Rowan Morkner
Jennings
ECON 431W
19 December, 2024

MACROECONOMIC FORECASTING: A CRITIQUE

Introduction:

Macroeconomic forecasting, and specifically inflation forecasting gives policy makers, firms, and individuals essential insight to inform their decisions. Without the ability to anticipate the future effects of our current actions, our capability to manage risk, make informed investments, and plan for various contingencies is significantly diminished. Reliable forecasting is crucial; without it, policymakers are like pilots trying to navigate through heavy fog without a clear flight plan. They can see their instruments, but if they fail to understand how the readings interact or predict the effects of their actions, they risk steering the economy off course. Economic forecasting serves as the policymaker's map, helping them interpret critical variables, chart a trajectory, and make informed adjustments to reach the desired economic outcomes, while accounting for both intended and unintended effects.

This paper explores the history and evolution of macroeconomic forecasting through the lens of inflation forecasting. I will go over the historically debated transmission mechanisms, causes of inflation, benefits and risks of inflation, and the reasons why inflation forecasting is important. Next I will review the Lucas Critique and the impact of this seminal paper on forecasting, and the Phillips curve. Additionally I will review Furman 2022, which analyzes the failures of the Phillips curve to predict the inflationary spike of 2021. Then I will explore the data set which I will be using for my analysis and estimate the Phillips curve, showing how the relationships have changed over time. Finally I will explore some of the modern techniques used for forecasting, including DSGE and machine learning models. Ultimately, I will examine the inherent challenges and limitations of attempting to predict the future with reliability, while acknowledging the usefulness of forecasting as a tool despite its shortcomings.

Thesis statement:

The evolution of inflation forecasting reflects significant shifts in economic theory, driven by unexpected events and limitations of prevailing models, such as the breakdown of the Phillips Curve and the challenges highlighted by the Lucas Critique. While advancements like

DSGE models and machine learning have improved predictive methods, this paper argues that the complexity of economic systems and unforeseen shocks render forecasting inherently unreliable, questioning its ultimate accuracy and utility.

Brief history of inflation forecasting:

This section utilizes OpenAI's ChatGPT for preliminary research and drafting support, ensuring alignment with APA ethical standards.

The origins of inflation have been debated among economists for centuries. During the 20th century, theories of inflation changed significantly in response to shifting economic realities and empirical evidence. Early Keynesian thought, which dominated in the first half of the century, emphasized demand driven inflation, advocating for active fiscal and monetary policies to stabilize prices and manage aggregate demand. The development of the relationship between inflation and unemployment was first examined by A.W. Phillips in his landmark 1958 paper. He established an inverse relationship between inflation and unemployment while looking at data from the 1860s- 1950s in the UK. This relationship which he described quickly became the foundation for a significant portion of the analysis around inflation. In the 70s, the monetarist school, led by Friedman, challenged the Keynesian view, arguing that inflation is primarily a monetary phenomenon, tied to excessive money supply growth, with long term neutrality of money. The stagflation of the period further spurred debates, as keynesian models struggled to explain high inflation and unemployment simultaneously. This led to the rise of neoclassical theories emphasizing rational expectations and policy credibility, alongside critiques of traditional econometric models like the Lucas Critique which directly pointed to the flaws in the prior methods for estimating the Phillips curve. By the late 20th century, New Keynesians integrated insights from both Keynesian and Neoclassical thought, emphasizing the roles of sticky prices, expectations, and supply shocks in inflation dynamics, while advocating for inflation targeting as a central policy tool. Together, these shifts reflect ongoing refinement of inflation theory driven by historical challenges and advancements in econometric modeling. (OpenAI, 2024).

After the 2008 financial crisis, the FED took drastic measures in an effort to stabilize the economy. Through open market operations and new quantitative easing measures, they expanded their balance sheet by more than 150% between July 2008 and July 2010. Notably during this period, inflation remained low, despite the dramatic increase in high powered money. This caused many economists to question the link between money supply and inflation with some economists even believing that the relationship had fundamentally changed. Due to this result, there was a belief that QE and rapid expansion of the money supply was a valid tool to stabilize output without any major consequences such as inflation. This was the belief in 2020 when the FED again responded to the crisis by creating an alphabet soup of lending programs, buying treasuries, cutting interest rates to near zero, and expanding the balance sheet by another 116%. In an interview with FED chair Powell in April 2020, he was asked if the expansion might be inflationary to which he responded 'if it didn't happen in 2008 it probably won't happen this time.' One year later CPI inflation was at nearly 5% and hit its peak of 8.5% the following year. This failure in forecasting is just one recent example which shows how difficult it is to accurately forecast the effects of monetary and fiscal policy, especially in the face of novel challenges. The challenge of knowing when to adapt our models, make exceptions, or draw new connections between economic variables can be challenging and failing to do so correctly can result in large economic consequences. To explore this evolution we will examine the Lucas Critique, as well as a more contemporary paper by Jason Furman. Our understanding of forecasting and its shortcomings have been greatly advanced by Robert Lucas' paper which became known as the Lucas critique. Additionally, more recent analysis in Furman 2021, has helped us understand some of the reasons for the discrepancies between forecasted inflation levels and the observed increase in prices post pandemic.

<u>The Lucas Critique (1979)</u>

Research Question:

The Lucas Critique is a seminal paper which criticized the contemporary methods of econometric forecasting. In the paper, Lucas questions if econometric time series models can reliably evaluate and forecast the long term effects of policy changes, when the models do not take into account the changes in people's behaviors based on those policies. The critique digs into the fact that these econometric models fail to incorporate how agents adapt to policy changes, which in turn will lead to meaningless or unreliable long term predictions.

Main Contributions & Theoretical Framework/ Hypotheses:

The paper introduces what is now known as the Lucas Critique, which has been very influential and has fundamentally changed how economists evaluate policy changes. The insight that the relationship between economic variables, such as inflation and unemployment in the case of the Phillips curve, may not remain stable over time due to policy changes, was unique at the time. Previous models often treated policy interventions as if they had no effect on people's expectations, behaviors, or the underlying structure of the economy. Additionally Lucas's focus on the importance of rational expectations as opposed to simply adaptive expectations, laid the groundwork for more realistic models that could better predict the response to policy changes. In his critiques, Lucas argues that the traditional econometric models fail to capture how individuals adjust their expectations and actions in response to anticipated policy changes which results in structural instability in these models and undermines their effectiveness for policy evaluation. Moreover, the paper spurred a shift towards rule based policy making such as inflation targeting, rather than a discretionary policy approach. Prior to Lucas's 1976 paper, economists typically used fixed parameter econometric models for forecasting and policy evaluation, assuming that the relationships between variables were stable over time, and would not change when policies changed. Lucas's approach made economists rethink the foundation of their econometric models, making the Lucas Critique one of the most important and influential critiques in macro theory. (Henderson, 2023).

Methodology:

First, Lucas plainly describes the theoretical framework for econometric policy evaluation which was used at the time. That being an economy at time: t+1 being described by the model:

$$Y_{t+1} = F(Y_t, X_t, \theta, \epsilon_t)$$

Where F is the specified functional form of the model, assumed to capture stable economic relationships. Here, Y_t are the "state" or endogenous variables, X_t are the "forcing" variables which are the exogenous factors that influence the economy (such as government policies), ϵ_t captures random shocks to the economy, and θ are the parameters which describe the relationships among variables.

He then examines adaptive forecasting, a method which updates θ dynamically, but argues that while this approach may improve short-term accuracy, it results in "infinite variance of the long-term operating characteristics of the system" (Lucas, 1976, p. 24), making reliable long-term forecasts impossible.

Finally, he supports his hypothesis by examining "specific decision problems underlying the major components of aggregative models." (ibid, p. 26), such as the consumption function, taxation and investment demand, and the Phillips Curve. He demonstrates that these issues are pervasive, affecting the reliability of econometric models for long-term policy evaluation.

Findings & Implications

Lucas suggests that instead of assuming policy changes have no long-term effects, we should model them as stochastic functions of the economy's current state. He allows θ to be dependant on λ, where λ is a fixed parameter that governs the policy setting and other external disturbances impacting the economy. Allowing for policy changes to systematically influence the parameter θ, giving the model the ability to account for how policy shifts alter the behavior of individuals and firms over time.

Lucas then argues for policy rules over discretionary actions, as a more predictable policy has greater consistency in its effects if it has predetermined rules rather than random, ad hoc decisions. He acknowledges that it is possible that delegating economic decision making to some group or individual could lead to "superior (by some criterion) economic performance than is attainable" (ibid, p.41) under a rules based system. Although he argues that even if this is true, policy evaluation is only possible under a rules based system and it only makes sense to compare the effects of different policy rules rather than arbitrary, authority-based decisions.

I do not think I could recreate his results.

Strengths & Limitations

This paper was a very powerful critique of the econometric policy evaluation status quo at the time of publishing and has had lasting effects on the methods and tools economists use to this day. The development of the dynamic stochastic general equilibrium or DGSE models which have "become one of the workhorses of monetary policy analysis in central banks" (Fernandes et al. 2016) is largely due to the influence of the Lucas Critique. Although the paper is highly compelling and demonstrates its propositions and conclusions clearly, there is evidence that certain models don't experience significant instability in their parameters given policy changes (Goutsmedt, 2017). This suggests that the application of the Lucas Critique is contextual and not necessarily applicable to every economic forecasting model.

<u>Why Did (almost) No One See Inflation Coming (Furman 2021)</u>

Research Question:

Furman's 2022 paper explores the many shortcomings of inflation forecasting during and after the COVID-19 pandemic. He asks, why did inflation forecasts from late 2020 and early

2021 significantly underestimate actual inflation, and what factors contributed to this discrepancy?

Main Contributions:

Furman adds to the literature by examining forecaster's underestimation of the Phillips curve relationship between inflation and unemployment. He critiques the limitations of relying on historical estimates of the Phillips curve slope, showing that low unemployment alone cannot explain the post-pandemic rise in inflation.

Theoretical Framework or Hypotheses:

The main relevant theory in this paper is that of the Phillips curve which relates unemployment to inflation. This general relationship can be shown through the following equation:

Inflation = expected inflation - θ*(unemployment - natural rate of unemployment) + error term

In this equation, expected inflation represents the long term forecasted level as well as an individual's perception of future inflation based on survey data. The estimated coefficient θ represents the slope or sensitivity of inflation to changes in unemployment. The term (unemployment - natural rate of unemployment) captures the difference between the theoretical long term "natural" rate at which the economy is balanced and the actual unemployment rate. Finally, the error term represents any unforeseen stochastic or exogenous factors or shocks that may influence inflation outside of the included variables.

This theory was developed by A.W. Phillips in his 1958 paper which looked back on nearly 100 years of data and observed an inverse relationship between unemployment and inflation in the UK. Since the publication of this seminal paper, this relationship has flattened (θ is near zero), meaning the tradeoff between the two is less pronounced than in the past.

Furman's main hypothesis is that policymakers should avoid relying on pre-pandemic statistical relationships, particularly the Phillips curve with anchored expectations, because it failed to anticipate the inflationary surge of 2021-22. The Phillips curve's limitations in predicting high inflation underscore the need to reassess models to account for factors beyond traditional labor market dynamics.

Methodology:

In his analysis, Furman uses a quantitative approach to highlight his hypothesis. He draws on forecasts of inflation and unemployment from institutions such as the Fed, CBO, and

Survey of Professional Forecasters. These forecasts assume stable relationships, such as the Phillips curve with anchored inflation expectations. His primary method is to compare these predicted values to actual observed data from 2021, highlighting the discrepancies between the forecasted and real inflation rates.

To evaluate the accuracy of these stable models, Furman considered both “normal” and “low” multipliers which are used to predict the effects of fiscal and monetary stimulus during the pandemic. By modeling expected changes in GDP and unemployment under these different scenarios, he shows that, even with extreme assumptions such as an impossibly low 1% unemployment rate, these models would only have predicted moderate inflation still below 3%. This prediction falls far short of the actual rates observed in 2021. Additionally, he examines a number of alternative hypotheses wherein he dissects the Phillips curve and offers modifications or possible explanations for why the Phillips curve got it so wrong. Lastly, he compares the post pandemic differences and similarities in responses and key macro variables between the US and Euro area. He shows that, despite differences in policy and GDP growth, both nonetheless experienced a similar unexpected shock to inflation levels.

Findings & Implications

Furman’s analysis highlights critical gaps in traditional economic models, particularly the Phillips curve, when predicting inflation during and after the COVID-19 pandemic. His findings reveal that the standard models, even under extreme assumption, failed to capture the inflation spike seen in 2021. The failure of these models demonstrates that the pre-pandemic models with linear, relatively flat assumptions about the shape of the Phillips curve as well as anchored long run expectations are insufficient for current conditions.

The implications are substantial for this finding. Furman suggests that policymakers need to reconsider their reliance on these assumptions and models, as they may no longer hold in economies with unprecedented fiscal and monetary stimulus such as those of the United States and Europe after the pandemic. Alternative models that focus on the nominal rather than real demand, bypassing the labor market intermediation, provide a better fit for the post-pandemic inflation we see. This means that policymakers should allow for more uncertainty in their predictions, as well as allow for more flexible, adaptable forecasting methods that incorporate broader economic variables and changing relationships, rather than leaning so heavily on historical models and relations that can change over time.

Strengths & Limitations

I would say this paper is very strong in its analysis and use of quantitative data, forecasts, and theoretical relationships to examine the shortcomings of the Phillips curve. The use of alternative hypotheses and explanations supports Furman's hypothesis by being exhaustive in his attempts to salvage the historic models and their relationships, pushing them to the reasonable limit of their assumptions and estimated values, only to show they are nonetheless incapable of explaining in full the extent of the observed inflation.

I don't see any major limitations to this paper currently, even his prediction at the end of the paper suggesting that "keeping inflation to 3%, especially in the United States, will be a challenge" seems to be a fairly accurate prediction, given that this paper was written in 2022.

Data paper:

Gathering and Cleaning

I am using a data set from the FED-QD set. The data contains observations of 127 variables observed on a quarterly basis from 09/01/1959 until 09/01/2024, for a total of 262 quarters of data. I did not need all these variables, so I cut it down to just the relevant variables I need for my analysis.

Modification and de-indexing.

Many of the variables I wanted to use, such as CPI and PCE are indexed variables; in order to extract information about inflation or growth rates, I had to create secondary variables which took the YoY growth rate of the variables. Here you can see the example of the conversion of key inflation variables from their indexed values to their YoY percentage change values.
In figure 1 you can see the variables graphed before de-indexing, and extracting the YoY change. Figure 2 shows three measures of inflation after extracting % change. The Great Inflation of the 70s is visible from the graph, as well as the recent inflationary spike post pandemic in 2021.

Figure 1: Inflation Measures Over Time

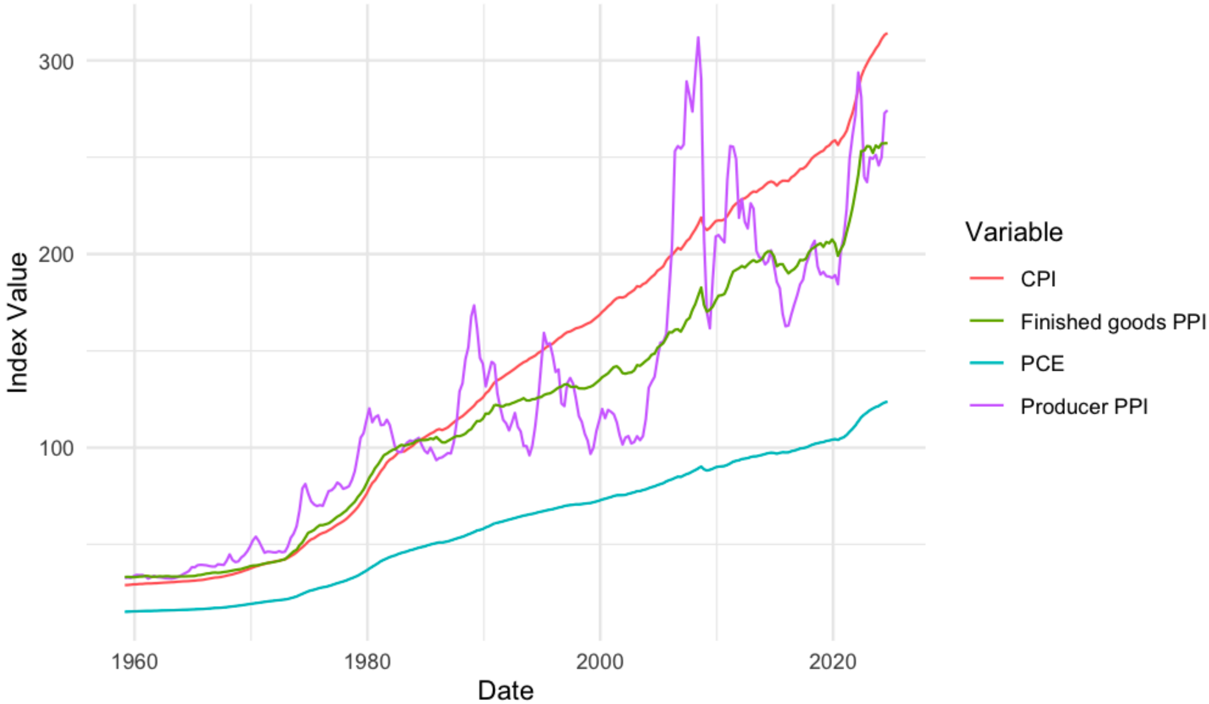
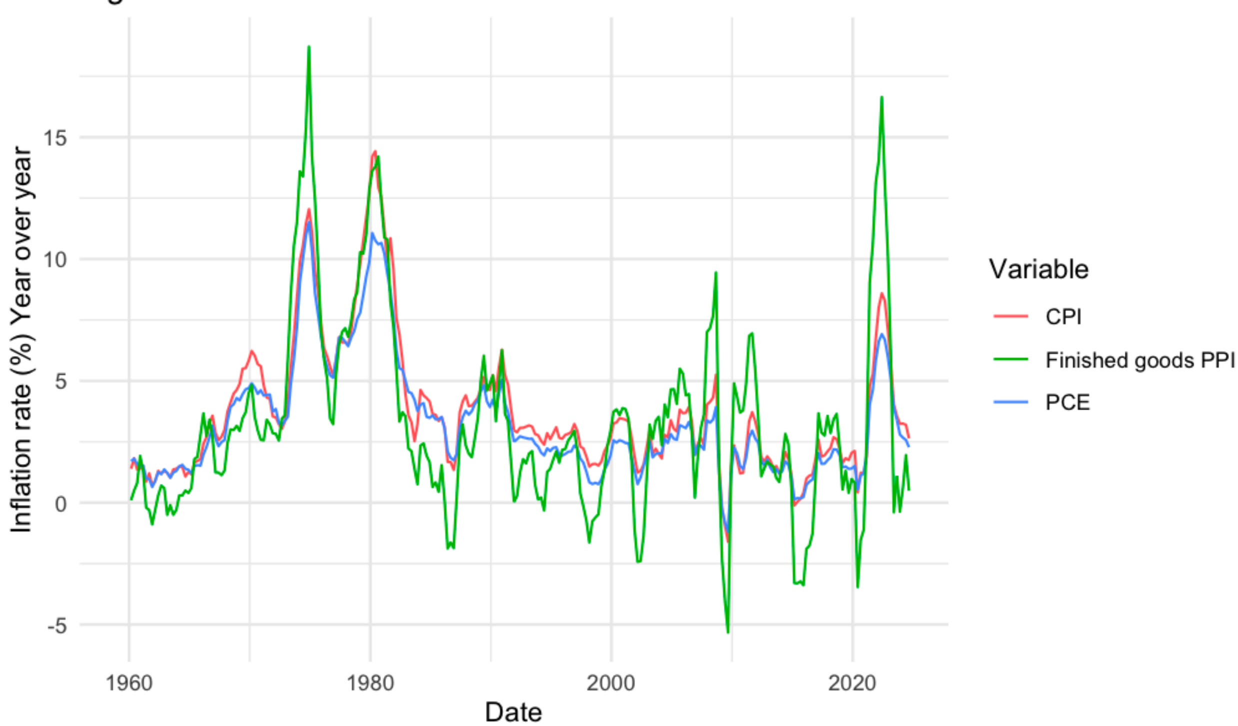
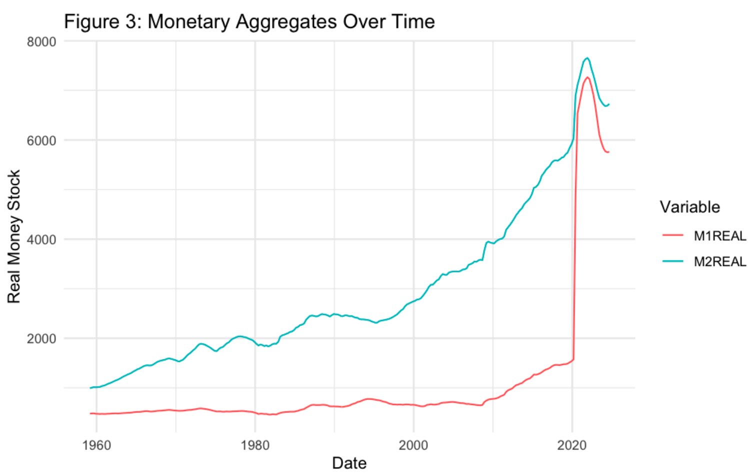
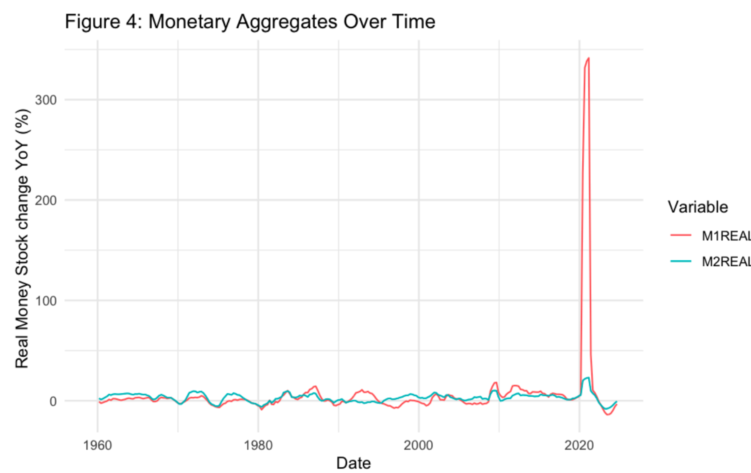


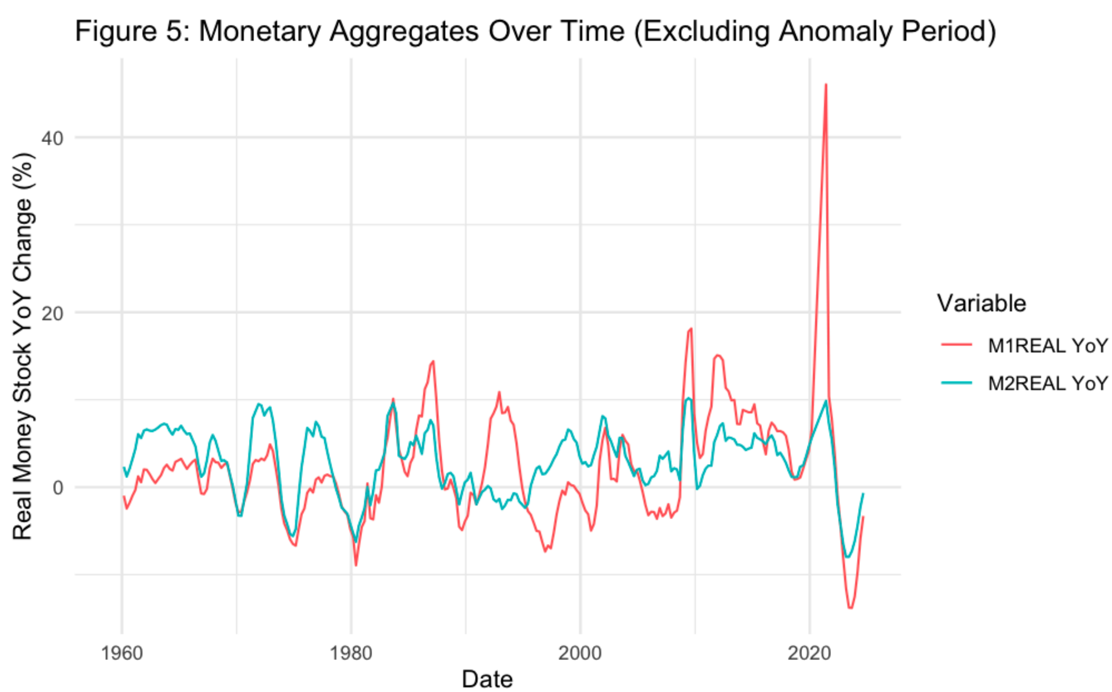
Figure 2: Inflation Measures Over Time



Another interesting case I looked at was M1. Due to the change in how M1 was measured in 2020, when I graph the YoY change in M1, the scale is completely messed up and hard to extract insight from. Below you can see the three versions of the data plotted, one without YoY change (Fig. 3), one YoY change with the anomaly period included (Fig. 4), and one with the period excluded (Fig. 5).



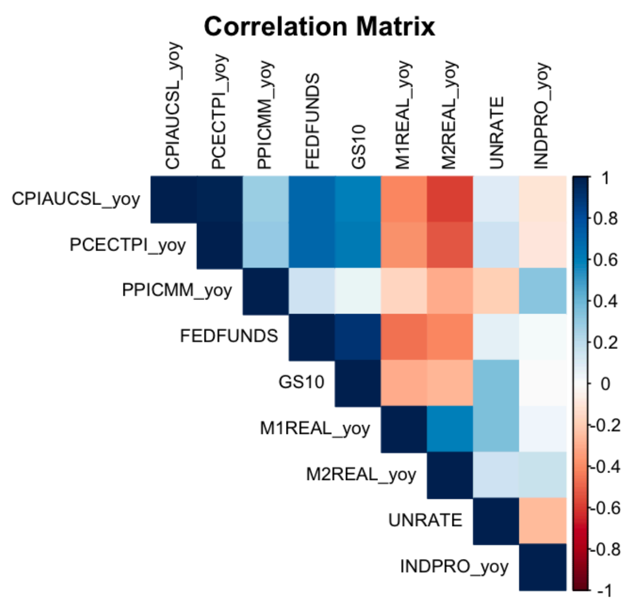




In order to perform rigorous econometric analysis or train machine learning models, corrections like this have to be made or adjusted for, otherwise the models will be influenced by and capture relationships between variables that do not exist.

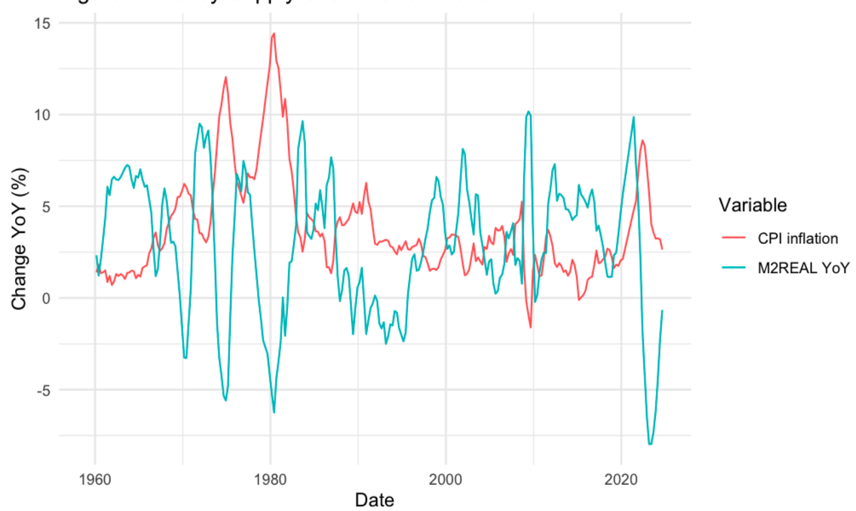
Visualizing Interesting relationships.

Next I created a correlation matrix to visualize which variables were correlated with each other:



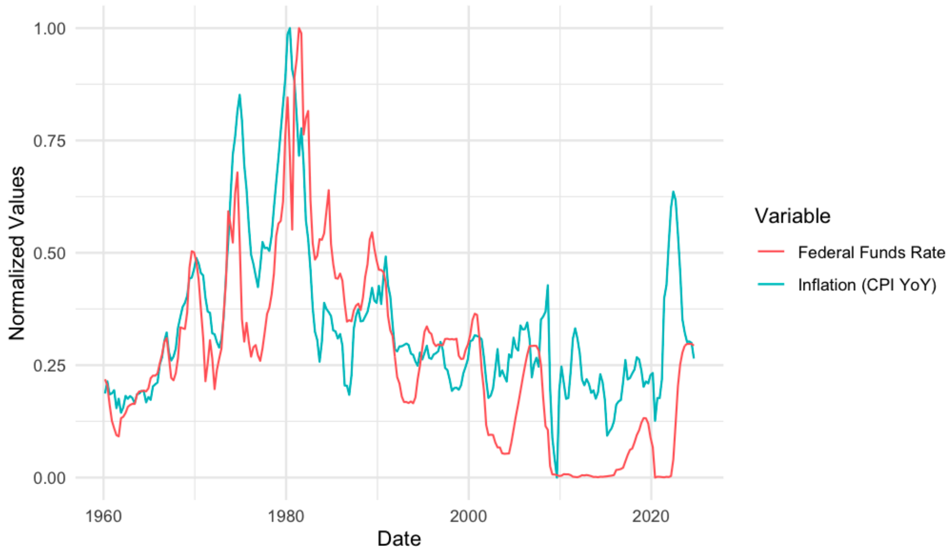
From the correlation matrix we can see that the M2 and M1 growth rate have a strong negative correlation with Inflation. I thought this was interesting because I assumed the effect would be opposite: years in which money supply grows significantly, would cause inflation to rise as well. This made me wonder if the relationship I am seeing is caused by the monetary policy response to recessions or inflationary periods. During recessions, when inflation is low, the FED boosts money supply, and during periods of high inflation, the FED cuts money supply. Then, due to somewhat slow economic transmission mechanisms, inflation lags behind M2 growth or cuts. Alternatively this could be due to the fact that fluctuations in M2 are affected by savings rates, and safe, low interest rate investment activity such as CDs and money market funds. When inflation is high, the real returns on these assets fall, prompting savers to pull their money out of these illiquid and low return assets in favor of higher growth assets which can better make up for inflation. To investigate further I graphed M2 against CPI to confirm my hypothesis. The results shown in figure 7, although expected, are still a little surprising.

Figure 7: Money Supply Growth and Inflation

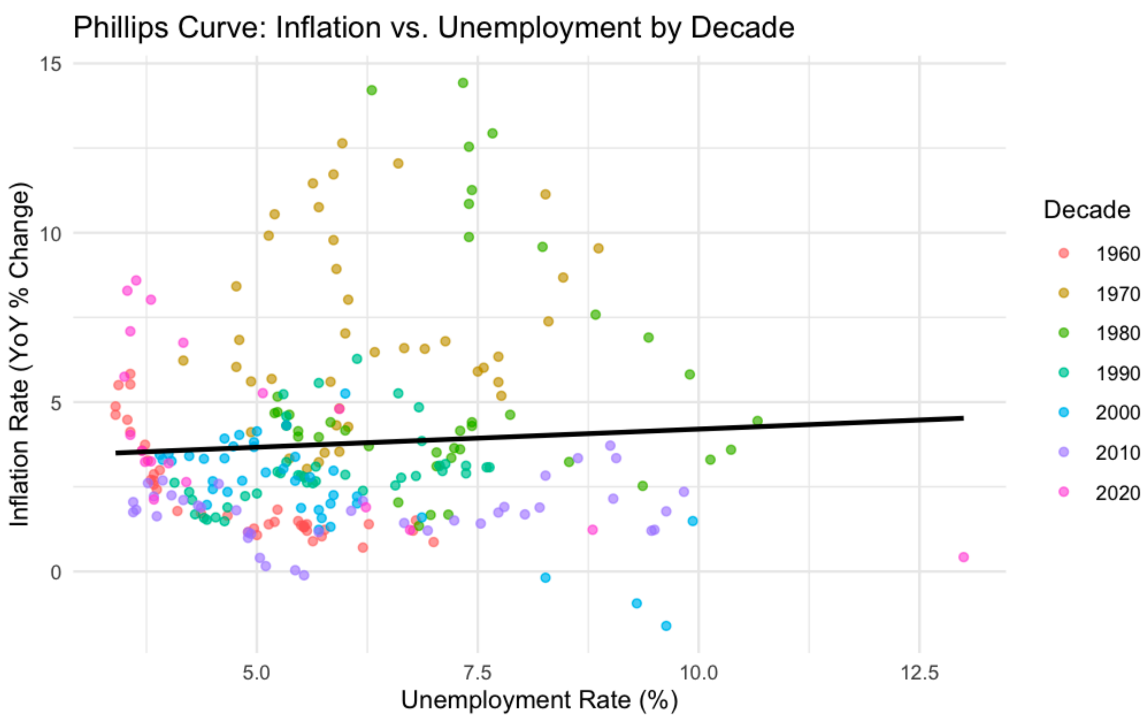


Additionally, I wanted to look at how the federal funds rate is related to inflation rates, so I normalized both the federal funds rate and CPI inflation so they can be graphed together and compared. We can see from the correlation matrix as well as figure 8 that the federal funds rate and inflation have a medium to strong positive correlation.

Figure 8: Normalized Inflation and Federal Funds Rate Over Time



Finally, I wanted to plot inflation versus money supply growth and examine the relationship described in the Phillips curve. Each different color represents a different decade, showing how the relationship Phillips observed may have changed over time. Notably if we exclude the 70s, there is a clustering which appears to be somewhat of a downwards sloping concave curve towards the bottom left of the graph. But as a whole, as shown by the trend line, there seems to be no strong negative relationship between the two as is suggested by the Phillips curve. This may be due to a misstep in my data conversion, or a poor understanding on my part of what the Phillips curve really is, but as it stands, it looks as if the relationship is not really there.



Examining Formulations of the Phillips Curve

For my econometrics class this semester I performed an in depth regression analysis and comparison between 3 different specifications of the Philllips curve, showing how its evolution and increased complexity in specification has improved over time. If you would like to read that paper for a deep dive specifically into the Phillips curve, follow the LINK. Otherwise I have included in this next section my model specifications, findings, and conclusions, pulled directly from that paper.

Models and Methods

Baseline Phillips Curve Equation (Samuelson & Solow 1960):

This is the most simple model for estimating the Phillips curve as it relates to price inflation. The model is specified as follows:

$$\pi_t = \beta_0 - \beta_1 U_t + \epsilon_t$$

where:

π_t = CPI_YOY$_t$: The observed percentage price change from a year ago, at time t.

U_t = UNRATE$_t$: The observed unemployment rate ate time T

The expected sign of β_1 is negative according to the relationship posited by Phillips and Samuelson & Solow, but this relationship has shifted over the decades, and recent estimates from Hazell et al. (2022) suggest it remains relatively flat.

Alternative Specifications:

Ball et al.

The second specification incorporates inflation expectations as well as the slack expectations and is copied directly from Ball et al. The specification is as follows:

$$\pi_t - \pi_t^e = \beta_0 - \beta_1 \tilde{u}_t + \epsilon_t$$

Where:

π_t^e = INFCPI10YR$_t$: $\pi_t - \pi_t^e$ = INFDIFF10$_t$ The difference between observed inflation and the 10 ahead year inflation prediction taken from the Survey of Professional Forecasters.

$\tilde{u}_t$ = SLACK$_t$: Difference between unemployment and the long-run, non cyclical rate of unemployment.

From now on, to simplify model specification and regression results , $\pi_t - \pi_t^e$ will be displayed simply as Δ_t.

The expected sign of β_1 is negative because when the employment gap ($\tilde{u}_t$) is positive, meaning unemployment is above its natural rate, it indicates slack in the labor market. This slack reduces inflationary pressures because firms face lower wage growth and cost pressures. Alternatively when $\tilde{u}_t$ is negative (unemployment is below its natural rate), tighter labor market conditions lead to upward wage pressures and higher inflation.

Lagged model.

The final specification is a lagged model which is based on the previous model but incorporates lags to the slack and inflation variables to capture lagged effects on inflation. The basic specification is as follows:

$$\Delta_t = \beta_0 - \beta_1 \tilde{u}_t - \beta_2 \tilde{u}_{t-1} + \beta_3 \Delta_{t-1} + \beta_4 \Delta_{t-2} + \epsilon_t$$

Where:

Δ_t = INFDIFF10

$\tilde{u}_t$ = SLACK

Including lags in this model is important because they account for the delayed effects of labor market slack (L1_SLACK) and capture inflation persistence through past values of inflation (L1_INFDIFF and L2_INFDIFF). This helps reflect the dynamic nature of inflationary pressures and improves the model's explanatory power. The expected signs of β_1 and β_2 are negative for the same reasons as described for the previous model. With the additional lags capturing the distributed, delayed effect labor market slack has on inflation at time t. The expected signs of β_3 and β_4 are positive because inflation in previous periods is likely positively correlated with observed inflation at time t.

Findings:

My regression analysis explores the relationship between unemployment rates and inflation through multiple model iterations. The initial baseline model revealed a weak and statistically insignificant positive relationship, with an almost negligible explanatory power (adj-R^2=0.0003), suggesting model bias from omitted variables. The scatter plot of which was displayed previously. Incorporating additional variables in the Ball et al. model improved the results, yielding a significant negative coefficient for unemployment and a higher adjusted adj-R^2=0.1269, although issues with heteroskedasticity and serial correlation persisted. Applying a Generalized Least Squares (GLS) approach dampened serial correlation, enhancing the Durbin-Watson statistic to 1.55 and maintaining a significant negative relationship, with a reduced adj-R^2=0.0907. The final lagged model introduced dynamic elements, substantially increasing the adjusted adj-R^2 to 0.7862 and demonstrates a strong fit with significant coefficients for both current and lagged variables. Diagnostic tests confirmed improvements in autocorrelation and multicollinearity, though some concerns with heteroskedasticity and residual normality remained, primarily due to structural economic events.

Conclusion of Regression:

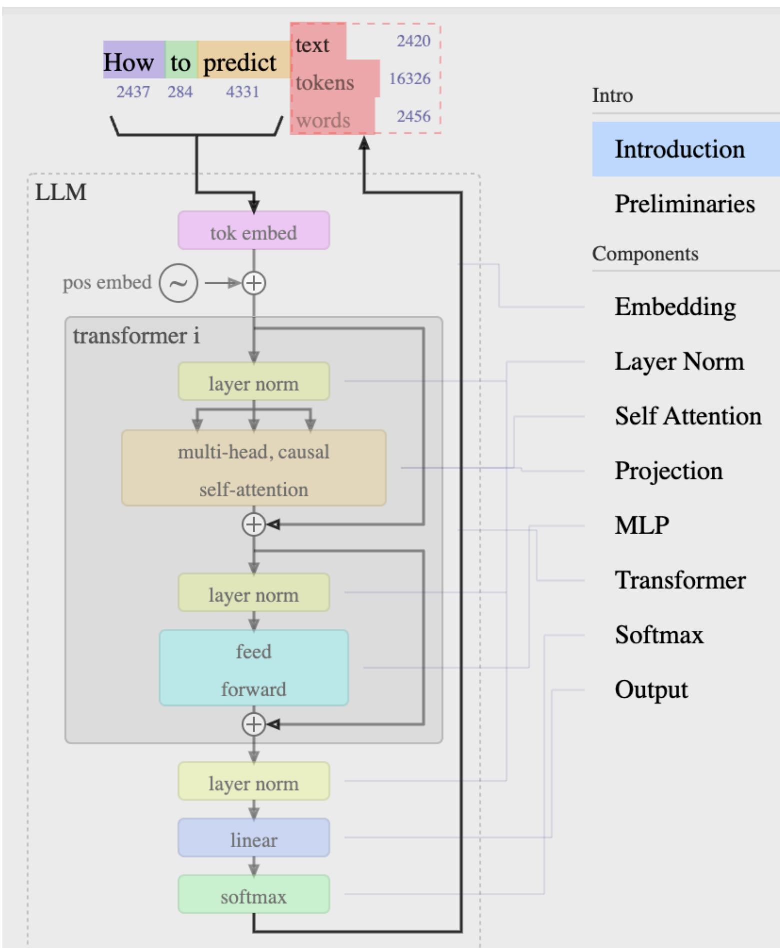
The step-by-step improvement of the regression models made them better aligned with economic theories and more reliable. We started with a simple model that didn't explain much, then added more factors and fixed statistical problems in each new version. This process led to a final model that effectively shows how unemployment and inflation influence each other over time. Although the last model works well and meets most standard assumptions, there are still small issues with data variability and distribution that could be improved in the future. Overall, the final model successfully explains the relationship between unemployment and inflation, highlighting how economic factors interact and change over time.

Modern techniques:

Over the past five decades, advancements in econometric modeling, influenced by critiques like Robert Lucas's 1976 paper, have led to the development of new forecasting methods. Dynamic Stochastic General Equilibrium (DSGE) models emerged as a response to the limitations highlighted by Lucas, emphasizing the importance of microeconomic foundations in macroeconomic analysis. In recent years, particularly following the 2017 publication of "Attention Is All You Need," which introduced the Transformer architecture, economists have begun exploring machine learning and neural network models to capture complex, nonlinear economic relationships.

DSGE models gained prominence in the 1980s with significant contributions from economists such as Finn Kydland and Edward Prescott, who developed Real Business Cycle (RBC) models within the neoclassical framework. These models serve to explain historical time-series data and are utilized for forecasting and policy analysis. These large, multivariate models are generally linearized and are built on a foundation of micro theory which allows them to be self consistent with the optimizing behavior of individual agents, market equilibrium, and rational expectations. The term ‘DSGE’ refers to a broad class of models which encompass many areas of macroeconomic forecasting. Their common feature is that they incorporate decision rules derived from “solving intertemporal optimization problems” (Negro, 2013) which means the agents are assumed to make decisions by looking at the trade-offs between today and the future. They can be tweaked to compare the assumption of frictionless acquisition of information with frictions which result in information asymmetries. They are designed to take into account rational expectations in their decision making and can be modified to add inputs such as long term inflation and interest forecasts which help ground the models and adjust for future changes. DSGE forecast estimation is generally done with Bayseian estimation methods which allow the models to project a probability density function into the future rather than simply a point estimate. The projected error bands give policy makers valuable information about the uncertainty in outcomes resulting from a given policy decision. Overall these models are greatly favored for policy forecasting because they address the Lucas Critique by implementing consideration of rational expectations, give policy makers detailed projections about the deviations of many macro variables from their non-stochastic steady state values, and can be adjusted to include real time "nowcasting" data to improve their forecast accuracy. Despite facing new competition from machine learning models, DSGE models remain a central tool in contemporary macroeconomic modeling and forecasting.

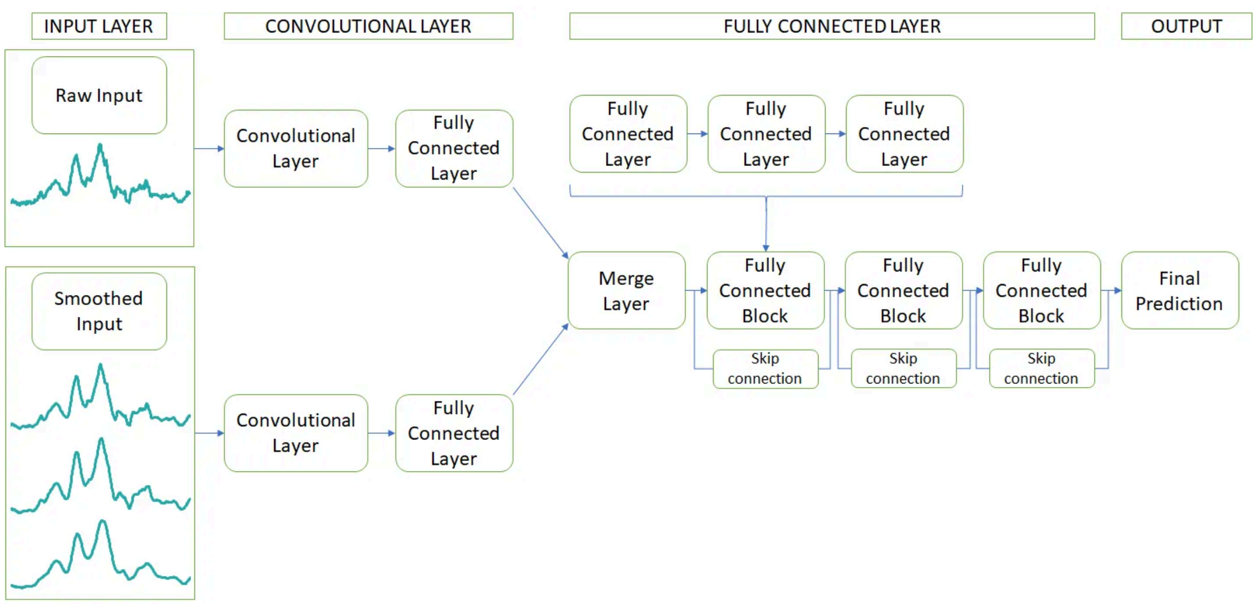
Neural networks and machine learning have been around in simple forms since the perceptron in 1957, and have been used for many tasks such as handwriting interpretation, image recognition and classification, and much more. Up until the early 2000s their applications were limited by the amount of computational resources available. Additionally, legacy methods of conditional programming were far more mature, capable, and required less compute power. Then, as the power and size of GPUs increased throughout the 2010s, programmers began to build larger and larger models and found applications for them in many tasks which had traditionally been performed by rigid conditional programs. Finally with the most recent revolution in neural networking, many can see a future in which generative pretrained transformer models may be able to replace the traditional, inflexible, programs we use today. This parallels a potential opportunity for an evolution in the methods we use in economics to model and forecast economic systems. Current econometric modeling is highly rigid with large amounts of human resources being devoted to developing models based on theoretical economic relationships. This represents a theory driven approach to modeling, in contrast neural models are data driven. Their magic is that the relationships between variables do not have to be specified in advance, the data is simply trained into the model, and the model is able to adjust its weights and biases in order to optimally fit the expected output. Additionally the number of input variables is incredibly flexible, large models can conceivably have an input layer of millions of variables. Moreover, with recent advancements in natural language processing, economic models may be able to take inputs that are not simply numeric or boolean dummy variables. Embeddings allow models to break down language, images, videos, audio and many other information types into tokens which can be used to greatly expand the model's input. For example, "in transformer models like GPT, embeddings are created using an embedding layer, which learns to represent words (or subword tokens) in a high-dimensional space during training. These embeddings are updated as the model learns to optimize its objectives, capturing semantic and syntactic relationships between tokens." (OpenAI, personal communication, Dec 2024).



https://bbycroft.net/llm

Here we can see a diagram of a single layer in a simple LLM transformer model. We can see how words are tokenized and embedded into a matrix of values, passed through the transformer, and output at the bottom in the softmax vector which represents the next most likely word in the sequence. It is possible to conceive of a pre-trained generative forecasting model which utilizes live news updates, stock market data, financial statements, in addition to traditional macroeconomic indicators as inputs to generate and dynamically update forecasts. Although this is a dream of a future which does not exist yet, some economists are experimenting with rudimentary forecasting models for inflation and other variables and seeing some tentative advantages over traditional methods due to the flexibility and predictive power of these models.

For example the economist and data scientist Nick Hallmark published an article in 2019 in which he utilizes a relatively simple, convolutional neural network for forecasting inflation. Convolutional models are generally used for image identification and classification, but neural models are incredibly flexible and can be easily modified to recognize and predict patterns in all types of data. Hallmark's model is a naive, univariate, model; it attempts to predict future inflation solely based on past inflation. The input layer takes the last 12 months of inflation data from the FRED-MD data set, and is tasked with predicting the next 12 months of inflation.



Source: Nick Hallmark, *Towards Data Science,* December 2019.

The model Hallmark created only has 7.4 million parameters, meaning each time data is passed through the model, there are 7.4 million matrix multiplications and addition operations performed before the output layer is finally calculated. This may seem like a lot, but even incredibly basic LLMs such as GPT 2, which was released in February of the same year, had 1.5 billion parameters, and newer models such as GPT 4 sit at 1.76 trillion according to The

Algorithmic Bridge. Even with a comparatively small, and non-task-specific model, Hallmark is able to show results which, at least on their surface, outperform the the accuracy of forecasts from the "Survey of Professional Forecasters (SPF) and the Federal Reserve of Philadelphia's DAR model (their highest performing benchmark for the SPF)." (Hallmark, 2019). Below we can see how the errors of his results compared to these two traditional forecasting benchmarks.

	Convolutional Model				DAR 2001:Q3 - 2019:Q3		SPF 2001:Q3 - 2019:Q3	
Time Horizon	Overall RMSE	Overall MAE	Test RMSE	Test MAE	RMSE	MAE	RMSE	MAE
T +1	0.122	0.044	0.217	0.100	1.774	1.447	1.521	0.950
T +2	0.093	0.042	0.159	0.084				
T +3	0.104	0.049	0.173	0.088	1.839	1.584	2.229	1.493
T +4	0.089	0.044	0.147	0.079				
T +5	0.127	0.053	0.220	0.108				
T +6	0.111	0.051	0.191	0.108	1.843	1.606	2.265	1.530
T +7	0.136	0.053	0.240	0.117				
T +8	0.140	0.053	0.247	0.116				
T +9	0.161	0.061	0.281	0.128	1.695	1.544	2.234	1.479
T +10	0.169	0.065	0.294	0.138				
T +11	0.138	0.062	0.236	0.133				
T +12	0.106	0.053	0.176	0.108	1.506	1.427	2.223	1.490
Average (T+1,3,6,9,12)	**0.121**	**0.052**	**0.208**	**0.106**	**1.731**	**1.522**	**2.094**	**1.389**

There are two main points of interest in this table. Looking at the overall RMSE and MAE from Hallmark's convolutional model, and comparing them to the RMSE and MAE for the DAR and SPF forecasts, we can see that the convolutional model's errors are much lower, by a factor of 10 - 20 or more. If we take this at face value, as Hallmark would like us to, his model is vastly superior to traditional forecasting. However, this interpretation is incorrect; his model is in fact much weaker than one might be led to believe. To understand why, we have to understand a little more about how neural models are trained and their predictions are validated.

When a model is trained, the data is split into two pieces: training data and validation or test data. The model is 'given full access' to the training data, and it is used to tweak the values of the parameters in order to more closely match the predicted output. The model does not directly train on the test data, instead, it is used to evaluate the accuracy of the model in a secondary testing process. This process tests to see if the model has truly learned and 'understands' deeper underlying patterns in the data, rather than simply memorizing the data points. In Hallmark's case the data set was split 70/30, with training data being 1959 - 2000, and the validation data 2001- 2019. When trying to evaluate the true predictive power of a model, and whether it has learned to identify deeper patterns, we have to compare the overall errors to the test errors; the smaller the difference, the better the model. A perfect model would have no difference in these two error measurements.

In this case when we compare the overall average to the test average, we can see that errors for MAE jump from 0.052 to 0.106 which is a 103% increase, and similarly with RMSE there is a 72% increase in errors. This represents a significant drop in performance, meaning the model generalizes poorly to the unseen test data, suggesting severe overfitting. This result means the model has likely 'memorized' the data rather than learning to predict the future. Indeed, when I examined Hallmark's code, he trained the model for an excessive 20,000 epochs (cycles) when most comparable models of a similar size can be trained in 500 or fewer epochs. Excessive training durations generally have two consequences: Overstatement of the accuracy of the model, and overfitting. It is disappointing that Hallmark did not make reference to this consequence in his article. However, in a subsequent article of his, *Deep Learning in Macroeconomic Forecasting — US Unemployment* published in May 2023, Hallmark explores and compares several different models for forecasting unemployment, and does in fact acknowledge the issues of overfitting more candidly. Overall, Hallmark's 2019 model is an interesting, experimental exercise in the application of machine learning in economic forecasting, but shows that there is still significant work to be done before these models become viable to replace traditional methods.

Is Accurate Forecasting Even Possible?

The argument I am about to make steps outside the branch of economics, but is relevant to understand the context in which the economy, and those who study it exist.

Over the history of economic theory, and specifically forecasting, we can identify a few recurring patterns. First there is some sort of crisis, a shock or event that we did not predict, and can't understand. Then economists retroactively look at all the information available to them and develop a new theory to explain the crisis or relationship. The new theory is almost always built on previous economic literature and relationships, but is also disruptive and challenges the status quo. If the theory is compelling enough, and the economist proposing the idea is well known, respected, and went to a good school, the theory may become a piece of macroeconomic gospel. For some amount of time, this way of thinking dominates the conversation, policy decisions are made, trade relationships change, we send jobs overseas or the FED conducts QE, or whatever the models suggest will result in the optimal outcome. Then, it turns out that there are serious, unforeseen consequences to the policy decisions which were informed by the theory, resulting in a new crisis, which we couldn't have predicted and don't understand. And the cycle repeats itself. With each iteration our models become more complex, our methods more refined, and our literature more dense, entangled, and opaque. My theory for why this cycle exists is due to a fundamental underlying principle of the universe: everything is interconnected and related, and even with near perfect information about every state and variable of every agent, complex relationships tend to devolve into chaotic, unpredictable, unstable states. To understand what I mean I will relate the issue to a simple principle pulled from physics: the 3-body problem.

This problem is a variation of the n-body problem which attempts to solve and predict the velocity and motion of objects which orbit and interact with each other gravitationally. With a simple 2-body model, the motions of the objects are stable, and can be described by a simple equation. When we add three or more objects, the model quickly devolves into unpredictable chaos, and even tiny, nearly immeasurable, variations in the starting conditions of each of the objects will result in massive variation as the system evolves. A similar behavior can be observed in the double pendulum experiment, and has been adapted into many popular interpretations such as the butterfly effect, and chaos theory. This unpredictability and model collapse exists even with the assumption of entirely deterministic physics, not even taking into account the indeterministic behavior of quantum physics. But I think that is well beyond the scope of an economics paper. The point being that even if we have perfect information about the initial conditions of all the agents in an economic system, which we don't, the complex interrelationships and interactions between those agents will eventually result in an entirely unpredictable state. That is not to say there are no rules for how agents interact; simple economic behaviors can be described, and important inferences can be drawn from these models. But as we project our predictions into the future, our models are hopelessly outmatched by the complexity of the real world. For instance, a routine visit to a wet market in Wuhan resulted in a global economic shutdown and shockwaves which are still being felt today and likely will be felt for decades to come. Nowhere in our models was there a variable for 'number of bats eaten.' All this means that stable models are useful, only while the economy IS stable.

Additionally, if we had this mythical, perfect forecasting model, which could reliably predict future economic events, such as bubbles, crashes, recessions, or trends, the model would become useless. Incorporating considerations for rational expectations, if agents are told with certainty that there will be a recession caused by an asset bubble in the next 12 months, they will adjust their behavior accordingly. Investors will sell off their portfolios to cut losses, businesses will lay off employees, consumers will save rather than spend, and the predicted recession will make its own bed 12 months earlier than predicted. If there ever was a predictive model which truly had this ability, it would have to be kept as a secret in order to preserve its own usefulness.

Summary and reflection

This is all to say that there are many fundamental challenges faced by the economic forecaster, many of which appear to be fundamentally impossible to overcome. And yet, we still devote significant resources to forecasting. If we did not do so, we would have no understanding or foresight of the future, and would not be able to consider future events whatsoever. The forecasting methods we have today are far more advanced than any we have had in the past. Our understanding of the inner workings of the macroeconomy, the volume of data we possess, and the computational power we hold in order to process it all has never been more substantial. Even models that are only effective for a short period of time, before they become outdated due to some fundamental underlying shift in the relationship between macro variables, are still useful during that time in order to plan and make decisions about the future. We can see throughout the history of macroeconomic forecasting, through the lens of inflation forecasting, how economists never fail to rise to the occasion and rethink how we model the economy and draw new conclusions from it. I only hope that this rich history of revelation and theoretical debate can continue into the future and not become stagnated over time as the economic literature expands.

Sources

Del Negro, M., & Schorfheide, F. (2013). DSGE model-based forecasting. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, Part A, pp. 57–140). Elsevier. https://doi.org/10.1016/B978-0-444-53683-9.00002-5

Dynamic stochastic general equilibrium. (n.d.). In *Wikipedia*. Retrieved December 19, 2024, from https://en.wikipedia.org/wiki/Dynamic_stochastic_general_equilibrium

Romero, A. (2024, September 13). OpenAI o1: A new paradigm for AI. *The Algorithmic Bridge*. https://www.thealgorithmicbridge.com/p/openai-o1-a-new-paradigm-for-ai

DSGE models. (n.d.). In *ScienceDirect Topics*. Retrieved December 19, 2024, from https://www.sciencedirect.com/topics/economics-econometrics-and-finance/dsge-model

Towards Data Science. (2019, October 10). Deep learning to predict US inflation. *Medium*. https://towardsdatascience.com/deep-learning-to-predict-us-inflation-70f26405bf76

Hallmark, N. (2019, October 10). Deep learning in macroeconomic forecasting: US unemployment. *Medium*. https://medium.com/@hallmark-nick/deep-learning-in-macroeconomic-forecasting-us-unemployment-cbcc5a33c7c9

Three-body problem. (n.d.). In *Wikipedia*. Retrieved December 19, 2024, from https://en.wikipedia.org/wiki/Three-body_problem

Lucas, R. E., Jr. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 1, pp. 19–46). North-Holland. https://doi.org/10.1016/S0167-2231(76)80003-6

Furman, J. (2022). Why did (almost) no one see the inflation coming? *Intereconomics, 57*(2), 79–86. https://www.intereconomics.eu/contents/year/2022/number/2/article/why-did-almost-no-one-see-the-inflation-coming.html