SSE RIGA

Phillips Curve in the Eurozone: Forecasting Inflation and the Output Gap Using Explainable Neural Networks.

Authors:

Igors Tatarinovs
Reinis Fals

Supervisor:

Boriss Siliverstovs

JEL Codes: C45, C53, C58, E31, E32

May 2024

Riga

COPYRIGHT DECLARATION AND LICENCE

Name(s) of the author(s) in full: Igors Tatarinovs, Reinis Fals

Title of the Thesis: *PC in the Eurozone: Forecasting Inflation and the Output Gap Using Explainable Neural Networks*

I (We) hereby certify that the above-named thesis is entirely the work of the persons named below, and that all materials, sources and data used in the thesis have been duly referenced. This thesis – in its entirety or in any part thereof – has never been submitted to any other degree commission or published.

In accordance with Section 1 of the Copyright Law of Latvia, the person(s) named below is (are) the author(s) of this thesis.

Pursuant to Article 40 of the Copyright Law the author(s) hereby agree(s) and give an explicit licence to SSE Riga to deposit one digital copy of this thesis in the digital catalogue and data base at SSE Riga Library for an unlimited time and without royalty. The licence permits SSE Riga to grant access to the duly deposited thesis to all users of the catalogue and data base without royalty and limitations to downloading, copying and printing of the digital thesis in whole or in part provided I am (we are) indicated as the author(s) of the thesis according to Clause 4 Section 1 Article 14 of Copyright Law. I (We) assert my (our) right to be identified as the author(s) of this thesis whenever it is reproduced in full or in part.

Signed:

______*/signed/*______ ______*/signed/*______

Igors Tatarinovs Reinis Fals

Date:

8 April, 2024

Abstract

This study introduces the use of Hemisphere Neural Networks (HNN) for forecasting the Inflation and output gap in the Eurozone within the macroeconomic model of PC, challenging the assumption that the region's economic complexity and fragmentation of data across the region fundamentally impede the use of neural networks to predict variables in question. With customization of the model and expanding it to account for inflation spillover effect, the study proves wrong the statement about fragmentation of data being a problem. Through advanced data preprocessing, ML techniques, and optimization algorithms this research proves the possibility to use and scale the HNN for forecasting output gap & inflation components, and also compares HNNs against conventional models, demonstrating HNNs' superiority in inflation forecasting and output gap estimation. The analysis highlights HNNs' enhanced interpretability and adaptability along with greater accuracy compared with benchmarks such as HP filtering, Linear Regression and ARIMA model based on RMSE, which proves the hypothesis about possibility to utilize explainable neural networks for the matter discussed.

The Github repository with the code produced in this research: Github Repository

Table of contents

Disclaimer & Acknowledgements

Throughout this research paper, for the sake of clarity and distinction between traditional econometric methods and Machine Learning (ML) approaches, a separation will be made between non-ML models and ML models. It is important to note that this differentiation is made to outline the varying degrees of sophistication in the methodologies and algorithms employed. However, it is imperative for readers to acknowledge that all statistical techniques, including even simple linear regressions, fall under the broad umbrella of ML, albeit of a different kind compared to gradient boosting or neural networks (NN).

While traditional econometric methods such as linear regression rely on explicit assumptions about the underlying data generating process and may require more manual feature engineering, ML techniques like NN leverage complex algorithms to extract patterns from data through the more sophisticated training process. Nonetheless, both approaches share the fundamental goal of modeling relationships within economic data to make predictions or inferences. Therefore, it is crucial for readers to maintain a clear understanding of the continuum of ML techniques, ranging from simpler linear models to more complex algorithms like NN. This acknowledgment ensures that discussions surrounding the application of ML in economic modeling remain grounded in a comprehensive understanding of the methodologies employed.

In this research paper we have used LLMs for paraphrasing, and avoiding grammar & style mistakes in the text, as well as to follow appropriate formal APA style of writing. No facts and conclusions were solely generated by LLMs

We would like to express our sincere gratitude to our supervisor Boriss Siliverstovs for the invaluable contribution to the methodology & theoretical reasoning of the thesis and the guidance throughout the writing period. Additionally, we would like to thank Dmitry Kulikov for the economical data provided for the research and Philippe Goulet Coulombe for the source code that served as a solid base for the beginning of our research.

1. Introduction

In the 1970s, criticism by Robert Lucas highlighted the Phillips Curve's (PC) oversight of inflation expectations, leading to the Expectations-Augmented PC (EAPC). This period also saw the emergence of rational expectations theory, challenging the Curve's traditional assumptions. With stagflation in the 1970s and 1980s prompting further reassessment, economists have since refined the model to incorporate factors like supply shocks and globalization, ensuring its continued relevance in macroeconomic policy.

The PC plays a pivotal role in the field of macroeconomic analysis, providing a framework for understanding and forecasting inflation (in that context understanding stands for determining drivers of the inflation). It outlines the relationship between unemployment and inflation, thereby aiding policymakers in formulating monetary policies that aim to balance economic growth and stability. The primary function of the PC is to offer insights into the factors driving inflation, which is crucial for managing an economy's overall health. This understanding enables the development of strategies to control inflation while minimizing its impact on employment levels, directly affecting economic policy decisions and their outcomes on the broader economy. The three primary components of the PC include inflation expectations, the output gap (OG), and supply shocks which play significant roles in influencing inflation and unemployment dynamics.

The OG is defined as the deviation between an economy's actual and potential output — serves as a critical component in understanding the dynamics of the PC. This relationship underscores the sensitivity of inflation to the levels of economic activity relative to the economy's capacity. Essentially, when actual output surpasses potential output (a positive OG), it indicates an economy operating beyond its sustainable capacity, often leading to inflationary pressures as demand outstrips supply. Beyond these components, other factors may influence the relationship captured by the PC. These may include: Labor Market Flexibility, Global Economic Conditions, Monetary Policy.

Traditional models, including the PC, have offered invaluable insights into the inverse relationship between unemployment and inflation. However, as economic environments evolve, the limitations of these models in capturing complex, dynamic relationships become apparent. This is especially true in the Eurozone, where economic diversity and integration add layers of complexity to inflation dynamics. One significant hurdle in traditional PC analysis is the modeling of key components such as inflation expectations and the OG — both of which are inherently unobservable and traditionally estimated through indirect methods. These estimation challenges introduce uncertainties and potential biases into the analysis, complicating the accurate prediction of inflation. Moreover, the influence of global economic conditions and supply shocks further complicates this task, necessitating a more sophisticated approach to understanding the multifaceted drivers of inflation.

The motivation for this research arises from the pressing need to overcome these obstacles through innovative methodologies aimed to capture nonlinearities specific for matter of the research. Recent advancements in ML offer promising avenues for enhancing the predictive accuracy and maintaining the interpretability of economic models. Specifically, Hemisphere Neural Networks (HNN) firstly introduced by Philippe Goulet Coulombe (2022) present a cutting-edge approach capable of deciphering the complex interplay between various economic variables and inflation. By leveraging the power of HNN, this study aims to provide a more nuanced understanding of the PC within the Eurozone, accounting for the region's unique economic characteristics, and specifically addresses the diversity of data across regions within Eurozone which was always considered as an obstacle using NN for economic predictions in Eurozone.

Furthermore, this research is driven by the desire to advance the application of ML in economics, moving beyond traditional econometric approaches for PC and OG. By comparing the HNN approach with traditional methods in the Eurozone context, we aim to clarify if the HNN approach is suitable for such a region and if so, how much more accurate it is compared with traditional approaches. Therefore, the research questions are following:

1. *How accurately, compared to the traditional approaches, is it possible to predict the inflation in the Eurozone using explainable NNs?*

2. How accurately, compared to the traditional approaches, is it possible to predict the OG in the Eurozone using explainable NNs?

The main hypothesis is that <u>*it is possible to predict inflation and OG with explainable NNs in the Eurozone, and the efficiency will be better than of the traditional approaches.*</u>

This research is aimed at exploring the application of HNN to estimate the PC and the OG within the Eurozone, an area marked by significant economic diversity and integration. The study seeks to address the complexities inherent in traditional PC analysis, such as the estimation of unobservable components like inflation expectations and the output gap. Additionally, the research is positioned to evaluate the effectiveness of HNNs in capturing the nuanced economic dynamics of the Eurozone, and if the flexibility of the HNN architecture can help to achieve the same accuracy as it does in the US despite insufficiency of the data in Eurozone compared to the US.

However, the complexity of the Eurozone's economy poses challenges in data collection and model training. The diversity of data across regions within the Eurozone may affect the model's generalizability, therefore data selection is the crucial point to consider (Šestanović & Arnerić, 2021). As with any ML model, the predictive accuracy of HNNs is contingent upon the quality and granularity of the available data. Limitations in data may constrain the model's performance.

Also, The comparison with traditional models is subject to the selection and implementation of these models, which may vary in their complexity and suitability for the Eurozone context and the general purpose (i.e. models aimed to predict the inflation as a single variable do not account for the OG, and select as a key benchmark the target variable only; either, OG filtering method it does not directly provide information about inflation).

Introduced in 2022, the HNN architecture represents a frontier in the integration of advanced ML for inflation and OG forecasting, offering a novel approach that has yet to be explored in the context of the Eurozone. This research stands at the intersection of economic theory and technological advancement, aiming to bridge the gap between traditional econometric models and the potential offered by ML to understand if it is possible to use HNN and its variations in the Eurozone. This study not only tests the adaptability and effectiveness of this novel NN architecture but also contributes to a deeper understanding of the Eurozone's economic conditions.

Also the research contributes to the ongoing discourse on making complex NN models more transparent and understandable, a crucial consideration for policymakers and economists relying on these models for decision-making. Lastly, the significance of this study extends beyond its immediate findings, laying the groundwork for future research in the application of ML to economic forecasting. By highlighting the challenges, opportunities, and potential strategies for integrating advanced computational techniques with economic analysis, this research opens avenues for further exploration and development within the field.

This thesis is structured to methodically explore the application of HNN for estimating the PC and the OG in the Eurozone. It begins with an introduction that sets the research context, followed by a literature review that situates the study within existing economic theories and models. The methodology section details the approach and techniques used to implement and assess the HNN model. Results are then presented and discussed, highlighting key findings and their implications for economic forecasting. The thesis concludes with a summary of contributions, limitations, and suggestions for future research, providing a comprehensive overview of the potential of advanced ML techniques in enhancing economic analysis within the Eurozone.

While similar practices have been successfully implemented in the United States, as pioneered by Philippe Goulet Coulombe of the Bank of Canada, the applicability of such methodologies to the Eurozone remains largely unexplored. Coulombe's innovative approach involves the use of interpretable NN, combining Long Short-Term Memory (LSTM) NN and an interpretable layer into a unified ensemble NN termed the Hemisphere. This methodology demonstrated success in the U.S., but its applicability and effectiveness of such a new approach in the Eurozone, with its difference from the US economic landscape and data characteristics, remain largely unexplored (Barbaglia & Consoli & Manzan, 2021). Also proof that the new concept immediately gained recognition in the industry is the interest of the Bank of England in using this method (Pierre-Etienne Caza, 2023)

The relevance of our research lies in the shift from traditional linear approaches to the utilization of cutting-edge, interpretable NN. The methodology proposed by Coulombe (2022) has the potential to revolutionize inflation forecasting, not only for the US, but potentially also for Europe. Success in our endeavor would provide a novel means of forecasting PC and OG, filling a void in the existing toolkit for economic analysts and policymakers in the region. Our study seeks to fill this research gap by addressing the research questions above.

2. Review of literature

2.1 Development of the PC

In the 1970s, Robert Lucas (1972), a prominent economist, criticized traditional linear regression estimates of the PC, arguing that they failed to account for the role of inflation expectations (Alogoskoufis & Smith, 1991). He proposed the EAPC, which included a term for expected inflation. The EAPC suggested that inflation expectations played a crucial role in determining the actual inflation rate. The 1970s and 1980s witnessed a period of high inflation and stagflation, challenging the traditional PC relationship (Alogoskoufis & Smith, 1991). Economists observed that the trade-off between inflation and unemployment seemed to have weakened or even disappeared, this led to a reassessment of the PC and its relevance in modern economies (Alogoskoufis & Smith, 1991). During this period of time the rational expectations theory was born, stating that past outcomes influence future outcomes, Sargent and Wallace (1976) were not only the founding fathers of this theory, but were also one of the first to model the PC in order to predict future inflation. In recent years, economists have continued to refine and extend PC models, incorporating additional factors such as supply shocks, globalization, and technological advancements. These models continue to be used to inform macroeconomic policy decisions.

While the EAPC introduced inflation expectations into the PC framework, fundamentally altering our understanding of inflation dynamics, it was the The New Keynesian PC (NKPC) that deepened this approach by integrating rigorous microfoundations and forward-looking elements. This shift not only provided a more robust theoretical basis for the relationship between inflation, expectations, and economic activity but also highlighted the importance of strategic price-setting by firms and the role of monetary policy in shaping expectations (Clarida, Gali & Gertler, 1998). The NKPC's emphasis on these forward-looking components has been instrumental in refining our comprehension of inflation processes, offering nuanced insights that have significantly influenced both academic research and practical policy-making in the contemporary economic landscape.

NKPC represents a significant advancement in macroeconomic theory, addressing some of the limitations of its predecessors by incorporating expectations and microfoundations into the analysis of inflation dynamics. Originating in the works of economists such as Clarida, Galí, and Gertler in the 1990s, the NKPC posits that current inflation is not only a function of the OG and past inflation but also critically depends on firms' and consumers' expectations of future inflation (Clarida, Gali & Gertler, 1998). This framework reflects a deeper understanding of how expectations about the future influence current economic decisions, a concept that has fundamentally altered monetary policy approaches worldwide.

These foundational studies underscore the NKPC's central premise that understanding inflation requires a nuanced appreciation of the interplay between expectations, economic slack, and price-setting behavior (Gali, Gertler, 2000). By integrating forward-looking elements, the New Keynesian PC offers a more comprehensive and realistic model of inflation dynamics, significantly impacting both economic theory and policy formulation.

2.2 Economical meaning of PC.

The interconnection between the PC, OG, and inflation from a microeconomic perspective delves into the intricate dynamics of economic variables and their impact on macroeconomic stability and growth. This analysis is rooted in foundational economic theories and models that elucidate the mechanisms through which these variables interact. In a microeconomic context, this relationship can be attributed to wage-setting behaviors and price adjustments in response to supply and demand imbalances. When unemployment is low, labor markets tighten, leading to wage increases as employers compete for a limited workforce. These wage pressures, in turn, translate into higher prices, contributing to inflation. The microeconomic theory behind this involves the wage-setting curve and the price-setting curve, which intersect to determine equilibrium employment and real wages.

The OG represents the difference between actual economic output and potential output - the level of output that an economy can produce at full capacity without inflating prices. From a microeconomic standpoint, the OG is a critical indicator of economic slack or overheating. When actual output exceeds potential output (a positive OG), demand pressures lead to inflationary tendencies, aligning with Keynesian perspectives on demand-pull inflation. Conversely, a negative OG indicates underutilized resources, typically associated with lower inflation or deflation (Galì, 2015).

NKPC provides a theoretical framework that links microeconomic foundations to the PC and OG dynamics. The NKPC incorporates expectations of future inflation and the concept of real rigidity to explain how current inflation is affected by the OG and anticipated future inflation:

$$\pi_t = \beta E_t[\pi_{t+1}] + k y_t + u_t$$

where π_t is the inflation rate at time *t,* $E_t[\pi_{t+1}]$ the expected inflation for the next period, y_t represents the OG, k is a coefficient capturing the sensitivity of inflation to the OG, and u_t denotes cost-push shocks (Romer & Mankiw & Ball, 1988). Theoretically, inflation expectations and the OG this integration acknowledges the dynamic interplay between these components and inflation, underscoring the importance of forward-looking elements in macroeconomic modeling. However, the empirical application of this integrated model faces the challenge of accurately capturing the essence of these unobservable components. The complexity of predicting inflation expectations and the OG stems not only from their non-observability but also from the multifaceted economic forces driving them (Mavroeidis, Møller & Stock, 2014). However, a nuanced understanding of this relationship necessitates the incorporation of two critical, yet inherently unobservable, components: inflation expectations and the OG.

Inflation Expectations: The anticipatory beliefs regarding future inflation significantly influence current economic behavior. These expectations shape consumer spending, wage negotiations, and business pricing strategies, thereby affecting actual inflation outcomes. The PC integrates these anticipations, acknowledging their pivotal role in the inflation-unemployment nexus. However, the subjective nature of expectations, coupled with their forward-looking characteristic, renders them challenging to measure and predict accurately (Mavroeidis, Møller & Stock, 2014).

Output Gap: Defined as the deviation of an economy's actual output from its potential output, the OG is a gauge of economic slack or overheating. It represents the capacity for an economy to grow without engendering inflationary pressures. The OG centrality to the PC model stems from its influence on inflation dynamics; a positive OG (where actual output surpasses potential output) typically exerts upward pressure on inflation (Cuerpo, Cuevas & Quilis, 2018).

Yet, the potential output is a theoretical construct, difficult to quantify due to its dependence on unobservable factors like technological advancements and labor market efficiencies (Cuerpo, Cuevas & Quilis, 2018).. The prediction of inflation expectations and the OG embodies a complex endeavor, primarily due to their intangible nature. Traditional methodologies employ various proxies and econometric models to estimate these components, albeit with inherent limitations:

- Survey-Based Methods for Inflation Expectations: Utilizing consumer and expert surveys to gauge inflation expectations provides direct insights but is subject to biases and may not accurately reflect collective market anticipations (Neissa & Nelson, 2002).
- Market-Based Measures: Instruments like inflation-linked bonds offer real-time market expectations of inflation, yet they encapsulate risk premia and liquidity conditions, potentially distorting pure inflation anticipations (Rudd & Whelan, 2005).
- Statistical Filters for OG Estimation: Techniques such as the Hodrick-Prescott (HP) filter and the Kalman filter are commonly applied to estimate the OG While these methods offer a mathematical approach to discerning cyclical economic fluctuations, they carry the risk of revision and are sensitive to the chosen parameters, questioning their reliability (Rudd & Whelan, 2005).

2.3 Models that are traditionally used to predict PC, OG, and inflation

The traditional modeling and estimation of the PC, the OG, and inflation within economic research have predominantly relied on linear regression models, Autoregressive Integrated Moving Average (ARIMA) models, Dynamic Stochastic General Equilibrium (DSGE) models, and HF filtering for OG (ECB, 2002), (Stock & Watson, 2008). These methodologies, grounded in classical statistical and econometric theory, have provided foundational insights into the macroeconomic relationships governing inflation dynamics, employment rates, and the OG. However, each approach comes with its inherent limitations, particularly when applied to complex, non-linear economic phenomena observed in real-world data (Stock & Watson, 2008).

In the evolving landscape of economic analysis, the quest for more accurate and insightful forecasting methods is perpetual. Traditional techniques like Linear Regression and Time Series Analysis, despite their widespread use in central banking, capturing approximately 40-50% adoption, are increasingly challenged by the complexity of economic phenomena (Smith et al., 2021). The emergence of ML methods, notably Random Forests and Gradient Boosting Machines with an adoption rate of 30-40% , represents a significant shift towards more dynamic and adaptable models (Smith et al., 2021). Yet, even as Support Vector Machines (SVM) find their footing in classification tasks with a 20-30% usage rate (Smith et al., 2021). Deep learning techniques, including NN, have seen a growing interest from about 15-25% of central banks, highlighting the sector's readiness to explore more complex, non-linear data relationships that traditional models can't easily unravel (Araujo et al., 2024). This trend underscores the critical need for research into ensemble NN, which combine the strengths of various ML approaches to offer unprecedented accuracy and interpretability in economic forecasting (Araujo et al., 2024).

2.3.1 Linear regressions

The Linear Regression Model has historically been a cornerstone in the empirical analysis of the PC, providing a statistical framework to explore the relationship between inflation and unemployment, and by extension, the OG (Ascari & Marrocu, 2003). The linear regression model, in the context of PC, is used to quantify this relationship, to estimate the trade-offs between inflation and unemployment. A simplified linear regression model for the PC can be expressed as:

$$\pi_t = \alpha + \beta U_t + \epsilon_t$$

The OG, defined as the difference between actual output and potential output, can be indirectly estimated through the PC by observing the impact of unemployment (a proxy for economic slack) on inflation. (Gross, Marco; Semmler & Willi, 2017). A linear regression model incorporating OG measures can be specified as (Gross et. al, 2017):

$$\pi_t = \alpha + \gamma(Y_t - Y_t^n) + \epsilon_t$$

The estimation of the linear regression model typically involves Ordinary Least Squares (OLS), which minimizes the sum of squared residuals between the observed and predicted values of inflation, providing unbiased and efficient estimates of α, β, and γ under classical linear model assumptions (Gross, Marco; Semmler & Willi, 2017). The statistical significance of the estimated coefficients is assessed through t-tests, while the overall model fit can be evaluated using the R-squared statistic.

Expanding on the foundational use of linear regression models to estimate the PC and derive the OG, we encounter significant complexities when incorporating multiple variables that contribute to the inflation component. These complexities necessitate a more nuanced approach to modeling and analysis, reflecting the multifaceted nature of economic relationships.

The basic linear regression model can be extended to include multiple explanatory variables that influence inflation, acknowledging that inflation is driven by a variety of factors beyond unemployment or the OG alone. This can be represented as:

$$\pi_t = \alpha + \beta_1 U_t + \beta_2 X_{t1} + \beta_3 X_{t2} + \ ...\ + \beta_n X_{tn} + \epsilon_t$$

Disaggregating the inflation component into different groups allows for a detailed analysis of how various factors contribute to overall inflation. For instance, core inflation (excluding volatile items like food and energy prices) can be distinguished from headline inflation to assess underlying inflation trends. The model thus evolves to capture these distinctions:

$$\pi_t = \alpha + \beta_1 U_t + \beta_2 CoreInflation_t + \beta_3 EnergyPrices_t + \ ...\ + \epsilon_t$$

This approach enables economists to isolate the impact of specific factors on inflation, enhancing the predictive power and relevance of the PC in policy analysis. However, extending the linear regression model introduces complications such as multicollinearity, where explanatory variables are highly correlated with each other, potentially distorting the estimation of coefficients. Additionally, the dynamic nature of the economy suggests that past values of inflation and other variables could influence current inflation, leading to the use of lagged variables in the regression model:

$$\pi_t = \alpha + \sum_{i=1}^{n} \beta_i U_{t-1} + \sum_{j=1}^{m} \gamma_j X_{t-j} + \epsilon_t$$

Incorporating lagged variables addresses the dynamic aspects of inflation but also increases model complexity, requiring careful selection of lag lengths based on criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Hill & Griffiths, 2019).

2.3.2 Autoregressive Integrated Moving Average (ARIMA)

Another option to predict the PC and estimate the OG is the ARIMA model, a statistical approach that extends beyond the simplicity of linear regression to capture the temporal dynamics inherent in economic data. The ARIMA model, by integrating autoregressive (AR) and moving average (MA) components with differencing to achieve stationarity, offers a robust framework for forecasting economic indicators (Hyndman & Athanasopoulos, 2021).

The ARIMA model is defined by three key parameters: *p, d, and q*. The p parameter specifies the number of autoregressive terms, d indicates the degree of differencing needed to render the time series stationary, and q denotes the number of moving average terms . Mathematically, an ARIMA*(p,d,q)* model for a time series Y_t can be expressed as:

$$\Delta^d Y_t = \alpha + \sum_{i=1}^{p} \phi \Delta^d Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

In the context of the PC and OG analysis, the ARIMA model facilitates the examination of inflation and unemployment rates as time series data, accounting for their temporal dependencies and trends. By modeling the dynamics of these variables, the ARIMA framework can yield forecasts for inflation based on historical patterns of unemployment and OG, thereby providing empirical insights into the PC's validity and behavior over time. The selection of ARIMA model parameters *(p,d,q)* is critical and typically involves iterative testing and validation, such as the Box-Jenkins methodology (Hyndman & Athanasopoulos, 2021). This process includes:

1. Identification: Analyzing the data's autocorrelation function (ACF) and partial autocorrelation function (PACF) to estimate initial values for p and q,
2. Estimation: Using techniques like Maximum Likelihood Estimation (MLE) to estimate the model parameters,
3. Diagnostic Checking: Assessing the model fit by examining residuals for autocorrelation or non-normality and adjusting parameters as necessary.

Expanding the ARIMA model to tackle the complexities of forecasting the PC and estimating the OG involves addressing several advanced considerations, especially when multiple variables contribute to inflation dynamics. The ARIMA framework can be extended to ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) to incorporate additional explanatory variables, enhancing the model's capability to forecast inflation more accurately by accounting for various inflation components(Hyndman & Athanasopoulos, 2021).

The ARIMAX model allows for the inclusion of external variables that influence inflation, enabling a disaggregated analysis of different inflation components. This is particularly useful for examining how various factors, such as energy prices, wage growth, and monetary policy changes, contribute to overall inflation. The ARIMAX model can be represented as:

$$\Delta^{d} Y_{t} = \alpha + \sum_{i=1}^{p} \phi \Delta^{d} Y_{t-i} + \sum_{j=1}^{q} \theta_{j} \varepsilon_{t-j} + \sum_{k=1}^{m} \gamma_{k} X_{tk} + \varepsilon_{t}$$

Economic data often exhibit seasonal patterns, which the basic ARIMA model does not account for. The Seasonal ARIMA (SARIMA) model extends ARIMA by incorporating seasonal differencing and seasonal autoregressive and moving average components, making it more adept at handling data with periodic fluctuations. The SARIMA model is denoted as ARIMA*(p,d,q)*(P,D,Q)*s*, where P, D, and Q represent the seasonal autoregressive, differencing, and moving average terms, respectively, and *s* is the length of the season.

Beyond ARIMA and its extensions, Dynamic Factor Models (DFMs) offer a sophisticated approach to estimating the OG by extracting common trends from a large panel of economic indicators. DFMs can distill the information contained in numerous time series into a few unobserved common factors, which can then be related to the OG, providing a comprehensive view of economic activity and potential output (Hamilton, 1994).

The application of these advanced time series models introduces several econometric challenges, including model identification, parameter estimation, and diagnostic testing. Selecting the appropriate model specifications and ensuring that the models are correctly identified requires careful analysis of the data's autocorrelation and partial autocorrelation functions, as well as consideration of information criteria such as AIC or BIC for model selection. Moreover, the inclusion of multiple variables and the decomposition of inflation into various components necessitate rigorous testing for stationarity, cointegration, and causality, ensuring that the relationships modeled are statistically valid and economically meaningful (Hamilton, 1994).

The ARIMA model represents a powerful tool in the arsenal of econometric techniques for analyzing and forecasting economic phenomena like the PC and OG. Its ability to model time series data with autocorrelation and non-stationarity attributes makes it particularly suited for capturing the complex dynamics that characterize macroeconomic indicators. However, the effectiveness of ARIMA models hinges on careful parameter selection, rigorous model diagnostics, and an understanding of their limitations in the face of economic volatility and structural changes.

While ARIMA models offer a versatile approach to time series forecasting, their reliance on past data trends may limit their ability to predict future changes influenced by external shocks or structural breaks in the economy. Moreover, the assumption of linearity and stationarity in ARIMA models may not always hold in real-world economic data, necessitating extensions such as ARIMA with exogenous variables (ARIMAX) or seasonal ARIMA (SARIMA) models (Hyndman & Athanasopoulos, 2021).

2.3.3 HP Filtering

The HP filter is a smoothing technique applied to time series data to separate the cyclical component from the trend component. This method is instrumental in estimating the potential output of an economy, thereby facilitating the calculation of the OG. The HP filter minimizes the following objective function:

$$min_{\tau}\{\sum_{t=1}^{T}(y_t - \tau_t)^2 + \lambda\sum_{t=2}^{T}((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2\}$$

The choice of λ is crucial; a higher value places more emphasis on smoothing the trend, making it less responsive to short-term fluctuations. Conversely, a lower value makes the trend component more sensitive to cyclical movements. For annual data, a common choice for λ is 100, while for quarterly data, 1600 is frequently used, which we will also use in the research (statsmodels, n.d.). In the context of our thesis, the HP filter is applied to disaggregate the cyclical fluctuations from the underlying trend in Eurozone GDP data, thus isolating the OG:

$$OG_t = y_t - \tau_t$$

Estimating the OG, we gain a more precise understanding of the economic slack or pressure within the Eurozone, which enhances the predictive accuracy of the PC regarding inflation dynamics (Grant & Chan, 2016). The cyclical component derived from the HP filtering process reflects the Eurozone's economic fluctuations, offering insights into the temporal dynamics influencing inflation (Grant & Chan, 2016).

2.4 ML approaches to predict inflation and OG

The landscape of economic modeling has changed dramatically over the years, from traditional linear approaches to more sophisticated ML approaches Nonlinear modeling at the end of the 20th century marked a major change in economics in the analysis of the PC and inflation. These nonlinear models recognized the complexity and dynamics of monetary policy and made it more flexible than the linear models that preceded them, leading to a better understanding of factors such as inflation and unemployment growth. The importance of ML in economics continues to grow, especially in its application to complex economic data such as the PC. Traditional approaches to modeling the PC have faced challenges, especially if its shape is to be captured evolving over time and under different economic conditions (Paranhos, 2023). The introduction of ML in economic modeling, particularly in understanding and predicting the PC, marks a significant departure from traditional econometric methods. According to Maccarrone, Morelli, Spadaccini (2021), traditional linear models, while providing a foundational understanding of economic relationships, often fall short in capturing the non-linear, dynamic nature of these relationships, while ML techniques, with their inherent flexibility and adaptability, excel in this regard (Maccarrone & Morelli & Spadaccini, 2021).

Theoretically, any ML algorithm that is suitable for time series forecasting can be used to forecast the PC. However, by design, some models are more suitable for analyzing economic data than others. Thus, out of the majority of the most widespread and known algorithms, there are 13 algorithms (7 of which can be considered as more advanced ML algorithms)[1] that can be used to predict the PC (interpretability decreases, and accuracy with sophistication increases every algorithm on): Random Forests, Ensemble Methods, Gradient Boosting Machines, Feedforward NN, Convolutional NN, Recurrent and LSTM, Autoencoders. For each one, it is possible to evaluate the theoretical advantages, disadvantages, as well as accuracy and explainability; however, the general principle maintains the same: the interpretability decreases as accuracy increases.

Random Forests: the most straightforward example from the whole area of more sophisticated algorithms. Although they can handle complex non-linear relationships, their ensemble nature makes them less interpretable. According to Coulombe (2019) in recent

[1] List of complexity & accuracy dynamics across models is listed in Figure 4 in the Appendix

advancements within economic modeling, particularly in the context of PC estimation, the Macroeconomic Random Forest (MRF) algorithm emerges as a notable development. This algorithm adapts the canonical Random Forest ML tool to macroeconomic contexts, yielding Generalized Time-Varying Parameters (GTVPs) which effectively accommodate various nonlinearities and offer direct interpretability. Notably, MRF has demonstrated proficiency in forecasting significant economic events, such as the 2008 unemployment surge, and shows robust performance in inflation modeling (Coulombe, 2019). The broader scope of Random Forest in economic analysis is further illuminated by its capability to handle complex nonlinearities and high-dimensional data with reduced overfitting and minimal tuning requirements. This versatility is particularly advantageous in the context of economic time series analysis, where Random Forest's alignment with common nonlinearities obviates the need for arbitrary specification searches. These attributes collectively underscore the significant role of Random Forest algorithms in enhancing economic modeling techniques and contributing valuable insights into macroeconomic phenomena.

Overall Random Forest can be classified by 4 parameters in the following way:

Advantages: Handles non-linear data well; robust to overfitting; can model complex relationships.

Disadvantages: Less interpretable compared to simpler models; can be computationally intensive.

Accuracy: High; can capture complex, non-linear relationships and interactions between economic variables effectively.

Interpretability: Moderate; individual decision trees are interpretable, but the ensemble model as a whole can be complex.

Gradient Boosting Machines (GBM) are the prominent example of more sophisticated, accurate and less explainable. According to Sprangers (2021) GBM achieves accurate probabilistic estimates in tasks with complex differentiable loss functions, showing improvements of up to 10% in point forecasting performance and up to 300% in probabilistic forecasting performance (Sprangers, 2021). The GBM is rarely used in direct inflation prediction and PC estimation due to being an intermediary between more accurate neural networks, and more explainable different nonlinear traditional models. Overall, GBM can be classified by 4 parameters in the following way:

Advantages: High predictive accuracy; handles different types of data effectively.
Disadvantages: Can be prone to overfitting; requires careful tuning of parameters.
Accuracy: Very high; among the best performers for predictive tasks with structured data.
Interpretability: Moderate; more interpretable than deep learning models but less so than simpler models.

According to Paranhos (2023), The study demonstrates that the LSTM network surpasses established benchmarks, encompassing various ML models, particularly in scenarios with extended time horizons and during phases of intensified macroeconomic uncertainty; nonetheless, LSTM model operates as a 'black box,' presenting challenges in terms of interpretability (Paranhos, 2023). Overall, LSTM NN can be classified by 4 parameters in the following way:
Advantages: Solves vanishing gradient problem; excellent for learning long-term dependencies.
Disadvantages: Complex architecture leading to longer training times; high computational costs.
Accuracy: Very high; especially effective in capturing long-term dependencies in time-series data.
Interpretability: Low; complex architectures make LSTMs less transparent.

The progression towards more advanced ML models like NN in PC estimation highlights a critical trade-off: as these models become more sophisticated, their predictive accuracy improves, but their interpretability diminishes. This is evident in algorithms such as Random Forests and LSTM networks, which, despite their enhanced forecasting abilities, are often critiqued for their opaque nature. Eventually, it leads us to the idea that an “ideal” instrument should combine the accuracy of NN and interpretability of linear regression models.

Neural networks, specifically LSTM models, represent a sophisticated approach to time series forecasting, offering significant advantages in estimating economic indicators like the PC. LSTMs are a class of recurrent NN capable of learning long-term dependencies in data, which is crucial for capturing the dynamic relationships between inflation and unemployment rates over time.

The mathematical foundation of LSTM networks hinges on their unique structure, which includes memory cells and multiple gates (input, forget, and output gates) (Zarzycki & Ławryńczuk, 2021). These components work together to regulate the flow of information, allowing the network to retain or discard data across long sequences. The general equations governing the operation of an LSTM unit can be represented as follows:

1. **Forget Gate** (f_t): Determines which information is discarded from the cell state.

   $f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f)$

2. **Input Gate** (i_t) and **Candidate Values** ($\overline{C_t}$): Decide which new information is stored in the cell state.

   $i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i$

   $\overline{C_t} = tanh(W_c \times [h_{t-1}, x_t] + b_c$

3. **Cell State Update**: Combines the old cell state (C_{t-1}) and the new candidate values to update the cell state:

   $C_t = f_t \times C_{t-1} + i_t \times \overline{C}_t$

4. **Output Gate** (o_t) and **Hidden State** (h_t): Determine the next hidden state, representing the LSTM's output.

   $o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o$

   $h_t = o_t \times tanh(C_t)$

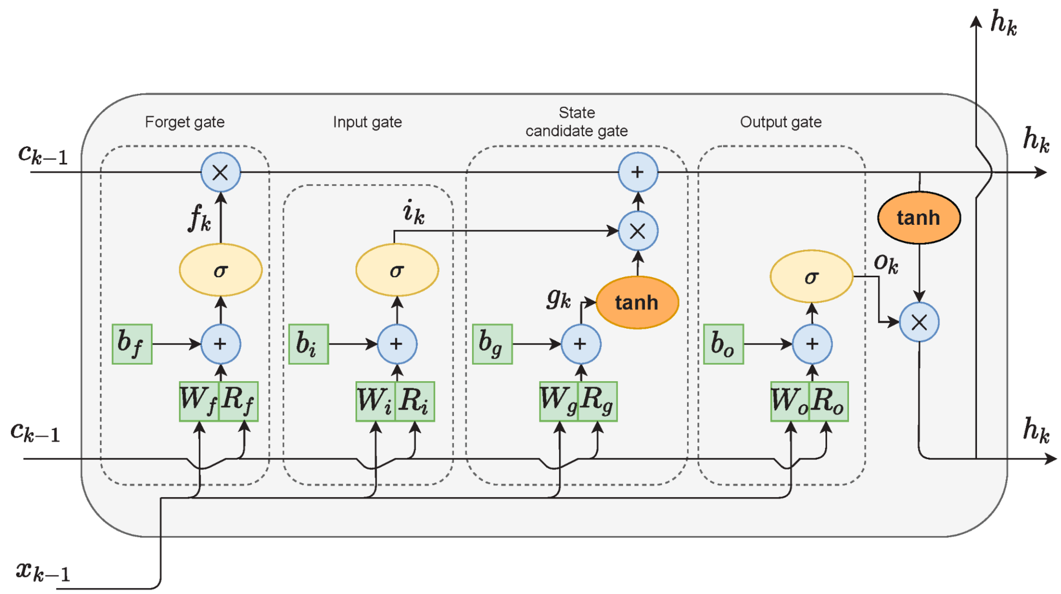
Where σ denotes the sigmoid activation function, tanh is the hyperbolic tangent function, W and b represent the weights and biases of each gate, respectively, h_{t-1} is the previous hidden state, is the input at time *t*, and C_t is the current cell state (Zarzycki & Ławryńczuk, 2021).

In the context of estimating the PC, LSTMs leverage this architecture to effectively model the intricate, time-varying relationship between macroeconomic variables. By capturing patterns over various time horizons, LSTMs provide a nuanced understanding of inflation dynamics and unemployment trends, facilitating more accurate and robust economic forecasts. This advanced modeling capability, rooted in the LSTM's mathematical structure, underscores the potential of NN to revolutionize the analysis of economic phenomena, offering a powerful tool for policymakers and researchers alike.

LSTM networks excel in capturing temporal dependencies in economic data, making them particularly suited for forecasting the PC, which involves understanding dynamic relationships over time. Unlike DSGE models, which are grounded in economic theory and rely on assumptions about the behavior of agents and markets, LSTMs can directly learn from data without pre-specified equations, offering flexibility and the ability to model nonlinear relationships.

However, DSGE models provide insights into the underlying economic mechanisms and policy implications, which purely data-driven approaches like LSTMs may not explicitly reveal. While LSTMs may outperform DSGE models in raw predictive accuracy, especially in complex and non-stationary economic environments, DSGE models offer explanatory power and theoretical insights into economic dynamics. The choice between these methodologies depends on the research objectives: whether the focus is on prediction accuracy or on understanding the economic theory underlying the relationships between variables.

Figure 1: LSTM Model Logic



Note: Visualization was made by Zarzycki & Ławryńczuk (2021)

The utilization of LSTM networks for predicting the PC and estimating the OG marks a significant advancement in the application of ML to economic forecasting. LSTMs, with their inherent capacity to capture long-term dependencies in time series data, present unparalleled predictive power in modeling complex economic phenomena. However, a notable limitation of

LSTMs, and indeed many deep learning models, lies in their lack of interpretability (Park & Yang, 2022).

LSTMs excel in handling the sequential nature of economic data, making them particularly suited for forecasting tasks where historical patterns and temporal dynamics play a crucial role. Their structure allows them to learn from long sequences of data without the risk of vanishing or exploding gradients, a common issue in traditional recurrent NN (RNNs). This capability ensures that LSTMs can effectively model the intricate relationships between macroeconomic indicators over time, potentially offering superior forecasts of inflation rates and unemployment figures in relation to the PC (Ozyegen et al., 2021).

Despite their strengths, LSTMs fall short in providing clear insights into how and why they arrive at specific predictions. The "black box" nature of these models obscures the causal relationships and economic rationale underlying their forecasts. This lack of interpretability is a significant drawback, especially in the field of economics, where understanding the mechanisms driving phenomena like the PC is as important as the accuracy of the forecasts themselves (Ozyegen et al., 2021). Policymakers and economists often require models that not only predict accurately but also elucidate the impact of various economic policies and shocks on inflation and unemployment .

The ideal economic forecasting model would combine the predictive accuracy of LSTMs with the interpretability of traditional econometric models. Achieving this blend would enable not only precise forecasts but also a deeper understanding of the economic dynamics at play. Recent research efforts have focused on enhancing the interpretability of ML models, including techniques such as model-agnostic methods, attention mechanisms, and explainable artificial intelligence (XAI) approaches. However, the complexity and non-linearities captured by LSTMs present ongoing challenges to these efforts.

In summary, while LSTMs offer state-of-the-art predictive capabilities for modeling the PC and the OG, their limited interpretability restricts their utility in economic analysis and policy-making.

2.5 HNN as a combination of best from LSTM and DSGE

HNN are a type of Artificial NN that were developed by Philippe Goulet Coulombe (2022) specifically for modeling economic relationships. HNNs have a unique architecture that consists of multiple hemispheres, each of which is responsible for modeling a different aspect of the economic relationship. The first hemisphere is responsible for modeling the relationship between inflation and inflation expectations. The second hemisphere is responsible for modeling the relationship between inflation and real activity. With the final hemisphere modeling the relationship between inflation and commodities.

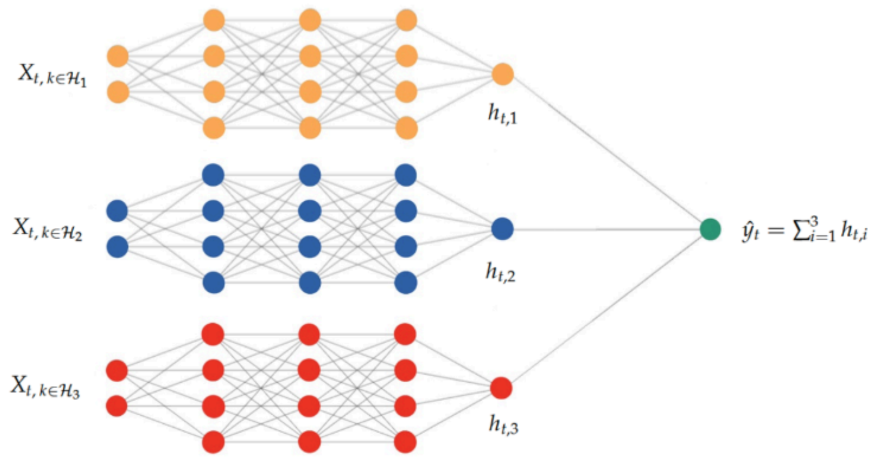
In his study, Coulombe (2022) applied HNNs to forecast inflation in the United States. His research employed quarterly data spanning from 1960 to 2021, encompassing a diverse range of economic conditions and structural changes. The HNN model outperformed conventional linear regression models, demonstrating superior accuracy and robustness.

Furthermore, Coulombe's (2022) study highlighted the interpretability of HNNs, a crucial aspect for gaining insights into the underlying economic dynamics. The model not only provided inflation forecasts but also decomposed the contributions of various factors, including inflation expectations, real activity, commodity prices and OG, contributing to the overall inflation prediction.

According to the Report on Monetary Polity by the Bank of England (2023), the policy makers and researchers employed the HNN approach addressing OG calculation (Bank of England, 2023). This fact supports the feasibility of such an approach applied not only to the US, but also to the Eurozone, since the US macroeconomic data collection practices and the data available are more similar to European. Also, that implies that the approach is relatively reliable and has sufficient external validity to be applied to different regions.

The research introduces HNN as a superior alternative to LSTM models for estimating the PC and the OG, emphasizing HNNs' unique blend of high predictive accuracy with enhanced interpretability. Unlike traditional LSTMs, which are powerful in prediction but lack transparency, HNNs incorporate an interpretable layer before the final output, enabling a clearer understanding of how input features influence predictions. This architectural innovation not only maintains the model's accuracy but also facilitates a level of interpretability that rivals that of DSGE models, offering a significant advancement in economic forecasting. The HNN's design, by providing insights into the contributions of various economic factors to inflation predictions, challenges the interpretability limitations commonly associated with deep learning models and positions HNNs as a formidable tool for economic analysis, balancing the need for accuracy with the demand for model transparency.

Figure 2: Hemisphere Neural Network's Architecture



Note: HNN illustration by Coulombe, P. G. (2022)

3. Methodology

3.1 Data-related preprocessing

3.1.1 Exploratory data analysis

Comprehensive exploratory data analysis is conducted with the data saved in the intermediary step of the pre-processing pipeline. This analysis involves the examination of data distributions and the application of Principal Component Analysis and Autoencoder Analysis to unveil latent structures within the dataset. According to Huang; Jiang; Li; Tong & Zhou (2022), PCA and Autoencoders are particularly suited for handling macroeconomic data due to their inherent ability to manage high-dimensionality and extract meaningful patterns from complex datasets (Huang; Jiang; Li; Tong & Zhou, 2022). Macroeconomic data, characterized by its multifaceted relationships among data and present a challenge for traditional analysis techniques due to its complexity and the potential for multicollinearity among variables, therefore dimension reduction is efficient in forecasting and extract meaningful patterns from complex datasets (Hauzenberger & Huber & Klieber, 2022). This step aims to uncover underlying patterns that may influence the inflation forecasting model. Based on this analysis, feature importance and selection can be hypothesized and later applied to the HNN model.

3.1.2 Data Transformation & Test/Train split

After collecting all of the necessary data, it needs to go through a data pre-processing pipeline, so that it is easily interpretable by NN. The initial step in this process is to clean the data by making sure all of the macroeconomic features have the same starting and end points, same granularity, as well as there are no missing values.

When the data is cleaned it is ready for the next step which is feature generation. For this research we decided to pursue with the same features that Coulombe (2022) used in his research, i.e., for every independent variable 4 new lagged features are added ranging from χ_{t-1} to χ_{t-4}. In addition to these newly generated lagged features, we incorporated Moving Average Rotation of X (MARX) variables (P. Columbe, M. Stevanovic, 2021) for all macroeconomic variables for *l =1* and *l=3*, summarized by this equation:

For each lag l and each variable v:

$$MARX_{l,v} = \frac{1}{l-1} \sum_{i=1}^{l-1} X_{lag_{i,v}}$$

Where: $X_{lag_{i,v}}$ represents the *i*-th lag of variable *v* in the original **mat_x** matrix.

Following the data split, a standardization process is implemented using *scikit-learn's* (n.d.) built-in StandardScaler module. This module standardizes each feature by subtracting the mean and then scaling it to unit variance. The formula representing this standardization is expressed as follows:

$$\chi_{scaled} = \frac{\chi_i - \mu}{\sigma}$$

The concluding stride in the data pre-processing pipeline involves the conversion of the data into the *tensor-data* (n.d.) format. Subsequently, the data is split into five distinct hemispheres, namely: Real Activity, Short-run Expectations, Long-run Expectations, Commodities, and Inflation Spillover Effects. This partitioning strategy is necessary for the HNN architecture.

In the process of compiling the dataset comprising EU macroeconomic variables for our analysis, we encountered a limitation in the temporal range of available data. Specifically, our dataset encompasses observations commencing from the first quarter of 1999, which imposes constraints on the longitudinal depth of our analysis. This temporal limitation becomes particularly salient when contrasted with the dataset utilized by Columbe (2022) in his seminal paper. The Federal Reserve Economic Data (FRED) dataset for the United States, as employed by Columbe, boasts a substantially larger corpus of observations, nearly tripling the volume of data points available in our European dataset.

Given this disparity in dataset comprehensiveness, we were compelled to implement strategic adjustments in our data preprocessing approach. One such adjustment pertained to the transformation of lagged values within our dataset. Recognizing that the creation of additional lags inherently constricts the observational space available for analysis, we opted for a more constrained grid of lagged value transformations. Specifically, whereas a moving average spanning 3 and 7 lags was utilized in Columbe's approach, we tailored our methodology to incorporate a narrower spectrum of 1 and 3 lags.

This methodological recalibration was informed by a deliberate trade-off between explanatory power and analytical robustness. The reduction in the breadth of lagged values was a strategic decision to mitigate the potential dilution of the dataset's informative capacity, a risk accentuated by our limited data range. This approach aims to optimize the balance between capturing relevant temporal dynamics and maintaining a robust observational framework, thereby ensuring the reliability and validity of our subsequent econometric analyses.

Another important point to note regarding data preprocessing is that our feature space is more narrow than the one defined by Columbe (2022) due to the reason that some of the macroeconomic variables available for US do not have a close EU proxy available on Eurostat.

Before pursuing modeling Neural PCs we made the necessary transformations to make our data stationary and ready for econometric analysis. We used the Augmented Dickey Fuller test to find the order of integration for each macroeconomic variable and defined inflation in line with Columbe’s (2022) paper from the consumer price index as:

$$\pi_{t+1} = \Delta log(CPI_{t+1})$$

A majority of the macroeconomic variables under study exhibited adherence to a 99% confidence interval upon first differencing, indicating statistical stationarity at this level.

However, six variables within the 'real activity' hemisphere—namely, Unemployment Rate (UR), Unemployment Gap (UnemplGap), Average GDP Growth (AVGDGP), Average Government Debt to GDP Ratio (AVGDGT), IMF World Economic Outlook Gap (IMFWEOgap), and Non-Accelerating Inflation Rate of Unemployment (NAIRU)—necessitated a second-order integration to attain stationarity. While Columbe (2022) does not explicitly delineate the procedures for data preparation for NN analysis in his study, it is plausible to deduce that he employed Augmented Dickey-Fuller tests for stationarity assessment or relied on the recommended order of integration as suggested by the FRED in more detail described by Labonne & Chen (n.d.).

Given the constrained dataset size, this study employs an 87.5% to 12.5% training and testing data split to maximize the training data available for the benchmark and HNN models. Subsequently, the dataset is partitioned into training and test sets. Given the comparatively diminished size of the European dataset vis-à-vis the United States, we opt for a reduced test set, constituting 15% of the entire dataset. This allocation ensures a validation period of approximately 2.5 years. All the models explored in this study are trained on quarterly macroeconomic data from 2000-Q1 to 2015-Q3, which equals 63 total observations in the training sample, leaving 9 observations for the test sample. This approach is particularly pertinent for deep learning models that necessitate substantial datasets to effectively discern complex patterns.

3.2 Model development

3.2.1 Linear model development

The benchmark OLS linear regression model serves as a foundational component in evaluating the performance of the proposed HNN model. This widely adopted linear model provides a baseline against which the HNN's inflation prediction capabilities can be compared, more precisely its RMSE. Linear regressions simplicity and interpretability make it a preferred choice in economic forecasting, allowing for a clear understanding of the relationship between input features and the target variable, in this case, inflation.

Results from the linear model serve as a critical baseline for comparing and understanding the HNN's forecasting capabilities. Discrepancies between the linear model's predictions and actual inflation values pinpoint areas where the NN may offer improvements. Additionally, insights gained from the linear model contribute to interpreting the HNN's findings, shedding light on the relative importance of different features in influencing inflation dynamics within the Eurozone.

3.2.2 ARIMA model development

In evaluating predictive models for inflation, an ARIMA model serves as an additional benchmark to compare against the HNN models. This section delineates the development and diagnostic evaluation of the ARIMA model employed in this study.

The process of ARIMA modeling commenced with the stationarity assessment of the time series data, which is a prerequisite for the application of ARIMA. The Augmented Dickey-Fuller (ADF) test was utilized for examining the presence of a unit root, which indicates if the series requires differencing to achieve stationarity. Conversely, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was employed to corroborate the stationarity of the differenced series, thereby affirming the order of integration, d, in our ARIMA model.

Subsequent to establishing stationarity, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were meticulously analyzed to ascertain the autoregressive order p and moving average order q. The ACF plot provided insights into the total correlation between different lags of the series, while the PACF plot isolated the direct effect of past lags. This analysis facilitated the identification of the appropriate lag values for the ARIMA model parameters.

3.2.3 NN development

Finally, a HNN architecture is developed, by following Columbe’s (2022) research, where 3 separate LSTM models with an exogenous time trend work in an ensemble with an intermediary contribution layer in which all of the produced contribution coefficients are summed up to get the final inflation value. (Appendix 1)

In this NN architecture each hemisphere processes its subset of input features through its respective NN layers, where the output of each hemisphere can be thought of as a contribution based on its specific set of input features. The final output $\hat{y}$, is the sum of the contributions from all hemispheres. Here the contributions from the hemispheres are adjusted (multiplied) by their respective varying time trends to modulate the output.

The novelty from Columbe's (2022) NN architecture design comes from the fact that it offers extra interpretability compared to traditional NN design, i.e., here we can predict not only inflation itself but also all of its drivers:

- **Components:** The adjusted outputs from all the hemispheres. These are the contributions from each hemisphere after being modulated by their trends.
- **Trends:** The trends computed from hemispheres
- **Gaps:** Raw, time unadjusted contributions of each hemisphere.

When it comes to the HNN's parameters we will employ a mix of parameters suggested by Columbe (2022) and a set of parameters that will come as a result of hyperparameter optimization, therefore achieving a tailored HNN architecture for the EU macroeconomic space.

Besides pure hyperparameter optimization, we advocate for the augmentation of the HNN architecture through the inclusion of an Inflation Spillover Effects Hemisphere. As delineated by Hall, Tavlas, and Wang (2023), the macroeconomic environment of the European Union is susceptible to the influence of inflation spillover effects emanating from the United States. In order to examine this hypothesis and offer a more comprehensive account of the variability in inflation, we propose the incorporation of an Inflation Spillover Hemisphere specifically designed to capture the effects cascading from economically significant regions such as North America and Asia (Appendix 2).

3.3 Model benchmarking

The section on "Model Comparison against Benchmarks" embarks on a critical examination of the performance of HNN models, in estimating the PC, against the backdrop of traditional econometric approaches. This analytical endeavor is pivotal, considering the PC's significant role in economic theory and policy-making, where it delineates the inverse relationship between inflation and unemployment rates. The juxtaposition of advanced ML methodologies against conventional models not only underscores the evolution of economic analysis techniques but also offers insights into their relative efficacy, adaptability, and interpretability in capturing complex economic dynamics.

In comparing these models, several metrics and methods stand out for their ability to provide a comprehensive evaluation. Among these, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Mean Absolute Percentage Error (MAPE) are paramount for assessing forecast accuracy. Additionally, the AIC and the BIC offer insights into model fit while penalizing model complexity, making them particularly relevant for evaluating the trade-off between model simplicity and predictive power. For this study, the focus will be primarily on the RMSE metric, given their direct interpretation in terms of forecast accuracy and their widespread acceptance in economic forecasting literature. These metrics are not only intuitive but also allow for a direct comparison between the predictive performances of NN and traditional econometric models. By evaluating the models on these grounds, the research aims to ascertain whether the increased complexity and computational demand of NN are justified by a corresponding enhancement in forecasting accuracy and insight into the economic phenomenon under investigation.

Furthermore, the comparison will extend beyond mere predictive accuracy to encompass aspects of model interpretability and computational efficiency. This holistic approach ensures that the adoption of advanced ML techniques, such as LSTMs and HNNs, is evaluated not just in the context of their immediate performance but also in terms of their broader applicability and contribution to economic analysis.

When all the out-of-sample (OOS) forecasts are made and collected by the benchmark models and all of the versions of the HNN model, they will be evaluated in terms of OOS RMSE:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\widehat{y}_i - y_i\right)^2}{n}}$$

Besides the RMSE measure, interpretability will also be compared by looking at the components, trends and gaps retrieved from the HNN model. Finally, after evaluating PC estimations in the EU, we will also compare our EU HNN model to the one developed by Columbe in 2022 for US data.

3.4 Training & Predicting pipeline

The training process can be defined as supervised learning: a model of outlined architecture assigns random values to weights between nodes, and using gradient descent adjusts them based on training data. When trained, the model is capable of predicting $\hat{y}$ (inflation) based on new input data (e.i., if the training period ends in 2015-Q3, then 2015-Q4 onwards input data (X_1, X_2, ... X_i) is "unfamiliar" for the model). The model performance measured on data from 2015-Q4 until 2017-Q4 since it is available to us, but not available for the model, therefore we can assess the model accuracy by comparing the model's prediction of 2015-Q4 onwards data (unfamiliar for the model) with the actual fully correct data which we know. The model itself forecasts the inflation one quarter ahead, and it is not retrained after it was initially trained, as by design LSTM NN trained by supervised learning, are supposed to retrain after new observations incoming due to complexity of training process. Otherwise, techniques of "teaching and preparing" model for out of sample data through the training act to avoid the necessity to retrain the model mentioned above.

If the model is trained successfully, and accuracy is sufficient, having necessary input variables for later periods (e.g. 2018-Q1 onwards), it will be predicting them. Since we do not have respective data from closed databases later than 2017-Q4, we are not able to test if the model would predict post-COVID or post Russia-Ukraine war inflation rise. Nevertheless, we measured the model accuracy based on the test-sample period.

3.5 Contribution to HNN & NN modifications: Hyperparameter optimization

Hyperparameters, distinct from model parameters that are learned during training, include the number of neurons in each layer, the number of hidden layers, learning rate, dropout rate, batch size, and the number of training epochs. The hyperparameters significantly influence the model's ability to capture complex nonlinear relationships inherent in economic data. Additionally, hyperparameter optimization is crucial to avoid potential overfitting and selecting the best[2] possible architecture for the model. In this context, the architecture of the proposed HNN is characterized by following hyperparameters:

- *Number of neurons and hidden layers (optimized)*
- *Number of inputs in each hemisphere*
- *Dropout rate in each layers (optimized)*
- *Learning rate*
- *Maximum number of training epochs (optimized)*
- *Number of epochs before stopping once your loss starts to increase*
- *Number of quarter the use in block bootstrap*
- *Number of bootstrap (optimized)*
- *In-sample sampling rate*
- *Early stopping criteria tolerance (optimized)*

The optimization of such hyperparameters is approached through methodologies that span from grid search, random search, to more sophisticated techniques like Bayesian optimization, Ray Tune, or Optuna optimization. For this study, the focus is on leveraging a systematic methodology that iterates over a predefined hyperparameter space to identify the configuration that yields the best performance in terms of predictive accuracy and model interpretability. This process is crucial for ensuring that the HNN model is not only adept at capturing the underlying economic phenomena but is also efficient and generalizable across different datasets. Mathematically, the general optimization process can be framed as:

[3] $$minimize(\theta):\ L(y,\ f(X;\theta))\ +\ \lambda R(\theta)$$

The objective function J(θ) quantifies the performance of a model given a set of hyperparameters θ. In the context of supervised learning, this often corresponds to the validation loss.

---

[2] The term "best" denotes a configuration that effectively generalizes across unseen data while avoiding overfitting. The goal is to identify a precise equilibrium, ensuring superior predictive capabilities without unnecessary complexity or overfitting.

[3] where L denotes the loss function measuring the discrepancy between the observed economic indicators y and the model predictions $f(X;\theta)$ given inputs X and hyperparameters θ, R represents a regularization term to prevent overfitting, and λ is a regularization coefficient.

$$J(\theta) = L(y_{val}, \widehat{y}_{val}(\theta))$$

The hyperparameter space Θ is defined as the multidimensional space wherein each dimension represents a hyperparameter to be optimized. The complexity of Θ is determined by the number of hyperparameters and their potential values, often resulting in a high-dimensional, non-convex landscape that HPO algorithms must navigate. The hyperparameter space Θ for HNN, particularly as defined in the provided model architecture, encapsulates a multi-dimensional landscape wherein each dimension corresponds to a specific hyperparameter. The hyperparameter space Θ can be mathematically described as a Cartesian product of the individual hyperparameter domains:

$$\Theta = \Theta_{n_{neurons}} \times \Theta_{n_{features}} \times \Theta_{d_{rate}} \times \Theta_{\alpha} \times \Theta_{e_{max}} \times \Theta_{p_{wait}} \times \Theta_{b_{options}}$$

Each domain $\Theta_{hyperparameter}$ represents the range of possible values for a given hyperparameter. For instance, $\Theta_{n_{neurons}}$ may include various configurations of layer sizes beyond the initially defined [400] × 3 and [100] × 3, whereas Θ_{α} spans a continuum of learning rates that could potentially enhance model performance. Given a set of hyperparameters $\theta \in \Theta$, the objective function *J(θ)* can be defined as the discrepancy between predicted and actual economic indicators, typically measured by the mean squared error *(MSE)* or a similar metric:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i(\theta))^2$$

where y_i are the actual values, $\widehat{y}_i(\theta)$ are the predicted values given hyperparameters θ and n is the number of observations. The optimization task seeks to solve:

$$\theta^* = argmin_{\theta\in\Theta} J(\theta)$$

This involves evaluating J(θ) across potentially vast regions of Θ, necessitating efficient optimization strategies such as Bayesian optimization or gradient-based methods, which can navigate the search space more effectively than exhaustive search methods.

In the context of HNNs, the complexity of the model architecture — comprising multiple LSTM layers tailored to capture temporal dependencies in economic data — requires careful consideration of hyperparameters such as the number of layers, number of neurons per layer, learning rate, and dropout rate. The interdependencies among these hyperparameters and their impact on model performance underscore the necessity for a nuanced optimization approach that can adaptively refine the search based on iterative feedback from model evaluations.

Incorporating regularization strategies, such as L1 or L2 regularization, into the optimization process can further mitigate the risk of overfitting, ensuring that the model generalizes well to unseen data. In Python, leveraging libraries such as PyTorch for NN implementation and Ray Tune or Optuna for hyperparameter optimization allows for an efficient and scalable search for optimal hyperparameters.

Optuna represents a significant leap forward in the domain of hyperparameter optimization, particularly in the context of advanced NN architectures such as the ensemble LSTM models encapsulated within HNN. This Python-based open-source framework is designed to automate the optimization process, offering an efficient and flexible approach to tuning the hyperparameters that govern the performance of ML models.

For HNNs, which involve an intricate ensemble of LSTM models tailored for comprehensive economic forecasting, Optuna provides a particularly adept solution. Its design philosophy emphasizes simplicity, efficiency, and the ability to scale across multiple dimensions of hyperparameters. By adopting a define-by-run style for optimization, Optuna allows for dynamic generation of the hyperparameter search space, enabling researchers to intricately tailor the optimization process to the unique architecture of HNNs.

Optuna's core functionality revolves around its efficient optimization algorithms, including Tree-structured Parzen Estimator (TPE), CMA-ES, and Random Search, among others. TPE, in particular, stands out for its applicability to complex models like HNNs. It models the probability of hyperparameters given the outcomes of previous trials to intelligently propose the next set of hyperparameters that are likely to yield improved results. Implementing Optuna within the HNN training workflow involves defining an objective function that encapsulates the model training and validation process, returning the performance metric to be optimized. Optuna then iteratively explores the hyperparameter space, guided by the outcomes of each trial, to find the configuration that minimizes the objective function.

The integration of Optuna with PyTorch models, such as those used for HNNs, is straightforward, benefiting from Optuna's comprehensive support for various ML frameworks. This compatibility ensures that researchers can focus on the nuances of model architecture and economic theory, rather than the technical complexities of hyperparameter tuning.

Given the complexity and computational demands of optimizing HNNs, Optuna's approach to hyperparameter optimization is not just suitable but highly applicable and recommended. Therefore, upon comprehensive consideration, Optuna is identified as the most advantageous framework for this research. The decision is predicated on Optuna's algorithmic efficiency and its proficiency in conducting a dynamic and informed exploration of the hyperparameter space. This is particularly relevant for HNNs, where the optimization process must be meticulously calibrated to enhance model accuracy without incurring prohibitive computational costs.

3.6 Contribution to HNN & NN modifications: Inflation spillover hemisphere

Theoretical models like the International Spillover of Inflation theory suggest that inflation in one country can affect another country's inflation through import prices, demand conditions, and monetary policy stances (Auer, Levchenko, Sauré, 2017). Incorporating this theory into the HNN model requires an architecture that can differentiate and learn from these diverse inflationary signals.

According to Khandokar, Aviral, Tiwari, Humaira & Kazi (2021), the domestic inflation figures have a statistically significant impact from the external inflation spillover even within G7 countries based on the largest trade partners (Khandokar, Aviral, Tiwari, Humaira & Kazi, 2021). The HNN, with its dual capability for precision and interpretability, offers a promising framework to incorporate the spillover effects into the general inflation forecasting model. Overall, there are significant grounds to include spillover effects for the Eurozone from the US, as the US being a largest trade partner, impacts the prices for a range of EU goods and services.

4. Data

Our analysis is based on a database, derived from a comprehensive paper issued by the researcher from the Bank of Estonia - Dmitry Kulikov with help of Nicolas Reigl (2019), that went mainly to the issue of how to estimate the PC with traditional models, but with the idea of grouping variables into groups similar to hemispheres. This database contains a wide range of economic variables, selected for the Eurozone, especially as it relates to PC analysis. To keep the highest granularity and the most frequent observations possible keeping the amount of data sufficient for model capturing multiple nonlinearities, the data is quarterly collected and used. The variables and their corresponding data sources are as follows:

Inflation Measures: The database includes the Harmonized Index of Consumer Prices (HICP) excluding energy and a more refined version excluding energy, food, alcohol, and tobacco, both observed quarterly from Q1 1999 to Q4 2017, with 76 observations each, sourced from Eurostat. These inflation measures are crucial for understanding the core inflation trends in the Eurozone, independent of volatile components like energy and food prices.

Inflation Expectations: A series of inflation expectations are included, sourced from the Survey of Professional Forecasters (SPF), ECB Statistical Data Warehouse (SDW), Consensus Economics, Eurozone Barometer, Bloomberg, and ECB. These range from short-term (one quarter ahead) to medium-term (two years ahead) expectations, covering periods from as early as Q1 1999 to Q4 2017. The incorporation of inflation expectations is vital, as they are a key component in understanding future inflation trajectories, which are central to the PC relationship.

Economic Slack Measures: Various measures of economic slack, such as the real GDP growth rate, different calculations of the OG from the ECB, AMECO, IMF, and specific studies like Lenza and Jarociński (2018), as well as measures of labor underutilization and unemployment rates, are included. These variables, observed from Q1 1999 to Q4 2017, are essential to evaluate the level of economic activity in relation to its potential, which is a fundamental aspect of the PC theory.

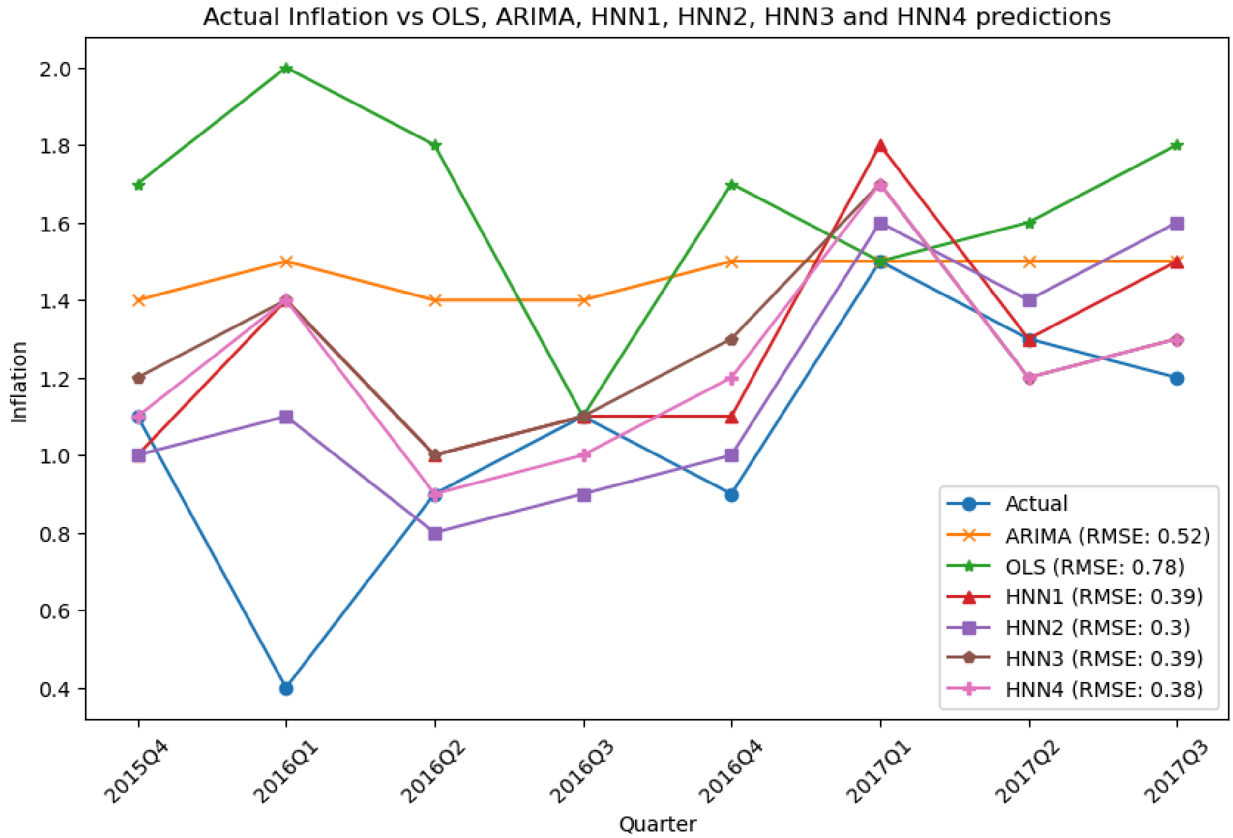
Inflation spillover: According to Fojtíková (2014), interdependence between the Eurozone and the United States, with the latter standing as the Eurozone's largest trade partner, it becomes imperative to consider the impact of US inflation on the Eurozone's economic dynamics (Fojtíková, 2014). This study incorporates an analysis of inflation spillover effects, examining how inflation trends in the United States influence inflation within the Eurozone by using Core CPI for the US on a quarterly basis for the same period, enriching the data to incorporate an additional hemisphere accounting for the inflation spillover.

5. Results

The application of the HNN in estimating the PC for the EU has yielded very promising results. The study entailed a comparative analysis encompassing 4 distinct HNN model configurations: the first being HNN1, which is a HNN adopting the hyperparameter specifications as delineated in Columbe's 2022 study; the second, HNN2, represents an iteration of the HNN wherein the hyperparameters have been meticulously optimized to align more closely with the specific economic dynamics of the EU; the third, HNN3 model with an additional inflation spillover effects hemisphere and default hyperparameters and HNN4 the final HNN model with an additional inflation spillover effects hemisphere and optimized hyperparameters. To better assess the predictive power of HNN models, we developed a simple Linear Regression Model and univariate ARIMA model, serving as a benchmark for traditional econometric inflation prediction approaches.

The evaluative focus of this comparative study was primarily centered on the models' proficiency in generating out-of-sample inflation predictions and interpretable components. These predictions were compared against the actual observed inflation rates within the EU, offering a tangible measure of each model's forecasting accuracy. The out-of-sample predictions of each model can be seen in the graph below:

Figure 5: Actual Inflation compared to Linear Regression, ARIMA, HNN1, HNN2, HNN3 and HNN4 out-of-sample predictions. (period from 2015-12-31 to 2017-09-30, one observation equal to one quarter)



Note: Created by the authors.

5.1 Hemisphere Neural Networks

The substantial disparity in the volume of macroeconomic observations between the EU and the US, as delineated in the preceding subsection, necessitates more than mere data transformation adjustments for accurate inflation forecasting within the European context. The efficacy of NN, and indeed other predictive models, is intrinsically linked to the quality and volume of the data at hand. This poses a substantial risk of informational insufficiency for the HNN model, potentially impeding its capacity to discern and interpret underlying data relationships effectively. Consequently, a NN architecture of reduced complexity might prove more efficacious in this scenario. Accordingly, we have embarked on developing multiple iterations of the HNN model, each variant characterized by a distinct configuration of hyperparameters. This iterative process aims to identify the most suitable architectural framework for the European macroeconomic landscape.

To address the issue of informational insufficiency due to the asymmetric data volumes between the EU and the US, an innovative approach was adopted in the form of an additional HNN variant. This model incorporates a specialized inflation spillover hemisphere designed to capture the transnational inflation spillover effects originating from significant external economies, particularly the US and China.

5.1.1 Base HNN1 Model

The base model HNN1 architecture was crafted to be in line with the one described by Columbe (2022) with 3 hemispheres containing macroeconomic variables related to real activity, short-run expectations, commodities and an additional long-run expectations trend hemisphere. For the three empirical hemispheres and the long-run expectations hemisphere we constructed 3 hidden layers with 400 neurons in each layer and added 3 hidden layers with 100 neurons in each layer to capture the trend for each of the three empirical hemispheres and an input layer equal to the number of features in the data.

We set the following list of hyperparameters for the HNN1 model:

- **Dropout rate = 10%** (one in ten inputs will be randomly excluded from each training loop)
- **Learning rate = 5%** (indicating the pace at which the gradient is descended)
- **Number of epochs = 500** (the number of training loop iterations of the training dataset)
- **Patience = 50 epochs** (number of epochs with no improvement after which training loop will be stopped)
- **Block size = 6** (number of quarters used in block bootstrap)
- **Number of bootstrap iterations = 300** (Number of iterative random samples taken from the data to generate information about the population)
- **Bootstrapping option = 2** (Option 1 refers to traditional and Option 2 refers to block bootstrapping)
- **Sampling rate = 87.5%** (In-sample sampling rate)
- **Tolerance = 1%** (Early stopping criteria tolerance (Mean Squared Error)

The HNN1 model generated a RMSE of 0.39. When training the model we also observed that during the 300 bootstrapping rounds the training process was always stopped early long before reaching the 500 epoch mark, meaning that when further developing a HNN model with optimized hyperparameters we can set the maximum amount of epochs much lower as in the first training case the maximum number reached was just 236.

5.1.2 Optimized HNN2 model

In the quest to adapt the HNN model more effectively to the European Union's economic data, we have developed an advanced iteration, designated as HNN2. This iteration represents a theoretically superior version of its predecessor, HNN1, achieved through a process of meticulous hyperparameter optimization.

Our approach to enhancing HNN2 involved a comprehensive evaluation of various hyperparameter optimization techniques. After an extensive comparative analysis, we elected to employ the Optuna framework, a decision influenced by its robustness and flexibility in handling complex optimization tasks. Within this framework, we crafted a custom objective function, enabling a more targeted optimization process. We embarked on a series of 100 studies to determine the optimal set of hyperparameters, ensuring a thorough exploration of the parameter space.

Given the resource-intensive nature of hyperparameter optimization, it was imperative to carefully select parameter ranges for optimization. Our strategy was to focus on parameters that would significantly impact the model's performance, avoiding adjustments to parameters with marginal effects or no effects at all. Consequently, we maintained specific parameters at fixed values: the maximum number of epochs was set at 350, the number of bootstrapping rounds at 300 (employing block bootstrapping), the tolerance level at 1%, and the sampling rate at 85%.

The core of our optimization process centered around several key hyperparameters, deemed critical for the model's efficacy. These included the number of neurons in each hidden layer, the dropout rate, the learning rate, and the block size. Adjusting these parameters was expected to substantially enhance the model's predictive accuracy and efficiency, particularly in the context of the complex and dynamic economic environment of the European Union.

Upon the culmination of an extensive hyperparameter optimization process, encompassing 100 individual studies, we have successfully refined the HNN2 model, achieving a RMSE of 0.3. The other optimal hyperparameters are as follows:

- **Hemisphere 1:** 403, 444, 255 neurons across three hidden layers
- **Hemisphere 2:** 163, 269, 210 neurons across three hidden layers
- **Hemisphere 3:** 449, 181, 355 neurons across three hidden layers
- **Trend Hemisphere 1:** 107, 153, 125 neurons across three layers
- **Trend Hemisphere 2:** 136, 92, 51 neurons across three layers
- **Trend Hemisphere 3**: 183, 109, 71 neurons across three layers
- **Dropout rate = 10%** (unchanged)
- **Learning rate = 5%** (unchanged)
- **Block size = 3** (lower than HNN1)

5.1.3 New Hemisphere HNN3 model

The evolution of the HNN models led to the development of an enhanced variant, denoted as HNN3, which integrates an additional inflation spillover hemisphere. This new hemisphere is aimed at capturing the inflation spillover effects from major global economies, specifically the US and China, and assessing their impact on European inflation forecasts.

The architecture of HNN3 mirrored that of the base HNN1 model, 400 neurons each for the empirical hemispheres, and 100 neurons each for the hidden layers to model the long-run trends. The input layer was once again matched to the number of features in the dataset and the set of hyperparameters were kept consistent with the HNN1 model.

The implementation of the HNN3 model with the additional inflation spillover hemisphere did not yield any significant improvements in forecasting accuracy, as evidenced by the same RMSE of 0.39.

5.1.4 New Hemisphere Optimized HNN4 model

Building on the advancements of the HNN3 model, we introduced the HNN4 model - an iteration further refined through hyperparameter optimization using Optuna. This optimization process led to the identification of an ideal set of hyperparameters that had previously proven to significantly enhance a model's predictive accuracy for European inflation rates.

The HNN4 model diverged from the previous homogeneous structure of hidden layers and neurons across hemispheres. Instead, it embraced a heterogeneous architecture tailored to the unique characteristics of each hemisphere, as determined by the optimization process. This approach resulted in the following configuration:

- **Hemisphere 1:** 492, 140, 406 neurons across three hidden layers
- **Hemisphere 2:** 158, 290, 273 neurons across three hidden layers
- **Hemisphere 3:** 483, 134, 429 neurons across three hidden layers
- **Hemisphere 4:** 109, 169, 107 neurons across three hidden layers
- **Hemisphere 5:** 373, 322, 277 neurons across three hidden layers
- **Trends for Hemispheres 1, 2, 3, and 5**: 71, 54, 120; 92, 79, 189; 133, 139, 160; and 179, 165, 194 neurons respectively across three layers

In this structure, the number of features (n_features) for each hemisphere was set to match the volume of input data, ensuring that each model component received the appropriate dimensionality of input data. The trends were added to hemispheres 0 through 4, with the aim of capturing temporal patterns and enhancing the model's ability to account for time-variant influences. The hyperparameters of the HNN4 model that were optimized are:

- **Dropout rate: 10%** (unchanged)
- **Learning rate: 5%** (unchanged)
- **Block size: 4** (lower than HNN3)

The result of the optimized HNN4 model was a further reduction in RMSE to 0.38, surpassing the performance of HNN3 and highlighting the effectiveness of the tailored hyperparameters. However, it is noteworthy that even with these improvements, the HNN4 model did not achieve the superior performance of HNN2, which remains the best-performing model with an RMSE of 0.3.

This slight improvement with HNN4 illustrates the diminishing returns of successive refinements in model complexity and the importance of balancing the architecture and hyperparameters. While HNN4 presents a more optimized version than its predecessor, it also emphasizes that there might be an upper limit to the predictive gains achievable solely through architectural and hyperparameter adjustments, especially when compared to a model like HNN2 that appears to be more suitably tuned to the dataset characteristics.

5.2 Benchmark Linear Regression Model (inflation-wise)

In the realm of empirical validation, it is important to establish a robust benchmark against which the efficacy of the HNN models can be measured. To this end, we have developed a simple linear regression model, serving as a preliminary benchmark, particularly in evaluating OOS RMSE for inflation predictions.

Given the expansive feature space inherent in our dataset, a key challenge in constructing the linear regression model was the need to mitigate the risk of multicollinearity. This phenomenon, wherein predictor variables in a multiple regression model are highly correlated, can significantly distort the reliability of the model's estimates. In the context of our dataset, this issue was particularly salient, as the variables within each hemisphere of the HNN model are designed to capture various dimensions of similar underlying economic phenomena, hence exhibiting a high degree of correlation.

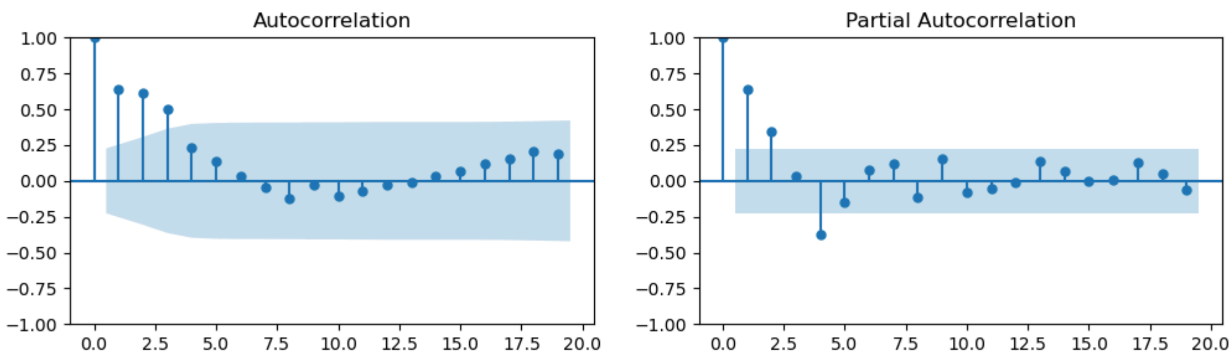
To address this challenge, we employed a methodical variable selection process, choosing only one representative variable from each hemisphere for inclusion in the linear regression model. This approach was aimed at simplifying the model and minimizing multicollinearity, thereby enhancing the interpretability and stability of the regression coefficients.

The simple linear regression model generated an out-of-sample RMSE of 0.78. However, it is crucial to acknowledge the limitations of this linear regression model as a benchmark. Its simplicity, while beneficial for reducing multicollinearity, also renders it somewhat inadequate as a comprehensive comparative standard for the HNN model. The linear regression model's inability to capture the complex, non-linear relationships that the HNN model is designed to identify, diminishes its utility as a practical benchmark. In recognition of this limitation, we envisage the development of a more advanced and nuanced benchmarking model in future research. This prospective model will aim to better encapsulate the intricacies of the economic relationships under study, providing a more fitting and rigorous yardstick against which the performance of the HNN model can be gauged.

5.3 Benchmark ARIMA

The ARIMA model was selected as a benchmark model to evaluate its forecasting accuracy for inflation against the HNN models. Through careful examination of the ACF, the PACF, and the KPSS tests, an ARIMA(2,0,3) model was identified as the most suitable fit for our time series data.

Figure 6: Autocorrelation and Partial Autocorrelation plots of inflation.

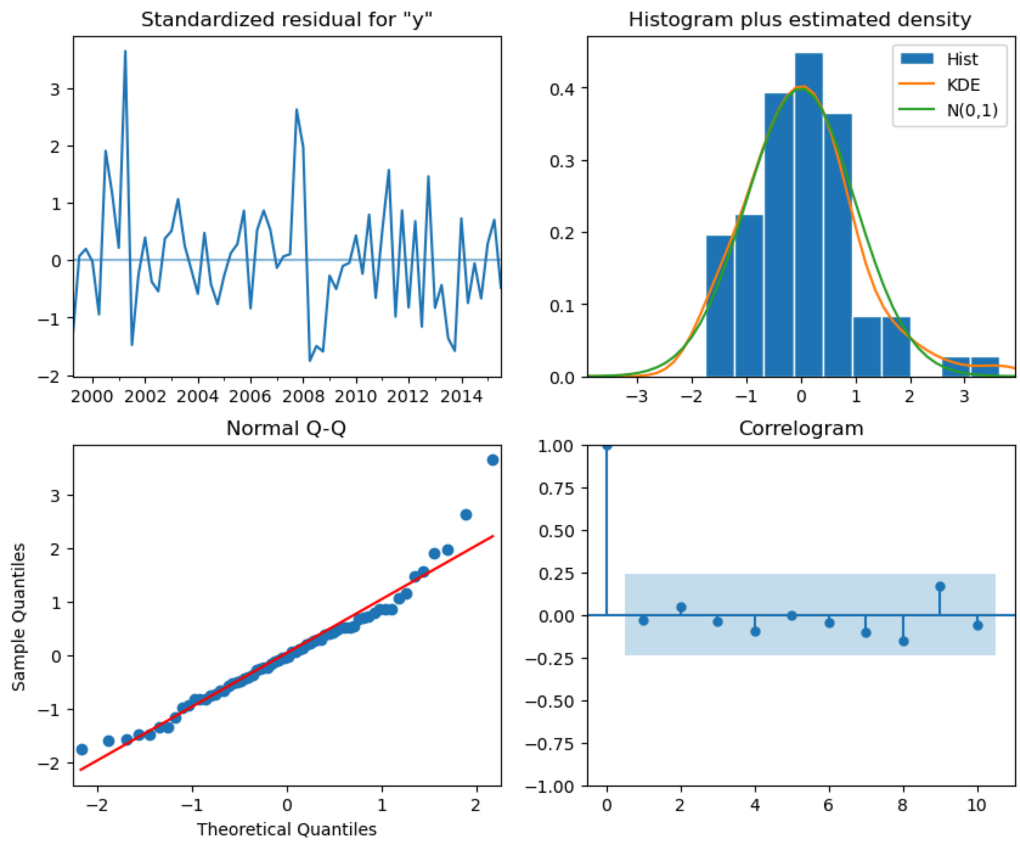


Note: Created by the authors.

The ARIMA(2,0,3) model specification implies that the best fit for the data was achieved with two autoregressive terms (p=2), no differencing (d=0), indicating that the series was stationary, and three moving average terms (q=3). This configuration was deduced to sufficiently capture the autocorrelations within the data while maintaining parsimony in the model's structure.

After fitting the ARIMA(2,0,3) model we also performed diagnostics checks to validate the robustness of the selected parameters. We opted for four total metrics that look at the standardization of residuals through time-series line charts, histograms, Q-Q plots and correlograms.

Figure 7: Standardized residual plot, residual histogram and density plot, Normal Q-Q plot and residual Correlogram.



Note: Created by the authors.

From the figure above we can observe that the time series of standardized residuals suggested no significant autocorrelations. This indicates that the model errors are white noise, which is an indication of a good fit. The histogram of standardized residuals, when compared with a normal distribution, showed a reasonable approximation to normality. This was supported by the Kernel Density Estimate (KDE) plot that closely followed the normal distribution curve N(0,1). The Normal Q-Q plot indicated that the ordered distribution of residuals closely follows the theoretical quantiles of a normal distribution, apart from a few outliers. This alignment suggests that the residuals of the model are approximately normally distributed. And, finally, the correlogram displayed a spike at lag zero, which is expected since the series is being correlated with itself, followed by correlation coefficients within the confidence bounds. This again indicates no significant autocorrelation and supports the adequacy of the model.

The out-of-sample inflation predictions of the ARIMA model, as visualized in the attached prediction plot, revealed an interesting comparison to the actual data. While the ARIMA model captures the general trend, it does not react swiftly to sudden changes in the actual series. This is expected, given the nature of ARIMA models which are typically more conservative in their predictions.

The out-of-sample RMSE of the ARIMA model was computed to be 0.52. This RMSE is superior to the simple linear regression model, which reported an out-of-sample RMSE of 0.78, indicating that the ARIMA model has a better predictive performance in this context. However, it is important to note that the RMSE of 0.52 for the ARIMA model did not outperform the HNN models, which suggests that while the ARIMA model has predictive validity, it may not be capturing some of the more complex patterns and relationships inherent in the data that the HNN models are able to exploit.

6. Analysis of the results

The empirical analysis conducted in this study revealed that all four HNN models: HNN1, HNN2, HNN3 and HNN4, exhibited significant predictive accuracy in forecasting future inflation rates, thereby outperforming the established benchmark linear regression model and the univariate ARIMA model. Among these, the HNN2 model demonstrated superior performance, proven by its achievement of the lowest Root Mean Square Error (RMSE) value of 0.3. This marked a notable improvement in predictive accuracy compared to the other HNN models. By contrast, the linear regression model and the ARIMA model, which served as conventional econometric benchmarks in this analysis, exhibited higher RMSE values of 0.78 and 0.52 respectively, underscoring the enhanced predictive capabilities of the advanced NN models.

As the HNN2 model proved to be the most successful model out of the four HNN models we will continue our analysis by further examining it and comparing it to the HNN model developed by Columbe (2022).

In the final step of our analysis of results we will subject the HNN2 and HNN4 models outputs, encompassing the identified gaps, trends, and contributions, to a comprehensive analysis by taking a closer look at the interactions between macroeconomic variable hemispheres and components. This subsequent phase of the study will delve into the granular aspects of the model's outputs, aiming to extract nuanced insights into the underlying dynamics of inflation. By dissecting these components, the study seeks to shed light on the complex interplay of economic variables and their collective influence on inflationary trends. This in-depth examination is anticipated to yield valuable contributions to our understanding of what the predicted inflation values are composed of, therefore allowing for the interpretability that is so crucial to macroeconomic analysis.

6.1 Comparing the EU data HNN model to the US data HNN model

In the process of benchmarking our HNN2 model against the model developed by Columbe (2022), which was applied to United States macroeconomic data, a critical initial consideration is the disparity in the out-of-sample periods between the two models. Specifically, our EU-based HNN2 model encompasses an out-of-sample period of merely two years, in stark contrast to the decade-long period utilized in the US model. This significant difference in the temporal scope of the datasets necessitates a methodological adjustment to facilitate a more equitable comparison. Accordingly, we have opted to compare our model against Columbe's HNN model with excluded data from the year 2020, to slightly mitigate the disparity in the evaluation time frames.

A comparative analysis of the Root Mean Square Error (RMSE) values between the two models reveals that our HNN2 model exhibits superior forecasting accuracy, with an RMSE of 0.3, as compared to the 0.866 RMSE of the US HNN model. However, it is vital to acknowledge that this comparison, despite the aforementioned adjustment, still confronts inherent limitations due to the fundamental differences in the out-of-sample periods. The US model's out-of-sample period is quintuple that of our EU model, a factor that could significantly influence the model's performance and predictive accuracy.

This discrepancy in the duration of the out-of-sample periods raises important considerations regarding the generalizability and robustness of NN-based forecasting models in macroeconomic contexts. A longer out-of-sample period, such as that in the US model, encompasses a wider array of economic cycles and market conditions, potentially challenging the model's adaptability and resilience. Conversely, the shorter period in our EU model may not fully capture the cyclical and structural complexities of the economy, possibly affecting the generalizability of the results.

In light of these considerations, we should further aim to extend the out-of-sample period of the EU-based HNN2 model to align more closely with the temporal breadth of the US model by collecting more recent data that spans further than 2017. Such an extension would enable a more comprehensive and fair comparison, offering deeper insights into the relative performance and scalability of the models across diverse macroeconomic landscapes.

6.2 Exploring other HNN outputs

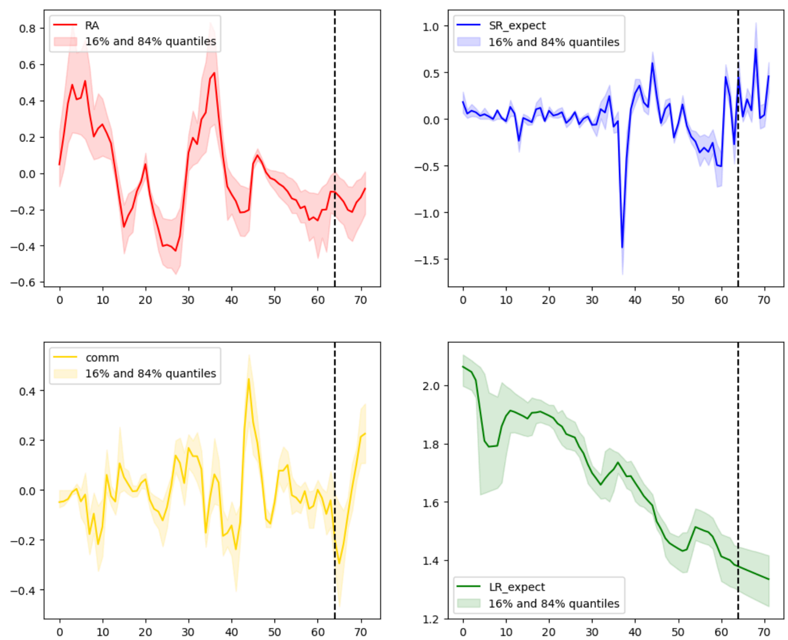
However, the utility of these models extends well beyond mere inflation prediction. A pivotal rationale underpinning the selection of this particular NN architecture is its capacity for enhanced interpretability, particularly in disentangling the multifaceted components that collectively constitute inflation. This architectural choice facilitates a granular analysis of inflation, transcending conventional metrics such as the Root Mean Square Error (RMSE). In addition to evaluating predictive accuracy, it is imperative to engage in a detailed interpretative analysis. This involves a thorough examination of the various elements, including gaps, trends, and contributions, as delineated by the outputs of the HNN models. Such a multifaceted approach not only quantifies the model's forecasting proficiency but also unveils the underlying economic dynamics it captures, thereby offering a more comprehensive understanding of the inflationary process.

To further refine our analysis, attention is now directed towards the performance evaluation of the four developed HNN models. In light of their distinct functionalities and outcomes, a selective approach will be adopted for deeper analysis. Specifically, the HNN2 model, identified as the most adept in our evaluations, will be the primary focus for extracting insights into the Contributions, Gaps, and Trends of the four empirical hemispheres. Meanwhile, the HNN4 model, particularly designed for the newly introduced fifth hemisphere addressing Inflation spillover effects, will be instrumental in interpreting the nuances and intricacies of this additional component. This targeted approach allows for a more nuanced understanding of the distinct roles and impacts of each hemisphere in our overall inflation analysis.

6.2.1 Contributions

In this analysis, we focus on the contribution plots (*see Figure 8.*) generated by the HNN model, which articulates the estimated impacts of distinct economic sectors on inflationary trends. The contributions of Real Activity (RA), Short-run Expectations (SR_expect), Commodities (comm), and Long-run Expectations (LR_expect) are examined to understand how each component influences the overall inflation rate. By scrutinizing these contributions from 1999 to 2017, we can pinpoint the immediate and sustained economic forces at play, offering a nuanced understanding of the inflation dynamics.

Figure 8: Contributions of Real Activity, Short-run Expectations, Commodities and Long-run Expectations hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



Note: Created by the authors.

Real Activity (RA): The contribution of RA, represented in the top left quadrant, oscillates significantly over the observed period, as denoted by the red line and its encompassing confidence interval. Notably, this measure reflects economic conditions such as output and employment levels, suggesting periods of procyclical and countercyclical inflationary pressures. The variance within the confidence intervals implies fluctuating certainty in the model's RA contribution estimates over time. From the RA contribution we can observe the 2008 financial crisis as around the 10 to 30 observation mark shows a disconnect between inflation and real activity, which afterwards seems to catch up and play a more pivotal role.

Short-run Expectations (SR_expect): In the top right quadrant, the contribution of Short-run Expectations demonstrates a relatively volatile nature, with the blue line indicating the model's median contribution estimate. The depicted fluctuations capture market participants' immediate reactions to news and economic events, which can influence short-term inflationary trends. We can also observe the GFC from the short-run expectations contribution as around the 35 observation mark which is the year 2009 we see a sharp decline, turning the contribution to inflation negative.

Commodities (comm): The bottom left quadrant showcases the contribution of commodity prices. The sharp peaks and troughs in the yellow line underscore the well-known volatility of commodity markets and their impact on inflation. The broad confidence intervals at certain points reflect the model's recognition of the inherent unpredictability in commodity price movements and their subsequent influence on the inflationary environment.

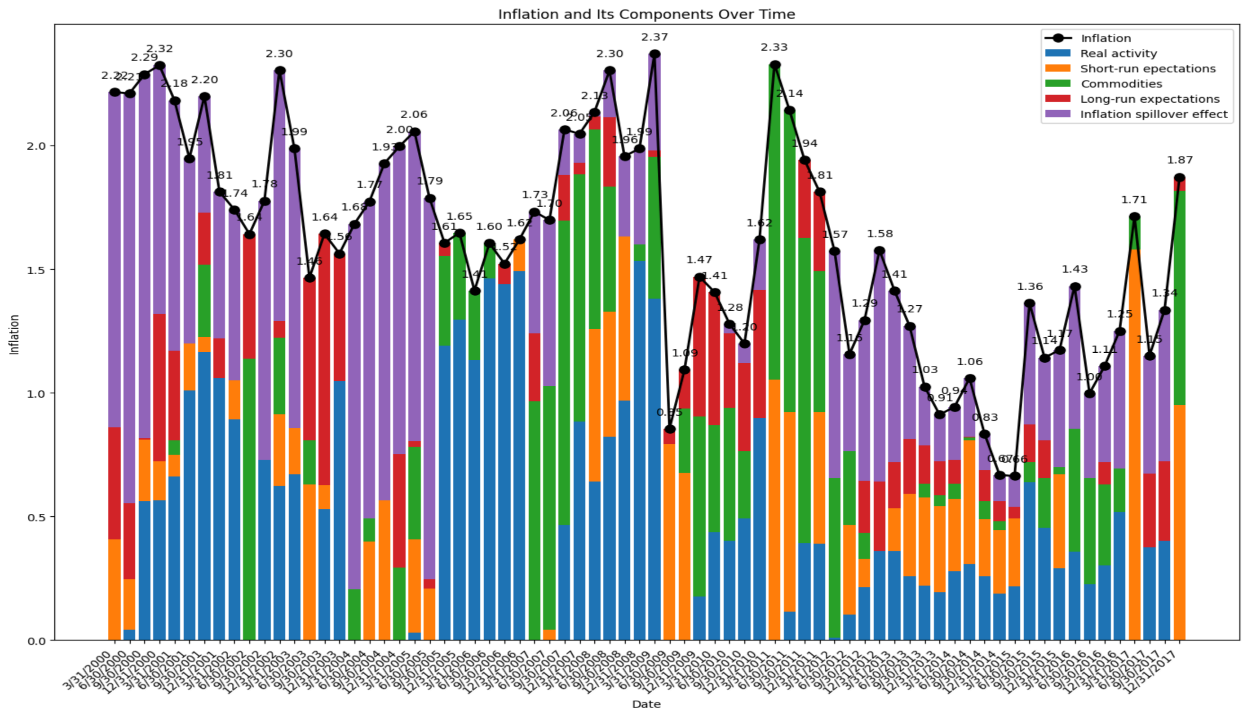
Long-run Expectations (LR_expect): Finally, the bottom right quadrant delineates the contribution of Long-run Expectations, presented as a relatively smooth declining trend by the green line. This component embodies the entrenched and structural expectations of inflation held by economic agents. The gradual downward trend suggests a diminishing contribution of long-run expectations to inflation, potentially indicative of improved anchoring of inflation expectations over time or a response to long-term monetary policy strategies.

6.2.2 Contribution Decomposition

In Figure 9, the decomposition of inflation into its contributing components over time is illustrated, which provides an aggregate perspective on the relative influence of each factor from 1999 to 2017. This bar graph segregates the cumulative effects of Real Activity, Short-run Expectations, Commodities, Long-run Expectations, and the newly introduced Inflation Spillover Effect.

Each bar in the graph corresponds to the observed inflation rate for a given time period, with the different colors representing the proportionate contribution of each component. By analyzing the variation in these contributions, we can discern the evolving economic landscape and the relative significance of each factor:

Figure 9: Contribution Decomposition of Real Activity, Short-run Expectations, Commodities, Long-run Expectations and Inflation Spillover hemispheres for HNN4 model.



Note: Created by the authors.

Real Activity (RA): Illustrated by the blue segments, the Real Activity's influence on inflation fluctuates, correlating with business cycle dynamics. For instance, during periods of economic expansion, such as the early 2000s, RA's contribution to inflation is pronounced, aligning with increased production and labor market tightening. Conversely, during recessions, its influence wanes, most notably during the 2008 financial crisis, where the RA's contribution dips, reflecting economic contraction.

Short-run Expectations (SR_expect): The orange segments for Short-run Expectations display rapid adjustments, reflecting market responsiveness to immediate economic events and policy decisions. A marked expansion in SR_expect's contribution is observed during the 2008 crisis, aligning with the drop in consumer confidence and investor sentiment, which in turn dragged on the inflation rate as all other factors diminished.

Commodities (comm): Commodities, shown in green, contribute noticeably to inflation's volatility. The sharp increases and decreases track major global commodity price swings, such as the oil price surge in 2008 and subsequent drop in 2009. This indicates commodities' significant, yet variable, pass-through effect on consumer prices.

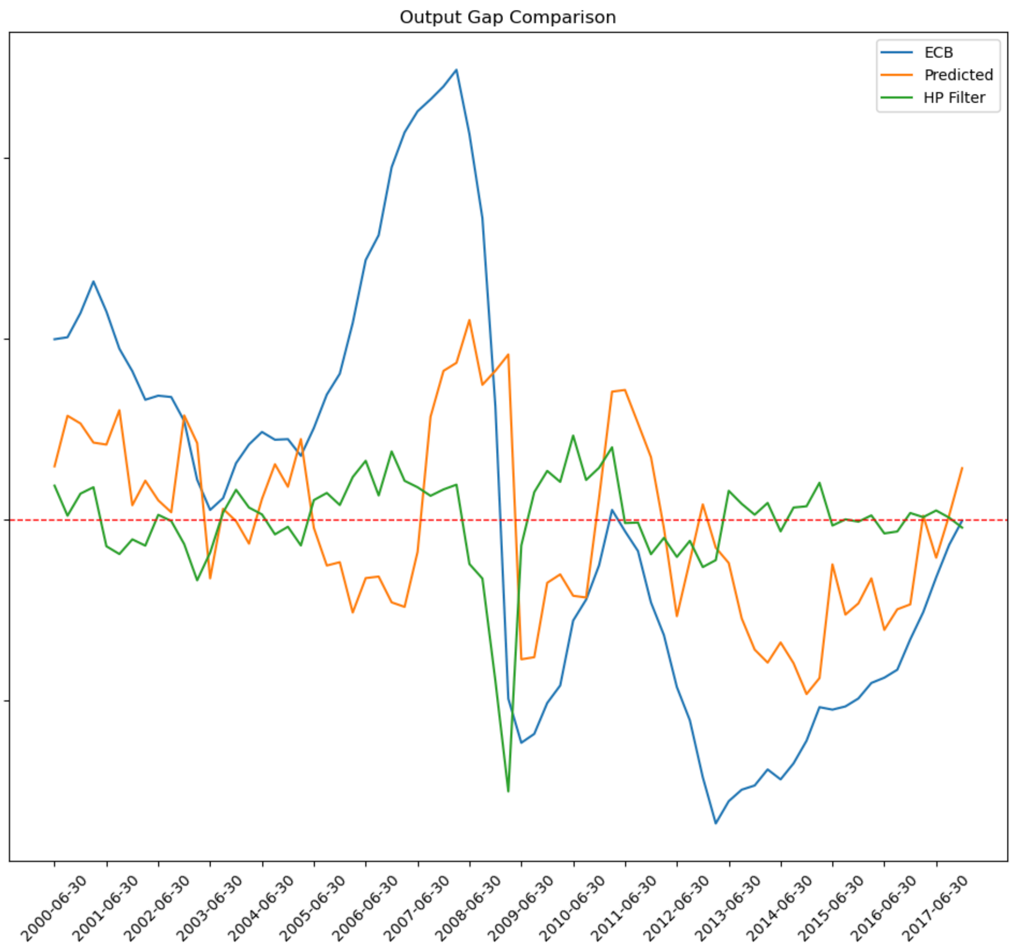
Long-run Expectations (LR_expect): The red segments represent Long-run Expectations. The shrinking contribution over time reflects the gradual stabilization of inflation expectations, potentially due to the credibility of central bank targets and effective communication strategies. This stabilization suggests that over the longer term, the public's views on inflation have become more anchored, reducing the impact of this component on actual inflation rates.

Inflation Spillover Effect (Infl_spill): The Inflation Spillover Effect, in purple, is introduced to account for external influences on domestic inflation. Its varying contribution over the years encapsulates the effects of globalization, such as the European debt crisis and China's economic expansion. Periods of increased global uncertainty or cross-border economic shocks are mirrored by a rise in this segment's size, indicating its significance in the overall inflation framework.

6.2.3 Output Gap

This section focuses on the comparative analysis of OG estimations derived from different methodologies. The OG represents the deviation of the actual economic output from its potential level, providing insights into the cyclical position of an economy. Two established estimations are compared to the OG derived from the HNN model.

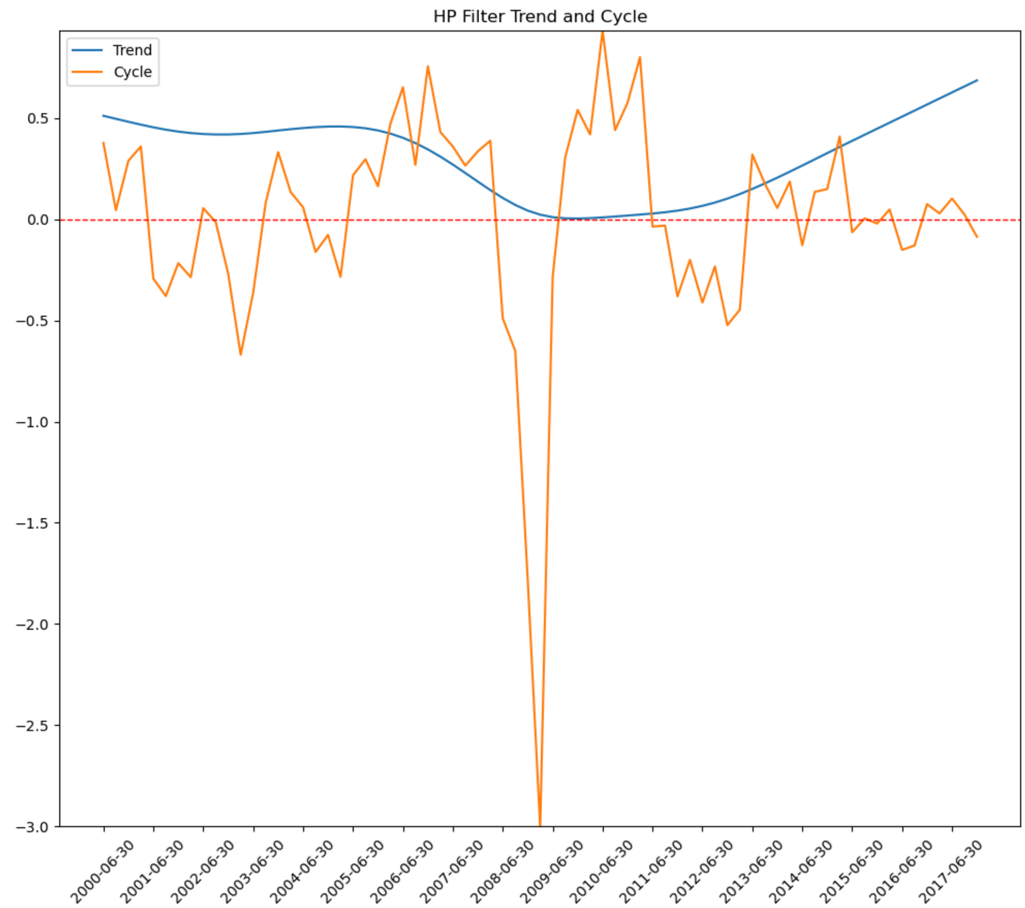
Figure 10: ECB Forecasted OG (ECB, 2018), HP Filter OG and HNN2 Model OG over time. (Quarterly data, 1999-2017)



Note: Created by the authors.

The first benchmark comes from the application of the HP filter, a widely used method in macroeconomic research to separate the cyclical component from the trend component of an economic time series. The HP filter OG is depicted by the green line in *Figure 10*. It is characterized by smoother fluctuations, suggesting a more moderated cyclical variation. Notably, the HP filter captures the broad economic downturn during the 2008 financial crisis with a pronounced trough, followed by a gradual return toward its potential output level. A closer look of the HP filter estimation and decomposition of cycle and trend components can be seen below in *Figure 11*:

Figure 11: HP Filter Trend and Cycle components over time, capturing inflation. (Quarterly data, 1999-2017)



Note: Created by the authors.

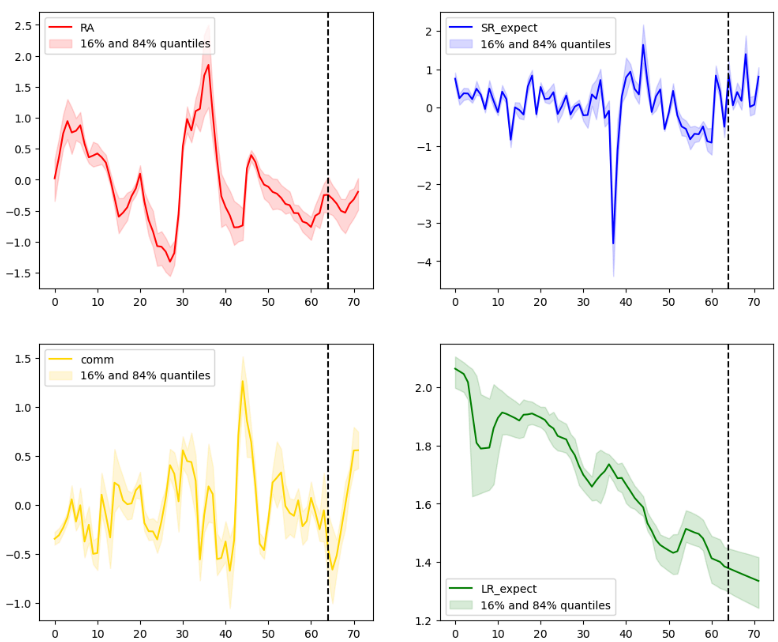
The second benchmark is the OG estimate as published by the ECB, represented by the blue line. The ECB's OG estimate aligns with major economic events, demonstrating sharper fluctuations than the HP filter. This could be attributed to the ECB's methodology, which may incorporate various economic indicators and judgments to provide a more responsive measure of the economy's cyclical position.

Our HNN2 model's predicted OG, shown in orange, offers a unique estimation. This line reflects a middle ground between the HP filter and ECB estimates, capturing the volatility of the economy while maintaining a degree of smoothness. It exhibits a strong correlation with the ECB's OG during periods of significant economic stress, such as the 2008 crisis and the subsequent European debt crisis. However, it diverges in periods of relative economic stability, which may suggest that the HNN2 model is detecting subtler cyclical factors not captured by the HP filter or incorporated into the ECB's estimation.

6.2.4 Output Gap Decomposition

The plot in *Figure 12.* offers a quantified analysis of the gaps for the four specified hemispheres derived from the HNN model outputs, critical for understanding the divergences from potential or equilibrium levels within the economic indicators related to inflation.

Figure 12: Gaps of Real Activity, Short-run Expectations, Commodities and Long-run Expectations hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



Note: Created by the authors.

Real Activity (RA): The top left graph illustrates the real activity gap, with the red line signifying the median gap from the potential output or equilibrium unemployment level. A positive value indicates an economy operating above potential, often associated with inflationary pressure, while negative values suggest slack in the economy, generally corresponding with disinflationary forces.

Short-run Expectations (SR_expect): The top right graph delineates the gap in short-run expectations. The blue line indicates the median gap in inflation expectations in the short run. The notably sharp downward spike represents an event with a significant negative impact on short-term inflation expectations, which in this case is the 2008 financial crisis.

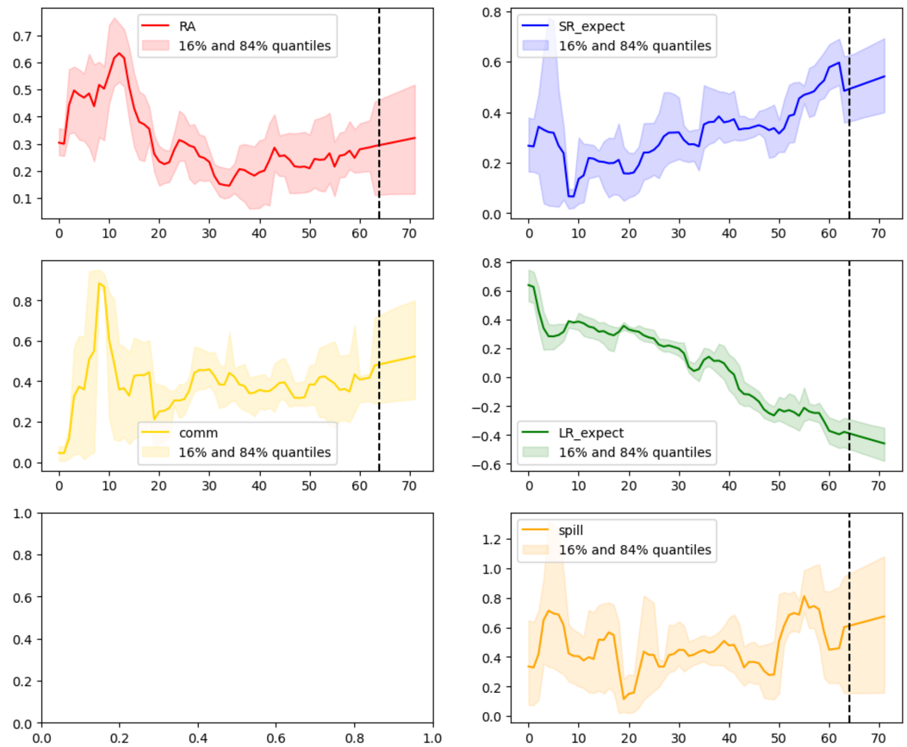
Commodities (comm): The lower left graph showcases the commodities gap. The yellow line illustrates the median gap. The fluctuations in this gap capture the impact of external shocks such as changes in oil prices or other commodity price fluctuations on the economy's inflationary stance.

Long-run Expectations (LR_expect): The bottom right graph features the gap in long-run expectations, where the green line depicts the median gap in long-term inflation expectations. The descending trend suggests a gradual alignment or reduction of long-term expectations towards the central bank's inflation target or a decreasing natural rate of unemployment over time. The tightening confidence intervals towards the end of the sample period may imply an increasing convergence of views about long-term economic trends.

6.2.5 Trends

The analysis of trend components within macroeconomic data serves as a critical tool for understanding the underlying momentum of various economic indicators over time. In this section, we examine the trend plots (*see Figure 13*.) derived from the HNN model, which provides a visual representation of the evolving tendencies of four key macroeconomic hemispheres: Real Activity, Short-run Expectations, Commodities, and Long-run Expectations. These trends are indicative of the persistent directional movements in the data that shed light on the protracted influences affecting inflation from 1999 to 2017. By dissecting these trends, we aim to gain a deeper comprehension of the long-term forces shaping the inflationary landscape.

Figure 13: Trends of Real Activity, Short-run Expectations, Commodities, Long-run Expectations and Inflation Spillover hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



Note: Created by the authors.

Real Activity (RA): The top left graph, delineated in red, displays a generally descending trend in the contribution from Real Activity to inflation, interspersed with periods of upturns. The plot suggests that the general influence of Real Activity on inflation may be diminishing over time, although shorter-term fluctuations are evident.

Short-run Expectations (SR_expect): The top right graph, marked in blue, exhibits a trend in short-run inflation expectations. The line represents the central trend of these expectations. This trend seems to oscillate without a clear long-term direction, indicating that short-run expectations have varied significantly throughout the period, possibly in response to cyclical economic conditions or policy changes.

Commodities (comm): The bottom left graph, colored in yellow, shows the trend in the impact of commodities on inflation. The trend line fluctuates, which captures the inherent volatility of commodity markets and their effects on inflation. The wide confidence intervals suggest a high degree of uncertainty about the magnitude of commodities' influence on inflation over time.

Long-run Expectations (LR_expect): The bottom right graph, in green, presents the trend in long-run inflation expectations. The descending line suggests that these expectations have been decreasing over time, which could be indicative of a stable or deflationary outlook among long-term economic agents. The narrowing of the confidence intervals over time could suggest an increasing consensus or reduced uncertainty about these expectations.

Inflation Spillover Effects (spill): The fourth quadrant, highlighted in orange, articulates the trends associated with the 'spillover' effects on inflation. The line indicates the central tendency of these spillover contributions to domestic inflation. It exhibits an oscillatory pattern without a definitive long-term trend, suggesting the episodic nature of these effects. Periods of heightened global economic uncertainty or international market volatility are reflected as spikes in the graph, denoting moments when external pressures exerted a more substantial impact on the domestic inflation landscape. The breadth of the confidence intervals also varies significantly, implying varying degrees of certainty within the model regarding the spillover contributions. Such intervals are notably wide during times of global financial stress, reflecting the increased complexity in forecasting inflationary pressures amid worldwide economic turbulence.

7. Conclusions

This thesis embarked on an ambitious journey to explore the efficacy of HNN in forecasting inflation and the OG within the Eurozone, by novel methodology and algorithm design created by P.G.Columbe in 2022. The rigorous analysis, structured around two primary research questions and a clear hypothesis, has resulted in a comprehensive evaluation of these advanced models against traditional econometric approaches.

Research Question 1: posited the potential of explainable NN to predict inflation in the Eurozone with a higher degree of accuracy than traditional models. The evidence gathered from the performance of four distinct HNN architectures, particularly the HNN2 model with optimized hyperparameters has resoundingly affirmed this potential. Not only did all the HNN models surpass the benchmark linear regression and univariate ARIMA models in predictive accuracy when predicting inflation, but they also enhanced interpretability, providing granular insights into the multifarious dynamics shaping inflation.

Research Question 2: addressed the capability of HNN models to estimate the OG in comparison to traditional methodologies. While HNN models did not outshine traditional approaches in the precision of OG estimation, they demonstrated a robust ability to track general economic trends. Moreover, they offered a more nuanced decomposition of the OG, shedding light on complex economic phenomena that otherwise would remain obscured by the simplicity of methods like the HP filter.

Our hypothesis posited the superiority of explainable NN over traditional models in predicting the inflation and OG in the Eurozone. The findings herein substantiate this hypothesis, revealing that, indeed, HNN models can provide more precise forecasts and deeper analytical insights. The outperformance of all HNN models over the benchmark models marks a significant advancement in the application of NN to macroeconomic forecasting.

In conclusion, this thesis stands as a testament to the transformative potential of ML in economic analysis. The developed HNN models, particularly HNN2 and HNN4, have not only pushed the boundaries of prediction accuracy but have also opened new vistas for interpretability and analytical depth. Additionally, we have proven that selected NN architecture is adjustable to data which is variable across regions, and the prediction accuracy can be improved via accepted optimization approaches; therefore, confirming the possibility to employ HNN on relatively small dataframes making it suitable for regions with no extensive historical macroeconomic data figures. The success of these models heralds a new era in economic prediction, one that embraces the complex, interconnected, and rapidly evolving nature of global economies. As the Eurozone continues to navigate the intricate tapestry of economic forces, the insights garnered from this research will undoubtedly contribute to a clearer understanding of the economic landscape.

8. References

The link to Google folder with supplemental materials: Link to Google folder

8.1 Acknowledgements of LLM use

In this research paper we have used LLMs for paraphrasing, and avoiding grammar & style mistakes in the text, as well as to follow appropriate APA style of writing. No facts and conclusions were solely generated by LLMs

8.2 References:

Alogoskoufis, G., & Smith, R. P. (1991). The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange-Rate Regimes. ResearchGate. https://www.researchgate.net/publication/4901121_The_Phillips_Curve_the_Persistence_of_Inflation_and_the_Lucas_Critique_Evidence_from_Exchange-Rate_Regimes

Ascari, G., & Marrocu, E. (2003). Forecasting inflation: A comparison of linear Phillips curve models and nonlinear time series models (Working Paper No. 2003/07). https://crenos.unica.it/crenos/sites/default/files/03-07.pdf

Auer, R. A., Levchenko, A. A., & Sauré, P. (2017, March). International inflation spillovers through input linkages (NBER Working Paper No. 23246). National Bureau of Economic Research. http://www.nber.org/papers/w23246

Bank of England. (2023). Monetary Policy Report - May 2023. https://www.bankofengland.co.uk/monetary-policy-report/2023/may-2023

Ball, L., Mankiw, N. G., & Romer, D. (1988). The New Keynesian Economics and the Output-Inflation Trade-off. https://www.brookings.edu/wp-content/uploads/1988/01/1988a_bpea_ball_mankiw_romer_akerlof_rose_yellen.pdf

Barbaglia, L., Consoli, S., & Manzan, S. (2021). Forecasting GDP in Europe with Textual Data. Social Science Research Network. https://doi.org/10.2139/ssrn.3898680

Caza, P. (2023, June 12). La Banque d'Angleterre utilise un outil créé à l'UQAM | UQAM. Actualités UQAM. https://actualites.uqam.ca/2023/la-banque-angleterre-utilise-un-outil-uqam/

Clarida, R., Galí, J., & Gertler, M. (1998). Monetary Policy Rules and Macroeconomic Instability: Evidence and Some Theory. https://www.nber.org/system/files/working_papers/w6442/w6442.pdf

Columbe, P. (2019). The macroeconomy as a random forest. Ar5iv. https://ar5iv.labs.arxiv.org/html/2006.12724

Coulombe, P. G. (2022). A Neural Phillips Curve and a Deep Output Gap. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2202.04146

Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2021, March 10). Macroeconomic data transformations matter. University of Pennsylvania; Université du Québec à Montréal.

Coulombe, P. G., Marcellino, M., & Stevanovic, D. (2021). CAN MACHINE LEARNING CATCH THE COVID-19 RECESSION? National Institute Economic Review, 256, 71–109. https://doi.org/10.1017/nie.2021.10

Cuerpo, C., Cuevas, Á., & Quilis, E. M. (2018). Estimating output gap: a beauty contest approach. Series (Berlin. Springer. Internet), 9(3), 275–304. https://doi.org/10.1007/s13209-018-0181-5

Douglas Araujo, Giuseppe Bruno, Juri Marcucci, Rafael Schmidt, & Bruno Tissot. (2021). Machine learning applications in central banking. https://www.bis.org/ifc/publ/ifcb57_01_rh.pdf

Douglas Araujo, Sebastian Doerr, Leonardo Gambacorta, & Bruno Tissot. (2024). Artificial intelligence in central banking. https://www.bis.org/publ/bisbull84.pdf

Eckert, F. A. (2021). Bayesian State Space Models in Business Cycle Research and Economic Forecasting [Doctoral Thesis, ETH Zurich]. https://doi.org/10.3929/ethz-b-000519702

ECB, Jarociński, M., & Lenza, M. (2016). An inflation-predicting measure of the output gap in the euro area. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1966.en.pdf

Fojtíková, L. (2014). Performance and growth of the eurozone export. Procedia Economics and Finance, 12, 154–163. https://doi.org/10.1016/s2212-5671(14)00331-1

Forecasting: Principles and Practice: Hyndman, Rob J, Athanasopoulos, George: 9780987507136: Amazon.com: Books. (n.d.). https://www.amazon.com/dp/0987507133?tag=otexts20

Gali, J. (2015). Monetary policy, inflation, and the business cycle : an introduction to the new Keynesian framework and its applications. https://ci.nii.ac.jp/ncid/BB18800063

Grant, A. L., & Chan, J. C. C. (2016, December). Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter. Research School of Economics, Australian National University.

Gross, M., & Semmler, W. (2017). Mind the output gap: the disconnect of growth and inflation during recessions and convex Phillips curves in the euro area. ECB Working Paper, No. 2004, ISBN 978-92-899-2726-0, European Central Bank (ECB), Frankfurt a. M., https://doi.org/10.2866/611955

Hamilton, J. D. (1994). Time Series analysis. In Princeton University Press eBooks. https://doi.org/10.1515/9780691218632

Hall, A. S., & Cook, T. R. (2017). Macroeconomic Indicator Forecasting with Deep Neural Networks. Social Science Research Network. https://doi.org/10.2139/ssrn.3046657

Hall, S. G., Tavlas, G. S., & Wang, Y. (2023). Drivers and spillover effects of inflation: The United States, the euro area, and the United Kingdom. Journal of International Money and Finance, 131, 102776. https://doi.org/10.1016/j.jimonfin.2022.102776

Hauzenberger, N., Huber, F., & Klieber, K. (2023). Real-time inflation forecasting using non-linear dimension reduction techniques. International Journal of Forecasting, 39(2), 901–921. https://doi.org/10.1016/j.ijforecast.2022.03.002

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2007). Principles of econometrics. http://www.gbv.de/dms/zbw/637121902.pdf

Huang, D., Jiang, F., Li, K., Tong, G., & Zhou, G. (2022). Scaled PCA: A new approach to dimension reduction. Management Science, 68(3), 1678-1695. https://ink.library.smu.edu.sg/lkcsb_research/6924

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: principles and practice, 3rd edition. OTexts: Melbourne, Australia. OTexts.com/fpp3

Istiak, K., Tiwari, A. K., Husain, H., & Sohag, K. (2021). The Spillover of Inflation among the G7 Countries. Journal of Risk and Financial Management, 14, 392. https://doi.org/10.3390/jrfm14080392

J. Gali, M. Gertler (2000). Inflation Dynamics: a Structural Econometric Analysis. https://www.nber.org/system/files/working_papers/w7551/w7551.pdf

Kulikov, D., & Reigl, N. (2019). Inflation Expectations in Phillips Curve Models for the Euro Area. In Eesti Pank Working Paper Series. https://doi.org/10.23656/25045520/082019/0171

Labonne, P., & Chen, B. (n.d.). fbi: Factor-Based Imputation for Missing Data. GitHub. https://github.com/cykbennie/fbi

Lucas, R. E. (1972). Expectations and the neutrality of money. Journal of Economic Theory, 4(2), 103–124. https://doi.org/10.1016/0022-0531(72)90142-1

Maccarrone, G., Morelli, G., & Spadaccini, S. (2021). GDP Forecasting: machine learning, linear or autoregression? Frontiers in Artificial Intelligence, 4. https://doi.org/10.3389/frai.2021.757864

Mavroeidis, S., Plagborg-Møller, M., & Stock, J. H. (2014). Empirical evidence on inflation expectations in the new Keynesian Phillips Curve. Journal of Economic Literature, 52(1), 124–188. https://doi.org/10.1257/jel.52.1.124

Neiss, K. S., & Nelson, E. (2002). Inflation dynamics, marginal cost, and the output gap: Evidence from three countries. Bank of England. https://www.frbsf.org/economic-research/wp-content/uploads/sites/4/nkpcnn.pdf

Paranhos, L. (2023). Predicting Inflation with Recurrent Neural Networks. https://arxiv.org/pdf/2104.03757.pdf

Park, S., & Yang, J. (2022). Interpretable deep learning LSTM model for intelligent economic decision-making. Knowledge-Based Systems, 248, 108907. https://doi.org/10.1016/j.knosys.2022.108907

Phillips, A. W. (1958). The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–19571. Economica, 25(100), 283–299. https://doi.org/10.1111/j.1468-0335.1958.tb00003.x

Proner, R. (2023). A Multi-Task deep learning model for inflation forecasting: Dynamic Phillips Curve Neural Network. Social Science Research Network. https://doi.org/10.2139/ssrn.4454118

Probabilistic gradient boosting machines for Large-Scale Probabilistic Regression. (Sprangers, 2021). Ar5iv. https://ar5iv.labs.arxiv.org/html/2106.01682

Rudd, J., & Whelan, K. (2005). Modelling inflation dynamics: A critical review of recent research (FEDS Working Paper No. 2005-66). Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board. https://www.federalreserve.gov/Pubs/FEDS/2005/200566/200566pap.pdf

Sargent, T. J., & Wallace, N. (1976). Rational expectations and the theory of economic policy. Journal of Monetary Economics, 2(2), 169–183. https://doi.org/10.1016/0304-3932(76)90032-5

Šestanović, T., & Arnerić, J. (2021). Can recurrent neural networks predict inflation in euro zone as good as professional forecasters? Mathematics (Basel), 9(19), 2486. https://doi.org/10.3390/math9192486

Sklearn.preprocessing.StandardScaler. (n.d.). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#sklearn.preprocessing.StandardScaler

Smets, F., & Wouters, R. (2002). An estimated stochastic dynamic general equilibrium model of the euro area. https://www.econstor.eu/handle/10419/152605

Steenkamp, D. (2023). Output gap uncertainty in SA. Codera Analytics. https://codera.co.za/output-gap-uncertainty-in-sa/

Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts (NBER Working Paper No. 14322). National Bureau of Economic Research. http://www.nber.org/papers/w14322

Statsmodels Developers. (n.d.). statsmodels.tsa.filters.hp_filter.hpfilter [Python]. https://www.statsmodels.org/stable/generated/statsmodels.tsa.filters.hp_filter.hpfilter.html

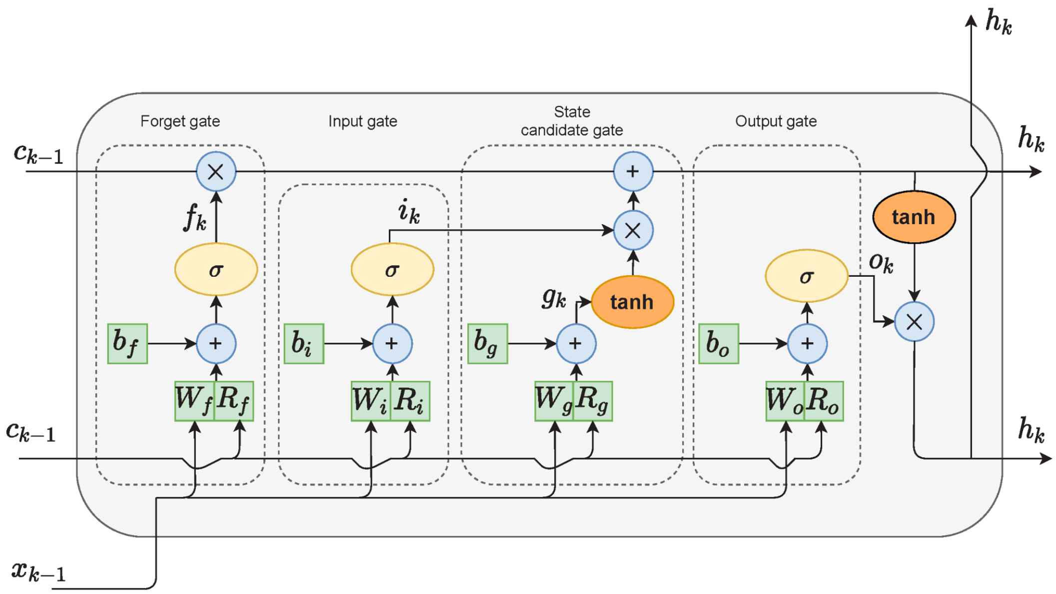
Šestanović, T., & Arnerić, J. (2021). Can recurrent neural networks predict inflation in euro zone as good as professional forecasters? Mathematics (Basel), 9(19), 2486. https://doi.org/10.3390/math9192486

Torch.utils.data — PyTorch 2.1 documentation. (n.d.). Retrieved from https://pytorch.org/docs/stable/data.html#torch.utils.data.TensorDataset

Zarzycki, K., & Ławryńczuk, M. (2021). LSTM and GRU neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors. Sensors, 21(16), 5625. https://doi.org/10.3390/s21165625

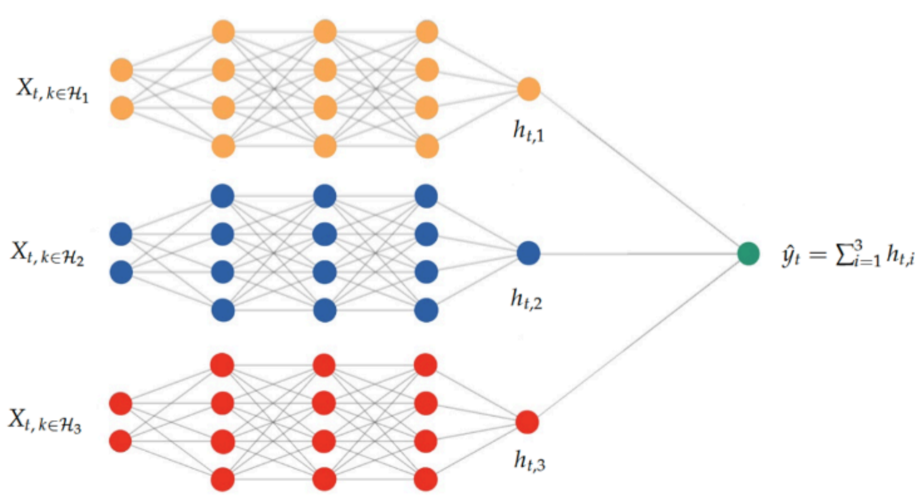
9. Appendix

Figure 1: LSTM Model Logic



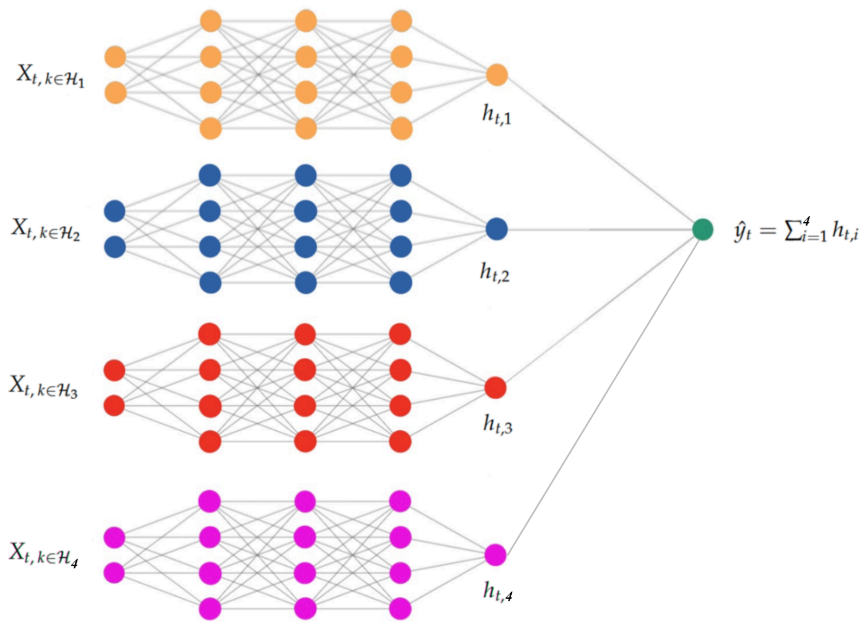
Note: Visualization was made by Zarzycki & Ławryńczuk (2021)

Figure 2. HNN Architecture



Note: HNN illustration by Coulombe, P. G. (2022)

Figure 3. Inflation Spillover Effect Hemisphere



Note: HNN illustration retrieved from Coulombe, P. G. (2022) and augmented with an additional hemisphere by the authors.

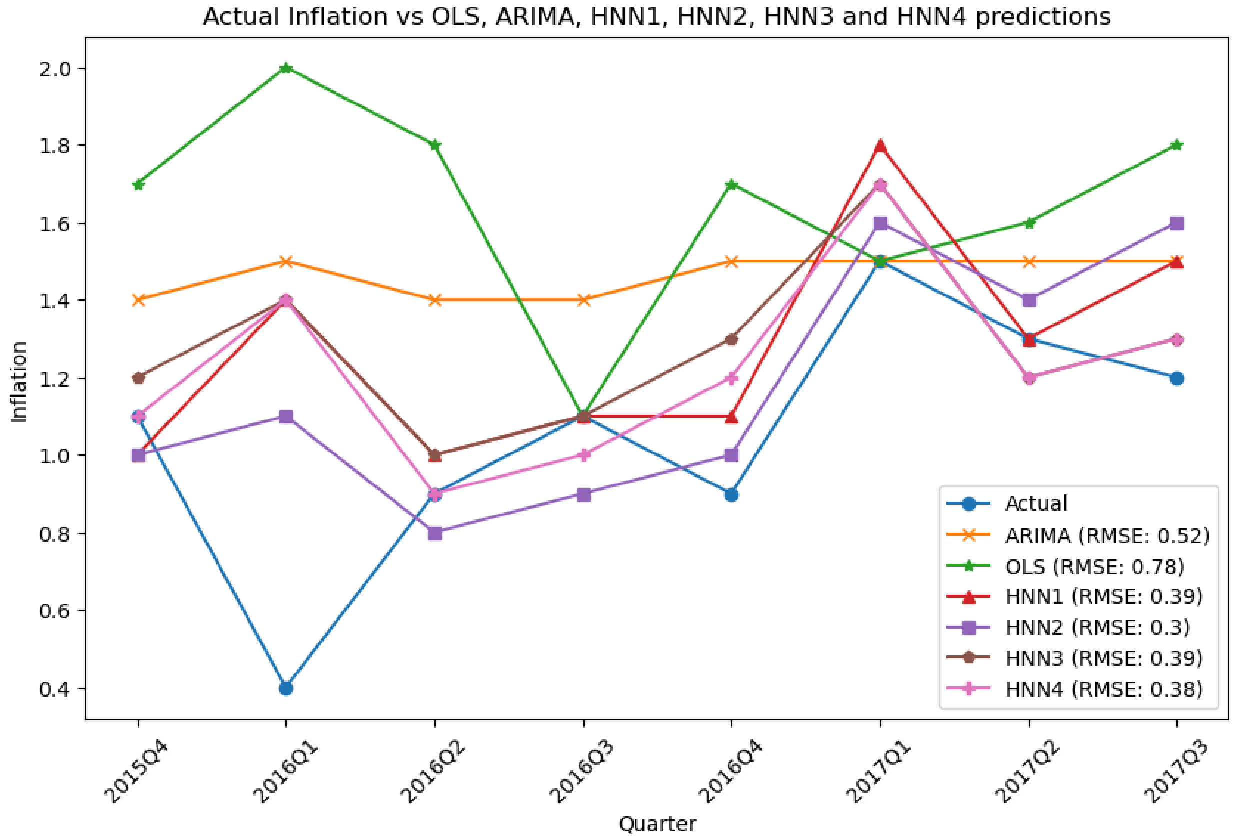
Figure 4. Comparison table of different models' accuracy and complexity with explanations.

Model	Accuracy	Complexity	Comments
Simple Linear Regression	Low	Low	Most basic form, often too simplistic to capture real-world dynamics
Multiple Linear Regression	Moderate	Low	Better than simple linear regression but still assumes a linear relationship among variables
Polynomial Regression	Moderate	Moderate	Captures non-linear relationships but can lead to overfitting
Time-Series Models (ARIMA, VAR)	High	Moderate	Account for temporal dependencies but may require large datasets for accurate predictions
Generalized Additive	High	Moderate	Allows for flexible relationships between variables but

Models (GAMs)			can be computationally intensive
Bayesian Models	High	High	Incorporates prior beliefs into the model, providing a more nuanced understanding but at the cost of computational intensity
Dynamic Stochastic General Equilibrium (DSGE) Models	High	High	Discovers the hidden state of the time-variant system over time and makes reliable predictions based on previous observations and new observations on each iteration.
Random Forests	High	High	Can capture complex non-linear relationships and interactions among variables but may overfit if not properly tuned
Neural Networks	Very High	Very High	Extremely flexible and can model complex relationships, but they are a "black box" and require a lot of data
Ensemble Methods (Stacking, Bagging, Boosting)	Very High	Very High	Combines multiple models to improve predictive accuracy but is computationally expensive and complex to implement
Deep Learning Models with interpretable layers	Maximum	Maximum	Resistant to non-gaussian noise and can capture more nonlinearities due to inheriting features of RNN. As a tradeoff they are less interpretable in default setting

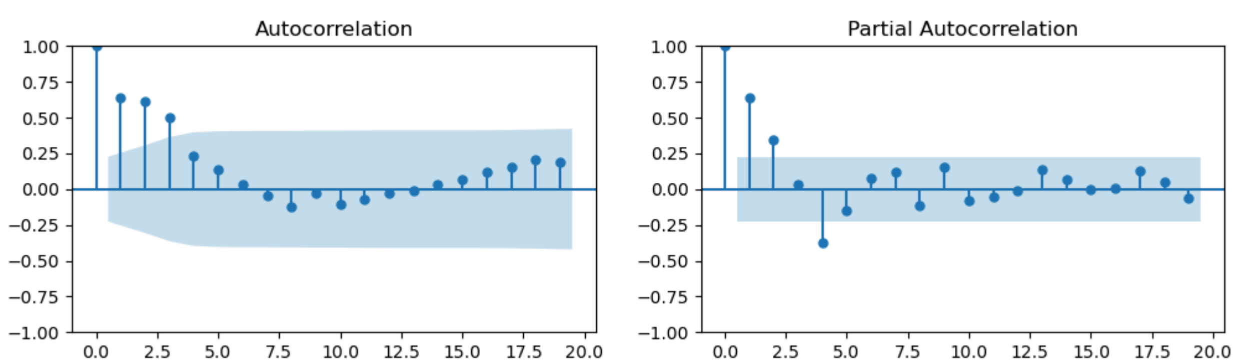
Note: Table created by the authors.

Figure 5: Actual Inflation compared to Linear Regression, ARIMA, HNN1, HNN2, HNN3 and HNN4 out-of-sample predictions. (period from 2015-12-31 to 2017-09-30, one observation equal to one quarter)



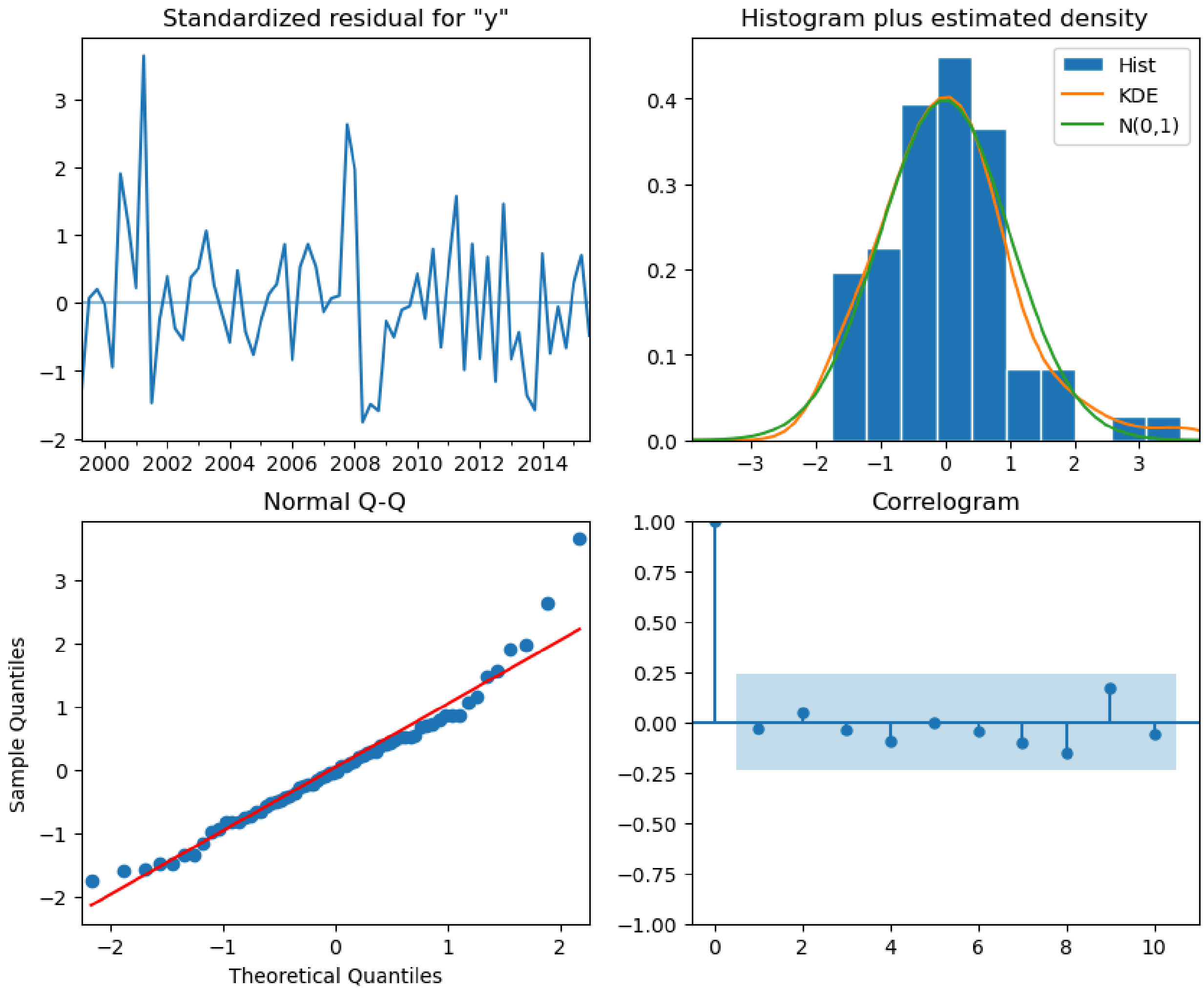
Note: Created by the authors.

Figure 6: Autocorrelation and Partial Autocorrelation plots of inflation.



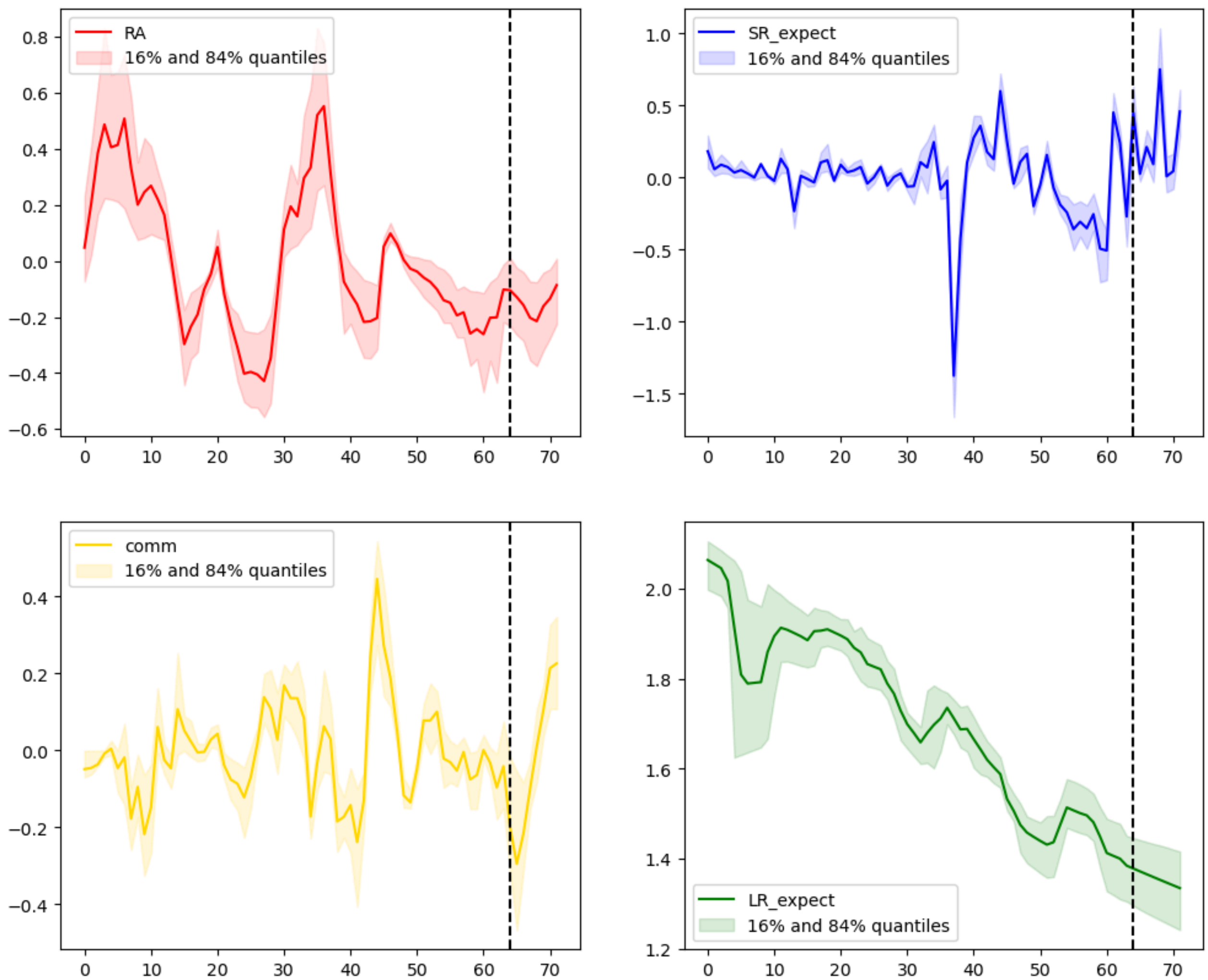
Note: Created by the authors.

Figure 7: Standardized residual plot, residual histogram and density plot, Normal Q-Q plot and residual Correlogram.



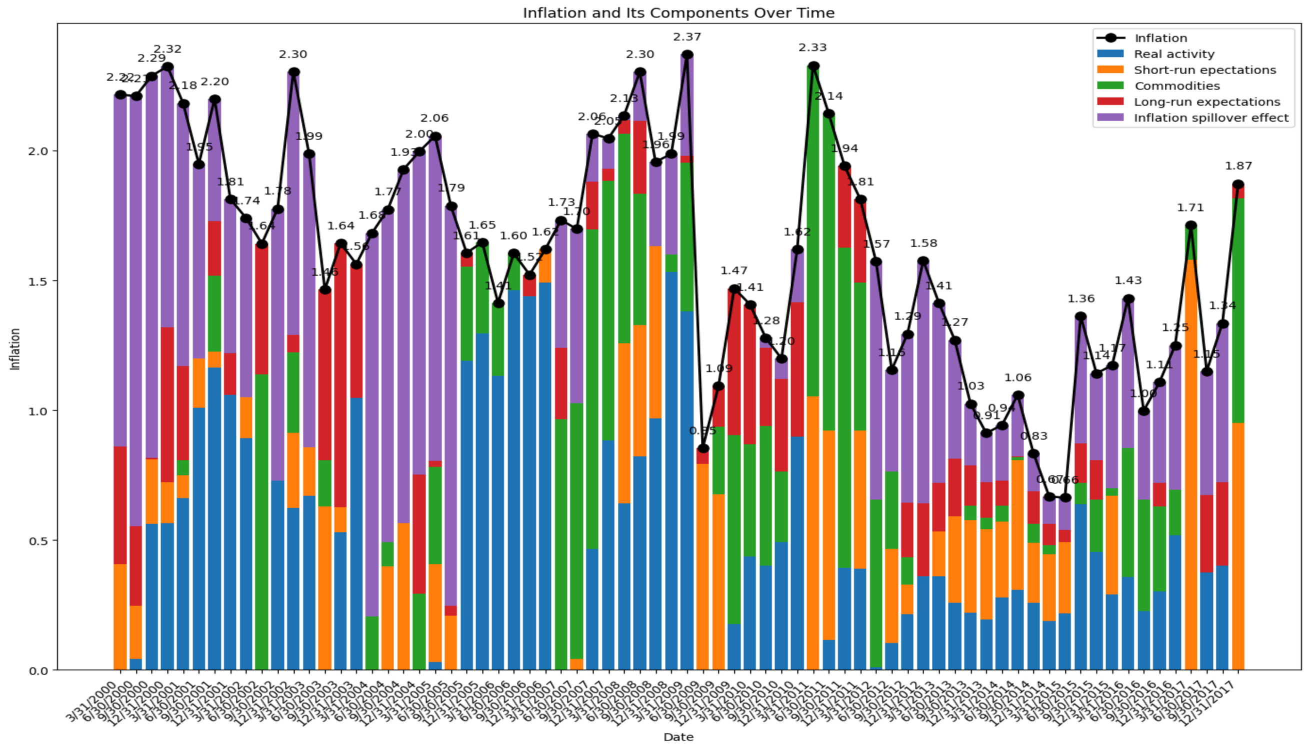
Note: Created by the authors.

Figure 8: Contributions of Real Activity, Short-run Expectations, Commodities and Long-run Expectations hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



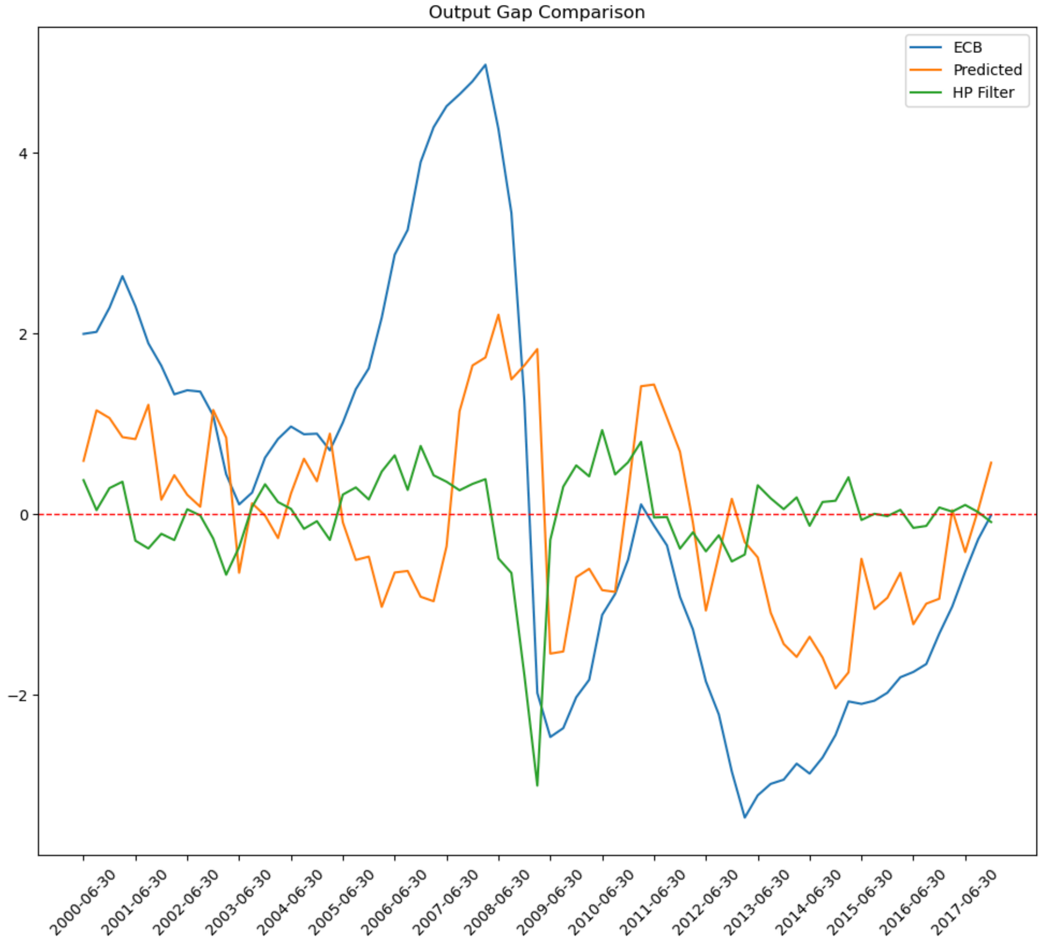
Note: Created by the authors.

Figure 9: Contribution Decomposition of Real Activity, Short-run Expectations, Commodities, Long-run Expectations and Inflation Spillover hemispheres for HNN4 model.(Quarterly data, 1999-2017)



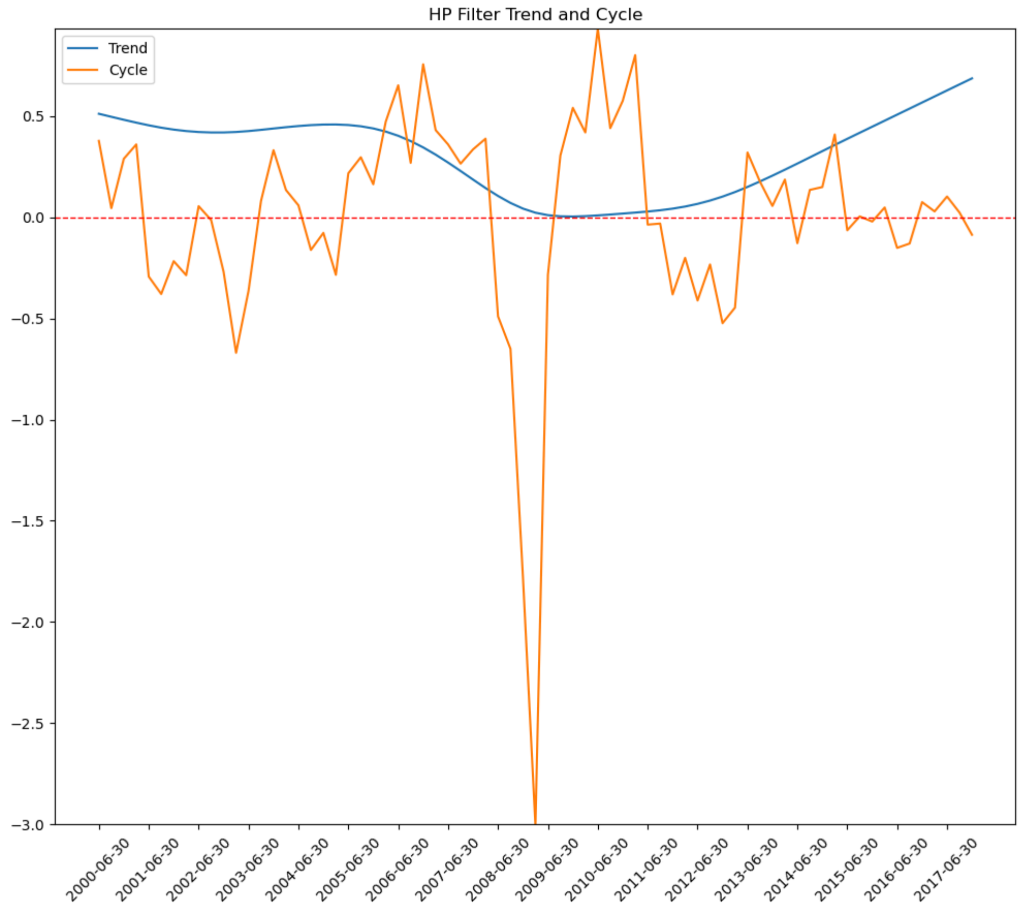
Note: Created by the authors.

Figure 10: ECB Forecasted OG (ECB, 2018), HP Filter OG and HNN2 Model OG over time. (Quarterly data, 1999-2017)



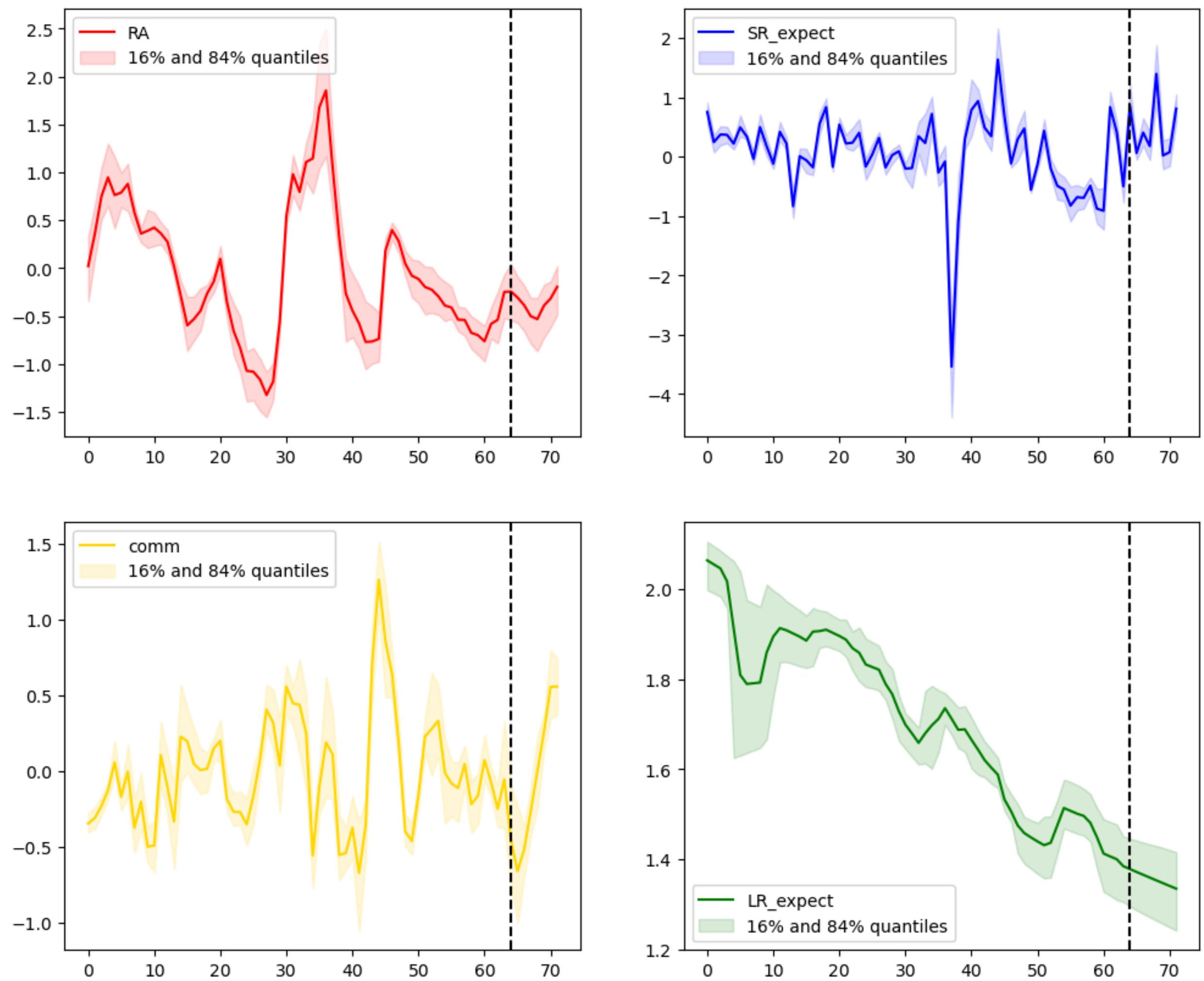
Note: Created by the authors.

Figure 11: HP Filter Trend and Cycle components over time, capturing inflation. (Quarterly data, 1999-2017)



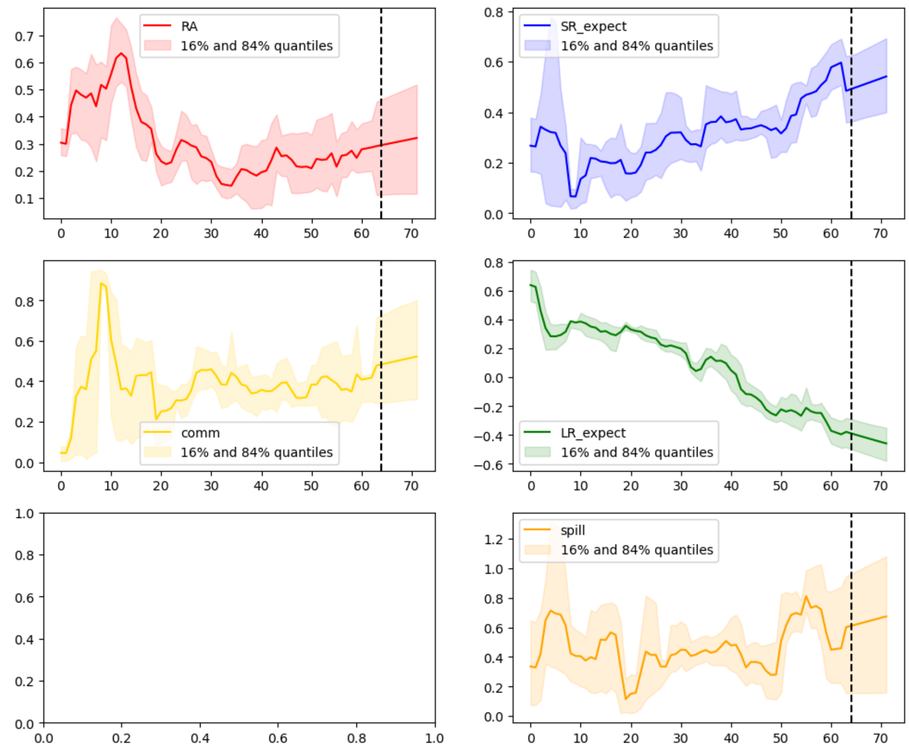
Note: Created by the authors.

Figure 12: Gaps of Real Activity, Short-run Expectations, Commodities and Long-run Expectations hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



Note: Created by the authors.

Figure 13: Trends of Real Activity, Short-run Expectations, Commodities, Long-run Expectations and Inflation Spillover hemispheres. (1999-2017, one observation equal to one quarter) Dotted line indicates the out-of-sample period.



Note: Created by the authors.