A Recreation of Vector Autoregression by James H. Stock and Mark W. Watson

Omer Abdelrahim

Sunday, November 26th 2023

1 Purpose

This project is a recreation of Stock and Watson's 2001 paper that focuses on the use of Vector Autoregressions (VAR) and how it can be applied to analyze macroeconomic data. Vector Autoregressions use 'lags' or previous values of a single variable to explain or predict the future values of that same variable.

The three variables we will be using are the Inflation Rate, Unemployment Rate and the Federal Funds Rate, with the data starting from 1960 going through today. The data was compiled on a quarterly basis, or once every three months. We will also be splitting the data into two groups. A training group composed of data from 1960-2009, and a testing group 2010-2023. We will use the groups accordingly to construct various AR(p) models and forecasts.

Just like in many other time series related processes, the data needs to be stationary and the variance needs to be mean reverting. This is why we modify the inflation rate (that is fixed to the year 1960 in the FRED data), to grow by a certain percent rate as explained by $IRATE_t = 400 \ln\left(\frac{P_t}{P_{t-1}}\right)$. We take the natural log value of the current inflation rate (in 1960 terms) over the inflation rate of the previous period and multiply that by 400. That being said, no data is perfect and the time series that are worked on in this project may have some issues with seasonality and trends (especially the Federal Funds Rate) but overall they do tend to have mean reverting variance in some form or fashion. We will now move on to predicting what type of AR(p) would be best used when forecasting future values of the Inflation Rate, Unemployment Rate and the Federal Funds Rate.

2 Partial Auto Correlation Functions

A Partial Auto Correlation Function is a function that shows the partial correlation between a variable and all its past lags or previous values. It is a process that is used on stationary time series, and gives guidance on up to how many lags we should expect to use when constructing an AR(p) model. Below we will see the significant lags for the three variables that we are working with.

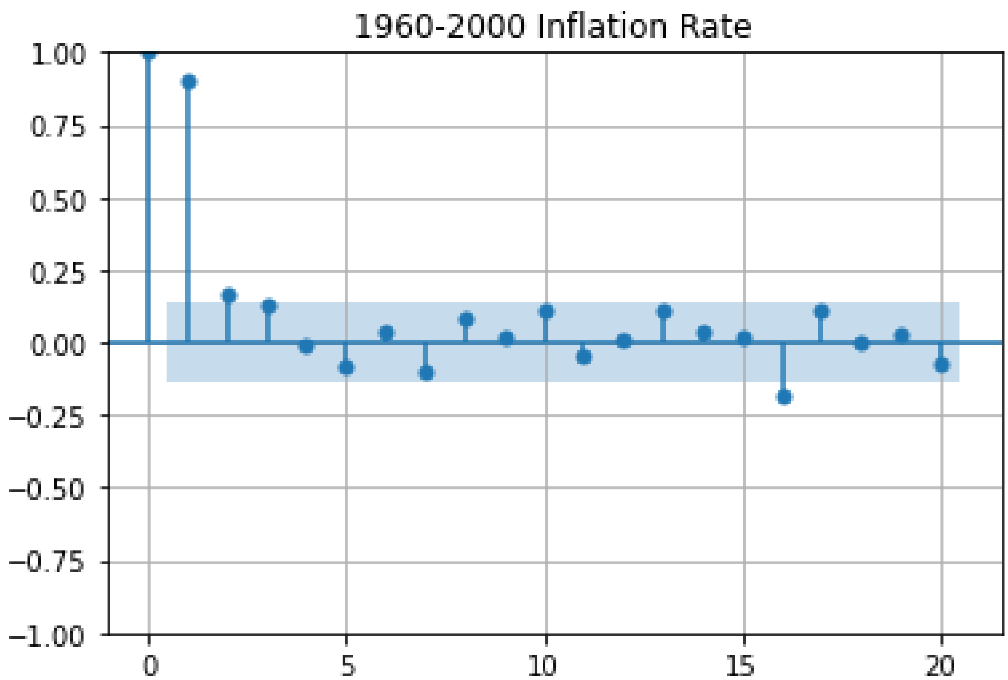

Figure 1: Partial Auto Correlation Function for the Inflation Rate from the years 1960-2009.

For Inflation Rate we see that that the first three lags pass the significance threshold (the area in blue). Beyond that there is a lag that is significant at 16, but considering that that lag is far away from any other significant lag we might want to ignore it if we can afford to. As such Inflation Rate seems to look like an AR(3) model. Since the data is compiled quarterly this means that the current inflation rate is partially correlated with the inflation rate from 9 months ago.

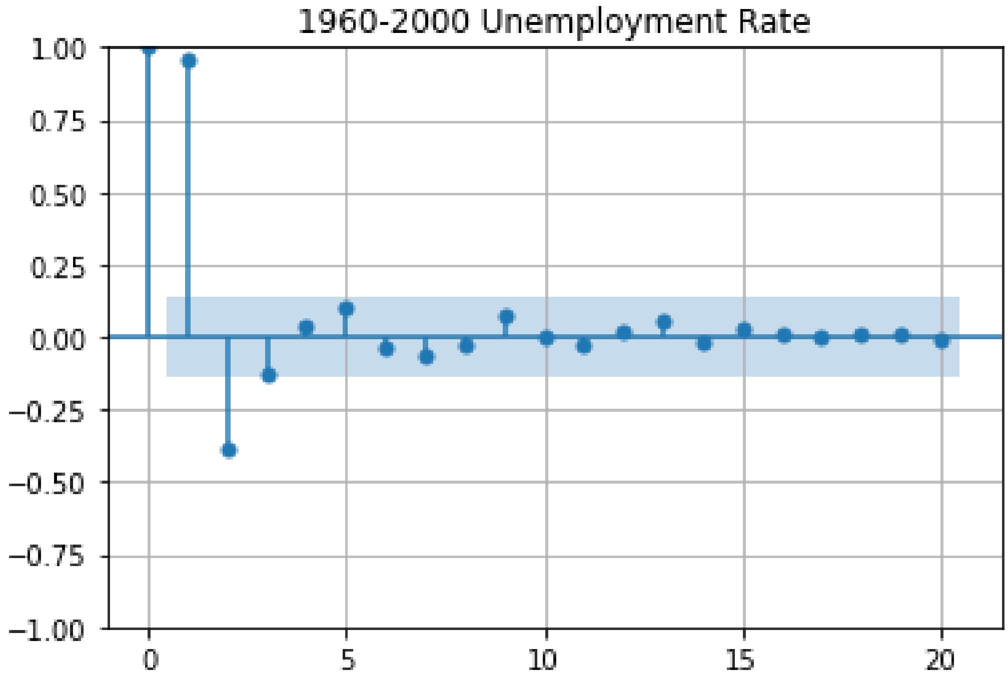

Figure 2: Partial Auto Correlation Function for the Unemployment Rate from the years 1960-2009.

For Unemployment rate we also see 3 significant lags, making it a candidate for an AR(*3*) model as well. Since the data is compiled quarterly it means that current unemployment rate is partially correlated with the unemployment rate up to 9 months or 3 quarters ago.

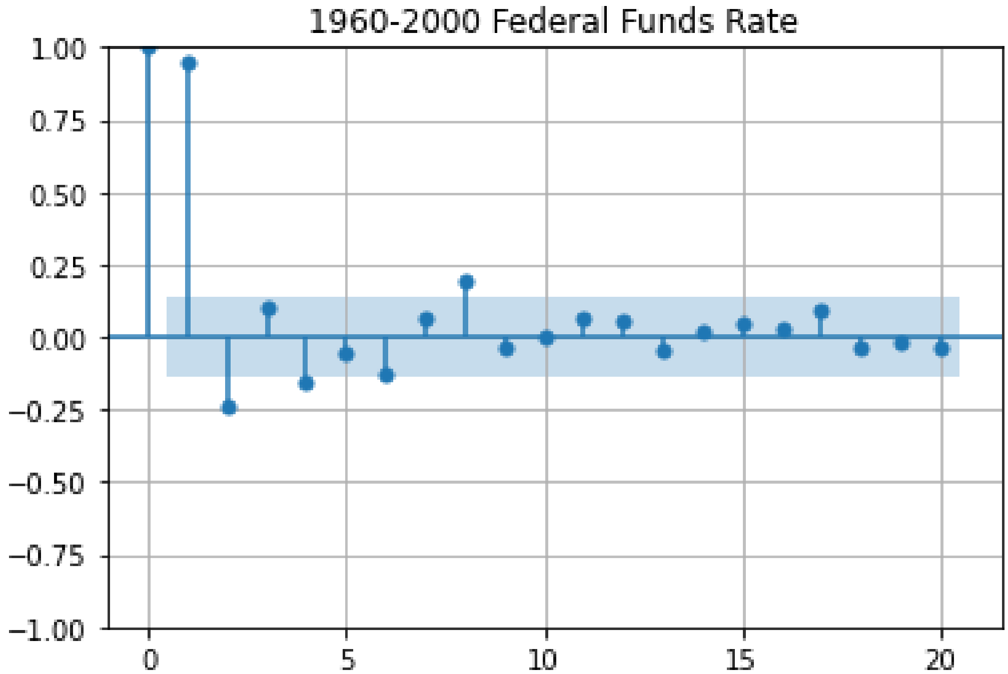

Figure 3: Partial Auto Correlation Function for the Federal Funds Rate from the years 1960-2009.

For the Federal Funds Rate we see significant lags up to the eight or so quarter (24 months). It makes sense that the Federal Funds Rate would have the most significant lags because it is a value that is controlled by the U.S. government which has predictable behavior relating to macroeconomic environment. The Fed considers past values when raising or lowering its basis points. As such we see a relatively higher AR value for the Federal Funds Rate AR(p) model standing at AR(8).

2.1 Optimal Lags

Using a function within the statsmodels.tsa.ar_model package called ar_select_order we are able to predict optimal amount of lags for an AR(p) model we receive the following output:

- Optimal lags for the Inflation Rate using AIC are:[1, 2, 3]
- Optimal lags for the Unemployment Rate using AIC are:[1, 2]
- Optimal lags for the Federal Funds Rate using AIC are:[1, 2, 3, 4, 5, 6, 7, 8]

3 Training Set AR Predictions (2010-2023)

Now we move onto using the lag values that we gleaned from our PACFs and the optimal lag function used in section 2. For Inflation Rate we will use 3 lags, Unemployment rate will use 2, and the Federal Funds Rate will use 8. Overall the values for the AR(p) models are pretty close to what we saw in our PACFs.

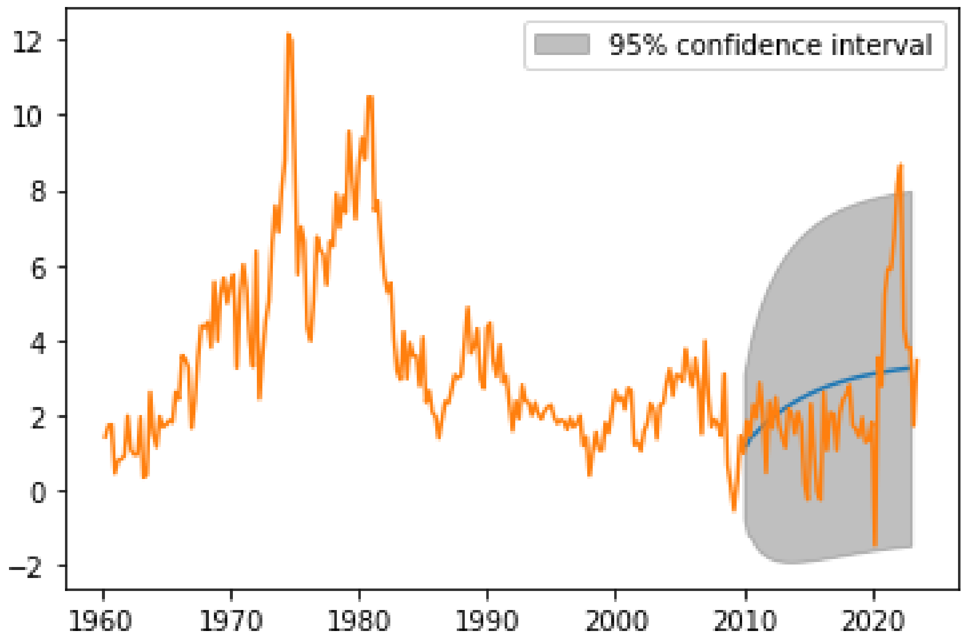

Figure 4: Time Series of the Inflation Rate from with a prediction of the Inflation Rate from the years 2010-2023 overlaid. Data from 1960-2009 is used as the training data for the AR(*3*) model prediction.

Inflation rate sees the prediction line from the years 2010-2023 follow or return the mean with the use of an AR(*3*) process. Most values fall within the .95 percent confidence interval, with the sole exception of the high inflation environment that we saw post-COVID. Since the pandemic was a statistical anomaly, we can expect to see more recent values, or at least the values from 2020, to fall outside normal expectations.

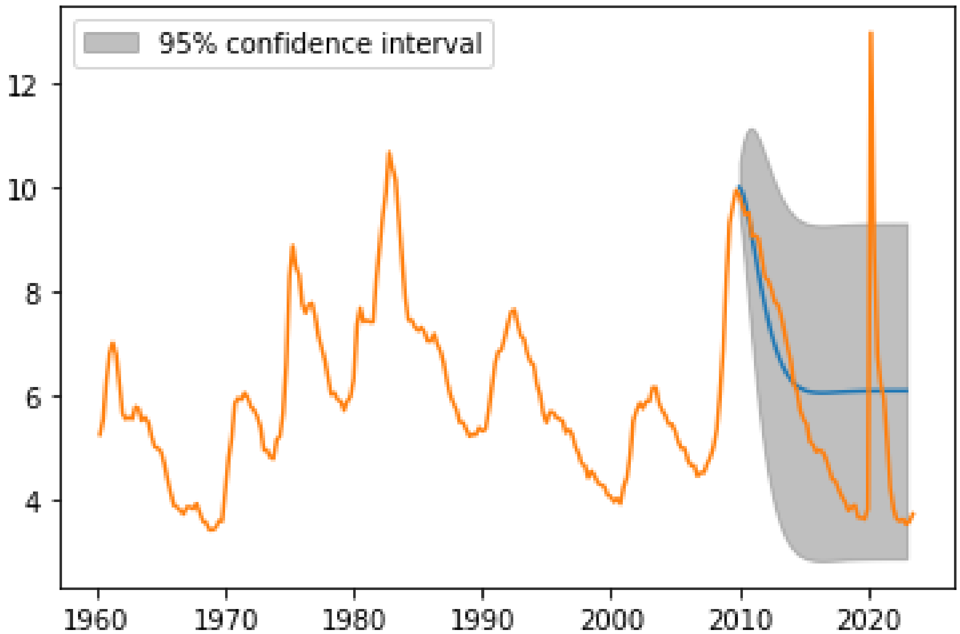

Figure 5: Time Series of the Unemployment Rate from 1960-2023 with a prediction of the Unemployment Rate from the years 2010-2023 overlaid. Data from 1960-2009 is used as the training data for the AR(*2*) model prediction.

Very similar to what we saw with the inflation graph although there is much more seasonality in the Unemployment graph. It might have been more appropriate to use a percent change from the previous value time series model for unemployment rather than just unemployment itself. We can see the beginning of the 2008 financial crisis having profound effects on the first parts of the forecast, and then normalizing quickly thereafter. Once again COVID and its unprecedented nature leads to actual values that lie outside the .95 confidence interval, before rocketing back down. The model has come to expect an increase back to the natural rate of unemployment within this model which lies just below 0.06 percent.

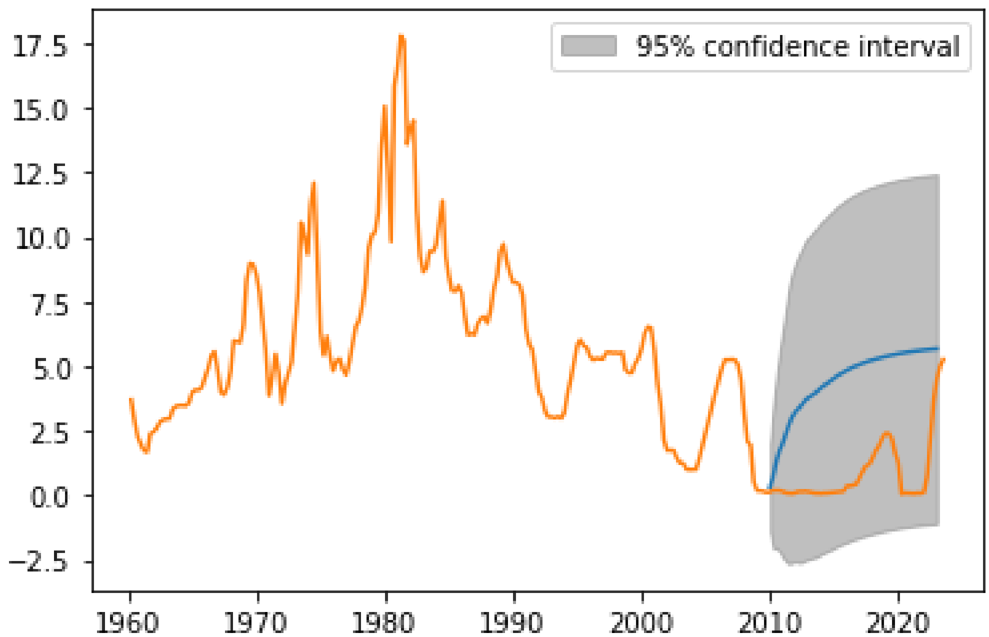

Figure 6: Time Series of the Federal Funds Rate from 1960-2023 with a prediction of the Federal Funds Rate from the years 2010-2023 overlaid. Data from 1960-2009 is used as the training data for the AR(8) model prediction.

Finally we see the Federal Funds Rate and its AR(8) prediction. There also is a concern of seasonality and trend with this dataset, which might lead to biased or incorrect predictions. We can see the prediction line revert to the mean, and coincidentally meeting up with the actual Federal Funds Rate value. Rates stayed lower for a substantially longer time than what the model expected due to macroeconomic conditions, such as the stagnated growth of the US economy following the 2008 crisis which perplexed many economists. Growth rates seem to have normalized relative to what economists were expecting post COVID.

4 Full Dataset (1960-2023) AR Prediction Extensions 2023-2025

Moving forward for the AR(p) models we use the entire model to train the forecast. Hopefully we can receive more accurate values for the prediction line and confidence interval.

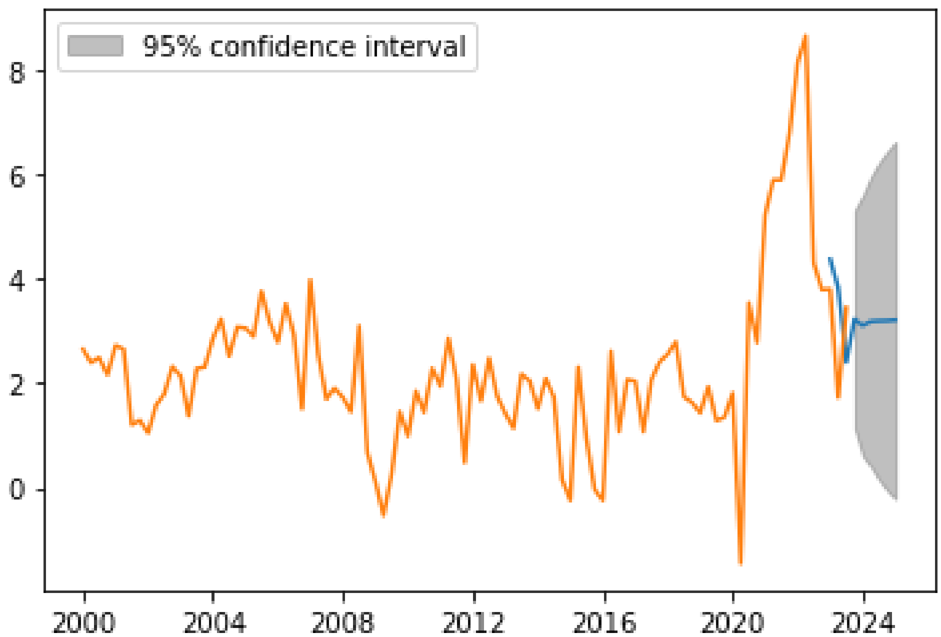

Figure 7: AR(*3*) Prediction extensions using the full dataset from 1960-2023 for the Inflation Rate. The AR(*3*) predictions themselves only go from 2023 to 2025.

Again we see the extension favoring a return to mean with a smaller confidence interval. Inflation Rate, after an initial shock is expected to remain stable around 3 percent of the next couple of years. Of course, this wouldn't be my personal guess of the inflation rate over the next couple of years as the Federal Reserve continues to increase the interest rate in an effort to drive the Inflation rate low, preferably around 2 percent or lower. One also needs to consider that within this sample there is heavy influence coming from the post COVID inflation.

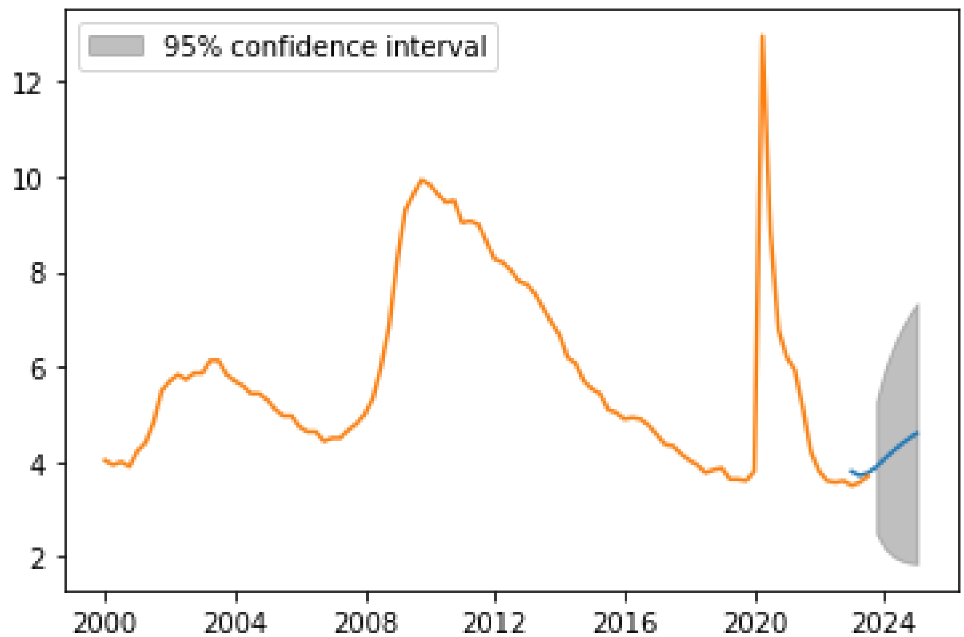

Figure 8: AR(*2*) Prediction extensions using the full dataset from 1960-2023 for the Unemployment Rate. The AR(2) predictions themselves only go from 2023 to 2025.

The seasonality of unemployment show itself in this smaller model. Even then, we see a slow return to mean, and a forecast that matches real life expectations with many expecting a rise in the unemployment rate. Still the confidence interval predicts either a decrease in the unemployment rate (very unlikely considering an increase in interest rates) or it staying around the same value/slightly increasing.

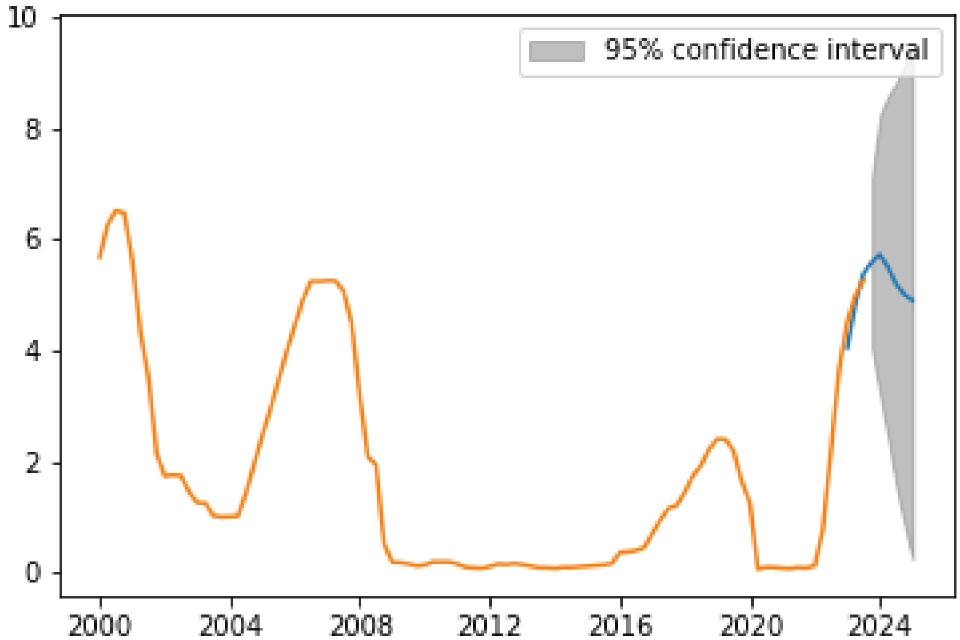

Figure 9: AR(8) Prediction extensions using the full dataset from 1960-2023 for the Federal Funds Rate. The AR(8) predictions themselves only go from 2023 to 2025.

Finally we see that the full dataset forecast for the Federal Funds Rate shows a decrease in the basis points, with a very large a confidence interval. Considering real ife expectations, I think we can only expect it to rise or stay the same. Otherwise, if there is a shock to the economy and unemployment rate rises or a recession happens, then the Federal Funds Rate might decrease.

5 Impulse Response Function and Variance Decomposition of Forecast Errors (1960-2009)

Next we will be viewing the Impulse Response Function dynamics between Inflation, Unemployment and the Federal Funds Rate. All the variables have been put into an AR(*4*) model in an effort to properly replicate the results of the Stock and Watson paper which used the same basis for each of the variables. We will also be seeing the Variance Decomposition of Forecast Errors and the dynamics between the variables in an AR(*4*) model.

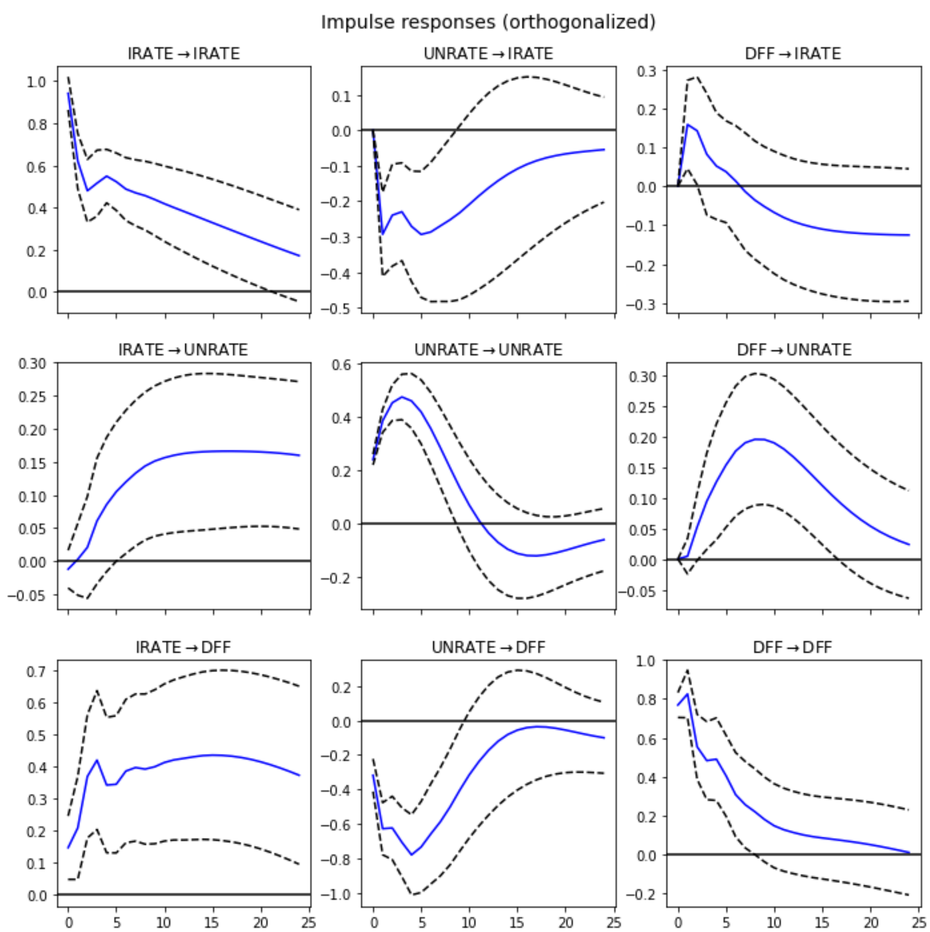

Figure 10: Impluse Response dynamics between IRATE (Inflation Rate), UNRATE (Unemployment Rate) and DFF (Federal Funds Rate) on data collected between 1960-2009.

With each x-axis tick representing a quarter and the y-axis representing the magnitude of that movement, we can see how much each variable responds to each other within the following periods. For example within the IRATE–¿DFF graph, we can expect to see a change in inflation to be meet with around a 30 percent increase in the Federal Funds Rate 5 quarters later. The IRF allows us to see the various shared dynamics between different values within the model, and the strength of response to each variable to another variable at a later or current date.

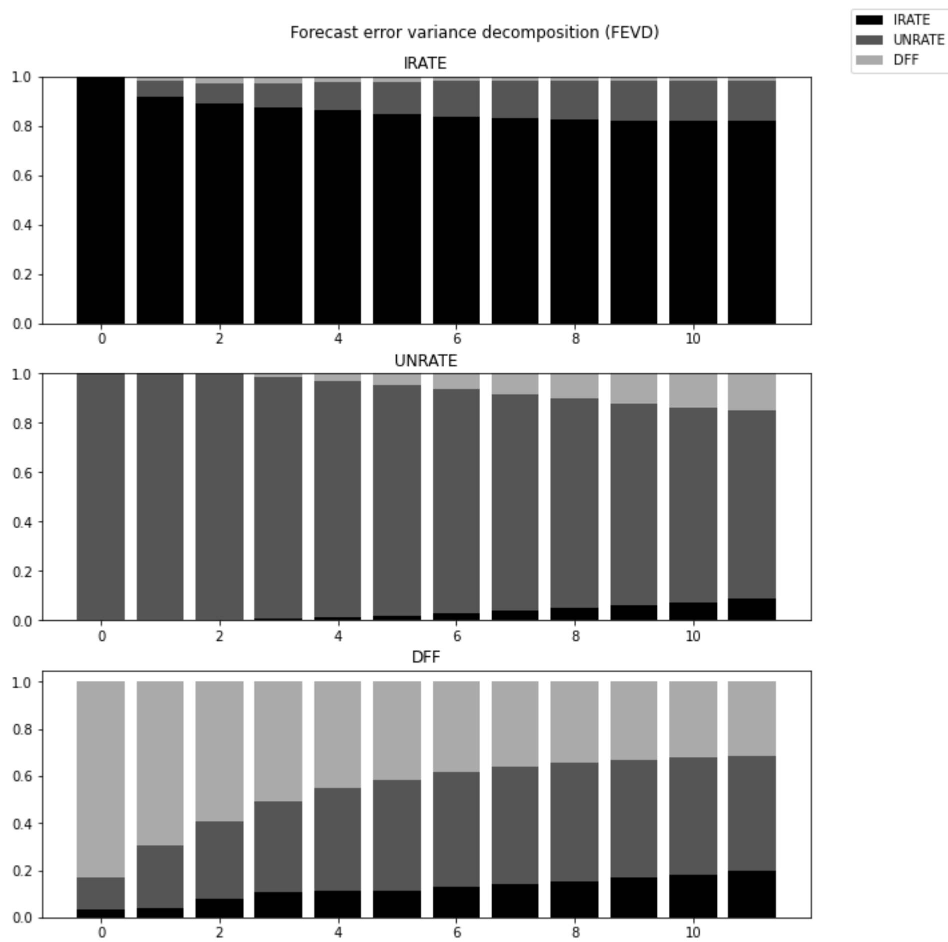

Figure 11: Forecast Error Variance Decomposition for IRATE (Inflation Rate), UNRATE (Unemployment Rate) and DFF (Federal Funds Rate) on data collected between 1960-2009. For actual values please reference Figure 12.

Variance Decompostions from Recursive VAR (For Training Data From 1960-2023) Ordered as π, u, R			
A. Variance Decompostion of π			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	***π***	***u***	***R***
1	100	0	0
4	88	9	2
8	83	15	2
12	82	16	2

*B. Variance Decompostion of **u***			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	***π***	***u***	***R***
1	1	99	0
4	1	97	2
8	4	88	8
12	9	76	15

*C. Variance Decomposition of **R***			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	***π***	***u***	***R***
1	3	14	83
4	11	38	51
8	14	50	36
12	20	49	31

Figure 12: A table illustrating the values from the plot in Figure 11.

Above you can see the values for the Forecaster Error Variance Decomposition (FEVD) plots, and their values at up to 12 quarters out. FEVD itself tells us the effect that each variable has upon itself and others, or in other words how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables. For example we can see Inflation is only affected by the Unemployment rate in terms of forecast error variance, and only by a relatively small amount (around 20 or so percent). Meanwhile Unemployment rate is affected by both the Federal Funds Rate and Inflation Rate, although both to relatively small degrees.

Finally we the Federal Funds Rate in which both the Unemployment and Inflation Rates have outsized influence on the forecast error variance. This can be explained by the fact that the Federal Funds Rate and its values are not random in the sense as to why the are raised or lowered. There is a conscious decision made to increase or decrease the rate depending on macroeconomic conditions which happen to indlude the Inflation and Unemployment rate.

6 Impulse Response Function and Variance Decomposition of Forecast Errors (1960-2023).

We will now be training the IRF on the full sample, and examine the differences between the graphs from the previous section.

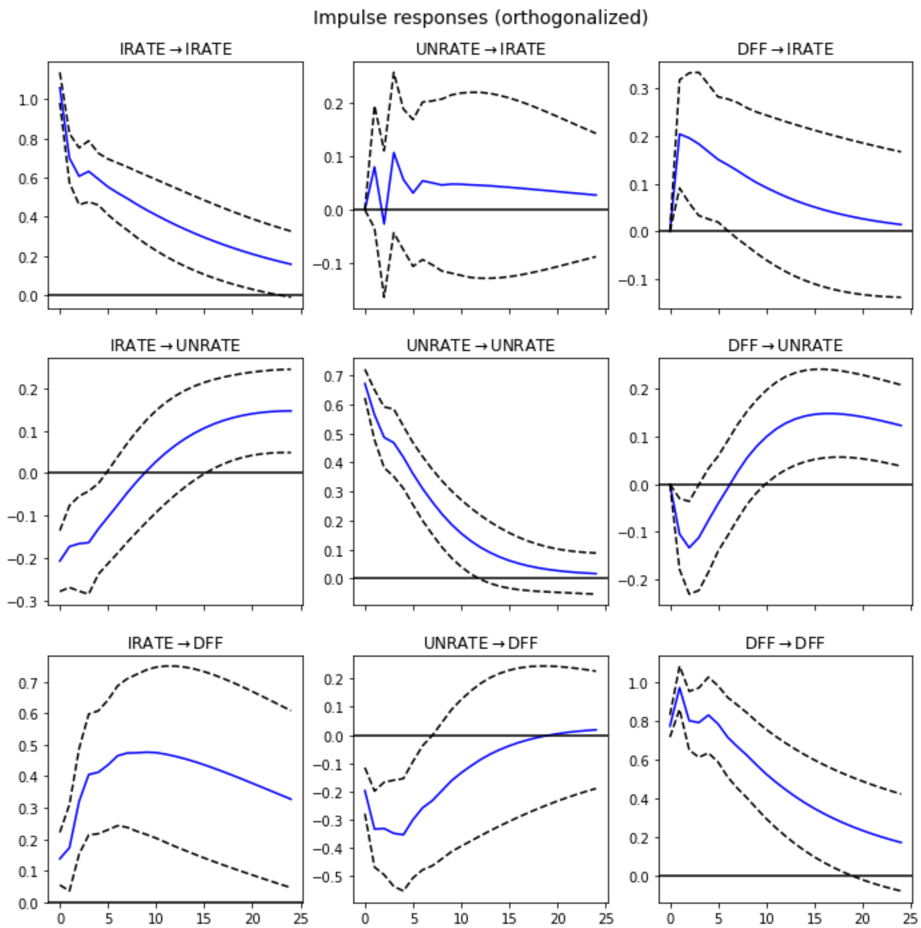

Figure 13: Impluse Response dynamics between IRATE (Inflation Rate), UNRATE (Unemployment Rate) and DFF (Federal Funds Rate) on data collected between 1960-2023.

Most of the graphs remain the same as compared to the initial IRF we ran on the training data. Yet there is one substantial change, with the association between Unemployment the Inflation Rate going from around a -30 percent shock to Inflation as Unemployment increased (something which is supported by the Phillips Curve) to there being very little to no shared dynamics. This might be another result of the sharp changes we see in both Unemployment and Inflation due to COVID.

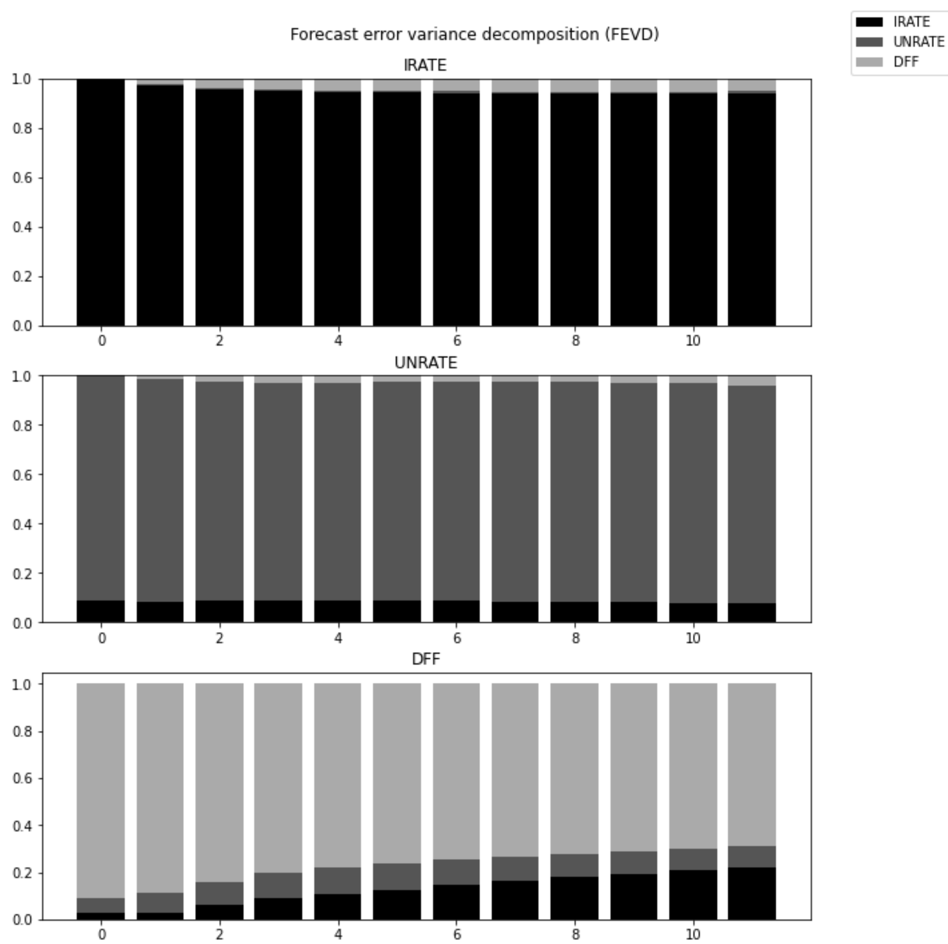

Figure 14: Forecast Error Variance Decomposition for IRATE (Inflation Rate), UNRATE (Unemployment Rate) and DFF (Federal Funds Rate) on data collected between 1960-2023. For actual values please reference Figure 12.

Variance Decompostions from Recursive VAR (For Training Data From 1960-2023) Ordered as π, u, R			
A. Variance Decompostion of π			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	π	u	R
1	100	0	0
4	95	1	4
8	94	0	5
12	94	1	5

B. Variance Decompostion of u			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	π	u	R
1	9	91	0
4	9	88	3
8	8	89	3
12	8	88	4

C. Variance Decomposition of R			
	Variance Decompostion (Percentage Points)		
Forecast Horizon	π	u	R
1	3	6	91
4	9	11	80
8	16	11	73
12	22	9	69

Figure 15: A table illustrating the values from the plot in Figure 14.

The variables also show weaker associations with each other when related to forecast error variance, especially the Federal Funds Rate. This is confirmed by the values we see in the table above. Once again this may be a result of the shocks related to COVID and how values of Inflation and Unemployment went from extreme to extreme, much before the Fed could react in an appropriate manner.

7 Granger Causality Tables

Finally we come to the Granger Causality Tables. It's not a strict causality test, but it looks at a time series and attempts to find causality via sharp changes in time series causing similar changes in the another at a later date. I.e. variable 'x' can 'Granger-cause' another variable 'y', 'z' periods after the initial shock in 'x'.

Granger Causality Tests (Training Data 1960-2009)			
	Dependent Variable in Regression		
Regressor	***π***	***u***	***R***
π	0	0.91	**2.39***
u	**2.721***	0	**6.72***
R	1.6	**2.473***	0

*Values in bold are statiscially significant

Figure 16: A Granger Causality Table for data from 1960-2009 showing the resulting t-test values. It tests Inflation Rate, Unemployment Rate and the Federal Funds Rate in that order.

As we can see in this initial table Inflation 'Granger-causes' the Federal Funds Rate and Unemployment also 'Granger-causes' the Federal Funds Rate. Unemployment also 'Granger-causes' inflation within this training data. Finally we see the Federal Funds Rate has a reverse causality with Unemployment. We once gain receive values that are in line with what we expect macro-economically. The reverse causality we see wtih unemployment and the Federal Funds Rate, as well as the links between inflation and the Federal Funds Rate, inflation and unemployment, and so on are explained by common trends we see such as the Phillips Curve and the Taylor Rule and so on.

Granger Causality Tests (Full Dataset 1960-2023)			
	Dependent Variable in Regression		
Regressor	***π***	***u***	***R***
π	0	0.3	**3.44***
u	1.84	0	0.97
R	**2.57***	**2.94***	0

*Values in bold are statiscially significant

Figure 17: A Granger Causality Table for data from 1960-2023 showing the resulting t-test values. It tests Inflation Rate, Unemployment Rate and the Federal Funds Rate in that order.

In the full sample Granger Causality table we see some changes such as the Federal Funds Rate now 'Granger-causing' inflation, and unemployment no longer 'Granger-causing' the Federal Funds Rate. We see the causalities acting stranger here once again because of the COVID shock, as the Federal Funds Rate should be a response to the unemployment and inflation rate, not the other way around. There needs to be more of a normalization period post COVID in order for the data to return to what we had seen in the previous 50 years, especially as we expect to see the Federal Funds Rate respond to changes in macroeconomc conditions, and not the other way around.