Calculating Correlations

Most of the economic releases were converted to monthly data. This is because CPI, along with many other economic numbers, is released monthly. Data series such as gas prices that can be observed daily were reduced to monthly data by taking the average of the daily prices over the month. In the case of economic releases that were released quarterly or semi-annually, CPI was converted to quarterly or semi-annual to match the frequency. An example is shown below demonstrating how to convert CPI to a quarterly frequency. If CPI (M_x) represents the CPI of any month x,

Let:

CPI (M_1) = 4%

CPI (M_2) = 7%

CPI (M_3) = 9%

Then, CPI ($M_1 + M_2 + M_3$) = 1.04 * 1.07 * 1.09 = 1.2130 = 12.13%

All the economic data series were then converted in terms of changes rather than the absolute level of data. Since financial markets are focused on volatility adjusted deviations from the mean, all data was standardized to a rolling Z-Score over a time period n (in months), defined by:

$$Z_{t,n} = (X_t - \bar{X}_{t-n-1,\ t-1}) / \sigma_{t-n-1,\ t-1}$$

The time periods over which rolling Z-Scores were calculated were 6 months, 12 months, and all data (the longest possible data is in the range of 10 to 50 years). The correlation was calculated between rolling Z-scores of economic indicators and CPI on both short-term and long-term periods, where the short-term period was defined as the last five years, and the long-term period was defined as all data (the longest data possible in the range of 10 to 50 years). The average of the correlations of the Z-scores over the last five years was the short-term correlation between CPI and economic variables. Similarly, the average of the correlations of the entire dataset was the long-term correlation between CPI and economic variables. We then took the short-term and long-term correlation measures to combine into a blended correlation number. If the short-term and long-term correlations differed by less than their average, then the blended correlation number would equal to the average of the two correlations. If the short-term and long-term correlations differed by more than their average, then we reduced the average of the short and long-term correlation measures by 50% to arrive at the blended correlation number. The mathematical process for deriving the blended correlation between an economic variable and CPI is described below:

We first used the Pearson Correlation Coefficient formula to find the correlation between two sets of data x and y:

$$\mathrm{r(x,y)} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}, \text{ where}$$

r = correlation coefficient
x_i = each individual value of the x-variable in a sample
y_i = each individual value of the y-variable in a sample
$\bar{x}$ = mean of the values of the x variable
$\bar{y}$ = mean of the values of the y-variable

To obtain our short-term correlation C_s, we found the average of the 5-year correlations of the 6-month Z-scores, 12-month Z-scores, and all data Z-scores.

$$C_S\,(\%) = \frac{r(Z_{6M(CPI)}, Z_{6M(\,variable)}) + r(Z_{12M,CPI}, Z_{12M,variable}) + r(Z_{all\ data\ (CPI)}, Z_{all\ data\ (\,variable)})}{3} \times 100$$

To obtain our long-term correlation C_L, we found the average of the correlations across the full length of the data of the 6-month Z-scores, 12-month Z-scores, and all data Z-scores.

$$C_L\,(\%) = \frac{r(Z_{6M(CPI)}, Z_{6M(variable)}) + r(Z_{12M(CPI)}, Z_{12M(variable)}) + r(Z_{all\ data(CPI)}, Z_{all\ data(variable)})}{3} \times 100$$

If $|C_S - C_L| \leq \left|\frac{C_S+C_L}{2}\right|$, the blended correlation (%) = $\frac{C_S+C_L}{2}$

If $|C_S - C_L| \geq \left|\frac{C_S+C_L}{2}\right|$, the blended correlation (%) = $\frac{C_S+C_L}{2} \times 0.5$

Finding Regression Model

We considered the seven variables in the highest correlation category to CPI (average gasoline price, average import price, producer price index (PPI), average export price, commodity index, 5-year market expected inflation, and oil price) in building a linear regression model. The goal was to find the regression model either with or without a linear constant that predicts CPI with the most accuracy on an out-of-sample basis, where out-of-sample dates implies using information known publicly before the release of the predicted CPI. The regression was based on a 12-month rolling dataset to better align with changing economic structure and be more relevant to the trading community. Inflation is a complex process and depends on many variables, both known and unknown, a lot of which are correlated with each other. To make this work relevant to the trading community, however, we tried to explain the most about inflation with the fewest number of economic variables as possible. In our research, we were able to narrow it down to three variables.

Using the formulas specified by our linear regression, we used a brute-force method—trying all the possible regressions of the seven variables—and found the model with the least amount of absolute error. Note that because we only had data for the 5-year market expected inflation variable since 2009, the final regression model spans from 2009 – 2021 while some variables can have data ranges from 10 – 50 years. The regression model was specified by the following equation, where Z is the 12-month rolling Z-score, and time t is in months:

$$Z_{\text{CPI chng, t}} = \alpha_{1,\,t} {}^{*}Z_{\text{GAS PRC chng, t}} + \alpha_{2,\,t} {}^{*}Z_{\text{IMPORT PRC chng, t}} + \alpha_{3,\,t} {}^{*}Z_{\text{5YR MARKET EXPECTED INFLATION chng, t}}$$

The prediction for the CPI change Z-score at time t+1 was then calculated by the following equation:

$$Z^{\text{Predicted}}_{\text{CPI chng, t+1}} = \alpha_{1,\,t} {}^{*}Z_{\text{GAS PRC chng, t+1}} + \alpha_{2,\,t} {}^{*}Z_{\text{IMPORT PRC chng, t}} + \alpha_{3,\,t} {}^{*}Z_{\text{5YR MKT EXPECTED INFLATION chng, t+1}}$$

The CPI for month t+1 is released in month t+2. Therefore, the gas price change and 5-year market expected inflation change at month t+1 are known before CPI is released for month t+1, so we could use the value of these two indicators at month t+1 in our prediction model. However, the import price change at month t+1 may not be known because it can be released after the CPI for month t+1. Because of this, we used the import price change from month t (thus with a one-month lag) in the prediction model. Using $Z^{\text{Predicted}}_{\text{CPI chng, t+1}}$, we could find the predicted CPI change at time t+1 through the equation:

$$\text{CPI chng}^{\text{Predicted}}_{\text{t+1}} = Z^{\text{Predicted}}_{\text{CPI chng, t+1}} {}^{*}\sigma_{\text{CPI chng, t-12, t}} + \text{CPI chng}_{\text{t-12, t}}$$

ACKNOWLEDGMENTS

We thank Dr. Randal Verbrugge (Cleveland Federal Reserve) for providing feedback on this study, Rachel Widra (Cleveland Federal Reserve) for providing advice on research techniques and methods, and Sumanth Hegde (NineAlpha Capital) for providing data for economic indicators.

Received: March 26, 2022
Accepted: July 11, 2022
Published: December 1, 2022

REFERENCES

1. *Consumer Price Index Frequently Asked Questions* : U.S. Bureau of Labor Statistics. www.bls.gov/cpi/questions-and-answers.htm. Accessed 22 Mar. 2022.
2. Verbrugge, Randal J., and Saeed Zaman. "Whose Inflation Expectations Best Predict Inflation?" *Economic Commentary (Federal Reserve Bank of Cleveland)*, 2021, pp. 1–7., doi:10.26509/frbc-ec-202119.
3. 2022. Bloomberg Professional Services. Bloomberg. Accessed 26 Nov. 2021.
4. Mitchell, Josh. "Soaring Energy Prices Raise Concerns About U.S. Inflation, Economy." *Wall Street Journal,* 10 Oct. 2021. www.wsj.com, www.wsj.com/articles/soaring-energy-prices-raise-concerns-about-u-s-inflation-economy-11633870800. Accessed 26 Nov. 2021.
5. *Import/Export Price Indexes : U.S. Bureau of Labor Statistics.* www.bls.gov/mxp/. Accessed 26 Nov. 2021.
6. Hulbert, Mark. "Gold as an Inflation Hedge: What the Past 50 Years Teaches Us." *Wall Street Journal.* www.wsj.com, www.wsj.com/articles/gold-as-an-inflation-hedge-what-the-past-50-years-teaches-us-11628283272. Accessed 22 Mar. 2022.
7. Sullivan, Becky. "Who Are the Winners and Losers of the Fed Hiking Interest Rates?" *NPR,* www.npr.org/2022/06/16/1105397522/who-are-the-winners-and-losers-of-the-fed-hiking-interest-rates. Accessed 22 Mar. 2022.
8. N, *et al.* "The Biden Administration Works to Curb Rising Inflation and Gas Prices." *NPR,* www.npr.org/2022/06/15/1105176519/the-biden-administration-works-to-curb-rising-inflation-and-gas-prices. Accessed 22 Mar. 2022.
9. Dezember, Ryan. "Natural-Gas Prices Plunge After Extended Outage at Texas LNG Facility." *Wall Street Journal.* https://www.wsj.com/articles/natural-gas-prices-plunge-after-extended-outage-at-texas-lng-facility-11655235895. Accessed June 14, 2022.
10. "Consumer Price Index: June 2009." *Bureau of Labor Statistics,* www.bls.gov/news.release/archives/cpi_07152009.pdf. Accessed 18 June 2022.
11. Foster, Sarah. "Quantitative Easing: Here's How It Works." *Bankrate,* www.bankrate.com/banking/federal-reserve/what-is-quantitative-easing/. Accessed 18 June 2022.
12. *Chart Book: Tracking the Post-Great Recession Economy. Center on Budget and Policy Priorities.* www.cbpp.org/research/economy/tracking-the-post-great-recession-economy. Accessed 26 Nov. 2021.