The Significance of Certain Demand-Pull Factors and Cost-Push Factors in the U.S. Inflation

CUNY SPS MSDS
Professor Arthur O'Connor
Euclid Zhang
Spring 2023

Table of Contents

Abstract

This study aims to find a small set of key indicators to guide us predict the future overall price levels. Previous studies on the inflation in England and Switzerland found that models with a few variables outperformed models with additional variables. While only a narrow range of Cost-Push factors were included in the prior studies, money supply was identified as the most significant variable. This study performs a similar analysis on the U.S. inflation with a wider range of Cost-Push factors such as the prices of certain raw materials. The results show that Gasoline plays a critical role in the U.S. inflation. Electricity and Labor Cost are also significant during many time periods. Cotton, Corn, Platinum, Polyethylene, and Milk did not have high significance in the past, but the significance is increasing dramatically in recent periods.

Introduction

During the recent House Rules Committee conference in discussion about the REIN IN Inflation Act, U.S. Representative Thomas Massie raised a question to his fellow Jamie Raskin: “Do you think creating new money and putting it in the economy has any effects on inflation?” (Forbes Breaking News, 2023). This was referencing the five trillion dollars printed by the Federal Government during the Covid 19 pandemic. Massie stated that when the government gave the money to people, the production of goods was not increased. Too many dollars chasing too few goods. That’s exactly what causes inflation. We may conclude the pandemic is over, but the impacts are still lasting. As the U.S. inflation had hit its 40-year high recently (Guilfoed, 2022), the concern of rising inflation is growing severely. High and variable rates of inflation affect the buying, selling, and planning decisions of businesses, workers, and consumers (Fernado, 2023). In order to mitigate the impacts of inflation on us as individuals, it is crucial to understand the influential factors of inflation.

There are countless factors that affect inflation, and it is impractical, and resource consuming to examine and monitor all of them. Finding a small set of key indicators would be a more appropriate option in this matter. While Massie stated that the increase of money in circulation is the primary source of inflation, he disagreed that it is a problem from the supply chain. He believed that the increase of the prices of a few commodities should not lead to the increased prices of all goods and services. However, the economy is interactive, the change of one essential input of production may have a chain reaction to other factors and eventually cause a considerable impact on the overall economy. Such essential inputs may be used as the key indicators of future inflation. We would verify this idea in this study by performing an analysis on the relative predictive power of certain Demand-Pull factors and Cost-Push factors on U.S. inflation.

The Federal Reserve utilizes monetary policies such as setting the interest rate to manage the money supply to control the demand of goods and service. The **Interest Rate** may be a good indicator of the Demand-Pull inflation as it is commonly used as a variable in inflation prediction models. In Keynesian economics, the aggregate demand is affected by the employment condition as people's wages (or disposable income) is determined by the labor market (Chen, 2022). The relation between inflation and the unemployment rate is also described in the Phillips Curve (Engemann, 2020). Hence, the **Unemployment** (or **Employment)** level and **Average Wage** level would be good candidate variables to be examined. As the Demand-Pull inflation is caused by too many dollars, the supply of money and credit may be a more direct measure of the change in demand. However, money is a broad definition. Based on the liquidity of the money, it is generally classified into **M0**, **M1**, or **M2**. Additionally, **Consumer Credit** and **Private Sector Credit** also affect the circulation of money. It would be beneficial to find out which of the variables are more suitable in forecasting the future inflation.

Cost-Push inflation occurs when higher costs of production decrease the aggregate supply and the price increases from production are passed onto consumers (Kenton, 2022). The **PPI** (Producer Price Index) is usually used to represent the costs of productions. While it is a good measure of the Cost-Push factors, this macro measurement could not help us find the components in productions that have significant impact on the overall inflation. Not much attention was given to the cost of a specific input of productions (raw materials, labor, energy, etc.) in the past inflation analysis. The increases of **Crude Oil** had played a significant role in the recent high inflations (Lioudis, 2022). Are there any other essential inputs that are key ingredients in the composition of the Cost-Push inflation? This may help us find strategic items that were not previously identified. It would provide great help in adjusting business or policy plans to mitigate the risk of inflation caused by these factors.

In this analysis, we will examine the relative predictive power of above-mentioned candidate variables using mathematical models. Based on its analysis, the study seeks to identify a relatively small group of variables that have high levels of predictive power in forecasting inflation.

Literature Review

From the paper *Time Series: Economic Forecasting* (2002) written by Stock, statistical approaches used in economic forecasting can be classified into two categories: the structural method and time-series method. Structural method uses economic theories to construct the relations between variables within a model. Time-series method uses economic theories solely in guiding variable selection. As we are not focusing on the relationships among variables, our choice will be the time-series method. Time-series models can also be separated into two types: the univariate models and the multivariate models. An inspiring finding from Stock's paper is that "among univariate models, simple linear models often forecast as well or better than more complicated nonlinear models; and the gains from moving from univariate to multivariate models are often (but not always) small." Though a univariate model may be performing better in forecasting, our main concern is to examine the relative predictive power of multiple variables. We will select the multivariate models to perform our analysis. The multivariate model discussed in Stock's paper and many other research is the **Vector Auto-Regression (VAR)** model.

The paper *VAR models of inflation* published in the quarterly bulletin (1993 Q2) by the Bank of England illustrates the use of VAR to show a small list of inflation indicators can be used in inflation forecasting. The study analyzed a list of 29 variables including the money supply, producer price index and world commodity price index in the U.K. The result of the research shows that the optimal monthly model consists of only five variables (including the inflation measure) and the optimal quarterly model consists of only three variables (including the inflation measure). The common variable other than the inflation measure included in both models is the changes in M0 money supply. Adding any of the additional variables does not improve the performance of the models with an exception that adding the exchange rate to the optimal monthly model does have a slight improvement in the 6-month and 12-month inflation forecasting but not the 1-month. This study shows that it is plausible that a small set of variables can be used as indicators of inflation effectively.

There is a similar finding in the paper *Forecasting Swiss inflation using VAR models* (2006) by Caesar Lack. Lack constructed his forecast using the combined results of the best performing models and the optimal forecasts are obtained from a pool of only seven variables. Adding more variables to the pool only increases the forecast error. Among the pool of variables, mortgage loans and M3 money supply are the most relevant ones. In addition to variable selection, the models are tuned using different length of training period and the number of lags in the models. According to Lack, a longer training period creates a more precise coefficient estimation, and a shorter training period allows the model to adjust faster to structural changes of the economy. An optimal length of training period should be found to achieve the best forecast performance. The number lags can also affect the performance of the models greatly as there is a trade-off between autocorrelation handling and the precision of coefficient estimates. These parameters must be considered for our study.

While both studies (*VAR models of inflation* and *Forecasting Swiss inflation using VAR models*) find that a small set of variables can be used to forecast inflation effectively, those analysis are done on the inflation in England and Switzerland, not the U.S.. Additionally, variables used in both analyses are mostly Demand-Pull factors such as the money supply, GDP and interest rate. Not much attention is given to the Cost-Push factors. Hence, it is worthy to perform an insightful analysis to examine if there are any similar or different key indicators of the U.S. inflation.

Research Question

Prior research show that a small set of variables can be used to forecast the inflation of England and Switzerland effectively. We would like to perform a similar study on the U.S. inflation and extend our analysis to include a wide range of essential production inputs (or Cost-Push factors). The following questions would be answered through our analysis:

- Are there a small set of variables that can be used as the key indicators of the U.S. inflation?
- If so, are there any common variables also found as the key indicators of the inflation in England and Switzerland?
- Are there any essential production inputs selected as the key indicators of the U.S. inflation?

Data and Variables

Monthly values from December 2009 to December 2022 of the following data are collected from multiple sources. For details of data sources, please see Appendix A.

- **CPI**: Consumer Price Index (U.S. city average, All items).
- **PPI**: Producer Price Index (Total manufacturing industries).
- **IPI**: Import Price Indexes (All Commodities).
- **Interest Rate**: Federal Funds Effective Rate.
- **M0 Money Supply**: Currency in circulation and money being kept by banks in reserves (in billions).
- **M1 Money Supply**: M0 plus remaining demand deposits not in reserves (in billions).
- **M2 Money Supply**: M1 plus savings accounts, time deposits (under $100,000), and retail money market funds (in billions).
- **Consumer Credit**: Outstanding credit flows extended to individuals for household, family, and other personal expenditures, excluding loans secured by real estate.
- **Private Sector Credit**: Outstanding credit flows extended to the private sector.
- **Real GDP Index**: An estimated measure of the inflation adjusted GDP. Since the real GDP is reported quarterly, this monthly measure is used as a proxy of the real GDP.
- **Average Wage**: Average hourly earnings of private employees.
- **Average Hours**: Average weekly work hours of private employees.
- **Employment Level**: A measure of the overall employment level (in thousands of people) from the current population survey.
- **Unemployment Level**: A measure of the overall unemployment level (in thousands of people) from the current population survey.
- **Price of Raw Materials** (multiple): Price of essential raw materials for productions. For example, crude oil, lumber, wheat, etc. Monthly closing prices of the materials

are used except for eggs as the data is incomplete from the same source. Average price of eggs is collected from a different data source.

- **Price of Electricity**: Electricity per KWH in U.S. city average

Since M0 Money Supply is included in the M1 measure and M1 is included in M2, these three variables may be highly correlated. Therefore, M1 and M2 are replaced by the following variables in our analysis:

- **M1 Minus M0**: represents the demand deposits not in reserves (in billions).
- **M2 Minus M1**: represents the total in savings accounts, time deposits (under $100,000), and retail money market funds (in billions).

One of the important inputs of production is the labor cost, which is the productivity of labor related to the cost of wages. The Unit Labor Cost reported by the U.S. Bureau of Labor Statistics is a good measure of the labor cost. However, the data is reported quarterly. The monthly value of the following calculation is added as a proxy measure of the labor cost:

- **Labor Cost =** (**Real GDP Index**) / (**Employment Level** * **Average Wage** * **Average Hour**)

This is the total value of (inflation adjusted) production divided by the total of wages spent in laborers.

To improve the performance of models and have the coefficients of the variables to be comparable, data is usually normalized (transforming each variable to have a mean 0 and standard deviation 1) before fitting to a model. In our analysis, we would like to keep the model output (CPI) to be in its original scale. Hence, the variables are scaled and shifted to have the same mean and standard deviation as CPI.

Statistical Methods

The statistical model used in this analysis is the Vector Auto-Regression (VAR) model. The model consists of a set of equations. Each equation predicts the next-period value of one variable using the past values of all variables (including the variable being predicted).

One parameter of the model is the number of lags, p. Since the degree of freedom (or the number of coefficients) of the model is increased by the multiplication of parameter p and the number of variables n, it would require a much higher number of observations to train the model if p and n are large. To keep the model in a reasonable complexity, we perform our analysis using models with p up to 3.

Additionally, in order to adjust the speed of the model's reaction to the structural changes of the economy, the models are also constructed using different length of training period, l. Based on the degree of freedom of the model, the required minimum length of training period may change.

As a note, the two parameters of our model are:

- p: the number of lags
- l: the length of training period

In this analysis, a list of VAR models is constructed using all valid combinations of p (up to 3) and l (up to 60 months). The selected variables in the best performing models are examined to determine the key indicator of inflation.

Back-testing is performed to compare the performance of the model and the **Root Mean Square Forecast Error (RMSFE)** is used for the performance evaluation.

The method **Forward Selection** is used for the variable selection in each model of specific p and$l$. The process starts by finding the best performing bivariate VAR model (constructed with CPI and one of the other variables) as our base model. Next, the model is re-constructed by adding one of the remaining variables. If the addition of any of the variables results in better performance, the model with the lowest RMSFE would be our new best performing model. The step of adding an additional variable is repeated until no improvement can be made to the model, or the total number of variables other than CPI reaches 10. Models including more than 10 inflation factors may contradict with the goal of finding several key indicators of inflation. However, if the models with a high number of variables have the best or near-best performance, the variable selection process should be continued, and any notable findings should be explained.

The algorithm is summarized as following:

For each option of number of lags (up to 3):

For each option of length of training period (up to 60 months):

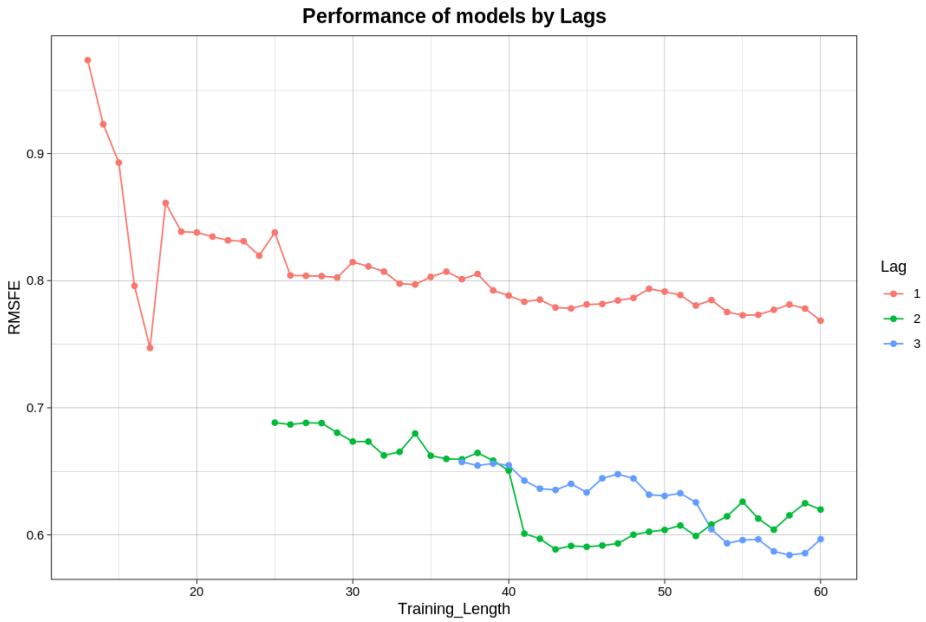
(a) Construct all possible bivariate VAR models with CPI and one of the other variables.
(b) Calculate the performance of the models using Back-testing.
(c) Identify the best performing model.
(d) Construct all possible models with variables from the previous best performing model and one additional variable.
(e) Calculate the performance of the new models using Back-testing.
(f) Identify the model that has the best improvement in performance.
(g) If no improvement can be made or the number of variables other than CPI reaches 10, stop the process of variable selection. Otherwise, go to step (d) and repeat.

An example of the performance outputs in the variable selection process is shown in the table of Appendix B. With number of lags = 1 and length of training period = 37, the combination of CPI and Palm Oil produces the best performing bivariate model. In the second step, the RMSEF is reduced from 0.812809 to 0.807304 when the variable Corn is added. The RMSEF is subsequently reduced to 0.801127 when the variable Nickel is added. After that, adding any fourth variable only increases the RMSEF. Therefore, the final model in this case consists of three variables (other than CPI): Palm Oil, Corn, and Nickel.

Finally, a group of best performing models is selected across all valid combinations of p and l. The appearance of the variables in the top models would give us a tentative identification of the key indicators of inflation. The significance of the variables would then be confirmed by checking the p-values of the coefficients in the best performing model. If the coefficient of a variable is statistically significant most of the time in the iterations of the Back-testing process, then the variable is confirmed as a key indicator of inflation.

Results

The following plot shows the performance of the models by different numbers of lags. The RMSFE measured the forecasting errors, the lower the number, the better the performance.

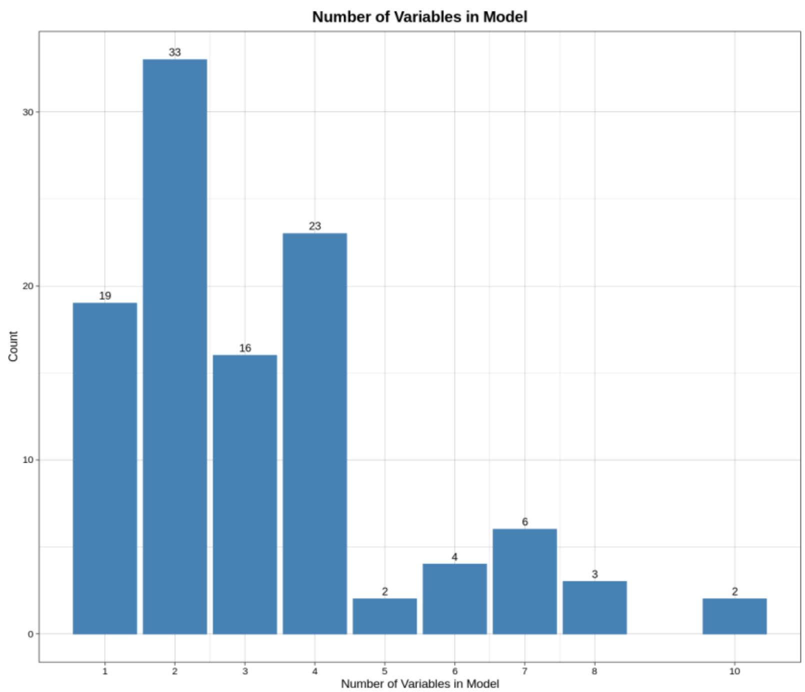


**The minimum length of training period increases as the complexity of the model increases

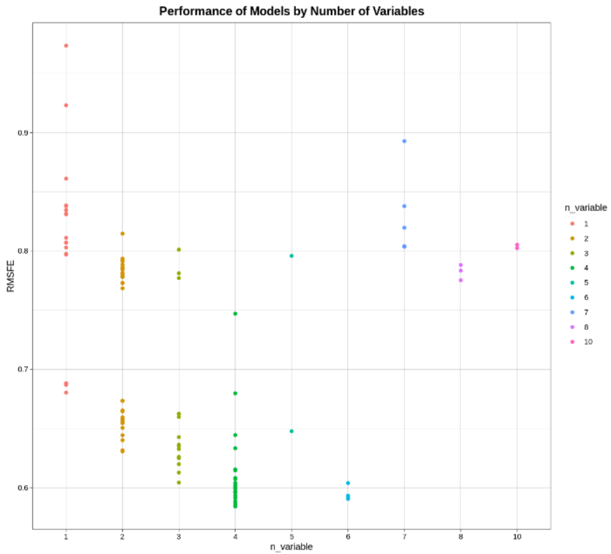
Models with p (number of lags) = 1 perform poorly. Models with p = 2 and models with p =3 have comparable performance. Setting the value p higher than 3 may produce models with slightly better performance. However, increasing the complexity of the models is costly and requires a much larger data set for the training. It may be done for further research but not included in this study.

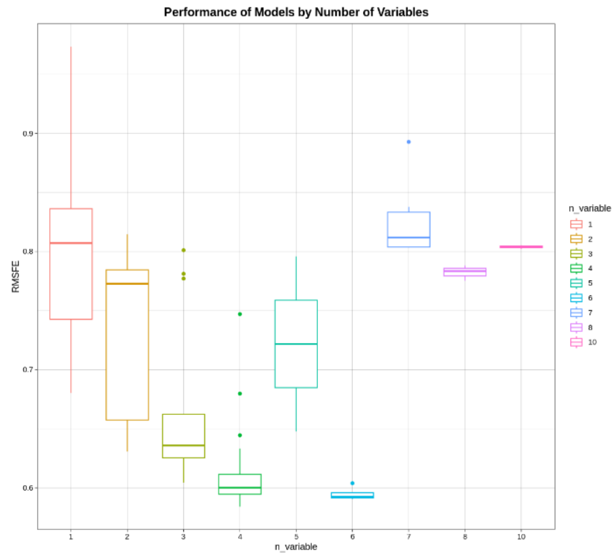
For all three values of p, the performance starts to improve generally as the l (length of training period) increases. However, when l reaches a certain value for $p = 2$ or $p = 3$, the performance does not have notable improvement and may go even lower. The optimal performance occurs when $l = 58$ for $p = 3$ and $l = 43$ for $p = 2$.

The number of selected variables in the models is summarized in the following plot. Across all combination of the parameters p and l, most of the models include 4 or less variables. Only two models include 10 variables. The performance of the two models may be improved if additional variables are added. We may check their performance to confirm if it is worthy to go beyond 10 variables.



From the following plots, models with 4 or less variables seem to have a better performance as the number of variables is higher. Models with 7 or more variables have significantly lower performance than the best performing models. It is not worthy to extend the models with 10 variables as that also increases the complexity of the models. Among all models, the ones with 4 or 6 variables have the best performance. This answers our first research question that a small set of variables can be used as key indicators of the U.S. inflation. We would look at the best performing models and confirm which are the key indicators.

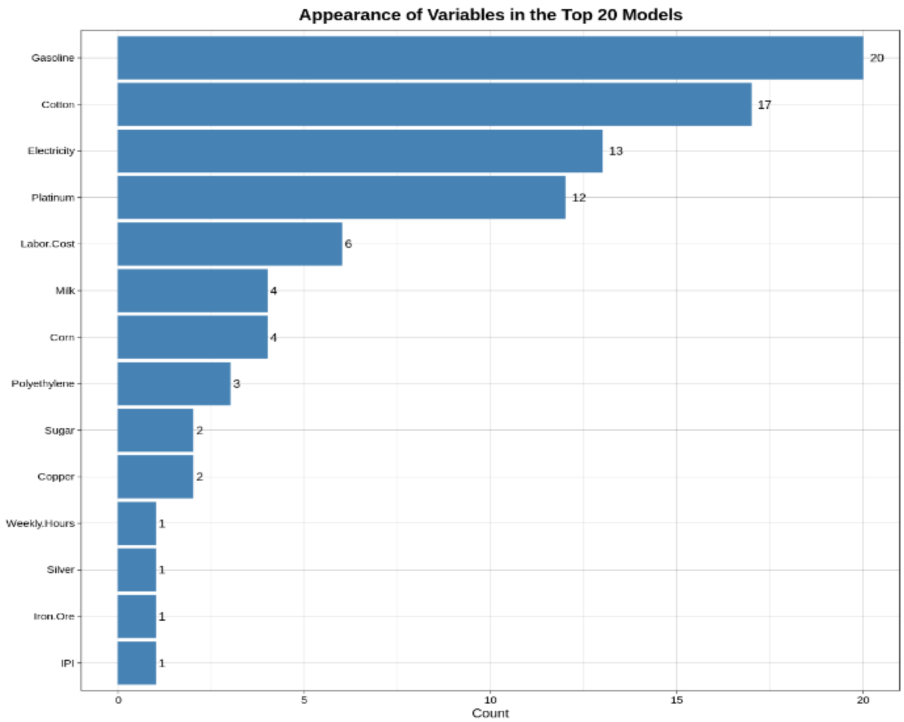




The following table shows the variables in the top 20 models.

	p	*l*	RMSFE	V1	V2	V3	V4	V5	V6
1	3	58	0.584211	Gasoline	Electricity	Cotton	Platinum	NA	NA
2	3	59	0.58564	Gasoline	Electricity	Cotton	Platinum	NA	NA
3	3	57	0.587074	Gasoline	Cotton	Electricity	Platinum	NA	NA
4	2	43	0.588683	Gasoline	Cotton	Corn	Labor.Cost	NA	NA
5	2	45	0.590681	Gasoline	Cotton	Platinum	Electricity	Polyethylene	Milk
6	2	44	0.591404	Gasoline	Cotton	Corn	Labor.Cost	NA	NA
7	2	46	0.591707	Gasoline	Cotton	Platinum	Electricity	Polyethylene	Milk
8	2	47	0.593314	Gasoline	Cotton	Platinum	Electricity	Polyethylene	Milk
9	3	54	0.59346	Gasoline	Electricity	Copper	Labor.Cost	NA	NA
10	3	55	0.595948	Gasoline	Electricity	Cotton	Platinum	NA	NA
11	3	56	0.596576	Gasoline	Electricity	Copper	Labor.Cost	NA	NA
12	3	60	0.596673	Gasoline	Cotton	Electricity	Platinum	NA	NA
13	2	42	0.597023	Gasoline	Cotton	Corn	Labor.Cost	NA	NA
14	2	52	0.599205	Gasoline	Cotton	Platinum	Electricity	NA	NA
15	2	48	0.600231	Gasoline	Cotton	Platinum	Sugar	NA	NA
16	2	41	0.601081	Gasoline	Cotton	Corn	Labor.Cost	NA	NA
17	2	49	0.602541	Gasoline	Cotton	Platinum	Sugar	NA	NA
18	2	50	0.603976	Gasoline	Cotton	Platinum	Electricity	Milk	Weekly.Hours
19	2	57	0.604169	Gasoline	Cotton	Iron.Ore	Silver	NA	NA
20	3	53	0.604447	Gasoline	Electricity	IPI	NA	NA	NA

It can be clearly seen that Gasoline is selected as the first variable in all models. Cotton, Electricity, and Platinum also appear frequently in the model. Surprisingly, the proxy variable Labor Cost also appears in 6 of the 20 models. The following plot shows the distribution of the appearance of the variables in the top 20 models.

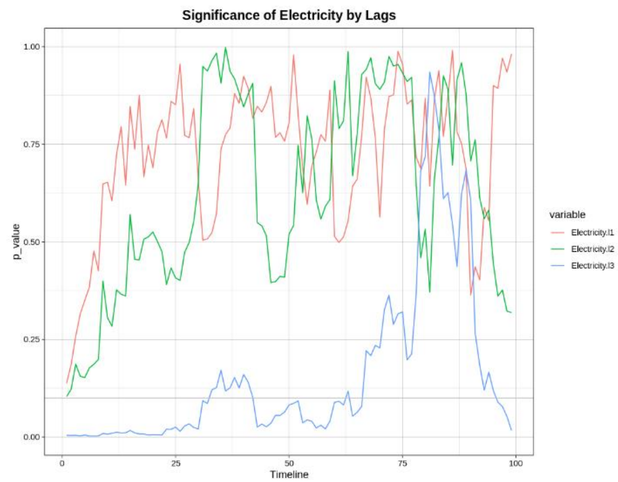


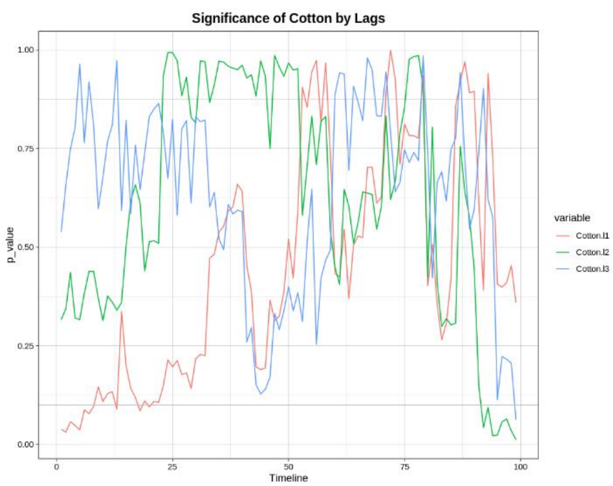
The top 3 models contain the same variables. We would focus on examining the highlight models that contain the most popular variables.

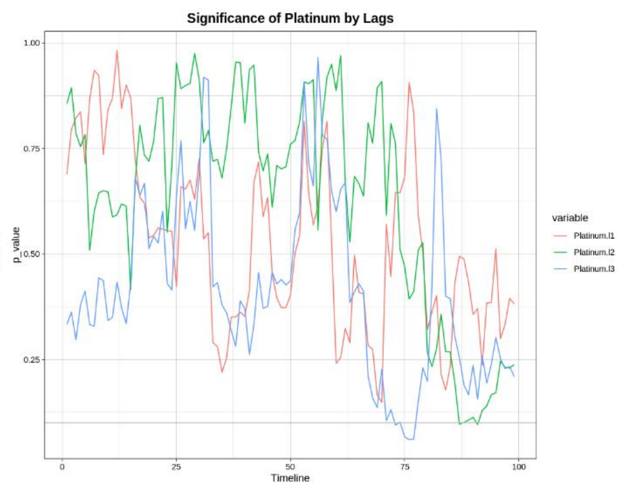
	p	*l*	**RMSFE**	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**
1	3	58	0.584211	Gasoline	Electricity	Cotton	Platinum	NA	NA
4	2	43	0.588683	Gasoline	Cotton	Corn	Labor.Cost	NA	NA
5	2	45	0.590681	Gasoline	Cotton	Platinum	Electricity	Polyethylene	Milk

Model 1

The following plots show the p-values of the coefficients of the variables in model 1. The p-value measures the probability that the coefficient equals 0. The lower the p-value, the higher the significance.



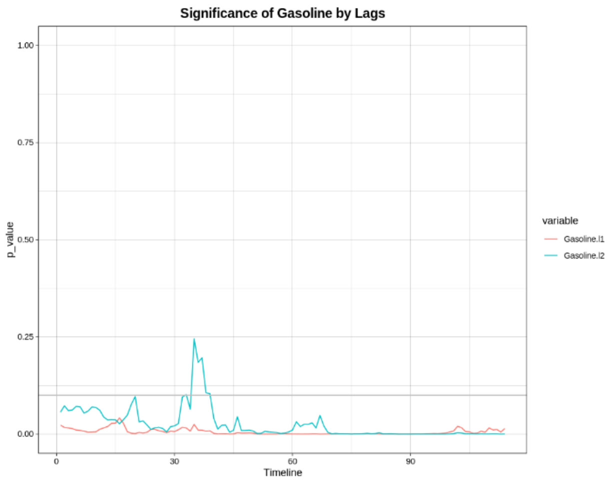


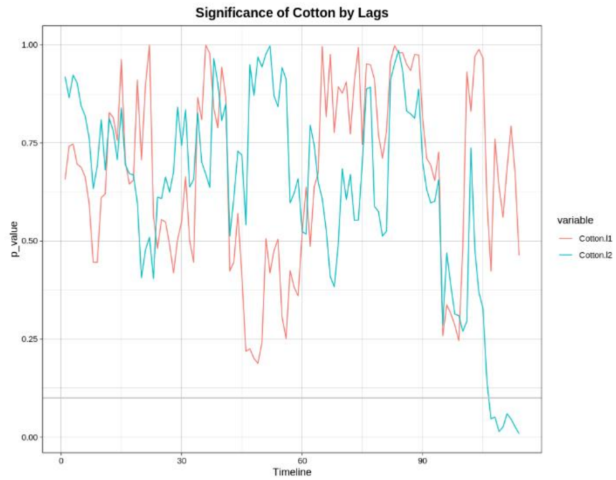


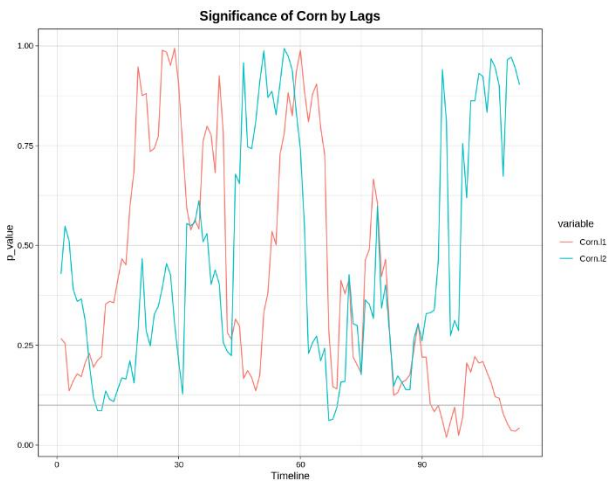
Gasoline is doubtlessly the most significant variable. The lag 1 and lag 2 values are strongly significant during all periods. The second variable, Electricity, has its lag 3 values significant most of the time. Its significance was decreasing from the beginning but then suddenly increased in the recent periods. The other two variables, Cotton and Platinum do not seem to be significant most of the time, but the significance for most of the lagged values are increasing dramatically in the recent periods.

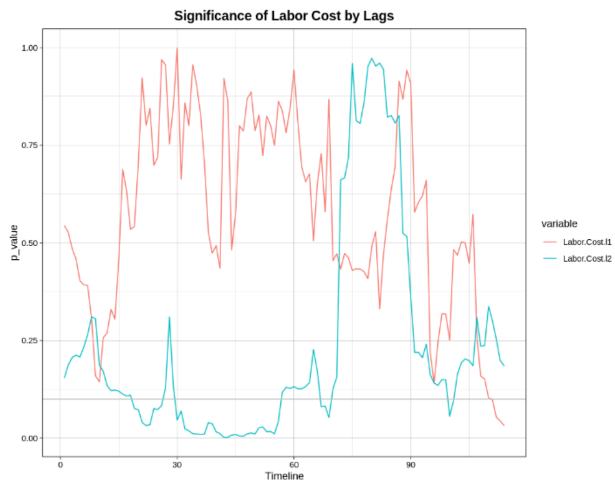
Model 4

The following plots show the p-values of the coefficients of the variables in model 4.





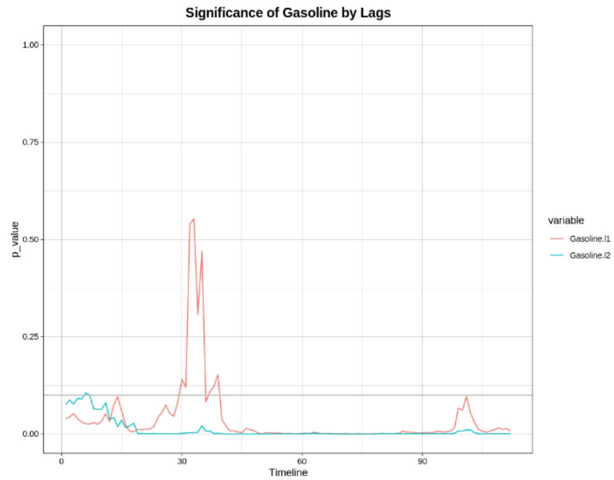


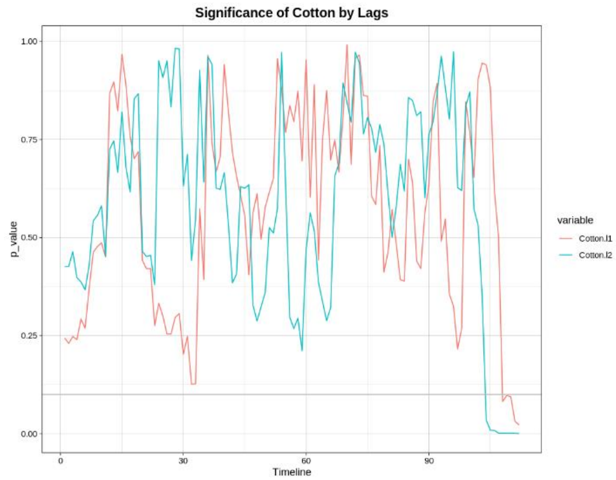


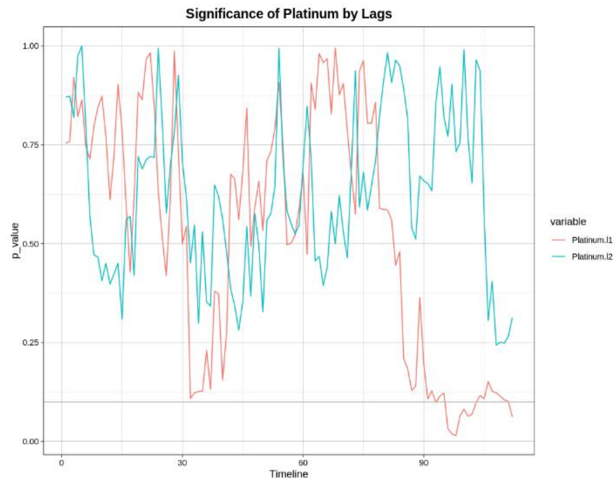
The lagged values of Gasoline are significant most of the periods. Cotton and Corn do not have high significance for most of the time, but some of the lagged values have increasing significance for the recent periods. Even though Labor Cost is a proxy variable, its lag 2 values play a somewhat important role in most of the periods.

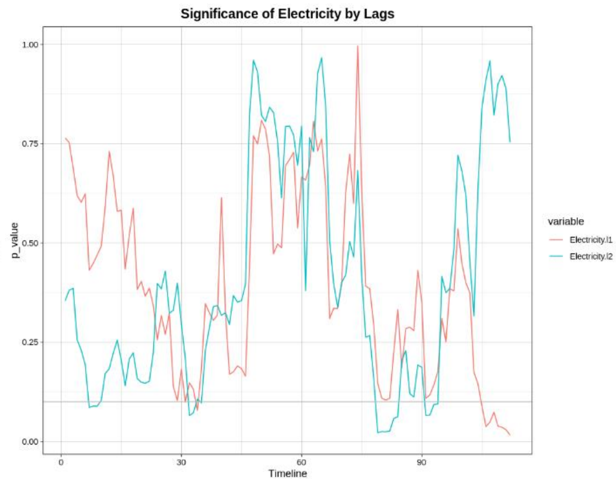
Model 5

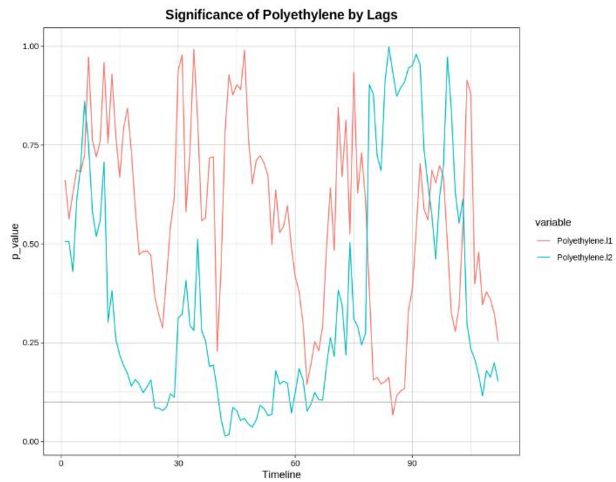
The following plots show the p-values of the coefficients of the variables in model 5.

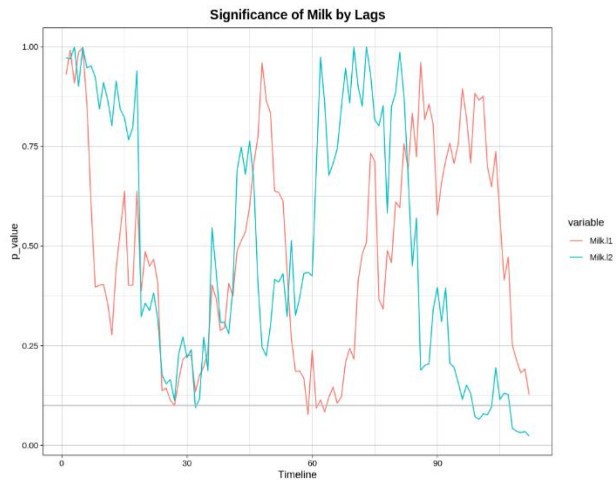






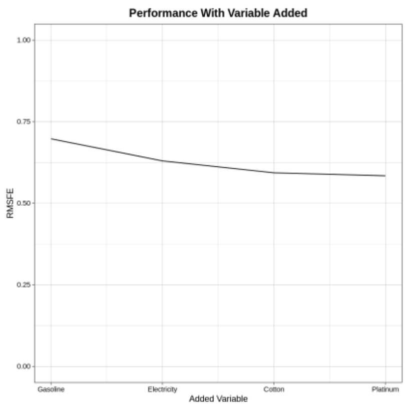




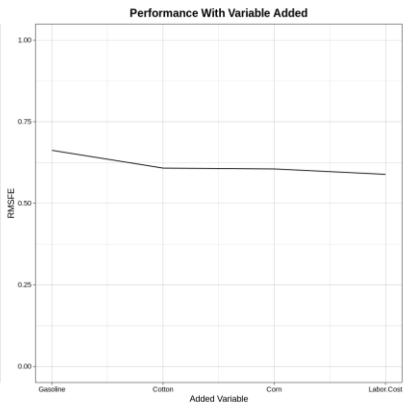


The results are similar to model 4, Gasoline is significant in most of the periods. All other variables do not have very high significance, but the significance is increasing dramatically in the recent periods for most of the lagged values.

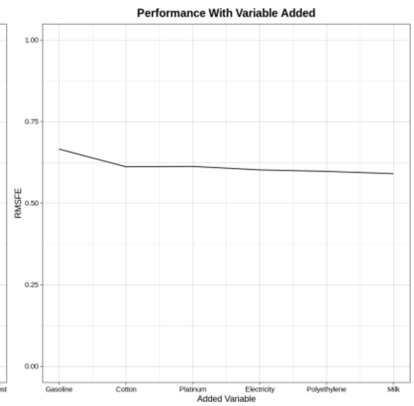
In all three models, only Gasoline is strongly significant in most of the periods. Other selected variables seem to be insignificant most of the time. Does including those variables actually improve the performance of the models? The following plots show the model performance when an additional variable is added.



Model 1



Model 4



Model 5

From the plots, only the second or third variable gives notable improvement to the model. Adding the remaining variables only produces a slightly better performance. A set of 3 or 4 variables may be sufficient to forecast the future inflation. However, the rapidly increasing significance of the remaining variables should not be ignored. Having a low significance in the past doesn't imply that they are not good indicators of future inflation as the economy is a changing environment. Including these variables is recommended as they also provide additional predictive power even though it's limited.

Conclusion and Discussion

From the results of the analysis, we find that a set of 6 or less variables can be used effectively as the key indicators of U.S. inflation. Models that include 7 or more variables have significantly lower performance. Among all models, we find the following three sets of variables produce the best forecasting results:

- Gasoline, Electricity, Cotton, Platinum
- Gasoline, Cotton, Corn, Labor Cost
- Gasoline, Cotton, Platinum, Electricity, Polyethylene, Milk

Among the variables in each set, the first two or three variables cover nearly all predictive power of the associated model. Though the other variables do not have significant contributions most of the time, their increasing significance in the recent periods must be aware of. They may potentially become the critical factors for future inflation.

The results show that Gasoline is strongly significant in most of the periods. This may be a function of whether this is a function of the study's timeframe (December 2009 to December 2022). As the government is promoting the use of renewable energy (Davis Polk, 2022), the significance of Gasoline may decline in the future. We can also see from the results that Electricity has increasing significance in the recent periods. Additionally, from the results, Gasoline appears to have a higher predictive power than Crude Oil. This may explain why Biden's Administration was pushing oil companies and gas stations to lower the retail price of Gasoline (Tayeb, 2022).

From the literature review, previous studies about the inflation of England and Switzerland include a narrow range of Cost-Push factors and found that Demand-Pull factors such as money supply are the key indicators of inflation. In this research of the U.S. inflation, we include a wider range of Cost-Push factors and find surprisingly that money supply and other macroeconomic measures such as the PPI and GDP are not selected by the models. Instead, all the selected variables in the top models are Cost-Push factors. The reason may be that a wider range of Cost-Push factors is included, or the economy of the U.S. is significantly different from the economy of England and Switzerland.

Though money supply is not included in the selected variables, it doesn't imply that money supply is not a significant factor of inflation. The results only suggest that the past values of the selected factors have relatively higher predictive power than past values of money supply for the forecast of future inflation. Moreover, the change in money supply may have an immediate effect on inflation so the past values of money supply may not carry much information about the future inflation.

Please note that the variables used in this analysis are only a small fraction of countless factors. Including additional variables may produce different results. Additionally, due to limited computing power, models of higher complexity are not tested. Further research may be done by expanding the number of candidate variables and raising the caps of the model parameters (number of lags and length of training period).

Reference

Chen, J., (2022, September 14). *What Is Demand-Pull Inflation?* Retrieved from: https://www.investopedia.com/terms/d/demandpullinflation.asp

Engemann, K. M., (2020, January 14). *What Is the Phillips Curve (and Why Has It Flattened)?* Retrieved from: https://www.stlouisfed.org/open-vault/2020/january/what-is-phillips-curve-why-flattened

Fernado, J., (2023 January, 18). *Inflation: What you need to know about the purchasing power of money and how it changes*. Retrieved from: https://www.investopedia.com/terms/i/inflation.asp

Forbes Breaking News, (2023 March, 4). *Massie Asks Raskin Point Blank: Did Recent Government Spending Have 'Any Effect On Inflation?'* [Video]. Youtube. https://www.youtube.com/watch?v=c4MMgHpWZdE

Guilfoed, G., (2022, October 13). *Inflation Sits at 8.2% as Core Prices Hit Four-Decade High*. Retrieved from: https://www.wsj.com/articles/us-inflation-september-2022-consumer-price-index-11665628037

Henry, S. G. B and Pesaran, B., (1993, May). *VAR models of inflation*. In: Bank of England Quarterly Bulletin. pp. 231-239. Retrieved from: https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/1993/var-models-of-inflation.pdf?la=en&hash=F2A1325C1342FED408F1F7F1287E6150964AAA07

Kenton, W., (2022, March 7). *Cost Push Inflation: When It Occurs, Definition, and Causes*. Retrieved from: https://www.investopedia.com/terms/c/costpushinflation.asp

Lack, C., (2006). *Forecasting Swiss inflation using VAR models*. In: Swiss National Bank Economic Studies. Retrieved from: https://www.snb.ch/n/mmr/reference/economic_studies_2006_02/source/economic_studies_2006_02.n.pdf

Lioudis, Nick., (2022, May 5). *What Is the Relationship Between Oil Prices and Inflation?* Retrieved from: https://www.investopedia.com/ask/answers/06/oilpricesinflation.asp

McKinsey & Company, (2022, August 17). *What is inflation?* Retrieved from: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-inflation

Davis Polk, (2022, August 16). *President Biden to sign largest ever U.S. investment in clean energy*. Retrieved from: https://www.davispolk.com/insights/client-update/president-biden-sign-largest-ever-us-investment-clean-energy

Stock, J. H., (2002). *Time Series: Economic Forecasting*. In: International Encyclopedia of the Social Sciences. pp. 15721-15724. Retrieved from: https://scholar.harvard.edu/files/stock/files/time_series_economic_forecasting.pdf

Tayeb, Z., (2022, September 27). *Biden tells gas stations to cut prices at the pump for American drivers: 'Do it now'*. Retrieved from: https://markets.businessinsider.com/news/commodities/biden-tells-gas-stations-lower-prices-reflect-falling-oil-costs-2022-9

Appendix

A. Data Sources

Variable	Data Source	Link
CPI	U.S. Bureau of Labor Statistics - Series Id "CUUR0000SA0"	https://data.bls.gov/
PPI	U.S. Bureau of Labor Statistics - Series Id "PCUOMFG--OMFG--"	https://data.bls.gov/
IPI	U.S. Bureau of Labor Statistics - Series Id "EIUIR"	https://data.bls.gov/
Interest Rate	ST. Louis FED	https://fred.stlouisfed.org/series/FEDFUNDS
M0	Trading Economics	https://tradingeconomics.com/united-states/money-supply-m0
M1	Board of Governors of the Federal Reserve System	https://www.federalreserve.gov/releases/h6/current/default.htm
M2	Board of Governors of the Federal Reserve System	https://www.federalreserve.gov/releases/h6/current/default.htm
Consumer Credit	Board of Governors of the Federal Reserve System	https://www.federalreserve.gov/releases/g19/current/default.htm
Private Sector Credit	Trading Economics	https://tradingeconomics.com/united-states/private-sector-credit
Average Wage	U.S. Bureau of Labor Statistics - Series Id "CES0500000003"	https://data.bls.gov/
Average Hours	U.S. Bureau of Labor Statistics - Series Id "CES0500000002"	https://data.bls.gov/
Employment Level	U.S. Bureau of Labor Statistics - Series Id "LNS12000000"	https://data.bls.gov/
Unemployment Level	U.S. Bureau of Labor Statistics - Series Id "LNS13000000"	https://data.bls.gov/
Monthly Real GDP Index	S&P Global Market Intelligence	https://www.spglobal.com/marketintelligence/en/mi/products/us-monthly-gdp-index.html
Crude Oil WTI	Trading Economics	https://tradingeconomics.com/commodity/crude-oil
Natural Gas	Trading Economics	https://tradingeconomics.com/commodity/natural-gas
Gasoline	Trading Economics	https://tradingeconomics.com/commodity/gasoline
Heating Oil	Trading Economics	https://tradingeconomics.com/commodity/heating-oil
Coal	Trading Economics	https://tradingeconomics.com/commodity/coal
Gold	Trading Economics	https://tradingeconomics.com/commodity/gold
Silver	Trading Economics	https://tradingeconomics.com/commodity/silver
Copper	Trading Economics	https://tradingeconomics.com/commodity/copper
Steel	Trading Economics	https://tradingeconomics.com/commodity/steel
Iron Ore	Trading Economics	https://tradingeconomics.com/commodity/iron-ore
Platinum	Trading Economics	https://tradingeconomics.com/commodity/platinum
Lead	Trading Economics	https://tradingeconomics.com/commodity/lead
Aluminum	Trading Economics	https://tradingeconomics.com/commodity/aluminum
Tin	Trading Economics	https://tradingeconomics.com/commodity/tin
Zinc	Trading Economics	https://tradingeconomics.com/commodity/zinc
Nickel	Trading Economics	https://tradingeconomics.com/commodity/nickel

Polyethylene	Trading Economics	https://tradingeconomics.com/commodity/polyethylene
Soybeans	Trading Economics	https://tradingeconomics.com/commodity/soybeans
Wheat	Trading Economics	https://tradingeconomics.com/commodity/wheat
Lumber	Trading Economics	https://tradingeconomics.com/commodity/lumber
Palm Oil	Trading Economics	https://tradingeconomics.com/commodity/palm-oil
Rubber	Trading Economics	https://tradingeconomics.com/commodity/rubber
Milk	Trading Economics	https://tradingeconomics.com/commodity/milk
Cotton	Trading Economics	https://tradingeconomics.com/commodity/cotton
Oat	Trading Economics	https://tradingeconomics.com/commodity/oat
Wool	Trading Economics	https://tradingeconomics.com/commodity/wool
Sugar	Trading Economics	https://tradingeconomics.com/commodity/sugar
Corn	Trading Economics	https://tradingeconomics.com/commodity/corn
Egg	U.S. Bureau of Labor Statistics - Series Id "APU0000708111"	https://data.bls.gov/
Electricity	U.S. Bureau of Labor Statistics - Series Id "APU000072610"	https://data.bls.gov/

B. Example of Model Performance in the Variable Selection Process

Variable	V1	V2	V3	V4
IPI	0.924901	0.860181	0.848479	0.842818
PPI	0.966925	0.868087	0.855043	0.852953
Unemp.Lv	2.095898	2.116108	2.017732	1.699902
Emp.Lv	1.369808	1.269166	1.221167	1.115166
Average.Wage	0.97061	0.840997	0.824087	0.821867
Weekly.Hours	0.932153	0.836102	0.841521	0.841599
FEDFUNDS	0.905848	0.852239	0.849992	0.860242
M0	0.953754	0.877596	0.854973	0.858719
Private.Credit	1.014199	0.905328	0.887328	0.878609
Consumer.Credit	0.930094	0.840296	0.826066	0.834336
Real.GDP	1.090207	0.951911	0.935071	0.917013
Crude.Oil.WTI	0.910617	0.840275	0.832724	0.843376
Natural.Gas	1.033293	0.880171	0.882079	0.885732
Gasoline	0.905615	0.846159	0.841172	0.847692
Heating.Oil	0.98818	0.853849	0.842732	0.849288
Coal	1.017444	0.85937	0.840092	0.839456
Gold	0.956815	0.845954	0.825242	0.83663
Silver	0.952321	0.874065	0.864647	0.87085
Copper	0.877653	0.837607	0.846121	0.847916
Steel	0.931122	0.868841	0.855852	0.857525
Iron.Ore	0.944445	0.858411	0.869546	0.879195
Platinum	0.908576	0.84036	0.842441	0.859164
Lead	0.952749	0.842054	0.835272	0.854002
Aluminum	0.891958	0.85471	0.840473	0.848751
Tin	0.900127	0.87773	0.887125	0.91536
Zinc	0.833103	0.809765	0.821354	0.825531
Nickel	0.923055	0.814546	0.801127	NA
Polyethylene	0.86368	0.821241	0.818445	0.833759
Soybeans	0.855594	0.820985	0.827116	0.830639
Wheat	0.865898	0.835408	0.835307	0.825773
Lumber	0.846903	0.81037	0.821464	0.840488
Palm.Oil	0.812809	NA	NA	NA
Rubber	0.88638	0.825486	0.835078	0.840294
Milk	0.990485	0.846375	0.841195	0.829787
Cotton	0.837822	0.825321	0.816961	0.827213
Oat	0.928128	0.855566	0.856563	0.858532
Wool	0.996504	0.837232	0.82701	0.817741
Sugar	0.942273	0.837151	0.8307	0.840845
Corn	0.875937	0.807304	NA	NA
Egg	0.947227	0.847689	0.839645	0.848194
Electricity	0.918401	0.814063	0.808752	0.804192
Labor.Cost	0.991986	0.838809	0.83623	0.830242
M1.Minus.M0	1.305071	1.237789	1.064101	1.067708
M2.Minus.M1	0.970983	0.879439	0.837858	0.913242

*** Number of lags = 1 and length of training period = 37