Nowcasting Inflation Indices Using Mixed Data Sampling (MIDAS) Time-Series Models

Gabriel Nappa

A dissertation submitted in partial fulfilment of the requirements of the Degree of MA in Banking, Finance & Investments Studies at the University of Malta

Department of Banking, Finance and Investments

Faculty of Economics, Management and Accountancy

University of Malta

October 2024

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL **Faculty of Economics, Management & Accountancy**

DECLARATIONS BY POSTGRADUATE STUDENTS

Student's Code [redacted]

Student's Name & Surname Gabriel Nappa

Course Master of Arts in Banking, Finance & Investments

Title of Dissertation

Nowcasting Inflation Indices Using Mixed Data Sampling (MIDAS) Time-Series Models

(a) Authenticity of Dissertation

I hereby declare that I am the legitimate author of this Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedures

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code FEMA-2024-00735.

As a Master's student, as per Regulation 77 of the General Regulations for University Postgraduate Awards 2021, I accept that should my dissertation be awarded a Grade A, it will be made publicly available on the University of Malta Institutional Repository.

[redacted]

Signature of Student

GABRIEL NAPPA

Name of Student (in Caps)

05/10/2024

Date

Abstract

Nowcasting Inflation Indices Using Mixed Data Sampling (MIDAS) Time-Series Models

This study's principal objective is to estimate and analyse an inflation nowcasting model using alternative data with mixed frequencies via a novel application of the Mixed Data Sampling (MIDAS) time series model. Through empirical application, this dissertation seeks to demonstrate the effectiveness and reliability of the MIDAS approach, offering valuable insight to market practitioners such as policymakers, investment analysts, and researchers who rely on real-time analysis of macroeconomic data.

Several time-series models were built to assess the robustness of the MIDAS regressions to nowcast two inflation indices, the CPI and the Core CPI, for the United States. Furthermore, the statistical significance of the parameters employed was used to indicate which high-frequency economic indicators and weighting functions were the most significant in nowcasting the inflation indices. Accordingly, by the employment of intra-month analysis, the robustness of the MIDAS regressions to nowcast inflation indices was evaluated and compared with another inflation nowcasting model with publicly available results.

The study finds that the employment of a large and varied sample of high-frequency data in models using the MIDAS approach allows for superior robustness when nowcasting inflation growth rates, especially when compared with the traditional Bridge equations used in nowcasting applications such as the inflation nowcast by the Federal Reserve Bank of Cleveland. Furthermore, the study also finds that the parsimony and parametric flexibility offered by the MIDAS approach allows the researcher to easily optimise the model regularly to realign the selection of parameters with the previous month's official estimate and recalibrating the lag-length and weighting gradient to be given to the high-frequency regressors.

This dissertation directly addresses the gap in the MIDAS model's application in nowcasting inflation indices in the United States. While there was scarce literature on inflation nowcasting, let alone the application of the MIDAS model in this regard, this dissertation suggests that this model enjoys considerable benefits which may be employed by market practitioners as a reliable inflation nowcasting model.

Keywords: Nowcasting, Inflation, MIDAS, Big Data, Mixed-Frequency Models

Acknowledgements

I would like to express my heartfelt gratitude to my tutor, Christian Manicaro, PhD, CFA, for his invaluable guidance, support, and remarkable patience throughout my dissertation and academic journey. It has been an honour working under your supervision, having fostered in me a culture of dedication and excellence, for which I am deeply thankful.

My gratitude also extends towards the rest of the Department of Banking, Finance & Investments, from whom it was my privilege to have had the opportunity to learn from your most distinguished staff. You have truly made a lasting impact on my perspective and enthusiasm towards the field of Finance.

Finally, I would like to extend my thanks to my parents, grandparents, and best friends for their support and patience throughout these past few academic years. This accomplishment would not have been possible without them.

Contents

List of Tables

List of Figures

List of Abbreviations

ADF – Augmented Dickey-Fuller Test

AIC – Akaike Information Criterion

API – Application Programming Interface

BLS – United States Bureau of Labor Statistics

BVAR – Bayesian Vector Autoregression

CPI – Consumer Price Index

DFM – Dynamic Factor Model

DMS – Deterministic Model Switching

ECB – European Central Bank

EIA – United States Energy Information Administration

FRED – Federal Reserve Economic Data (Online database created and maintained by the Research Department at the Federal Reserve Bank of St. Louis)

GDP – Gross Domestic Product

IC – Information Criteria

IMF – International Monetary Fund

KPSS – Kwiatkowski-Phillips-Schmidt-Shin Test

MAPFE – Mean Absolute Percentage Forecast Error

MIDAS – Mixed Data Sampling

OLS – Ordinary Least Squares

P-Value – Probability Value

R – Programming Language

RMSFE – Root Mean Squared Forecast Error

SBIC – Schwarz Bayesian Information Criterion

US – United States

VIF – Variance Inflation Factor

Chapter 1 – Introduction

1.1 Background to the Study

Nowcasting is the specialized practice employed by market practitioners such as federal reserves, central banks, financial institutions, and investment managers to estimate movements in low-frequency macroeconomic variables in real time using alternative data from a variety of sources and complex econometric techniques (Stock & Watson, 2020; Bańbura, Giannone, Modugno, and Reichlin, 2013). The popular approach to nowcasting macroeconomic variables employs what is referred to as Bridge models (Dauphin, et al., 2022). This means that nowcasting models such as the Atlanta Fed's "*GDPNow*" estimate the movements in each separate subcomponent of the low-frequency macroeconomic variable before aggregating them to create a global nowcast (Higgins, 2014).

While literature is abundant on nowcasting GDP growth, such as Bańbura, Giannone, Modugno, and Reichlin, (2013), Higgins, (2014), Marcellino, (2021), and Andreini, Hasenzagl, Reichlin, Senftleben-König, & Strohsal, (2023), nowcasting inflation growth rates have remained particularly unresearched even though it is one of the most monitored macroeconomic variables (Modugno, 2013). Furthermore, most studies about inflation nowcasting, such as the case of Knotek II & Zaman, (2023) and Modugno, (2013) discard the Mixed Data Sampling (MIDAS) approach due to the suboptimization and poor results achieved compared to models employing the Dynamic Factor approach and Bridge equations.

1.2 Objectives and Rationale of the Study

This study's principal objective is to estimate and analyse an inflation nowcasting model using alternative data with mixed frequencies via the Mixed Data Sampling (MIDAS) approach. While this model provides numerous benefits such as parsimony and parametric flexibility, it also allows for the exploitation of high-frequency data to induce recency and increase the robustness of the nowcasts produced (Asimakopoulos, Paredes, and Warmedinger, 2013). This makes it a prominent alternative to the currently used Bridge models to nowcast macroeconomic indices (Schumacher, 2014). This also makes the mixed frequency model a significant alternative for researchers with mixed availability of data, since the option to employ different frequencies allows them to exploit more data that is available to them (Barbaglia, et al., 2024).

Nowcasting inflation growth rates will therefore help tackle the issues arising from the ragged edge problem by providing direct estimates of the current movements in the inflation rates before the official estimates are made available (Bańbura, Giannone, Modugno, and Reichlin, 2013). This will also provide an indispensable tool to market practitioners who rely on real-time analysis for decision-making or conducting investment analyses. Moreover, this tool will aid researchers in interpreting forecast errors and alleviate macroeconomic surprises (Reichlin, 2019).

1.3 Research Questions & Importance of the Study

The research question which this study attempts to answer is the following:

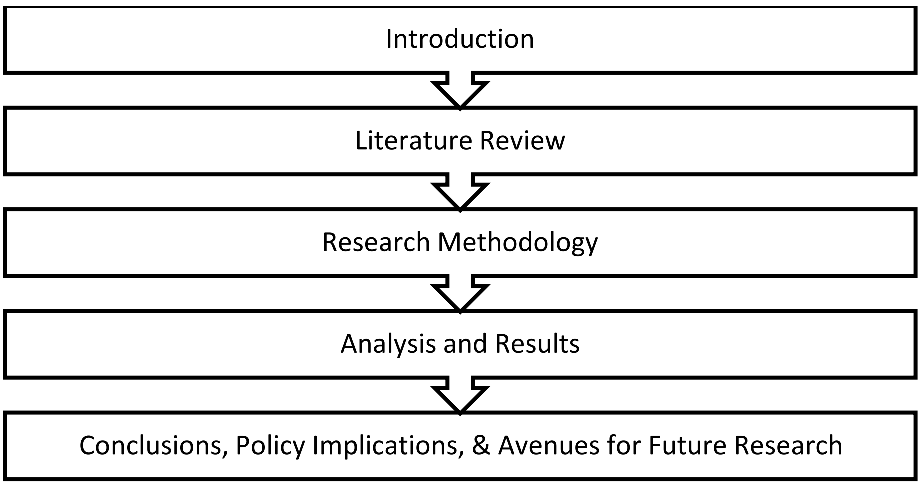
"How robust is the Mixed Data Sampling (MIDAS) time-series model to nowcast low-frequency macroeconomic variables such as inflation indices?"

This research question will be addressed by constructing several MIDAS time-series models to nowcast inflation growth rates for the United States economy. This will involve combining various categories of economic indicators with mixed frequencies and unsynchronised publication lags. Upon estimating the model, the original prediction will describe the statistical significance of the regressors employed, indicating which method of parameterisation was the most significant in forecasting the inflation indices. Accordingly, by the employment of empirical application, the robustness of the MIDAS regressions to nowcast inflation indices will be evaluated.

By answering this research question, the MIDAS model will be proposed as a suitable alternative to the traditionally used Bridge equations to nowcast inflation indices. The MIDAS model can therefore be employed by market practitioners who rely on real-time analyses for decision-making, thereby exploiting high-frequency data to increase informational efficiency in the financial markets (Giannone, Reichlin, Small, 2006).

1.4 Dissertation Structure

To provide an answer for this research question, this study will be structured in five chapters:



The first chapter in this study will introduce the reader to the subject of interest, providing a background of the topic of research and outlining the research question that this study intends to answer.

The second chapter will deal with the existing literature on the nowcasting practice and the models currently in use by market practitioners. A discussion will take place towards the literature gap and this study's contribution to the existing literature in this field.

The third chapter will describe the methodology which will be employed to tackle the research question. Namely, the theoretical model employed will be described and the process which will be undergone to statistically evaluate and optimize our dataset will be discussed. The process undergone for estimation will be examined, and an interpretation will be done to confirm the feasibility of our models.

The fourth chapter will include a discussion of the results obtained and an analysis of the robustness of our models. The economic indicators selected, the results of the statistical tests, and a detailed interpretation of the forecast errors obtained will be discussed. The results of our model obtained from the nowcasting application will then be compared with another model with publicly available results.

The fifth chapter will conclude this study with a brief discussion of the results obtained considering the limitations encountered, and policy implications and recommendations for further research.

Chapter 2 – Literature Review

2.1 Background to Nowcasting

Nowcasting, a term coined in meteorology, is the practice of estimating a macroeconomic variable's very near state. Macroeconomic variables which may be nowcasted include GDP and inflation growth rates (European Commission; Kapetanios, Marcellino, & Papailias, 2018). Nowcasting these variables is important since their official estimates are published at a low frequency. To illustrate, in the United States GDP estimates are typically issued quarterly whereas inflation indices are issued monthly (Federal Reserve Banks of St. Louis). There is also the presence of a publication lag which is the delay between the publication date and the end of the reference period. Furthermore, since the official figures are estimates they are also subject to regular re-estimations and revisions. This creates what is known in literature as the 'ragged edge' problem (Bouwman & Jacobs, 2009; Wallis, 1986). This means that different important macroeconomic variables are not available at similar time horizons making proper analysis of the macroeconomic environment difficult and impractical.

Nowcasting techniques tend to incorporate a variety of high-frequency data and complex econometric techniques to estimate these key macroeconomic variables, and therefore produce real time estimates of their current state before any official estimates are published (Ghysels & Marcellino, 2018). Therefore, nowcasting these variables helps mitigate the consequences which arise from the 'ragged edge' problem by mitigating the incompleteness of important macroeconomic data, bridging the gap between high and low-frequency data, enabling for the incorporation of information from different sources and time horizons.

Several institutions actively research and use nowcasting in their operations, using various models, namely the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of St. Louis, the Federal Reserve Bank of Cleveland, the European Central Bank (ECB), the International Monetary Fund (IMF), the Federal Reserve Bank of New York, and the Bank of England, amongst others. One of the most notable macroeconomic nowcast with publicly available results is the Federal Reserve Bank of Atlanta's *GDPNow*™ model. This model by the Atlanta fed employs several techniques including a Dynamic Factor Model (DFM) and Bridge equations to nowcast US GDP growth. What this means is that the Atlanta Fed nowcasts the thirteen individual subcomponents which make up GDP growth and aggregates them to create their real GDP growth nowcast.

Nowcasts with publicly available results are not so common outside of the United States. The reason is that the United States is commonly used as the sample to test nowcasting models since it is characterised by its availability of recent high-frequency data (Hopp, 2022). Furthermore, the United States enjoys several research institutions which produce robust economic surveys providing as good leading indicators of the state of the US economy. Moreover, countries with restrictions in availability of data, inaccessible data, or large delays in the publication of data, struggle to produce robust econometric models to nowcast important macroeconomic variables (IMF, 2021). Hence why, econometric models such as the Mixed Data Sampling (MIDAS) model are crucial in allowing for the employment of mixed frequency data and offering parametric flexibility to allow developed countries to have practical nowcasting models.

2.1.1 Dynamic Factor Model

Due to the nature of unsynchronised publication lags in economic data, time series models may incur missing observations for specific points in time. The Dynamic Factor Model is commonly employed since the individual indicators are driven by a few factors which are common to each variable. This means that the economic variables tend to co-move mostly due to sharing the same common factors and cyclical fluctuations (Ghysels & Marcellino, 2018). Examples of these common factors may include the business cycle, monetary policy, productivity, and global economic conditions. The problem arises since these few 'common factors' tend to be unobservable.

2.1.2 Bridge Model

An inflation nowcast for the United States is made publicly available by the Federal Reserve Bank of Cleveland. The Cleveland Fed uses a Bridge model which similarly nowcasts the individual subcomponents which make up the Consumer Price Index (CPI) and aggregates them to create a global nowcast. These subcomponents may include a combination of core inflation, inflation for food, and inflation for energy.

A Bridge model for the United States CPI may take the form:

$$\pi_{cpi} = W1(\pi_{core\ cpi}) + W2(\pi_{food}) + W3(\pi_{energy})$$

Where: $\sum W_n = 1$

While Bridge models are used widely due to their robustness, the consequences of using such a model becomes immediately visible, being that a degree of complexity arises when having to nowcast a set of different subcomponents before arriving to the final nowcast. Furthermore, the Bridge models suffer from partly-fixed weights, being that the weighting given to the different components of the CPI are partly fixed over time (Schumacher, 2014).

2.1.3 Forecasting and Nowcasting

While time-series forecasting provides reasonably reliable outputs, it is vulnerable to exogenous events and loses predictive accuracy with an increasing forecasting horizon. Furthermore, econometric forecasting struggles to predict and are subject to break down around turning points (Brooks, 2019). In fact, a characteristic of forecasting is that they increase in accuracy the closer they get to the forecast date and update themselves with new, recent information. This is not the case with nowcasts, since nowcasts estimate for a very near time horizon instead. The challenge when building nowcasting models is therefore to employ suitable high-frequency regressors so that the short-term movements in the low-frequency macroeconomic variable may be effectively measured (Bańbura, Giannone, Modugno, and Reichlin, 2013). This does not however make econometric forecasting a substitute for judgement.

Therefore, nowcasting is less vulnerable to exogenous events as it uses current alternative data available to estimate what the low-frequency variable stands at in its very near state. Recency is therefore induced by including data with high frequencies and released at a low publication lag. A robust nowcast may therefore also be a significant component of a forecasting model. This is since it will help the researcher interpret forecast errors and in turn alleviate macroeconomic surprises (Reichlin, 2019). Consequently, nowcasting models also allow the user to regularly monitor the factors which are driving the forecast of the change in the growth rate of the inflation indices.

2.1.4 Artificial Intelligence Approach

An innovative approach to nowcasting macroeconomic indices, apart from time-series prediction, is by the employment of Artificial Intelligence. Traditional time-series forecasting models attempt to predict the dependent variable using pre-defined linear relationships between the information set, being the independent variables and the dependent (forecasted) variable (Jung, Patnam & Ter-Martirosyan, 2018). Machine learning and Deep Learning algorithms, which are now becoming an increasingly popular approach at predicting low-frequency information, make no assumptions regarding the underlying relationships between the features and the target variable (Hopp, 2022). The drawback, however, is that these methods lack explanatory power by failing to identify what factors are driving the forecasts.

2.2 Overview and Importance of Nowcasting Inflation Indices

The inflation growth rate is approached as an epistemic uncertainty. This means that the real-time movement in the inflation growth rate is not easily available, however, can be roughly determined when using the correct tools and resources. To accomplish this, complex econometric models and a varied information set are utilised comprising of leading and lagging alternative economic indicators (Modugno, 2013).

Leading indicators can be defined as variables which are significant due to their ability at uncovering future trends. Examples include construction permits and average weekly hours of all employees. The explanation is that an increase in construction permits shows a positive sentiment and an expectation of future economic growth (Coble and Pincheira, 2017). Furthermore, when there is negative economic sentiment and firms cut costs by laying off workers, the first visible signal would be a reduction in average weekly hours of all employees. This is since layoffs are a slow and costly process, and firms would avoid having the need to rehire in the future, so the first step taken by firms would be to cut on overtime hours. This would appear as a reduction in the average weekly hours worked by the firms' employees (Glosser and Golden, 1997).

On the other hand, lagging indicators can be defined as variables of which the significance in their variation is maximised when considered in conjunction to previous economic performance. Unemployment is a perfect example of a lagging indicator. A short-term increase in the unemployment rate may be perceived as a positive signal of economic recovery. This is since in periods of negative economic sentiment, unemployed persons may become discouraged and therefore no longer accountable for in the unemployment rate since they stop seeking employment. This period between the first increase in the unemployment rate and the subsequent decrease can be identified as the timing lag of the economic indicator (Eduardo André Costa, Maria Eduarda Silva and Ana Beatriz Galvão, 2024).

Leading indicators are therefore not necessarily superior to lagging indicators since they are employed for their different functions being that leading indicators are typically used for discovering trends whereas lagging indicators have the function of confirming trends and changes in trends.

2.2.1 Consumer Price Index

The Consumer Price Index (CPI), an index which is published monthly, provides a measure of the rate of change of inflation of a representative basket of goods and services (U.S. Bureau of Labor Statistics, 2023). In addition to this, the Core CPI is an index which exclude food and energy prices. Since food and energy prices tend to be distinguishably volatile, core indices provide reliable measures which help market practitioners better identify inflationary or deflationary periods.

2.2.2 Cost-Push and Demand-Pull Mechanisms

These inflationary pressures may be categorised between cost-push and demand-pull mechanisms which explain the rising prices in an economy (Schwarzer, 2018). Namely, cost-push pressures are those stemming from supply-side factors, resulting in increasing production costs being passed on to consumers. These may be due to increasing raw material costs, increasing wages, taxes, and regulations. On the other hand, demand-pull pressures are caused due to an economic shortage resulting from increased consumer spending, excess government spending, monetary stimulus, and foreign demand.

2.2.3 CPI as a Measure of Poverty

Despite many attempts by researchers, the concept of poverty has proven difficult to scale and consequently measure. This is since the needs of people continuously evolves, creating new forms of dependencies, especially with the development of new technologies. Nobody would have expected a few decades ago that not having access to the internet at home would indicate that a person is economically disadvantaged (Pope Francis, 2020). Furthermore, the concept of poverty tends to be relative. While some people may have the necessities of life, such as clothing, food, and access to affordable housing, these same people may be deprived economically to match the median standard of living enjoyed by society.

Amongst the macroeconomic variables, inflation indices such as the CPI may prove to be our most useful tool in measuring poverty due to its direct implication on the lowest classes of income in society (Institute for Research on Poverty, 2024). This is since the CPI measures the loss of purchasing power from a currency occurring from an increase in the prices of a pre-defined basket of goods. The CPI therefore also measures the deterioration of a person's income and his cash savings.

2.2.4 Applications of Inflation Nowcasting Models

In the context of investments, the inflation premium is a principal factor employed in pricing financial assets. Namely, the valuation of fixed income securities is heavily reliant on the expectation of the direction of future inflation (Cieslak and Pflueger, 2023). The presence of information asymmetry, due to the publication delay inherent in the inflation indices, may result in pricing shocks affecting the yields and valuation of financial assets, especially when the official estimates released deviate significantly from the market's expectations. The employment of nowcasting models may therefore guide investment managers on mitigating these macroeconomic surprises, providing preliminary estimates of the inflation indices before their official release.

Furthermore, the inflation rate is a principal component monitored by central banks and federal reserves in guiding monetary policy decisions, particularly those regarding changes in interest rates. Inflation surprises may disrupt the expectations of the central bank regarding their planned future monetary policy decisions (Federal Reserve Bank of Cleveland, 2023). Moreover, central banks may also struggle to monitor the effects of their inflation reduction mechanisms, being most prominent in scenarios of crises where the central bank pushes for rapid decisions aimed at alleviating severe inflationary pressures.

2.3 Mixed Data Sampling Model

The Mixed Data Sampling (MIDAS) model, proposed by Eric Ghysels and co-authors Santa-Clara, & Valkanov, (2004), provide for time-series regressions which allow for the use of variables sampled at different high frequencies, including real-time data, to directly estimate or forecast the lower frequency variable (Ghysels & Marcellino, 2018). In other words, it relates variables measured at a frequency to current and lagged values of other highly correlated variables measured at different frequencies (Mogliani & Simoni, 2020). The model was not originally intended for nowcasting however the application eventually became obvious.

The key features of the MIDAS model are its parsimony and flexibility, since it can flexibly deal with a large amount of data sampled at mixed frequencies to provide direct forecasts of the dependent variable (Kohns & Potjagailo, 2022; Mogliani & Simoni, 2020). This model is explained in greater depth in the research methodology section of this study.

A MIDAS regression with a single high-frequency regressor for m-steps ahead forecasting, with high-frequency data available up to $x_t^{(m)}$ is given by:

$$y_t = \beta_0 + \beta_1 B\left(L^{\frac{1}{m}}; 0\right) x_{t-1}^{(m)} + \varepsilon_t^{(m)}$$

The frequency of the dependent variable is referred to as the interval of reference. Furthermore, the timing lag employed allows for the tailoring of forecasting horizons, allowing for the employment of the model for nowcasting applications. It is not an autoregressive model since autoregression implies that the data is sampled at the same frequency in the past (Ghysels, Santa-Clara, & Valkanov, 2004).

2.3.1 Parameterization

Parameterization of the MIDAS equations determine the weighting that the high-frequency regressors are given within the model. It is a critical component as it ensures that the weights have positive coefficients and that the weighting of the regressors will always sum up to one. In this model, the parameters are highly flexible meaning that they can take on various shapes with few parameters (Ghysels & Marcellino, 2018). The weighting function used depends solely on the data employed.

Two of the most popular non-linear finite polynomial specifications employed for the MIDAS model are what are referred to in the literature as the Exponential Almon lag weighting function and the Beta lag specification (Ghysels, Sinko and Valkanov, 2006).

2.4 Academic Literature Related to Nowcasting

2.4.1 Bańbura, Giannone, Modugno, and Reichlin, 2013

This working paper introduces the main themes surrounding the nowcasting practice, such as the models commonly employed, the monitoring of real-time data, and the reaction of policy makers to deviations from the expectation of the state of the economy following macroeconomic data releases. Throughout this study, the authors review several different statistical approaches to nowcasting. Particularly, this study is based on multivariate dynamic models which can be represented in the state space representation, solving the information incompleteness present through the employment of the Kalman filter. Finally, this study also reviews the application of Dynamic Factor Models, mixed-frequency VAR models, Bridge equations, and single equation models such as MIDAS.

The results from this study suggest that high-frequency data availability positively contributes to the accuracy of US GDP nowcasting models, exhibiting surveys as a notably significant predictor of the state of the US economy. Furthermore, this study projects the granular fluctuations in the US GDP which are typically not available in the official releases.

2.4.2 Modugno, 2013

This paper proposes an inflation nowcasting model, exploiting the availability of weekly and daily data within the United States economy. This study adopts a Dynamic Factor Model, discarding the MIDAS model due to treating the high-frequency variables as exogenous, allowing the researcher the possibility to forecast all the regressors individually. Moreover, the model employed by the researcher forecasts the sub-components of the inflation indices before aggregating them to create the global nowcast.

The results exhibit a significant impact on the model performance following the inclusion of high frequency energy and raw material price data within the data set. Furthermore, financial variables were noted to not demonstrate any considerable improvement to the accuracy of the models.

2.4.3 Schumacher, 2014

This discussion paper compares the application of the single-equation approach against the traditionally used Bridge models in nowcasting GDP growth. Namely, this paper evaluates the difference between the MIDAS model and the Bridge equations used in GDP nowcasting models. Moreover, this paper analyses how these models tackle multi-step ahead nowcasting, the difference in weighting functions between these models (estimated versus time-aggregation weights), and how current data is treated within the models.

The results of this discussion paper indicate that the MIDAS model produces direct multi-step nowcasts, whereas the Bridge equations are based on iterated forecasts of the subcomponents of the macroeconomic variable. Furthermore, the study finds that the Bridge equations suffer from partly-fixed weights of the economic indicators employed due to time aggregation. Finally, this paper suggests that there is no single dominant approach to nowcasting macroeconomic indices, being that the robustness of the results mainly stem from the indicators selected and the evaluation sample, with model averaging yielding to more consistent nowcasts over time.

2.4.4 Higgins, 2014

This working paper describes the employment of Bayesian Vector Autoregression (BVAR) and Bridge equations to nowcast the 13 subcomponents which make up the Federal Reserve Bank of Atlanta's proprietary *GDPNow* Real GDP nowcast. The methodology borrows from several authors utilizing diverse approaches to forecasting the different subcomponents of the real GDP index. Namely, the approaches of Banbura, Giannone, and Reichlin (2008), Stock and Watson (2002), and Chin and Miller (1996) are utilised. Through modelling these subcomponents, an underlying factor representing the business cycle is derived reflecting national economic activity. Finally, aggregating the subcomponents results in an estimate of current real GDP.

The results of this working paper suggest that forecasting performance of nowcasting models improve upon the release of new, real-time macroeconomic data. Furthermore, the researcher recommends that the Mixed Data Sampling (MIDAS) model may be a more effective approach around business cycle turning points due to its ability to exploit higher-frequency (weekly or daily) data.

2.4.5 Marcellino, 2021

This study reviews the employment of mixed-frequency models, such as the Unrestricted-Mixed Data Sampling (U-MIDAS) approach, Dynamic Factor Model, and machine learning techniques (Neural Networks) in nowcasting GDP growth in a small open economy.

The results of this paper suggest that during turbulent and volatile crisis periods, the Mixed-Frequency DFM model, Neural Networks, and Three-Pass Regression Filter performed best in nowcasting the GDP growth of Luxembourg. Furthermore, simpler specifications such as Autoregressive models were proven sufficient in 'calm' periods, finding that nowcasting error is largest during deep recessions and fast recoveries. Finally, the study supports the usefulness of including surveys of expected future economic conditions and employment indicators as principal indicators for reliable nowcasting models.

2.4.6 Dauphin, et al., 2022

This working paper employs unstructured data and nontraditional variables, such as google search and air quality, to nowcast GDP growth across European economies during calm and crisis times.

This study finds that the DFM performs better during calm periods whereas ML techniques perform the strongest at identifying turning points. The researcher concludes by stating that the analysis of back-testing results and human judgement is crucial in identifying which model can be trusted to provide the most robust and informative nowcasts depending on the country and time horizon in question.

2.4.7 Andreini, Hasenzagl, Reichlin, Senftleben-König, & Strohsal, 2023

This paper proposes a GDP nowcasting model for the German economy. Employing the DFM, the model performs forecasts not only for GDP growth, but also for the other key variables present within the model. This model, therefore, produces real-time updates for the state of the macroeconomic variable along with short-term forecasts of the included economic indicators.

This paper, like the findings of previous authors, finds that financial variables tend to not contribute significantly to the nowcasting performance of real GDP nowcasting models. Furthermore, the study also indicates that while stock prices tend to reflect the state of the business cycle, they do not possess any leading information which is useful for nowcasting models.

2.4.8 Knotek II & Zaman, 2023

Building on Higgins work on GDP nowcasting, this working paper proposes a nowcasting model for inflation indices within the United States which is in use and is published daily by the Federal

Reserve Bank of Cleveland. Highlighting the growing interest in nowcasting inflation indices, this paper reviews the mixed-frequency models commonly employed in nowcasting platforms. Furthermore, empirical studies are conducted on the Deterministic Model Switching (DMS) and Bridge equations proposed by Knotek & Zaman (2017).

A key feature of this model is that it employs only ten data series, one of which being weekly, and another of daily frequency. The results from this paper show that this model outperforms competing mixed-frequency inflation nowcasting models, making it the benchmark inflation nowcasting platform for the United States economy.

2.5 Contribution to the Literature

Following the analysis of previous literature related to the nowcasting practice, it became clear that nowcasting inflation growth rates have remained particularly unresearched even though it is one of the most monitored macroeconomic variables (Modugno, 2013). Furthermore, most studies about inflation nowcasting, such as the case of Knotek II & Zaman, (2023) and Modugno, (2013) discard the Mixed Data Sampling (MIDAS) approach due to the suboptimization and poor results achieved compared to models employing the Dynamic Factor approach and Bridge equations. This research proposes an inflation nowcasting model using alternative data with mixed frequencies via a novel application of the Mixed Data Sampling (MIDAS) time series model.

Chapter 3 – Research Methodology

3.1 Introduction

To nowcast the inflation growth rates for the United States, several time-series models will be constructed which will involve combining various categories of economic indicators with unsynchronised publication lags and mixed frequencies, being monthly and weekly.

The premise behind these models is the assumption that there is a significant co-movement between the economic indicators employed and the inflation indices. Furthermore, it is also assumed that this co-movement can be observed, measured, and applied in an out-of-sample test to produce robust nowcasts of the corresponding inflation growth rates. For estimation, testing, model selection, and nowcasting, the Mixed Data Sampling (MIDAS) estimation functions from the R package, MIDASR, will be employed.

The methodology starts with the formulation of an estimable theoretical model. In this section, a discussion of why the MIDAS model was chosen over other models will be provided. Furthermore, the implications of the different possible parameterizations of the model will also be examined.

The next step in our methodology will describe the data collection process. Namely, the sources from which we will source our data, and the criteria used to select our economic indicators will be identified. An explanation will be provided regarding the treatment of the data, including the necessary seasonal adjustments and transformations. The statistical tests which will be employed to ensure the optimization of our model will also be discussed.

Once an indicative dataset has been constructed, an estimation of the model will take place. A comparison of the three different weighting functions employed for the MIDAS model will be provided. The MIDAS model using the Almon Polynomial Distributed Lag structure and two non-linear weighting schemes, the MIDAS Exponential Almon and the MIDAS Beta specifications will be evaluated. Furthermore, the process used for lag selection will be discussed.

Forecasting error measurement will be employed to measure the robustness of these models. If the models are found to be statistically adequate, they will then be used for analysis and extrapolation. The work done in this research will then contribute towards economic and financial theory.

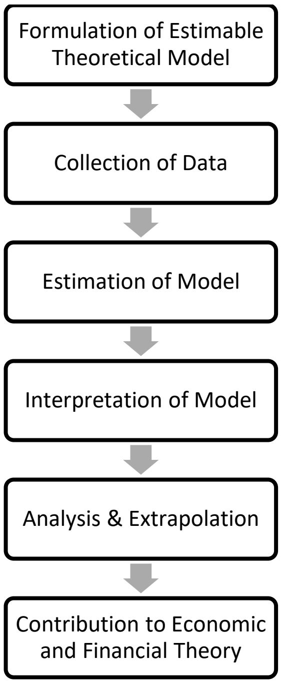


Figure 1: Formulation of Econometric Model (Brooks, 2019)

3.2 Theoretical Model

The Mixed Data Sampling (MIDAS) model developed by Eric Ghysels and co-authors Santa-Clara, & Valkanov, (2004), was selected as the approach to nowcast the inflation growth rates due to this model having several major advantages over the commonly used Bridge models being:

1) This model offers parsimony due to being a tightly parameterised single-line equation model, allowing for direct forecasts of the low-frequency variable instead of the iterated approach employed by the Bridge models (Schumacher, 2014),

2) This model allows for the exploitation of different frequencies of data including high-frequency data, and

3) This model is characterised by the flexibility of its parameters. This is because the model can flexibly deal with a large amount of data sampled at mixed frequencies to provide direct forecasts of the dependent variable (Kohns & Potjagailo, 2022; Mogliani & Simoni, 2020).

Thereby this model will allow for an accurate answer to our research question, being to assess if the MIDAS model produces robust enough results to be considered an alternative to the traditionally used Bridge equations. Unlike Bridge models, however, by using the MIDAS approach we will not be restricted by fixed weights since the parameters of the MIDAS equations are highly flexible and able to take upon various parametric shapes. Furthermore, optimization of the time-series models will be possible since the use of mixed frequencies of economic indicators will be permitted. Therefore, since this model allows the exploitation of high-frequency data, it will provide the advantage of recency, which is expected to increase the robustness of the time-series models.

3.2.1 Mixed Data Sampling Model

The Mixed Data Sampling (MIDAS) regressions, proposed by Eric Ghysels and co-authors Santa-Clara, & Valkanov, (2004), provide for time-series regressions which allow for the use of variables sampled at different high frequencies, including real-time data, to directly estimate or forecast the lower frequency variable (Ghysels & Marcellino, 2018). In other words, it relates variables measured at a frequency to current and lagged values of other highly correlated variables measured at different frequencies (Mogliani & Simoni, 2020).

A MIDAS regression with a single high-frequency regressor for m-steps ahead forecasting, with high-frequency data available up to $x_t^{(m)}$ is given by:

$$y_t = \beta_0 + \beta_1 B\left(L^{\frac{1}{m}}; 0\right) x_{t-1}^{(m)} + \varepsilon_t^{(m)}$$

Where:

Y = the dependent variable,

x = the regressor, and

m = the frequency.

ε = the error term, and

$B\left(L^{\frac{1}{m}}; 0\right)$ is a lag distribution.

Therefore:

$B\left(L^{\frac{1}{m}}; 0\right) = \sum_{k=0}^{kmax-1} B(k; 0) L^{\frac{k}{m}}$ must sum to one, and $L^{\frac{1}{m}}$ is a lag operator.

The model would incur missing observations for specific points in time since some of the data releases are not synchronised. Therein, not all variables can be observed at all dates, especially since some of them would be released monthly whereas others would be released weekly. Therefore, a problem of time misalignment would usually exist; however, the model already solves this by employing a timing lag being the exact timing difference between the higher frequency variables using the lag operator $L^{\frac{1}{m}}$. The frequency is then denoted within the model using the superscript of the regressor $x_{t-1}^{(m)}$ where *m* denotes the frequency of the regressor (Ghysels, Sinko and Valkanov, 2006). The frequency of the dependent variable will therein be referred to as the interval of reference.

Parameterization of the MIDAS equations determine the weighting that the high-frequency regressors are given within the model. It is a critical component as it ensures that the weights have positive coefficients and that the weighting of the regressors will always sum up to one. In this model, the parameters are highly flexible meaning that they can take on various shapes with few parameters. For example, weightings of the regressors with different lags can take on shapes exhibiting a slow decline, a rapid decline, or a hump shape, meaning that the weighting of the high-frequency regressors demonstrate a brief increasing period before they decline (Ghysels & Marcellino, 2018).

Lag selection in this model is purely data-driven, therefore the use of the information criterion, such as the Akaike information criterion and the Schwarz Bayesian information criterion is not required. Estimation of the parameters using the data employed will determine the shape given to the weights, meaning that the optimal number of lags will be derived from the regression (Asimakopoulos, Paredes and Warmedinger, 2013). Therefore, parameterization tackles the issues which may arise related to lag selection.

An important consideration is that the selection of the parameters of the model is horizon-specific. Therefore, the model would need to be estimated for a specific forecasting horizon (Ghysels & Marcellino, 2018). For any other forecasting horizon, the model would need to be re-estimated and different parameters employed.

Two of the most popular non-linear finite polynomial specifications employed for the MIDAS model are what are referred to in the literature as the Exponential Almon lag weighting function and the Beta lag specification (Ghysels, Sinko and Valkanov, 2006). This is since these methods of parameterisations allow for the minimization of parameters while preserving the parametric flexibility of the model.

The Exponential Almon lag weighting function which is based on the Almon lag polynomial specification derived from the distributed lag model is expressed as follows:

$$B(k;\theta) = \frac{\exp\left(\theta_1 k + \cdots + \theta_Q k^Q\right)}{\Sigma_{k=0}^{k}\ \exp\left(\theta_1 k + \cdots + \theta_Q k^Q\right)}$$

In contrast to the Exponential Almon lag function, the Beta Lag function has also two parameters theta $\theta = (\theta_1, \theta_2)$, that is:

$$B(k;\theta_1,\theta_2) = \frac{f\left(\frac{k}{K},\theta_1;\theta_2\right)}{\Sigma_{k=0}^{K} f\left(\frac{k}{K},\theta_1;\theta_2\right)}$$

Where:

$$f(x,a,b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

and

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1}\, dx$$

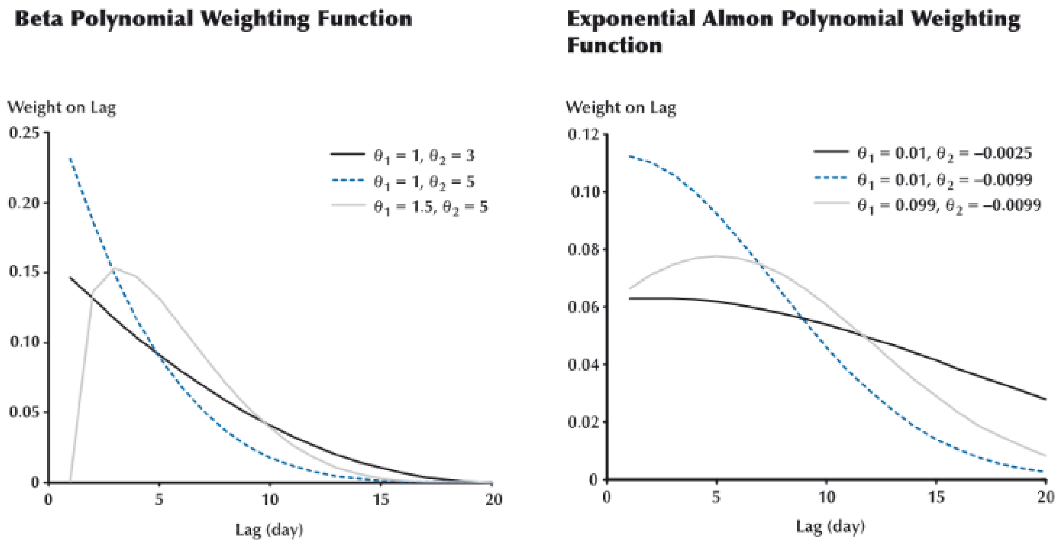


Figure 2: Beta and Exponential Almon Polynomial Weighting Functions (Armesto, Engemann and Owyang, 2010)

The weighting function used depends solely on the data employed. The parameterisation which allows for the most efficient use of parameters within the model should always be preferred. There are penalties associated with having to estimate too many parameters within the model. Namely, the model may be prone to distraction from the sample noise, resulting in inferior performance when being exposed to unseen data during the out-of-sample test (Stock & Watson, 2020). Furthermore, estimating many parameters can be computationally intensive and may result in a complex model, making it difficult to produce a meaningful interpretation.

On the other hand, increasing the efficiency of the model's parameters is beneficial as it contributes to the robustness of the results, making the model less prone to overfitting. Furthermore, interpretability is increased and the risk of spuriosity within the model is reduced.

The MIDAS model, as proposed by Ghysels, Santa-Clara and Valkanov, (2004), makes several assumptions including:

1. Stationarity, therefore, the variables employed must be tested for and exhibit weakly stationary characteristics,

2. No serial correlation is present in the model's residuals. This means that the error terms are not correlated over time,
3. A constant, homoscedastic variance across the model's residuals,
4. No other measurement errors, such as perfect multicollinearity, are present in the regressors employed, and
5. That the high-frequency and low-frequency observations are correctly aligned in time.

3.2.2 Nowcasting Platform

Nowcasting models depend on the frequent updating of continuously releasing data which is needed to maintain accuracy. This is because old data may fail to capture current economic trends and result in forecasting errors or macroeconomic surprises (Reichlin, 2019). Since data is being published in real-time, the models must also be updated frequently to produce nowcasts up to the point of the release of the official estimates (Kapetanios, Marcellino, & Papailias, 2018; Armesto, Engemann, & Owyang, 2010; Bańbura, Giannone, Modugno, & Reichlin, 2013).

We can define a vector of stationarized monthly variables at time *t* as

$$y_t^m = (y_{1,t}^m, y_{2,t}^m, y_{N,t}^m).$$

Similarly, we can define similar vectors for weekly variables at time *t* as

$$y_t^w = (y_{1,t}^w, y_{2,t}^w, y_{N,t}^w).$$

This data can be collected in the information set:

$$\Omega_v = (y_t^m, y_t^w).$$

Therefore, Ω_v is defined as an information set comprising a variety of highly correlated variables taking monthly and weekly frequencies. However, since the data is released in a non-synchronous manner, we can say that a ragged edge is present in our information set (Bańbura, Giannone, Modugno and Reichlin, 2013).

The nowcasting platform will adhere to the following process. Input data, being the information set originally collected, Ω_{v-1}, will be used to produce what we will refer to as the original model prediction $\mathbb{P}(y_t|\Omega_{v-1})$. This original prediction will determine what method of parameterisation will be employed and the optimal lag selection. The original nowcast can therein be considered our training sample.

The dependent variable in our model, y_t, which releases at a significant publication delay and which we will attempt to nowcast, will be of the monthly frequency and therefore that will be considered our interval of reference. The problem of time misalignment within our information set will be solved by the employment of timing lags within our theoretical models.

The employment of the FredR API will allow for an automated feed of data, I_v, from several external sources. This will allow the data set to automatically update with the most recent releases and revisions every time the model is run. Ω_v and Ω_{v-1} will represent a consecutive information set whereby I_v is the news reflected in the latter period Ω_v.

This means that we must also create revisions for our original prediction using the new information, creating what will be referred to as our revised model prediction $\mathbb{P}(y_t|\Omega_v)$. The updating of this model prediction is what will be referred to as the out-of-sample test.

$$\mathbb{P}(y_t|\Omega_v) = \mathbb{P}(y_t|\Omega_{v-1}) + \mathbb{P}(y_t|I_v)$$

Therefore, the revised model prediction, or the nowcast, can be defined as the sum of the original model prediction and a factor I_v representing news.

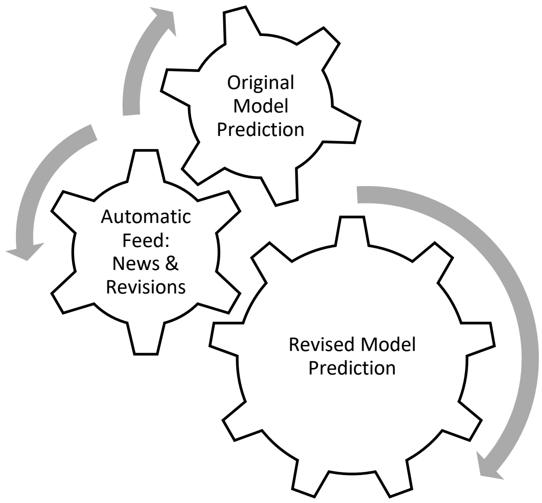


Figure 3: Nowcasting Platform (Produced for this study)

3.3 Data Collection

3.3.1 Description of Data

The dependent variables in our models will be the inflation growth rates which we will nowcast. Within the United States model, our dependent variables will be the growth rates of the Consumer Price Index (CPI) and the Core CPI. The Consumer Price Index (CPI), an index which is published monthly, provides a measure of the rate of change of inflation of a representative basket of goods and services (U.S. Bureau of Labor Statistics, 2023). In addition to this, the Core CPI is an inflation index which excludes food and energy prices. Since food and energy prices tend to be distinguishably volatile, core indices provide reliable measures which help market practitioners better identify inflationary or deflationary periods.

The economic data used for our models will be time-series, structured macroeconomic indicators. There are six categories of data which will be employed in the models:

Prices	This catergory includes the CPI series and prices of various commodities
National Income & Product Accounts	This category includes national accounts such as Personal Consumption Expenditure (PCE) and federal spending
Money, Banking & Finance	This category includes monetary variables such as M2 and market yields on treasury securities
Population, Employment, & Labor Markets	This category includes employment and labour productivity figures
Production & Business Activity	This category includes data related to production, construction, and manufacturing
Surveys	This category includes surveys related to consumer sentiment, business conditions, and inflation expectations

These categories of data were selected due to their expected high correlation with inflation (Ghysels, Santa-Clara, & Valkanov, 2004; Kapetanios, Marcellino, & Papailias, 2018). Furthermore, each category involves easily accessible data which is released at different frequencies ranging from weekly to monthly. This makes them suitable for the experiment since they satisfy the condition of recency.

3.3.2 Data Sources

Data selection for the model will be based on availability and is sourced from the Federal Reserve Bank of St. Louis's Federal Research Economic Database (FRED) at 'https://fred.stlouisfed.org/'. This database was selected since it efficiently redistributes different data releases and revisions from various national, international, and private sources in a single platform, which can easily be extracted from the FRED Application Programming Interface (API).

The main data sources for the United States model include:

- Board of Governors of the Federal Reserve System (US) at 'https://www.federalreserve.gov/'
- Federal Reserve Bank of Atlanta at 'https://www.atlantafed.org/'
- Federal Reserve Bank of Cleveland at 'https://www.clevelandfed.org/'
- U.S. Bureau of Labor Statistics at 'https://www.bls.gov/'
- U.S. Census Bureau at 'https://data.census.gov/'

- U.S. Energy Information Administration at 'https://www.eia.gov/'

The use of the FredR API will allow for the automatic updating of our model using the latest data releases or revisions from these sources every time the program is executed.

3.3.3 Data Selection Criteria

Employing good data selection criteria is important to ensure that the economic indicators selected positively impact the forecasting ability of the model. We would not want to include variables in the model which are insignificant and create noise, thereby distracting our model from producing robust forecasts. Our selection criteria will reflect findings by other authors which used similar economic data to nowcast different low-frequency macroeconomic variables.

In any time-series model, the use of the highest frequencies of data possible should induce robustness due to the element of recency which they introduce within the model. For our model, variables that are of the frequencies ranging from monthly to weekly will be selected. The study by Marcellino and Sivec, (2021), demonstrates how an Unrestricted Mixed Data Sampling (U-MIDAS) approach to nowcasting Real GDP can be used when the frequency mismatch is minimal. However, since we are nowcasting variables with a monthly frequency, being the inflation indices, the basic restricted MIDAS model will be sufficient given that we have a considerable frequency mismatch to benefit from.

When selecting economic indicators, variables are encountered such as Gross Domestic Product (GDP) which are made up of a set of sub-components which can be included individually. Similarly, in our model, the individual sub-components of the CPI, being the CPI for food and energy amongst other measures, can be included. The study by Andreini, Hasenzagl, Reichlin, Senftleben-König, and Strohsal, (2023), demonstrates however that models using non-segregated indicators outperform those using segregated indicators due to implicitly including the same information set and variability while preserving the parsimony of the model.

Furthermore, the paper by Giannone, Reichlin, and Small, (2006), exhibits the difference between using real against nominal economic variables when nowcasting inflation indices. While both variables are significant in estimating inflation, nominal variables tend to be released at a higher frequency due to requiring fewer price adjustments. For this reason, nominal variables are characterised by timeliness and will be included within our model since we expect them to outperform real variables when nowcasting inflation indices.

3.3.4 Publication Lag

Publication lag can be defined as the difference in time, expressed in days, between the publication date of a release of data and the end of the reference period to which it pertains (Bańbura & Rünstler, 2007).

$$Publication\ Lag\ =\ Publication\ Date\ -\ End\ of\ Reference\ Period$$

While this publication lag may vary month by month, it would be wise to include variables in the models which have historically reflected a low publication delay. Furthermore, it is expected that the high-frequency regressors will have a lower publication lag as the updating of data becomes more frequent (Giannone, Reichlin, and Small, 2006).

Minimizing the publication lag is crucial for the time-series models since timeliness is a key factor in producing robust estimates for inflation indices. Employing data which is released late after the release of the inflation indices will have minimal effect on the accuracy of our nowcasts.

3.4 Data Adjustments and Transformations

3.4.1 Seasonal Adjustment

Once the data is collected, necessary adjustments and transformations must be made. Unless a data series is readily provided with adjustments for seasonality the series will have to undergo seasonal adjustment. This is to remove effects in the data series related to seasonality.

Seasonal patterns which we aim to mitigate from our time series include fluctuations which repeat themselves regularly and are typically brought about by predictable weather patterns and holiday effects. For this X-13ARIMA-SEATS by the U.S. Census Bureau will be employed.

This method requires that the data series be in a monthly format and that negative or null values are not present. It also assumes that a data series can be decomposed into three factors which are the Trend component, the Seasonal component, and the Random component. This takes the form of:

$$Y_t\ =\ T_t\ \times\ S_t\ \times\ I_t$$

If the data series has negative values present, however, the additive approach can be used instead of the multiplicative approach. This takes the form of:

$$Y_t\ =\ T_t + \ S_t + \ I_t$$

Seasonal adjustment is therefore the process whereby these three elements are estimated to omit the seasonal component from the data series, thereby producing what is referred to as a seasonally adjusted series.

3.4.2 Returns Transformation

The raw variables employed must then be transformed into a difference series. There are two different ways to go about this. Logarithmic difference or first difference will be employed as appropriate to ensure that our model is well-calibrated (Brooks, 2019). The logarithmic difference tends to be employed for financial and economic data where the data may experience exponential growth or decay. On the other hand, the first difference is typically used for macroeconomic indicators where the data experiences constant or linear growth and the variance is constant over time.

The formula for the first difference returns:

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100\%$$

The formula for logarithmic difference returns:

$$R_t = ln\left(\frac{p_t}{p_{t-1}}\right) \times 100\%$$

Where:

R_t = the return at time t,

p_t = the data point at time t, and

ln = the natural logarithm

The use of logarithmic difference returns has the characteristic that the data series can be interpreted as a continuously compounded return series. These returns can be added up to derive the return of a variable for a specified time horizon. However, this cannot be applied to all variables since a limitation of using logarithms is that the log of the sum is not equal to the sum of logs.

A disadvantage of using returns for a data series is that every difference reduces the number of observations by one.

3.5 Statistical Testing

In this section, we will evaluate and validate the assumptions which are associated with the time-series models employed in this study.

3.5.1 Hypothesis Testing

To test any hypothesis, we will determine what we refer to as the null hypothesis (denoted as H_0) and the alternative hypothesis (denoted as H_1). The null hypothesis H_0 is the statement that we will be testing. If the test rejects the null hypothesis, then the alternative hypothesis will remain as the outcome of interest.

Hypothesis testing can take the form of either a two-tailed test or a one-tailed test. For example:

$$H_0: \beta = 1;$$

$$H_1: \beta \neq 1.$$

This is what we refer to as a two-tailed test. On the other hand, a one-tailed test would take place when we possess prior information about the direction of the effect. For example:

$$H_0: \beta = 1;$$

$$H_1: \beta > 1 \text{ or } H_1: \beta < 1.$$

This is what we refer to as a one-tailed test.

Using the Test of Significance Approach, we will calculate the test statistic using the formula:

$$\text{Test statistic} = \frac{\hat{\beta}-\beta^*}{SE(\hat{\beta})}$$

Where β^* is the value of β under the null hypothesis. This approach is modelled after the student's t-distribution with T – 2 degrees of freedom. Furthermore, levels of significance must be determined to define the region where we will reject or will not reject our null hypotheses.

Using the t-tables, we will be able to determine the critical values which we will use to compare to the test statistic. If our test statistic is determined to lie inside of the rejection region, we will reject the null hypothesis H_0. Otherwise, if the test statistic is determined to lie outside of the rejection region, we will not reject the null hypothesis H_0.

3.5.2 P-Value

The probability value, also referred to as the exact significance level, represents a marginal significance level at which we would be indifferent to rejecting or not rejecting the null hypothesis H_0.

We typically define three levels of significance at which we aim to reject the null hypothesis H_0. The larger the test statistic is in absolute value, the smaller the p-value will be. Therefore, if the p-value is lower than the predefined levels of significance, we can reject the null hypothesis H_0 (Brooks, 2019).

Levels of significance can be for example 1%, 5%, and 10%. If our p-value is therefore less than 1%, for example, p-value = 0.001, we can say that we reject the null hypothesis H_0 at all levels of significance.

3.5.3 Spurious Regression

Spurious relationships occur when we employ regressors which we assume have a statistical relationship with the dependent variable without having the necessary theoretical justification. Especially when employing many regressors, there is always the probability that a few of them would randomly exhibit a relationship even though they would be completely independent of each other (Brooks, 2019).

Statistical relationships should therefore be backed up by theoretical justification and by using non-trending stationary variables. We should therefore possess apriori information about how we expect a regressor to impact the dependent variable.

Out-of-sample testing or de-trending can be employed to eliminate the risk of spurious relationships going undetected in the in-sample tests. Furthermore, correlation analysis can be employed to test for insignificant variables within the categories of indicators. The variables employed must therefore be proven to be correlated with the inflation indices to ensure their significance as an inflation indicator.

3.5.4 Linearity

Time series models typically assume that the underlying relationships between the regressors and the dependent variables do not change over time. Therefore, linearity must be present within the data employed. This is assessed by examining residual analysis and diagnostic plots. Therefore, the employment of pairwise scatter plots allows for the examination of linear relationships between the regressors and the inflation indices.

3.5.5 Stationarity

Covariance stationarity in a data series can be defined as a process where a data series has a constant mean, a constant variance, and its autocovariance is constant across lags but not in time (Brooks, 2019).

A stationary series satisfies the three equations:

1. $E(y_t) = \mu$, t = 1, 2, …, ∞

2. $E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty$

3. $E(y_{t1} - \mu)(y_{t2} - \mu) = y_{t2-t1}\ \forall t_1, t_2$

If a data series must be differenced *d* times to become stationary, then we can say that that series is integrated of order *d*. We define an I(0) series as a stationary series. Furthermore, we define series I(1) as containing one unit root and series I(2) as containing two unit roots, therefore requiring to be differenced twice to achieve stationarity. The presence of any unit roots in our series means we do not satisfy stationarity criteria.

We test for unit root processes using Dickey-Fuller tests:

H_0: unit root is present;

H_1: unit root is not present (stationary).

$$\Delta y_t = a_0 + \delta y_{t-1} + u_t$$

We compare the results from this test with critical values derived from the Dickey-Fuller stat table.

The DF tests are only valid when u_t is white noise. Otherwise, we employ the Kwiatkowski-Phillip-Schmidt-Shin (KPSS) test and compare the result against the Augmented Dickey-Fuller test.

We test for stationarity using KPSS tests:

H_0: series is stationary;

H_1: series is not stationary.

$$KPSS_N = \frac{1}{N^2\hat{\sigma}_N^2} \sum_{n=-1}^{N} S_n^2 = \frac{R_N}{\hat{\sigma}_N^2},$$

Where $\hat{\sigma}_N^2$ represents the estimator of the long-run variance of the residuals σ^2.

The KPSS tests provide a test statistic and a set of critical values. When the test statistic is smaller than the critical values, we do not reject the null hypothesis H_0 and consider the series stationary.

We compare the previous results against the ADF test:

H_0: series is not stationary;

H_1: series is stationary.

$$\Delta y_t = \gamma y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta\, y_{t-j} + \varepsilon_t$$

The ADF test differs from the KPSS test since it uses opposite null hypotheses. The ADF tests produce a p-value which determines the significance of the result. When the p-value is less than 0.01, we reject the null hypothesis H_0 and consider the series stationary.

When the results of the KPSS and ADF tests conflict, we typically stick with the results of the KPSS test. The main difference between both is that while the ADF test proves stationarity by rejecting the null hypothesis H_0, in the KPSS test stationarity is proved by not rejecting the null hypothesis H_0. This makes the KPSS a more robust test for stationarity characteristics.

3.5.6 Multicollinearity

Multicollinearity occurs when high correlation is present between two or more regressors. This results in inflated standard errors in the regressors employed. An early indication of multicollinearity is given when examining the correlation between the regressors using a correlation matrix. As a rule of thumb, no two regressors should have correlation coefficients higher than 0.8 between themselves. However, this is not sufficient.

The Variance Inflation Factor (VIF) is used to quantify the impact of multicollinearity on a regressor.

$$VIF_j = \frac{1}{1 - R_j^2}$$

The higher the VIF of a given regressor, the more redundant it is given that it can be accurately predicted by the complementing independent variables. Typically, a VIF value greater than five indicates that a multicollinearity problem may be present and would therefore warrant further investigation into the given regressor. However, a VIF value greater than ten would indicate a serious multicollinearity problem that would require remedy.

The remedy for multicollinearity includes transforming the highly correlated variables into ratios, using higher frequencies of data, and if the problem persists, excluding one of the highly correlated variables from the model.

3.5.7 Normality of Residuals

The regression assumes that the error term follows a normal distribution. A mean of zero and a constant variance is assumed. The Jarque-Bera test is employed to test whether a distribution would exhibit normality characteristics by depicting zero skewness and no excess kurtosis, like the normal distribution.

H_0: normal (skewness = 0, kurtosis = 3);

H_1: not normal.

$$JB = \frac{n}{6} \times \left[s^2 + \frac{(k-3)^2}{4} \right]$$

Where:

S = the skewness of the series,

K = the kurtosis of the series, and

n = number of observations.

As a remedy for non-normality, we can eliminate outliers where appropriate from a date series by using dummy variables. However, any use of dummy variables must be backed up by theoretical reasoning.

3.5.8 Homoscedasticity of Errors

Heteroscedasticity occurs when the residuals experience non-constant volatility. In other words, the variance of the errors is time-varying. On the other hand, homoscedasticity occurs when the variance of the errors is constant (Brooks, 2019).

In contrast to other tests for heteroscedasticity, such as the Breusch Pagan test which tests specifically for linear forms of heteroscedasticity, the White test tests for more generic forms of heteroscedasticity.

Therefore, we test for heteroscedasticity using the White test:

H_0: Homoscedasticity is present;

H_1: Variance is not constant across time.

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

We reject the null hypothesis H_0 that the error term is constant across time if the test statistic is greater than the critical values derived from the chi-square distribution table.

3.5.9 Residual Autocorrelation

We assume that and would have to validate that there are no patterns of serial correlation in our residuals. Having positive autocorrelation within a model will result in a pattern showing cyclicality in the residuals over time. This means that returns would tend to trend by demonstrating repeating positive returns, and after repeating negative returns. On the other hand, negative autocorrelation would result in an opposite pattern of the return series crossing the time axis more frequently than what would have happened had the series been distributed randomly (Brooks, 2019).

The Durbin-Watson test will not be employed since it tests primarily for first-order serial correlation, including both positive and negative autocorrelation. Instead, we will validate this assumption using the Breusch-Godfrey test, which is a more general test for autocorrelation.

H_0: $\rho_1 = 0$ and $\rho_2 = 0$ … and $\rho_r = 0$;

H_1: $\rho_1 \neq 0$ or $\rho_2 \neq 0$ … or $\rho_r \neq 0$.

The model for the errors when using this test is given by;

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \cdots + \rho_r u_{t-r} + v_t,$$

$$v_t \sim N(0, \sigma_v^2)$$

The test statistic is given by;

$$(T - r)R^2 \sim x_r^2$$

Where:

T = Total number of observations, and

u_t = residual error.

The test statistic produced is thus compared with the critical values obtained from the Chi-squared statistical table, which if exceeded, will mean the rejection of the null hypothesis H_0 that no autocorrelation is present.

3.6 Estimation of the Model

Estimation of the model will be done using the native MIDAS estimation functions from the MIDASR package. Selection of the weighting method will take place following sensitivity analyses, assessing the robustness of the model under different parameterizations. The weighting functions assessed will be the MIDAS Almon Polynomial Distributed Lag structure, and the non-linear MIDAS Exponential Almon and MIDAS Beta weighting functions.

3.6.1 Estimation of Parameters

The estimation technique employed for estimating the coefficients of the parameters of the model will be that of Ordinary Least Squares (OLS). This is because this method selects the coefficients which minimise the sum of the squared residuals between the dependent variable and the linear function of the independent variables of our models. This provides the minimum variance among all linear unbiased estimators (BLUE – Best Linear Unbiased Estimator) under the Gauss-Markov theorem. Furthermore, OLS always outperforms any non-linear estimator when the assumption of normality in residuals can be satisfied (Stock & Watson, 2020).

3.6.2 Training and Validation Horizon

The training and validation horizon will employ sixty (60) low-frequency observations, starting from the month of April 2019 (2019M4) and spanning to the month of April 2024 (2024M4). This five-year training horizon was chosen since the past few years were described by macroeconomic volatility and rapidly changing economic conditions (Armesto, Engemann, & Owyang, 2010; Bańbura, Giannone, Modugno, & Reichlin, 2013). Furthermore, due to the complexity of the models, longer time horizons may result in a lack of interpretability and transparency (Modugno, 2011).

3.6.3 Lag Selection

The selection of the optimal number of lags used and the shape of the weighting distribution for the MIDAS models will be determined when estimating the parameters using R and is solely determined by the data employed. Since the selection of lags does not impact the number of parameters in the MIDAS model, the information criterion, such as the Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (SBIC) will not be used (Ghysels & Marcellino, 2018).

Selecting the maximal lag K is another story. Selecting K too small is problematic, resulting in a loss of statistical significance in the model due to the omission of important dynamics in the relationship between the regressor and the dependent variable. On the other hand, the only consequence of picking K too large is that more high-frequency data is required at the beginning of the sample as the weights of the model would quickly vanish to zero.

3.6.4 Constant Term and Beta Coefficients

In our regression, the term β_0 will be the coefficient of the constant term (the term which we would have referred to as alpha, α, in a simple univariate regression).

Upon estimation of the model, a set of partial slope coefficients, or beta coefficients will be produced describing the relationship between the regressor and the dependent variable. A positive beta coefficient means that the regressor has a positive impact on the dependent variable. Therein, a positive increase in the regressor would result in a positive impact on the dependent variable. Oppositely, a negative beta coefficient means that the regressor would negatively impact the dependent variable. The size of the beta coefficient describes the magnitude of this impact.

3.6.5 Random Term

A disturbance term is included in the regression to capture what may be neglected determinants of the dependent variable. This random term also captures errors in the measurement of the dependent variable and random outside influences which both cannot be modelled.

When nowcasting macroeconomic variables, some form of measurement error is unavoidable since the official estimate itself which we will compare our nowcast to has an inherent error, being the subject of regular revisions and re-estimations.

3.6.6 Regressor Significance

Upon estimation of the model, an output will be produced displaying the coefficient and standard error of each regressor within the model. The significance of each regressor can be measured by dividing the coefficient of a regressor by its standard error. The T-statistic obtained may then be compared with the critical value obtained from the T-table, from which the probability values will be produced.

As a rule of thumb, regressors with probability values higher than 0.1 may be omitted from the model as they would be considered statistically insignificant. On the other hand, regressors with a low p-value, typically under 0.05, are considered statistically significant and should remain included within the model. Therefore, each model should be re-estimated to exclude insignificant regressors.

3.7 Empirical Application

Empirical application will assess the model's ability to adapt to unseen data. Forecasts for the inflation growth rates will be produced at a monthly interval at the end of each reference period. This will allow for comparison with the official estimates which release with a publication delay after the end of the reference period. Furthermore, a Nowcasting exercise will take place to provide Intra-month estimates of the inflation indices in advance of the official release. Intra-month analysis will determine our ability to produce preliminary estimates of the inflation indices during the month.

3.7.1 Forecasting Error Measurement

We will compare the different parameterizations of our model by measuring the stability of errors over time. Forecasting error can be defined as the difference between our nowcast and the actual values. This error will be measured using the Root-Mean-Square Forecast Error (RMSFE) and Mean Absolute Percentage Forecasting Error (MAPFE) methods. Our objective is to minimize both criteria.

After comparing our results with the official estimates, error results will be produced to evaluate the accuracy of the models. The model with the lowest RMSFE and MAPFE after testing will be selected as the weighting function for our models.

The Root-Mean-Squared method calculates the forecasting error by summing the squared differences between the forecasts and the actual value. The difference between the forecast and the actual value is referred to as the residual (or the error). The average error is then calculated, and the square root is taken to produce the RMSFE result.

Root-Mean-Squared Forecasting Error:

$$RMSFE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Similarly, the Mean Absolute Percentage forecasting error sums the results of the individual residuals divided by the actual value. An average is then taken to produce an average of the percentage errors. The benefit of using a percentage unit is that it is easier to interpret since it is what is referred to as scale-independent.

Mean Absolute Percentage Forecasting Error:

$$MAPFE = \frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i}$$

3.7.2 Intra-month Analysis

Intra-month analysis will be conducted for the month of May 2024. This is done by revising the original prediction, $\mathbb{P}(y_t|\Omega_{v-1})$, produced for the month of April 2024 with new high-frequency data. During the month of May, new data arising from the unsynchronised weekly releases of the high-frequency regressors will be employed to revise the original prediction each week providing for weekly nowcasts of the inflation indices. This will be done up to the end of the month where the low-frequency regressors will also be published. The utilization of recent low-frequency regressors will allow for the most robust estimate of the inflation indices at the end of the month.

The benchmark used to compare our MIDAS model will be publicly available results from the Federal Reserve Bank of Cleveland's inflation nowcasting model.

Chapter 4 – Analysis and Results

4.1 Introduction

The analysis and results section will start with the presentation of the variables selected for our United States time series models. This presentation will include the treatment of the variables, including the necessary seasonal adjustments and transformations. There are six categories of data which will be employed in the models:

1. Prices,
2. National Income & Products Accounts,
3. Money, Banking & Finance,
4. Population, Employment, & Labor Markets,
5. Production & Business Activity, and
6. Surveys.

The next section of this chapter will discuss some descriptive statistics followed by the results achieved from the statistical tests employed to ensure the optimization and validation of the assumptions of our model.

Once the dataset has been constructed, several estimations of the model will take place. Different lag selections will be employed to simulate different shapes for the weighting functions of the models. Therefore, three estimations will be conducted for the MIDAS model per inflation index. Namely, two estimations will be done using the Almon Polynomial Distributed Lag structure, two will be done using the Exponential Almon weighting function, and another two will be done using the Beta specification.

Moreover, the output from the estimation of the equation will allow for the identification of the regressors that were the most significant, and which if any were insignificant in describing the inflation indices. This will allow for the exclusion of the insignificant variables and re-estimation of the models to ensure optimization. Furthermore, an out-of-sample test will be conducted to assess the robustness of the models. This will be done by comparing forecasting error results, enabling the analysis of the statistical adequacy of each regression.

Finally, the intra-month nowcasting application will allow for an interpretation to be conducted and used towards analysis and extrapolation.

4.2 Data Set

Table 1: United States Model Data Set

No	Name	Frequency	Category	Transformation	Source
1	Consumer Price Index for All Urban Consumers: All items in U.S. City Average	Monthly	Prices	Seasonal Adj, YoY Change, Difference (1)	U.S. Bureau of Labor Statistics
2	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average	Monthly	Prices	Seasonal Adj, YOY Change, Difference (1)	U.S. Bureau of Labor Statistics
3	Export Price Index (End Use): All Commodities	Monthly	Prices	Difference (1)	U.S. Bureau of Labor Statistics
4	US Diesel Sales Price	Weekly	Prices	Seasonal Adj Difference (1)	U.S. Energy Information Administration
5	Real Disposable Personal Income	Monthly	National Income & Product Accounts	Seasonal Adj Difference (1)	U.S. Census Bureau
6	Trade Balance: Goods and Services, Balance of Payments Basis	Monthly	National Income & Product Accounts	Seasonal Adj Difference (1)	U.S. Census Bureau
7	Federal Surplus or Deficit	Monthly	National Income & Product Accounts	Seasonal Adj Difference (1)	U.S. Census Bureau
8	Commercial Paper Outstanding	Weekly	Money, Banking, & Finance	Seasonal Adj Difference (1)	Board of Governors of the Federal Reserve System (US)
9	Unemployment Rate	Monthly	Population, Employment, & Labour Markets	Seasonal Adj Difference (1)	U.S. Bureau of Labor Statistics
10	Average Hourly Earnings of All Employees, Total Private	Monthly	Population, Employment, & Labour Markets	Seasonal Adj Difference (2)	U.S. Bureau of Labor Statistics

11	Average Weekly Hours of All Employees, Manufacturing	Monthly	Population, Employment, & Labour Markets	Seasonal Adj Difference (1)	U.S. Bureau of Labor Statistics
12	Job Openings: Total Nonfarm	Monthly	Population, Employment, & Labour Markets	Seasonal Adj Difference (1)	U.S. Bureau of Labor Statistics
13	Total Nonfarm Private Payroll Employment	Weekly	Population, Employment, & Labour Markets	Seasonal Adj Difference (1)	Automatic Data Processing, Inc
14	Total Vehicle Sales	Monthly	Production & Business Activity	Seasonal Adj Difference (1)	U.S. Bureau of Economic Analysis
15	New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units	Monthly	Production & Business Activity	Seasonal Adj Difference (1)	U.S. Census Bureau
16	Total Construction Spending: Total Construction in the United States	Monthly	Production & Business Activity	Seasonal Adj Difference (2)	U.S. Census Bureau
17	Manufacturers New Orders: Durable Goods	Monthly	Production & Business Activity	Seasonal Adj Difference (1)	U.S. Census Bureau
18	Current General Activity, Diffusion Index for Federal Reserve District 3: Philadelphia	Monthly	Surveys	Seasonal Adj Difference (1)	Federal Reserve Bank of Philadelphia
19	University of Michigan: Consumer Sentiment	Monthly	Surveys	Seasonal Adj Difference (1)	University of Michigan

4.3 Descriptive Statistics

Table 2: United States Model Descriptive Statistics

No	Name	Ticker	Mean	Std. dev.	Skewness	Kurtosis	JB Test
1	Consumer Price Index for All Urban Consumers: All items in U.S. City Average	CPIAUCSL	0.00062	0.46172	-0.52145	8.53086	386.74
2	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average	CPILFESL	0.00399	0.18080	1.816028	14.1415	1676.5
3	Export Price Index (End Use): All Commodities	IQ	0.16667	1.21825	-0.27589	9.30647	490.93
4	US Diesel Sales Price	GASDESW	0.00193	0.05677	2.655311	32.9974	49612
5	Real Disposable Personal Income	DSPIC96	24.6548	364.969	2.489042	66.8072	50178
6	Trade Balance: Goods and Services, Balance of Payments Basis	BOPGSTB	-167.686	3995.19	0.278118	6.07741	119.4
7	Federal Surplus or Deficit	MTSDS133FMS	-543.089	104875.08	0.517042	51.53834	28776
8	Commercial Paper Outstanding	COMPOUT	-0.28374	19.25305	-0.42500	8.09166	1365.7
9	Unemployment Rate	UNRATE	0.00102	0.663042	12.892402	207.2722	521068
10	Average Hourly Earnings of All Employees, Total Private	CES0500000003	-9.13242	0.141991	-4.006611	74.69599	47491
11	Average Weekly Hours of All Employees, Manufacturing	AWHAEMAN	0.00045	0.227453	-4.448732	49.45632	20509
12	Job Openings: Total Nonfarm	JTSJOL	10.9787	296.5811	-0.520479	6.098904	125.57
13	Total Nonfarm Private Payroll Employment	ADPWNUSNERSA	37305.04	137969.63	-8.334528	98.93444	297870
14	Total Vehicle Sales	TOTALSA	-0.00685	0.974102	-0.524900	13.78192	1442.5
15	New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units	PERMIT	-0.77891	68.47258	-0.293549	4.880447	47.539
16	Total Construction Spending: Total Construction in the United States	TTLCONS	-6.41980	13974.08	-0.115054	4.124655	16.088
17	Manufacturers New Orders: Durable Goods	DGORDER	230.5952	10384.71	-0.512160	12.28937	1069.9
18	Current General Activity, Diffusion Index for Federal Reserve District 3: Philadelphia	GACDFSA066MSFRBPHI	-0.00102	10.93629	0.139014	11.41431	871.21
19	University of Michigan: Consumer Sentiment	UMCSENT	-0.14949	4.30125	-0.434821	4.017661	21.876

4.4 Statistical Testing

Observations from 2000M01 to the most recent observations as at time of testing were included so to not discard any underlying trends which would not have been captured had a short testing horizon been used.

Seasonality adjustment was conducted where applicable using X13-ARIMA-SEATS by the U.S. Census Bureau. This ensured that our data was free from any effects brought about by seasonality in our various data series. Furthermore, the necessary transformations were conducted, and tests, namely the ADF and KPSS tests, were done to ensure that our data series exhibit weakly covariance stationarity characteristics.

Heatmaps were employed to demonstrate the pairwise correlation between the categories of economic indicators and the inflation indices. From the heatmap one can start identifying the significance of the relationships we expect to uncover between the economic indicators and the inflation indices. The regressors exhibiting low correlation were not discarded since they may still be significant for the model when used in conjunction with the other independent variables. This is since the correlation heatmap only uncovers linear forms of correlation and may fail to reveal any underlying non-linear association between the regressors. On the other hand, correlation heatmaps provide for the perfect tool to detect early signs of multicollinearity within the regressors of our model. Early signs of multicollinearity may be detected by monitoring for pairs of regressors which exhibit more than 0.8 in pairwise correlation. The most common remedy for multicollinearity would then be to exclude one of the highly correlated regressors from the model (Brooks, 2019).

Similarly, the pairwise scatterplot matrix further examines the relationship shared between the categories of economic indicators and the inflation indices. Through pairwise scatterplot examination, we can demonstrate that linear associations are present between the regressors and the inflation indices and can start identifying the polarity of this relationship (Stock & Watson, 2020). Some distinctive shapes may arise when complex relationships are present between the regressors and the inflation indices. Namely, the nature of leading or lagging indicators may mean that a time-lagged relationship may be visible on the scatterplot resulting in patterns that are shifted or spread out compared to a simultaneous relationship. Furthermore, the scatterplot may show a delayed relationship where the points might form a pattern that reflects past movements rather than current correlations. Finally, even though linearity may not be immediately present in the pairwise scatterplot, a statistically significant linear association for a regressor may be estimated when used in conjunction with the other regressors in our model.

4.4.1 Correlation Analysis

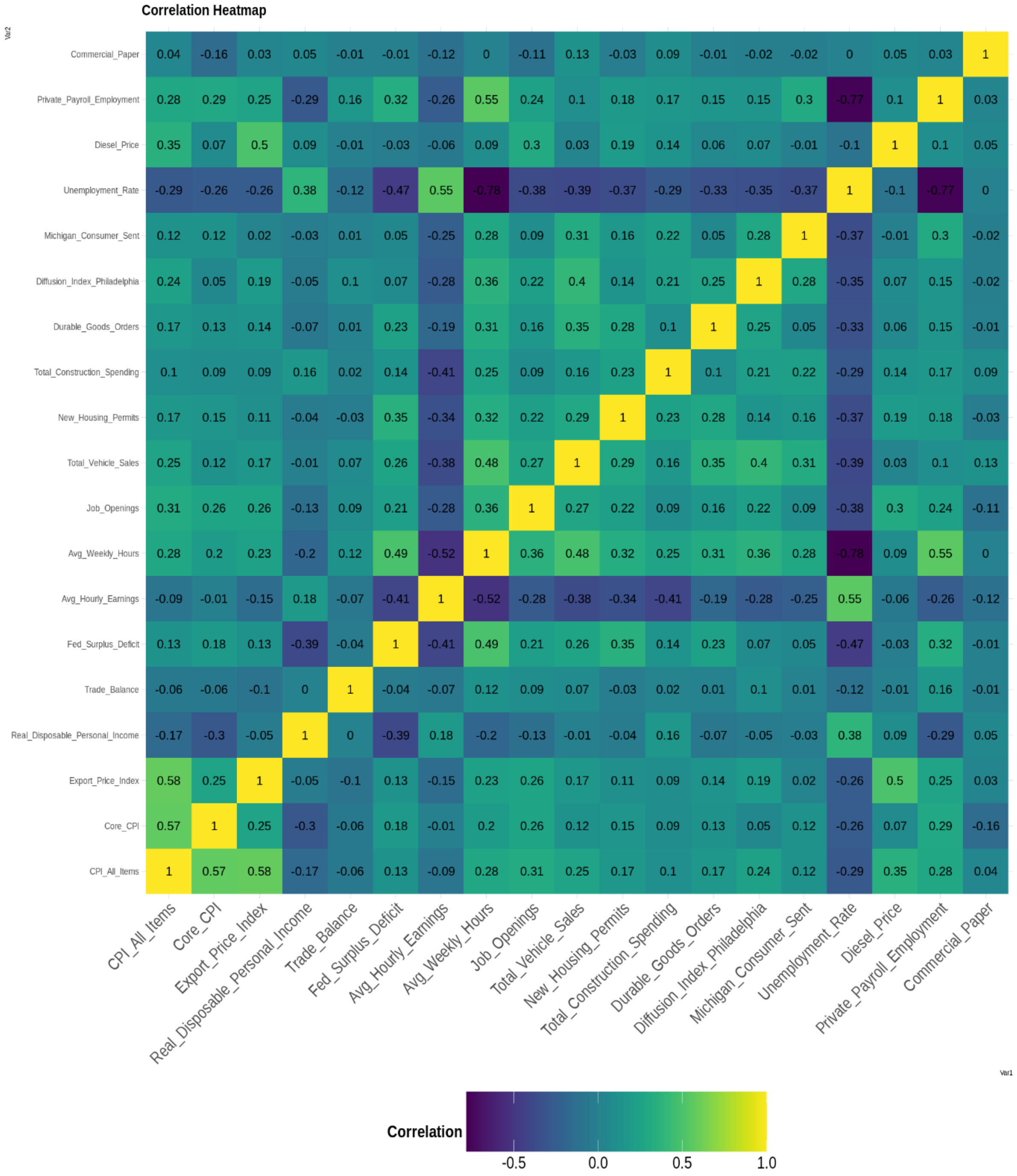


Figure 4: United States – Correlation Heatmap

4.4.2 Linearity

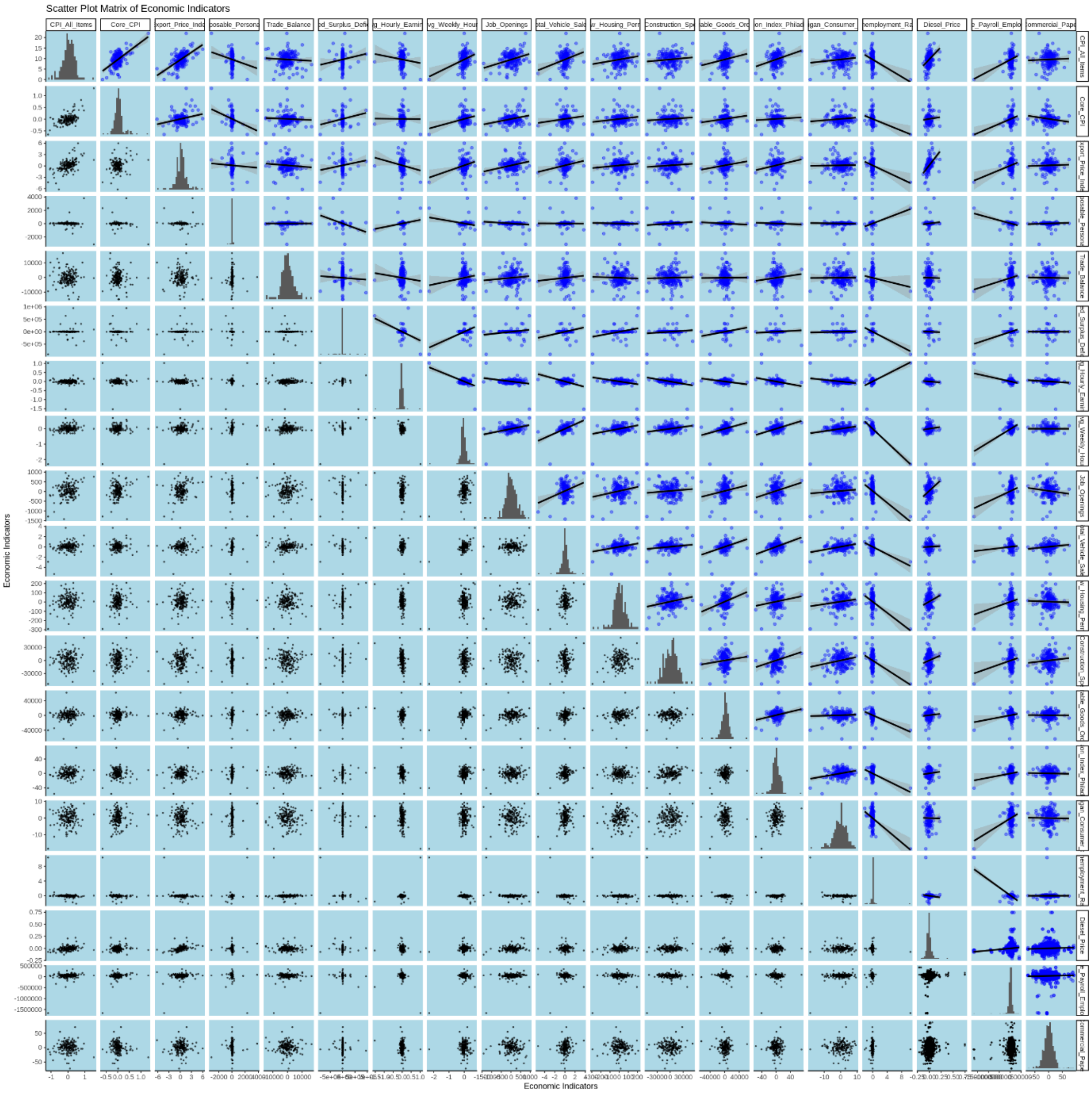


Figure 5: United States – Pairwise Scatterplot Matrix

4.4.3 Stationarity

Stationarity testing is conducted to confirm that the regressors employed exhibit weakly covariance stationarity characteristics as assumed by the econometric models employed in this study. This is done by conducting Augmented Dickey-Fuller (ADF) Tests and confirmed using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Tests.

The difference between the ADF and KPSS tests is that they have opposite null hypotheses H_0, since whereas the KPSS test has a null hypothesis that the data series exhibits stationarity characteristics, the ADF has an opposite null hypothesis that the data series does not exhibit stationarity characteristics.

Table 3: Augmented Dickey-Fuller Tests for the United States Model

No	Name	ADF Statistic	ADF P-Value
1	Consumer Price Index for All Urban Consumers: All items in U.S. City Average	-5.6207	< 0.01
2	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average	-4.6127	< 0.01
3	Export Price Index (End Use): All Commodities	-5.1146	< 0.01
4	US Diesel Sales Price	-9.6881	< 0.01
5	Real Disposable Personal Income	-9.2611	< 0.01
6	Trade Balance: Goods and Services, Balance of Payments Basis	-6.3149	< 0.01
7	Federal Surplus or Deficit	-9.116	< 0.01
8	Commercial Paper Outstanding	-11.019	< 0.01
9	Unemployment Rate	-7.6328	< 0.01
10	Average Hourly Earnings of All Employees, Total Private	-11.602	< 0.01
11	Average Weekly Hours of All Employees, Manufacturing	-6.5129	< 0.01
12	Job Openings: Total Nonfarm	-4.4593	< 0.01
13	Total Nonfarm Private Payroll Employment	-6.9452	< 0.01
14	Total Vehicle Sales	-8.3233	< 0.01
15	New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units	-5.5136	< 0.01
16	Total Construction Spending: Total Construction in the United States	-8.9115	< 0.01
17	Manufacturers New Orders: Durable Goods	-5.8766	< 0.01
18	Current General Activity, Diffusion Index for Federal Reserve District 3: Philadelphia	-7.7726	< 0.01
19	University of Michigan: Consumer Sentiment	-7.5199	< 0.01

ADF Probability Values of less than 0.01 rejects the Null Hypotheses H_0 that the data series tested does not exhibit a stationary process.

Table 4: Kwiatkowski-Phillips-Schmidt-Shin Tests for the United States Model

No	Name	KPSS Statistic	KPSS P-Value
1	Consumer Price Index for All Urban Consumers: All items in U.S. City Average	0.030708	> 0.1
2	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average	0.065561	> 0.1
3	Export Price Index (End Use): All Commodities	0.065601	> 0.1
4	US Diesel Sales Price	0.060147	> 0.1
5	Real Disposable Personal Income	0.024555	> 0.1
6	Trade Balance: Goods and Services, Balance of Payments Basis	0.073664	> 0.1
7	Federal Surplus or Deficit	0.012247	> 0.1
8	Commercial Paper Outstanding	0.11021	> 0.1
9	Unemployment Rate	0.066291	> 0.1
10	Average Hourly Earnings of All Employees, Total Private	0.014166	> 0.1
11	Average Weekly Hours of All Employees, Manufacturing	0.052644	> 0.1
12	Job Openings: Total Nonfarm	0.17556	> 0.1
13	Total Nonfarm Private Payroll Employment	0.12248	> 0.1
14	Total Vehicle Sales	0.047059	> 0.1
15	New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units	0.20176	> 0.1
16	Total Construction Spending: Total Construction in the United States	0.020317	> 0.1
17	Manufacturers New Orders: Durable Goods	0.048578	> 0.1
18	Current General Activity, Diffusion Index for Federal Reserve District 3: Philadelphia	0.019902	> 0.1
19	University of Michigan: Consumer Sentiment	0.083057	> 0.1

KPSS test statistic values lower than the 10% critical value means that we do not reject the Null Hypotheses H_0 that the data series tested exhibits a stationary process.

4.5 Estimation of the Model

Estimation of the coefficients employed sixty (60) low-frequency (monthly) observations of the monthly regressors and three hundred (300) high-frequency (weekly) observations of the high-frequency regressors. This represents a time horizon of 5 years. The number of low-frequency observations is an important specification since it provides for time alignment with the high-frequency observations and restricts the number of parameters we can specify in our model.

Nowcasting models work by using high-frequency data that is available up to the present moment. This real-time data stream is represented as being the lag value of 0. Furthermore, the economic indicators increase in significance the lower its publication delay compared to the inflation indices if its release has an immediate effect on the estimate of the inflation indices.

The Constant or Intercept term shows the inflation growth rate expected if all other regressors remain unchanged. As often happens, this term tends to be statistically insignificant. This is due to the elimination of trend effects when the differencing of the inflation series occurs, leaving the intercept term to capture only minor residual effects.

The estimated coefficients describe the expected relationship between the current and lagged values of the regressors and the inflation indices. The employment of non-linear weighting functions allowed us to create parametric shapes which exhibit rapidly declining or hump-shaped weights for our high-frequency regressors. This is especially evident for the weekly frequency regressors as the weighting function of a lagged regressor is constructed to increase before declining towards zero.

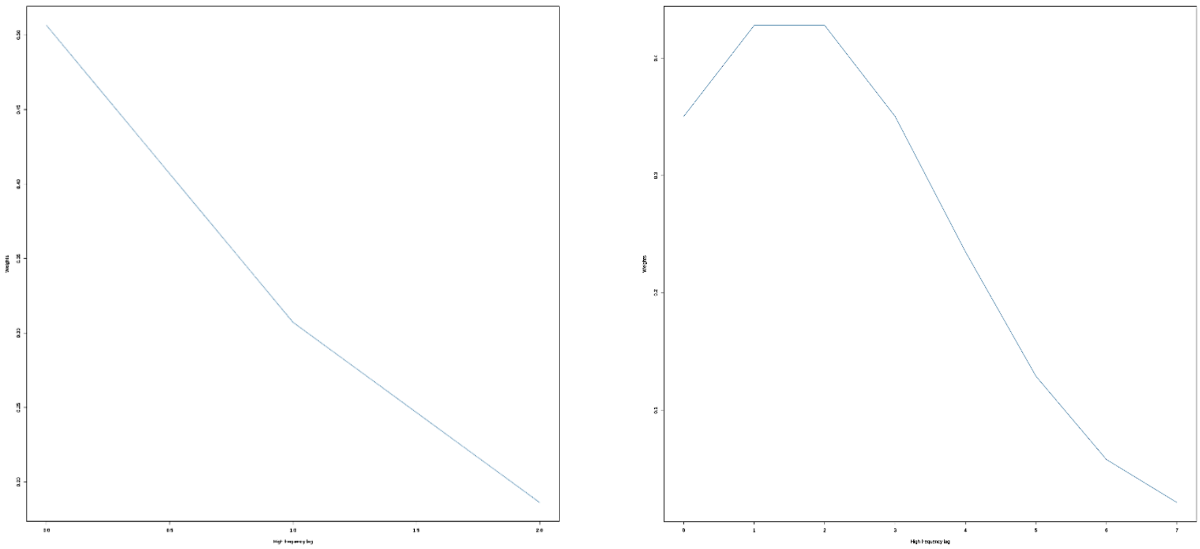


Figure 6: High-Frequency polynomial coefficient of Monthly and Weekly regressors (Produced for this study)

Finally, the random term captures the deviation of our forecasts from the official results. Error in our model exists due to several reasons, namely due to the factors of randomness, unexplained influences, ignorance, model misspecification, and errors inherent from official results such as revisions and re-estimations. While a margin of error is always expected, statistical tests are done on the residual term to confirm its validity. These tests include tests for normality, autocorrelation, and homoscedasticity. The tests conducted on our models show that our residuals were approximately normally distributed. Furthermore, they exhibited freedom from the negative effects of multicollinearity, residual serial correlation, and heteroscedasticity.

4.6 Forecasting Error Measurement

Forecasting error is employed to measure the in-sample predictive ability of the models. Three methods of parameterisation were evaluated being the Almon Polynomial Distributed Lag structure and the non-linear Exponential Almon and Beta weighting schemes. The model with the least error does not necessarily imply that it is the most efficient model, or that it will perform positively during an out-of-sample test. Another factor to be considered is therefore the quantity and significance of the parameters employed. This is since complex models may perform well in the training sample, however very poorly when exposed to new data. Furthermore, the effectiveness of the economic indicators is most prominent during forecasting applications since the presence of non-linear relationships is difficult to measure using traditional probability values. The most efficient model was found to be the model employing the Exponential Almon weighting function resulting in the lowest RMSFE and MAPFE measures for both United States inflation indices.

Figure 7: United States CPI Exponential Almon Model Forecasting Error (Produced for this study)

Table 5: Forecasting error measurement for the United States CPI models

US Consumer Price Index	**RMSFE**	**MAFE**	**MAPFE**
MIDAS Almon Polynomial Distributed Lag Model	0.14671	0.11907	118.04%
MIDAS Exponential Almon Model	**0.0943**	**0.07079**	**78.50%**
MIDAS Beta Lag Model	0.1065	0.08177	100.66%

Table 6: Forecasting error measurement for the United States Core CPI models

US Core Consumer Price Index	**RMSFE**	**MAFE**	**MAPFE**
MIDAS Almon Polynomial Distributed Lag Model	0.13527	0.10284	353.85%
MIDAS Exponential Almon Model	**0.12987**	**0.09877**	**296.68%**
MIDAS Beta Lag Model	0.15104	0.12378	389.39%

4.7 Intra-Month Analysis

The Intra-Month Analysis represents the nowcasting empirical application that this study intends to achieve. This is done by producing estimates of the inflation indices at specific points within the month corresponding to new weekly releases of data, providing for preliminary estimates of the inflation rate up to the official estimate we expect to be released at a publication delay later for that corresponding month.

To produce estimates, optimisation of the regression equation is conducted every month to ensure that the original prediction aligns with the inflation growth rate of the previous month. The major adjustments are then made to the lags of the weekly regressors since these would be the only regressors being updated during the duration of the month.

Upon the release of new data, any high-frequency regressor may be determined to have a non-significant or distortionary effect on our ability to produce estimates for the inflation indices. Visible signs for this error are weird shapes or highly volatile movements which may appear highly unrealistic. In this case, since we benefit from freedom of weights, we are easily able to add or remove regressors to always ensure that we are estimating the best optimised regression. In our model, this was the case with commercial paper, of which its removal allowed us to produce superior results compared to the model employed by the Federal Reserve Bank of Cleveland. This represents a major advantage over the Bridge equations since the removal or addition of a regressor from such model would be too complex considering that major changes would be required to the structure of the equation, due to involving fixed weights and specific economic relationships. This also means that using a single-line equation model, we were more correctly reflecting the sensitivity which the high-frequency economic indicators had to the inflation indices during the inflationary period analysed of May 2024.

Finally, a brief understanding of working with percentages is crucial in understanding the nowcasts produced by our model. Through subject of the formula, the final nowcasts were produced of Year-over-Year change format. This means that the percentage change estimated refers to the change in the inflation index from a year's month compared to where that index stood from that same month a year ago. The nature of inflation rates, since we are using percentage against level data, build upon previous years of (high) inflation. Therefore, having low inflation growth rates after extended periods of high growth rates is expected and does not directly imply a good or positive economic result. It is easy to demonstrate low inflation rates when building upon already highly inflated prices. Using level data or considering previous years of inflation rates is ideal when interpreting current results to avoid being misled by the illusion of percentages.

Table 7: Nowcasting results for the United States CPI

Date	Actual US CPI	MIDAS Nowcasts	Cleveland Fed
1st May 2024	3.36%	3.361%	3.430%
8th May 2024		3.339%	3.560%
15th May 2024		3.384%	3.388%
22nd May 2024		3.330%	3.371%
1st June 2024	3.25%	3.262%	3.361%

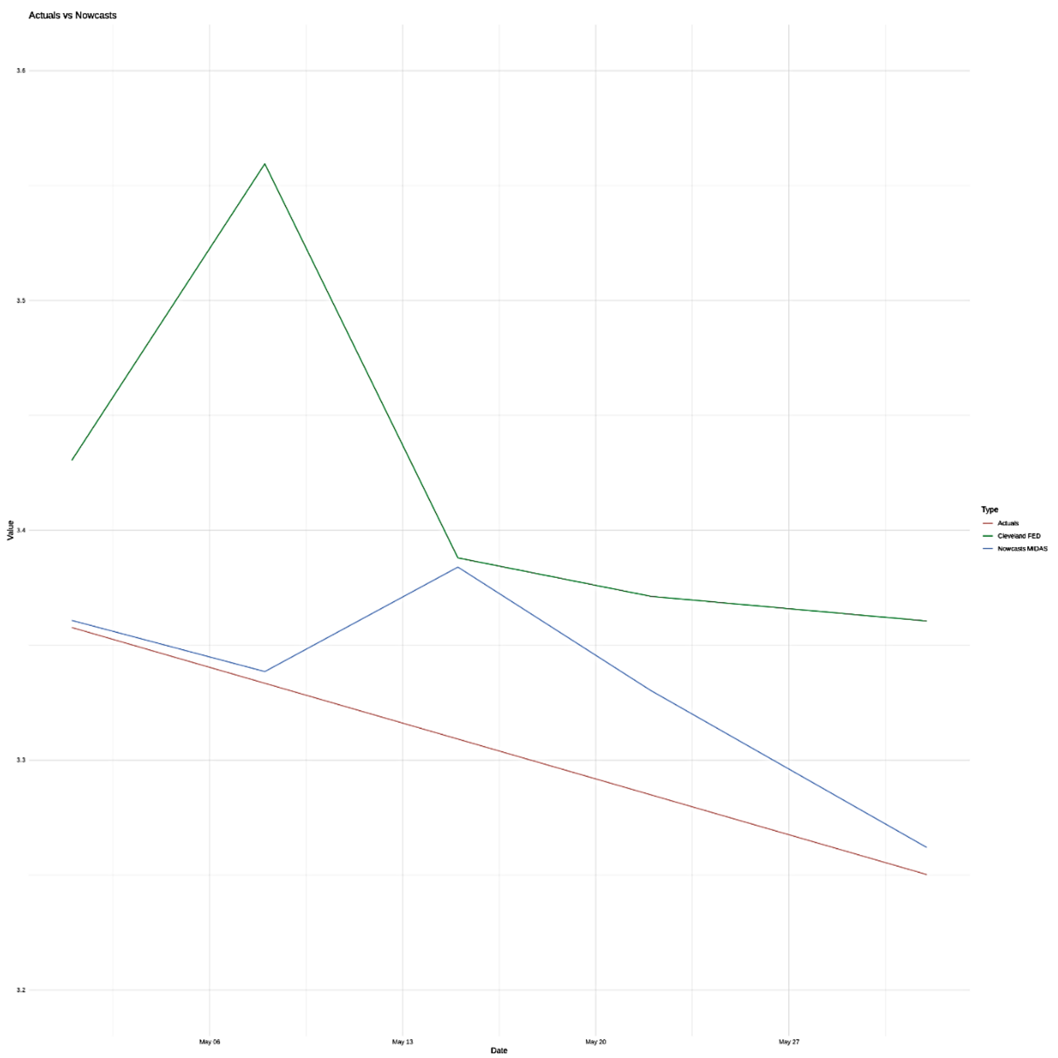


Figure 8: United States CPI Intra-Month Analysis (Produced for this study)

Figure 8 represents the results obtained from the United States CPI nowcasting application during the month of May 2024. The results obtained using the MIDAS equations are represented in blue whereas the results extracted from the Federal Reserve Bank of Cleveland are represented in green. The official estimates are represented in red.

Table 8: Nowcasting results for the United States Core CPI

Date	**Actual US Core CPI**	**MIDAS Nowcasts**	**Cleveland Fed**
1st May 2024	3.62%	3.594%	3.594%
8th May 2024		3.585%	3.594%
15th May 2024		3.518%	3.547%
22nd May 2024		3.554%	3.547%
1st June 2024	3.41%	3.391%	3.547%

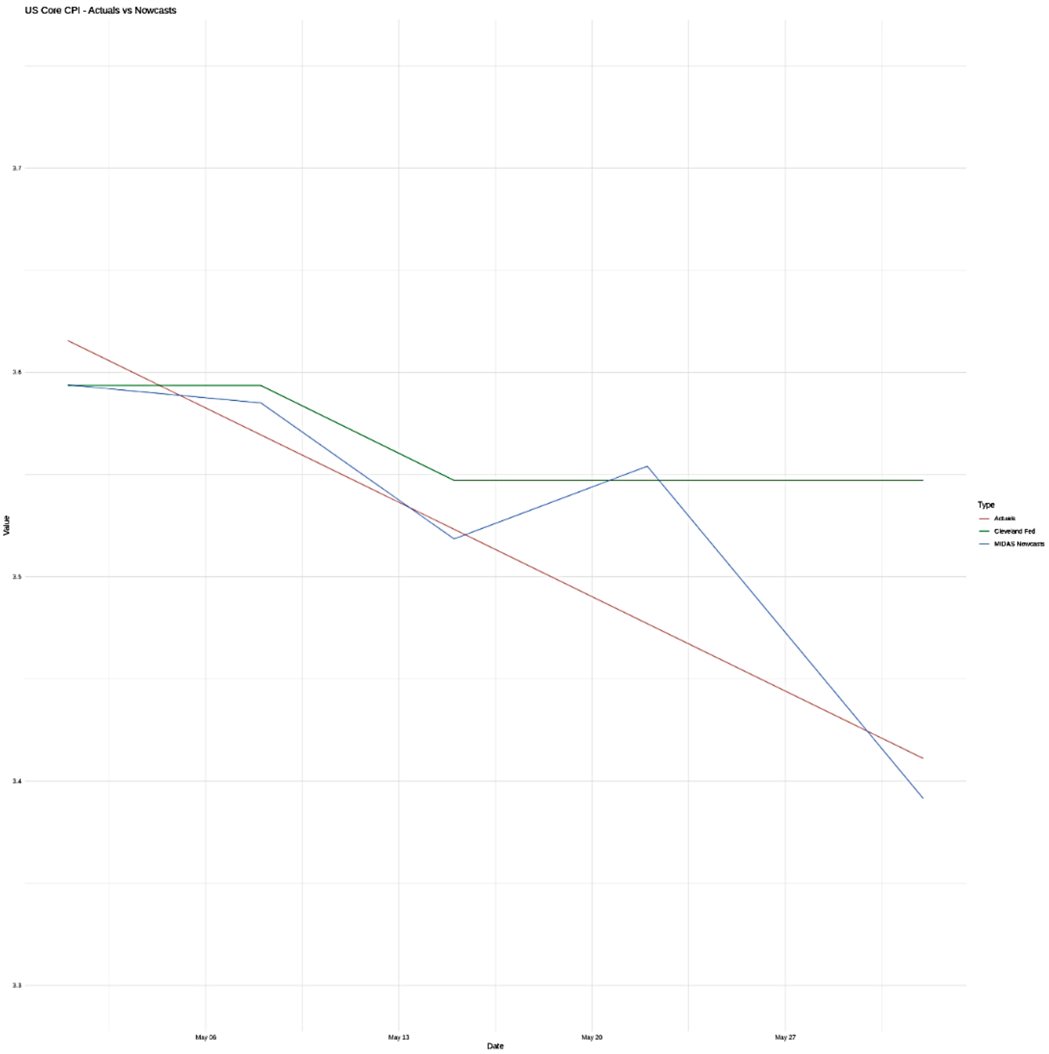


Figure 9: United States Core CPI Intra-Month Analysis (Produced for this study)

Figure 9 represents the results obtained from the United States Core CPI nowcasting application during the month of May 2024. The results obtained using the MIDAS equations are represented in blue whereas the results extracted from the Federal Reserve Bank of Cleveland are represented in green. The official estimates are represented in red.

4.8 Analysis

Nowcasting, a term coined in meteorology, is the practice of estimating a macroeconomic variables near state. Applied econometric time-series forecasting is a beneficial approach for nowcasting macroeconomic indices since it provides the necessary tools to understand which and how economic factors are driving our forecasts. This is an advantage over other techniques such as machine learning which do not always provide theoretical backing as to what is driving the forecast (Hopp, 2022). In economics, prediction should be backed up by explanation.

Nowcasting inflation is important since it is one of the most monitored macroeconomic variables by researchers and policymakers. This will also provide an indispensable tool to market practitioners who rely on real-time analysis for decision-making or conducting investment analyses. Moreover, this will aid researchers in interpreting forecast errors and alleviate macroeconomic surprises (Reichlin, 2019).

Six categories of economic data were selected due their expected impact on the inflation indices. These categories were 1) Prices, 2) National Income & Products Accounts, 3) Money, Banking & Finance, 4) Population, Employment, & Labor Markets, 5) Production & Business Activity, and 6) Surveys. Furthermore, these categories allowed us to take advantage of monthly and weekly frequencies of data as allowed by our models.

The MIDAS (Mixed Data Sampling) regression model addresses the challenge of different frequencies by using specialized weighting schemes (like Exponential Almon or Beta lag polynomials) to incorporate the high-frequency data into the low-frequency regression. Instead of aggregating, MIDAS effectively "distributes" the impact of the high-frequency data over the low-frequency periods, allowing it to capture the more granular dynamics. This proves the parsimonious benefit of employing a single equation model.

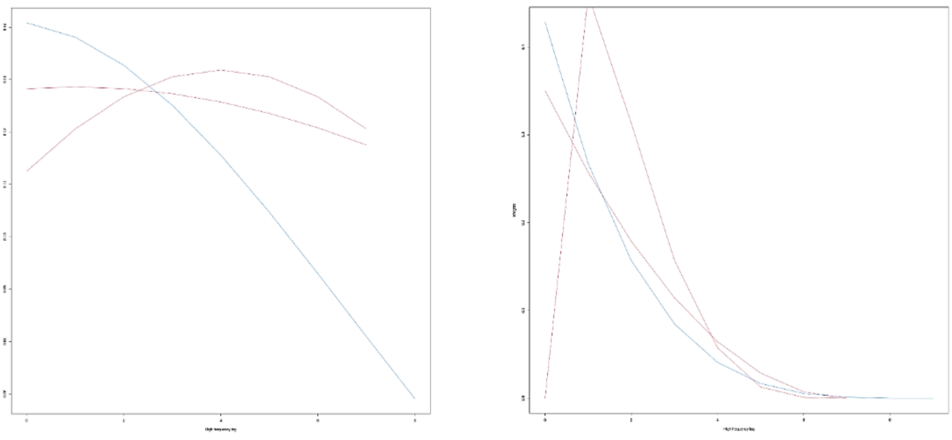


Figure 10: Exponential Almon and Beta polynomial weighting functions (Produced for this study)

Parameter optimisation was shown to be horizon specific. Therefore, minor calibration to the parameters may be required with every new release of data to preserve the accuracy of the models. Furthermore, since estimation is done for specific time horizons, the Mixed Data Sampling models are shown to not suffer from the curse of fixed weights. This is crucial since the sensitivity of the inflation indices to certain regressors may change over time and a single equation that would work indefinitely would not be suitable. Therefore, the estimation of the equations is always done using a near sample where the older observations are eventually phased out and replaced by more recent releases.

Optimal model selection was done by comparing the various forecasting error results of the different regressions. Since official estimates are only available at a monthly frequency, the error is measured by comparing our estimates with the actual estimates which are released at the end of the month following a publication delay. While the employment of high-frequency data is expected to provide robust estimates of the Intra-month change in the growth rate of the inflation indices, it is also expected that the nowcasts increase in accuracy as the more up-to-date information is released during the periods closer to the end of the month. This may be a limitation since regular calibration would be required to ensure that the models are optimized and producing robust nowcasting results. Furthermore, this may require the constant attention of econometric experts to adjust the selection of parameters within the models.

The MIDAS model and the employment of high-frequency data were noted to increase the robustness of the forecasts produced. This is due to the MIDAS model being better able to incorporate the granular dynamics of high-frequency data that lead to more accurate forecasts of the lower-frequency variable. The objective of nowcasting is to have a preliminary estimate of the direction and magnitude of the movement in the inflation indices. Comparison of our model with the CPI Nowcasting model of the Federal Reserve Bank of Cleveland shows that the MIDAS models produced competitive results in nowcasting applications. This is since while both models correctly predict the change in the direction of the inflation rate, the efficient use of parameters in the MIDAS model also allows it to predict the magnitude of this change more robustly.

It is the employment of high-frequency data which allows for intramonth analysis to be produced. While intramonth estimates cannot be measured for accuracy due to the lack of official data, we can still show robustness when the nowcasts manage to reflect the movement in the end-of-month official estimates. The timeliness of high-frequency data can provide early signals of changes in the economy, allowing for more responsive decision-making and mitigation of macroeconomic surprises. Furthermore, while core inflation indices do not account for fuel prices, energy was shown to remain a strong predictor due to the spillover effect it has on other prices in the economy, making it a significant indicator to predict changes also in core indices.

Both Nowcasting applications evaluated showed that, similarly to the Cleveland FED, we were accurately able to predict the direction and magnitude of the inflation indices before their official release for the month of May 2024. Furthermore, while a model may produce superior results to others, model averaging may reduce the bias to a single model, producing more consistent nowcasts.

Chapter 5 – Conclusions, Policy Implications & Avenues for Future Research

5.1 Summary of Mixed Data Sampling Results

The Mixed Data Sampling (MIDAS) model was used to nowcast the CPI and the Core CPI for the United States economy. This enabled us to approach the research question, being to measure the robustness of the model to nowcast inflation indices:

"How robust is the Mixed Data Sampling (MIDAS) time-series model to nowcast low-frequency macroeconomic variables such as inflation indices?"

Utilizing different categories of economic indicators enabled us to benefit from the non-linear relationships present in alternative economic data within the United States economy. Upon examining different methods of parameterisation for this model, the Exponential Almon weighting function was found to make the most efficient use of parameters, enjoying the lowest RMSFE and MAPFE across the training horizon employed (2019M4 – 2024M4). Furthermore, this parameterisation allowed for the discovery and measurement of several non-linear associations between the varied categories of economic indicators employed, allowing for a better understanding of the driving factors behind the forecasts produced.

The results suggest that this model enjoys several benefits over the traditionally used Bridge models, such as parsimony and parametric flexibility, allowing it to achieve competitive nowcasting results when compared to the publicly available results from the Federal Reserve Bank of Cleveland's inflation nowcasting model during the testing period of May 2024. This occurred because the MIDAS model does not suffer from partly-fixed weights, allowing the selection of parameters and weighting gradient to be regularly calibrated so that the inflation indices always reflect approximately the true sensitivity to the categories of economic indicators employed. On the other hand, the calibration of Bridge models tends to incur more complexity requiring structural changes to allow for the adjustment of parameter selection.

Finally, the proposed model can be easily estimated daily using a continuous stream of recent information. The employment of mixed frequency nowcasting models using diverse categories of alternative data is proven to exhibit strong predictive power allowing its user to uncover turning points and macroeconomic trends as they occur. Furthermore, the results obtained from these models are generally good supplements for human judgement, allowing for more informed decision-making when faced with the information incompleteness of using traditional low-frequency macroeconomic data.

5.2 Policy Implications and Recommendations

There are positive implications to increasing the availability of macroeconomic data. For researchers, the employment of various nowcasting models mitigates the ragged edge problem by making macroeconomic data available at a single point in time. The availability of robust preliminary estimates of the inflation indices provides for an early warning system allowing for the mitigation of macroeconomic shocks and surprises exacerbated by lengthy publication delays (Reichlin, 2019). Furthermore, the employment of high-frequency macroeconomic nowcasts allows researchers to understand the factors driving the forecasts, helping to alleviate long-term forecasting errors.

For policymakers, the availability of robust high-frequency nowcasting data allows for the anticipation of inflationary pressures resulting in the proactive adjustment of monetary policy. Namely, if the nowcasting model anticipates a surge in the growth rate of the inflation indices, central banks may consider pre-emptive interest rate hikes or tapering the purchasing of assets from the open market to curb further growth in the inflation indices (Federal Reserve Bank of Cleveland, 2023). The benefit of this model is best endowed when used in conjunction with other macroeconomic nowcasts, namely, those for Real Gross Domestic Product and Personal Consumption Expenditure, allowing the central bank to gain a holistic picture of the macroeconomic environment. Finally, the nowcasting model allows the central bank to study and confirm the impact that these monetary measures would have on the growth rate of the inflation indices upon taking inflation reduction measures.

Governments may use nowcasting models to target fiscal measures aimed at alleviating the negative impact on consumers brought about by inflation driven from surging energy and food prices. This study finds that high-frequency fuel prices were significant even for nowcasting core indices due to the contagion effect it has, being a major factor of production, on other product prices in the economy (Modugno, 2013). The nowcasting model therefore measures the effect that fiscal measures aimed at decreasing energy prices would have on controlling the inflation growth rates, allowing the government to target this sector to alleviate inflationary pressures in the economy. Furthermore, governments may employ inflation nowcasting models to anticipate changes in monetary policy, allowing for better public debt management.

Inflation nowcasting models have significant implications for market practitioners in the field of portfolio and risk management. The availability of real-time insights in the movement of the inflation rate allows the investment manager to make better informed decisions using up-to-date information, inducing timeliness and mitigating economic uncertainty. Using robust preliminary estimates of the inflation indices, portfolio managers may better adjust asset allocations, hedge positions, and manage exposure to inflation-sensitive securities, solving the information asymmetry present when relying on low-frequency macroeconomic data with a significant publication delay (Cieslak and Pflueger, 2023). Namely, upon expecting an increase in the growth rate of the inflation indices, portfolio managers could shift away their asset allocations from inflation sensitive assets and instead hold on to inflation-protected assets such as inflation linked indices, inflation protected securities (TIPS), and commodities.

Moreover, arbitrage profits may be produced when nowcasting models produce statistically and economically significant results, deviating drastically from the expected official estimates which are released with a publication delay. These surprise results may induce shocks in the price movements of various asset categories, including commodity futures, gold, treasury bonds, and stock indices, such as the S&P 500, which react heavily to surprise changes in the growth rate of inflation. The mitigation of information asymmetry brought about by the application of nowcasting models may reduce these arbitrage opportunities, increasing market transparency by enabling more accurate pricing of assets, leading to more efficient and less volatile financial markets.

The employment of nowcasting models can improve communication with investors providing more precise and timelier explanations of portfolio performance, especially in periods of high volatility. This can enhance trust and confidence within the financial system as investment decisions are grounded in the most recent and relevant data. This may also help mitigate some of the behavioural effects brought about by cognitive biases such as overreaction in the financial markets. Furthermore, increasing informational efficiency in the financial markets may contribute towards mitigating the effects of leptokurtic (extreme) returns which characterize most financial markets.

Finally, the benefits of employing nowcasting models shows that an effort should be undertaken by more economic authorities and statistical agencies to achieve better and more efficient dissemination of macroeconomic data. This will allow for similar inferences to be conducted in different economic areas such as the nowcasting of various important macroeconomic indices. Furthermore, handing public access of high-frequency nowcasting results may foster greater coordination between policymakers and regulatory authorities ensuring a cohesive response to inflationary pressures.

5.3 Comparison to Existing Literature, Limitations, and Avenues for Future Research

This research enhances the existing literature by estimating and analysing an inflation nowcasting model using alternative data with mixed frequencies via a novel application of the Mixed Data Sampling (MIDAS) time series model. The results of this dissertation suggest that this model enjoys considerable benefits which may be employed by market practitioners as a reliable inflation nowcasting model.

This study addresses the omission of the MIDAS model present in the existing literature, as seen in works like Knotek II & Zaman, (2023) and Modugno, (2013), which pertain to the nowcasting of inflation indices. The findings of this study indicate that the MIDAS model is a parsimonious and robust alternative to the current Bridge and Dynamic Factor models commonly employed in nowcasting macroeconomic variables such as inflation growth rates.

Furthermore, this study finds that the employment of a large and varied dataset comprising of high-frequency data in models using the MIDAS approach allow for superior robustness when nowcasting inflation growth rates, especially when compared with the traditional Bridge equations used in the inflation nowcasting platform published daily by the Federal Reserve Bank of Cleveland.

Limitations encountered during this study include that:

- The model depends on the frequent updating of continuously releasing data which is needed to maintain accuracy. Since data is being published in real-time, the models must also be frequently updated to produce nowcasts up to the point of the release of the official estimates. The researcher might therefore alleviate this by setting a nowcasting frequency, such as weekly or daily, or else decide to conduct the nowcasts in irregular intervals pertaining to significant releases of data only (Kapetanios, Marcellino, & Papailias, 2018; Armesto, Engemann, & Owyang, 2010; Bańbura, Giannone, Modugno, & Reichlin, 2013),

- The model may not produce meaningful results if the data fed into them is insignificant in explaining the variation in the inflation indices. A reason may be that the indicators employed may fail to reflect current economic trends, resulting in forecasting errors and macroeconomic surprises (Reichlin, 2019). To ensure that the data used to produce the nowcasts is significant, the researcher must ensure that an observable relationship exists between the regressors and the dependent variable, and that the indicators employed exhibit the necessary characteristics of volume (scale of data), velocity (analysis of streaming data), variety (different forms of data), and veracity (uncertainty of data) (Kapetanios, Marcellino & Papailias, 2018), and

- Regular calibration may be required to ensure that the models are optimized and producing robust nowcasting results. Furthermore, this may require the constant attention of econometric experts to adjust the selection of parameters within the models.

While this study, like most literature found, employs data from and proposes nowcasting models for the United States economy, it is imperative that similar experiments are conducted to produce feasible nowcasting models for other economic jurisdictions. Namely, similar experiments may be conducted for countries which are characterised by a substantial accessibility of mixed and high-frequency data, such as Germany, the United Kingdom, or a monetary union such as the Eurozone.

Finally, it is also imperative that studies are conducted to measure the impact which nowcasting models could have on the informational efficiency of financial markets. This may involve the impact of mitigating uncertainty and macroeconomic surprises, and the implications that such models could have on crisis management, portfolio management, and risk management, such as taking advantage of potential arbitrage opportunities and hedging risk through the adjustment of asset allocations.

Bibliography

Andreini, P., Hasenzagl, T., Reichlin, L., Senftleben-König, C. & Strohsal, T. 2023, *Nowcasting German GDP: Foreign factors, financial markets, and model averaging*, Elsevier BV.

Armesto, M.T. (2010). Forecasting with mixed frequencies. *Review*, [online] 92(Nov), pp.521–536. Available at: https://ideas.repec.org/a/fip/fedlrv/y2010inovp521-536nv.92no.6.html.

Bańbura, M., Belousova, I., Bodnár, K. & Tóth, M.B. Nowcasting employment in the euro area, European Central Bank.

Bańbura, M., Giannone, D., Modugno, M. & Reichlin, L. 2013, Now-Casting and the Real-Time Data Flow, Elsevier, European Central Bank.

Bańbura, M. and Rünstler, G. (2024). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, [online] 27(2), pp.333–346. Available at: https://econpapers.repec.org/article/eeeintfor/v_3a27_3ay_3a_3ai_3a2_3ap_3a333-346.htm.

Barbaglia, L. (2024). *Nowcasting economic activity in European regions using a mixed-frequency dynamic factor model*. [online] Papers. Available at: https://ideas.repec.org/p/arx/papers/2401.10054.html.

Barnett, W.A., Chauvet, M. & Leiva-Leon, D. *Real-Time Nowcasting of Nominal GDP Under Structural Breaks*, SSRN.

Brooks, C. (2019). *Introductory econometrics for finance*. Cambridge, Uk Cambridge University Press.

Bouwman, K.E. & Jacobs, J.P.A.M. 2009, "Forecasting with real-time macroeconomic data: The ragged-edge problem and revisions", *Journal of macroeconomics,* vol. 33, no. 4, pp. 1–19.

Cieslak, A. and Pflueger, C. (2023). Inflation and Asset Returns. *Annual Review of Financial Economics*.

Claudio, J.C., Heinisch, K. & Holtemöller, O. 2020, "Nowcasting East German GDP growth: a MIDAS approach", *Empirical economics,* vol. 58, no. 1, pp. 29–54.

Coble, D. and Pincheira, P.M. (2017). Now-Casting Building Permits with Google Trends. *SSRN Electronic Journal*. doi: https://doi.org/10.2139/ssrn.2910165.

Dauphin, J., Taheri Sanjani, M., Suphaphiphat, N., Dybczak, K., Zhang, H., Maneely, M. & Wang, Y. 2022, Nowcasting GDP - A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies, International Monetary Fund (IMF), International Monetary Fund.

Degiannakis, S. 2023, "The D-model for GDP nowcasting", *Schweizerische Zeitschrift für Volkswirtschaft und Statistik,* vol. 159, no. 1, pp. 7–33.

Eduardo André Costa, Maria Eduarda Silva and Ana Beatriz Galvão (2024). Real-time nowcasting the monthly unemployment rates with daily Google Trends data. *Socio-economic planning sciences*, pp.101963–101963. doi: https://doi.org/10.1016/j.seps.2024.101963.

European Commission a, *Euro Area Macroeconomic Real-time Monitoring*. Available: https://web.jrc.ec.europa.eu/rapps/pub/ea-nowcasting/ [2023, 20/11/].

European Commission b, *Nowcasting*. Available: https://joint-research-centre.ec.europa.eu/scientific-activities-z/macroeconomic-monitoring-and-fiscal-surveillance/nowcasting_en [2023, 20/11/].

Federal Reserve Bank of Cleveland, *Inflation Nowcasting*. Available: https://www.clevelandfed.org/indicators-and-data/inflation-nowcasting [2023, 26/11/].

Federal Reserve Bank of St. Louis, *Nowcast - Economic Data Series - FRED* [Homepage of Federal Reserve Bank of St. Louis], [Online]. Available: https://fred.stlouisfed.org/tags/series?t=nowcast [2023, 20/11/].

Ghysels, E. and Marcellino, M. (2018). *Applied economic forecasting using time series methods*. New York, Ny: Oxford University Press.

Ghysels, E., Santa-Clara, P. & Valkanov, R. 2004, *The MIDAS Touch: Mixed data sampling (MIDAS) regression models*, NORTH-HOLLAND PUBL CO.

Ghysels, E., Sinko, A. & Valkanov, R. 2006, *MIDAS Regressions: Further Results and New Directions **.

Giannone, D., Reichlin, L. and Small, D.H. (2006). Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases. *SSRN Electronic Journal*. doi: https://doi.org/10.2139/ssrn.873658.

Glosser, S.M. and Golden, L. (1997). Average work hours as a leading economic variable in US manufacturing industries. *International Journal of Forecasting*, 13(2), pp.175–195. doi: https://doi.org/10.1016/s0169-2070(96)00725-x.

Grover, S.P., Kliesen, K.L. & Mccracken, M.W. 2016, *A Macroeconomic News Index for Constructing Nowcasts of U.S. Real Gross Domestic Product Growth*, Federal Reserve Bank of St. Louis.

Higgins, P. 2014, *GDPNow: A Model for GDP "Nowcasting"*, Federal Reserve Bank of Atlanta, Federal Reserve Bank of Atlanta.

Hopp, D. (2022). Benchmarking Econometric and Machine Learning Methodologies in Nowcasting. *SSRN Electronic Journal*. doi: https://doi.org/10.2139/ssrn.4101997.

IMF. (2021). *Impact of COVID-19: Nowcasting and Big Data to Track Economic Activity in Sub-Saharan Africa*. [online] Available at: https://www.imf.org/en/Publications/WP/Issues/2021/05/01/Impact-of-COVID-19-Nowcasting-and-Big-Data-to-Track-Economic-Activity-in-Sub-Saharan-Africa-50296.

INSTITUTE FOR RESEARCH ON POVERTY. (2024). *What is the consumer price index and how is it used?* [online] Available at: https://www.irp.wisc.edu/resources/what-is-the-consumer-price-index-and-how-is-it-used/#:~:text=Relevant%20to%20the%20topic%20of.

Jung, J.-K., Patnam, M. and Ter-Martirosyan, A. (2018). *An Algorithmic Crystal Ball: Forecasts-Based on Machine Learning*. [online] Ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3297651.

Kapetanios, G., Marcellino, M. & Papailias, F. 2018, *Big Data and Macroeconomic Nowcasting*, Editorial Universitat Politècnica de València.

Knotek, E.S. & Zaman, S. 2023, *A real-time assessment of inflation nowcasting at the Cleveland Fed*, Federal Reserve Bank of Cleveland.

Knotek, E.S. & Zaman, S. 2015, *Nowcasting U.S. Headline and Core Inflation*, Federal Reserve Bank of Cleveland.

Kohns, D. & Potjagailo, G. 2023, *Flexible Bayesian MIDAS: time-variation, group-shrinkage and sparsity*, Bank of England.

Kohns, D. & Potjagailo, G. 2022, *A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting*, Bank of England.

Marcellino, M. 2021, *Nowcasting GDP Growth in a Small Open Economy*, Cambridge University Press, London.

Modugno, M. 2013, *Nowcasting Inflation Using High Frequency Data*, SSRN, European Central Bank.

Mogliani, M. & Simoni, A. 2020, *Bayesian MIDAS Penalized Regressions: Estimation, Selection, and Prediction*, SSRN, Banque de France.

Paredes, T., Warmedinger, T., Paredes, J. and Asimakopoulos, S. (2013). *Forecasting fiscal time series using mixed frequency data*. [online] Repec.org. Available at: https://econpapers.repec.org/paper/ecbecbwps/20131550.htm.

Pettenuzzo, D., Timmermann, A.G. & Valkanov, R.I. 2014, "A Bayesian Midas Approach to Modeling First and Second Moment Dynamics", *SSRN Electronic Journal*, pp. 1–46.

Pope Francis (2020). *Fratelli Tutti*. [online] www.vatican.va. Available at: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html.

Reichlin, L. 2019, *Nowcasting*, London Business School.

S. Knotek II, E. & Zaman, S. *Inflation Nowcasting: Frequently Asked Questions*, Federal Reserve Bank of Cleveland, Federal Reserve Bank of Cleveland.

Schumacher, C. (2014). A comparison of MIDAS and bridge equations. International Journal of Forecasting, 32(2), pp.257–270. doi: https://doi.org/10.1016/j.ijforecast.2015.07.004.

Schwarzer, J.A. (2018). Retrospectives: Cost-Push and Demand-Pull Inflation: Milton Friedman and the ‘Cruel Dilemma’. *Journal of Economic Perspectives*, [online] 32(1), pp.195–210. doi: https://doi.org/10.1257/jep.32.1.195.

Slotman, J. *Forecasting and nowcasting at EAPD*, UN/DESA - Economic Analysis and Policy Division, UN/DESA - Economic Analysis and Policy Division.

Stock, J.H. and Watson, M.W. (2020). *Introduction to econometrics*. Harlow, Uk I Pozostałe: Pearson Education Limited.

Wallis, K.F. 1986, *Forecasting with an Econometric Model: The 'Ragged Edge' Problem*, Journal of Forecasting, Journal of Forecasting.

Appendix

A Estimation Results

This appendix includes the results obtained from the estimation of the regressions. To estimate the regression, the following programs were executed on R. Parameter selection was based on sensitivity analysis, providing the most efficient model by minimising the forecasting error with the least number of parameters possible. Furthermore, the estimated coefficients obtained describe the expected relationship between the parameters of the model and the inflation indices. Finally, the p-value exhibits the statistical significance of the parameters employed within the models. However, while some parameters may appear to be non-significant, the presence of non-linear relationships within the different categories of economic indicators provides for improved forecasting performance within the models.

<u>United States CPI – MIDAS Exponential Almon Regression Model</u>

Start = 2019(7), End = 2024(4)

```r
# MIDAS Equation 2
# Nealmon

CPI_US_EQ2 <- midas_r(CPIAUCSL_ts ~
  mls(CPILFESL_ts, 0:1, 1, nealmon) +
  mls(IQ_ts, 0:1, 1, nealmon) +
  mls(DSPIC96_ts, 0:2, 1, nealmon) +
  mls(BOPGSTB_ts, 0, 1, nealmon) +
  mls(MTSDS133FMS_ts, 0:1, 1, nealmon) +
  mls(CES0500000003_ts, 0:1, 1, nealmon) +
  mls(AWHAEMAN_ts, 0:1, 1, nealmon) +
  mls(JTSJOL_ts, 0:2, 1, nealmon) +
  mls(TOTALSA_ts, 0:2, 1, nealmon) +
  mls(PERMIT_ts, 0, 1, nealmon) +
  mls(TTLCONS_ts, 0, 1, nealmon) +
  mls(DGORDER_ts, 0:1, 1, nealmon) +
  mls(GACDFSA066MSFRBPHI_ts, 0:1, 1, nealmon) +
  mls(UMCSENT_ts, 0:2, 1, nealmon) +
  mls(UNRATE_ts, 2, 1, nealmon) +
  mls(GASDESW_ts, 0:4, 5, nealmon) +
  mls(ADPWNUSNERSA_ts, 0:5, 5, nealmon),
  #mls(COMPOUT_ts, 0, 5, nealmon),
  start = NULL,
  weight_gradients = list(nealmon = nealmon_gradient)
)

summary(CPI_US_EQ2)
```

Figure 11: United States CPI - MIDAS Exponential Almon Regression Model

```
 Parameters:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)            -1.137e-02  4.061e-02  -0.280 0.782991    
CPILFESL_ts1            1.061e+00  1.795e-01   5.910 2.20e-05 ***
CPILFESL_ts2           -1.424e-01  2.297e-01  -0.620 0.544034    
IQ_ts1                  2.609e-02  2.218e-02   1.176 0.256736    
IQ_ts2                  7.787e-02  2.421e-02   3.217 0.005385 ** 
DSPIC96_ts1             1.661e-04  4.137e-05   4.014 0.001002 ** 
DSPIC96_ts2             1.499e-04  5.660e-05   2.648 0.017541 *  
DSPIC96_ts3             7.669e-05  3.293e-05   2.328 0.033324 *  
BOPGSTB_ts             -1.193e-05  9.045e-06  -1.319 0.205620    
MTSDS133FMS_ts1         1.135e-06  2.480e-07   4.577 0.000310 ***
MTSDS133FMS_ts2         2.728e-07  1.462e-07   1.865 0.080573 .  
CES0500000003_ts1       8.917e-01  5.489e-01   1.624 0.123821    
CES0500000003_ts2      -3.975e-01  3.629e-01  -1.095 0.289550    
AWHAEMAN_ts1            4.475e-01  3.298e-01   1.357 0.193625    
AWHAEMAN_ts2            5.165e-01  1.943e-01   2.659 0.017169 *  
JTSJOL_ts1              3.548e-04  8.489e-05   4.180 0.000708 ***
JTSJOL_ts2             -2.590e-04  1.119e-04  -2.314 0.034257 *  
JTSJOL_ts3             -4.232e-04  6.087e-05  -6.952 3.26e-06 ***
TOTALSA_ts1             8.417e-02  2.272e-02   3.705 0.001923 ** 
TOTALSA_ts2            -1.004e-01  3.822e-02  -2.626 0.018345 *  
TOTALSA_ts3            -2.051e-01  8.619e-02  -2.379 0.030140 *  
PERMIT_ts              -1.418e-03  4.789e-04  -2.960 0.009214 ** 
TTLCONS_ts              5.091e-06  1.031e-06   4.937 0.000149 ***
DGORDER_ts1            -9.515e-06  3.468e-06  -2.744 0.014418 *  
DGORDER_ts2             8.603e-06  3.604e-06   2.387 0.029676 *  
GACDFSA066MSFRBPHI_ts1 -4.269e-03  3.088e-03  -1.382 0.185870    
GACDFSA066MSFRBPHI_ts2 -8.197e-03  3.652e-03  -2.244 0.039297 *  
UMCSENT_ts1             1.159e-02  9.202e-03   1.260 0.225882    
UMCSENT_ts2             1.983e-02  9.430e-03   2.103 0.051647 .  
UMCSENT_ts3             6.642e-03  5.595e-03   1.187 0.252514    
UNRATE_ts              -6.358e-02  7.771e-02  -0.818 0.425239    
GASDESW_ts1             4.160e+00  1.210e+00   3.437 0.003385 ** 
GASDESW_ts2            -5.986e-01  8.139e-01  -0.735 0.472730    
GASDESW_ts3             9.212e-02  7.371e-01   0.125 0.902102    
GASDESW_ts4            -1.058e+00  7.904e-01  -1.339 0.199258    
GASDESW_ts5            -5.587e-01  4.821e-01  -1.159 0.263486    
ADPWNUSNERSA_ts1        7.256e-08  3.832e-07   0.189 0.852205    
ADPWNUSNERSA_ts2        1.534e-07  6.401e-07   0.240 0.813651    
ADPWNUSNERSA_ts3       -1.206e-06  5.165e-07  -2.336 0.032845 *  
ADPWNUSNERSA_ts4        6.724e-08  6.559e-07   0.103 0.919628    
ADPWNUSNERSA_ts5        1.876e-06  3.637e-07   5.159 9.51e-05 ***
ADPWNUSNERSA_ts6       -1.067e-06  2.478e-07  -4.307 0.000543 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Residual standard error: 0.1795 on 16 degrees of freedom
```

Figure 12: United States CPI - MIDAS Exponential Almon Estimation Results

United States Core CPI – MIDAS Exponential Almon Regression Model

Start = 2019(8), End = 2024(4)

```r
# MIDAS Equation 2
# Nealmon

CORE_CPI_US_EQ2 <- midas_r(CPILFESL_ts ~
  #trend +
  mls(CPIAUCSL_ts, 0:1, 1, nealmon) +
  mls(IQ_ts, 1, 1, nealmon) +
  mls(DSPIC96_ts, 0:1, 1, nealmon) +
  mls(BOPGSTB_ts, 1:2, 1, nealmon) +
  mls(MTSDS133FMS_ts, 0:1, 1, nealmon) +
  #mls(CES0500000003_ts, 0, 1, nealmon) +
  mls(AWHAEMAN_ts, 1, 1, nealmon) +
  mls(JTSJOL_ts, 1:2, 1, nealmon) +
  #mls(TOTALSA_ts, 0, 1, nealmon) +
  #mls(PERMIT_ts, 0, 1, nealmon) +
  mls(TTLCONS_ts, 3, 1, nealmon) +
  mls(DGORDER_ts, 0:1, 1, nealmon) +
  mls(GACDFSA066MSFRBPHI_ts, 1, 1, nealmon) +
  mls(UMCSENT_ts, 0, 1, nealmon) +
  mls(UNRATE_ts, 1, 1, nealmon) +
  mls(GASDESW_ts, 0:3, 5, nealmon) +
  mls(ADPWNUSNERSA_ts, 0:4, 5, nealmon),
  #mls(COMPOUT_ts, 0:6, 5, nealmon),
  start = NULL,
  weight_gradients = list(nealmon = nealmon_gradient)
)

summary(CORE_CPI_US_EQ2)
```

Figure 13: United States Core CPI - MIDAS Exponential Almon Regression Model

Parameters:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.234e-03	2.862e-02	0.288	0.775659	
CPIAUCSL_ts1	6.567e-01	7.276e-02	9.026	6.39e-10	***
CPIAUCSL_ts2	1.459e-01	7.890e-02	1.849	0.074663	.
IQ_ts	-8.965e-02	2.181e-02	-4.110	0.000296	***
DSPIC96_ts1	-1.792e-04	3.191e-05	-5.616	4.59e-06	***
DSPIC96_ts2	-2.374e-04	5.161e-05	-4.600	7.70e-05	***
BOPGSTB_ts1	1.141e-05	6.882e-06	1.658	0.108172	
BOPGSTB_ts2	2.378e-05	6.872e-06	3.460	0.001691	**
MTSDS133FMS_ts1	-7.792e-07	1.846e-07	-4.221	0.000219	***
MTSDS133FMS_ts2	-1.025e-06	2.221e-07	-4.615	7.39e-05	***
AWHAEMAN_ts	4.802e-01	2.561e-01	1.875	0.070905	.
JTSJOL_ts1	1.465e-04	8.058e-05	1.819	0.079315	.
JTSJOL_ts2	3.789e-04	6.511e-05	5.820	2.62e-06	***
TTLCONS_ts	1.057e-05	2.302e-06	4.591	7.90e-05	***
DGORDER_ts1	9.298e-06	3.845e-06	2.418	0.022098	*
DGORDER_ts2	7.909e-06	2.987e-06	2.648	0.012955	*
GACDFSA066MSFRBPHI_ts	2.265e-03	1.568e-03	1.444	0.159453	
UMCSENT_ts	-2.019e-02	8.223e-03	-2.455	0.020313	*
UNRATE_ts	2.024e-01	6.764e-02	2.993	0.005600	**
GASDESW_ts1	-2.384e+00	9.911e-01	-2.405	0.022768	*
GASDESW_ts2	-4.473e-01	3.460e-01	-1.293	0.206284	
GASDESW_ts3	3.495e-01	3.926e-01	0.890	0.380751	
GASDESW_ts4	-9.610e-01	6.240e-01	-1.540	0.134388	
ADPWNUSNERSA_ts1	5.282e-07	3.765e-07	1.403	0.171306	
ADPWNUSNERSA_ts2	-8.529e-07	4.929e-07	-1.730	0.094189	.
ADPWNUSNERSA_ts3	8.290e-07	3.307e-07	2.507	0.018042	*
ADPWNUSNERSA_ts4	2.765e-07	3.002e-07	0.921	0.364617	
ADPWNUSNERSA_ts5	-6.743e-07	2.538e-07	-2.657	0.012684	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1821 on 29 degrees of freedom

Figure 14: United States Core CPI - MIDAS Exponential Almon Estimation Results

B Residual Testing

This appendix includes the results obtained from the execution of:

- VIF Multicollinearity tests
- Descriptive statistics on the model residuals
- Jarque-Bera test for residual normality
- White test for residual heteroskedasticity
- Breusch-Godfrey test for residual autocorrelation

VIF tests were conducted to examine the properties of multicollinearity within the parameters of the model. In these estimates, none of them exhibited a VIF value greater than 5 representing freedom from the negative effects of multicollinearity. Furthermore, the Jarque Bera test on the model residuals indicates how similar the residuals are distributed compared to the normal distribution. The White test, which on R is represented as the BP test with non-studentized residuals, tests for general forms of heteroskedasticity within the residuals of the model. Finally, the Breusch-Godfrey test confirms the absence of residual serial correlation within our models.

<u>United States CPI – MIDAS Exponential Almon Regression Model</u>

Start = 2019(7), End = 2024(4)

A matrix: 17 × 3 of type dbl

	GVIF	Df	GVIF^(1/(2*Df))
mls(CPILFESL_ts, 0:1, 1)	14.96	2	1.97
mls(IQ_ts, 0:1, 1)	12.89	2	1.89
mls(DSPIC96_ts, 0:2, 1)	43.33	3	1.87
mls(BOPGSTB_ts, 0, 1)	2.75	1	1.66
mls(MTSDS133FMS_ts, 0:1, 1)	49.09	2	2.65
mls(CES0500000003_ts, 0:1, 1)	343.37	2	4.30
mls(AWHAEMAN_ts, 0:1, 1)	208.71	2	3.80
mls(JTSJOL_ts, 0:2, 1)	78.17	3	2.07
mls(TOTALSA_ts, 0:2, 1)	382.19	3	2.69
mls(PERMIT_ts, 0, 1)	4.10	1	2.03
mls(TTLCONS_ts, 0, 1)	4.73	1	2.18
mls(DGORDER_ts, 0:1, 1)	29.46	2	2.33
mls(GACDFSA066MSFRBPHI_ts, 0:1, 1)	32.33	2	2.38
mls(UMCSENT_ts, 0:2, 1)	44.72	3	1.88
mls(UNRATE_ts, 2, 1)	16.94	1	4.12
mls(GASDESW_ts, 0:4, 5)	137.75	5	1.64
mls(ADPWNUSNERSA_ts, 0:5, 5)	197.49	6	1.55

Figure 15: United States CPI - Exponential Almon Model VIF Results

```
[1] "Mean:  -8.85499926272711e-18"
[1] "Standard Deviation:  0.0951272893691348"
[1] "Skewness:  0.830976633262672"
[1] "Kurtosis:  3.52449497119613"

        Jarque Bera Test

data:  a
X-squared = 7.3399, df = 2, p-value = 0.02548

NULL

        Breusch-Pagan test

data:  residuals(CPI_US_EQ2) ~ fitted(CPI_US_EQ2) + I((fitted(CPI_US_EQ2))^2)
BP = 0.16547, df = 2, p-value = 0.9206

        Breusch-Godfrey test for serial correlation of order up to 1

data:  residuals ~ fitted
LM test = 0.22055, df = 1, p-value = 0.6386
```

Figure 16: United States CPI - Exponential Almon Model Residual Testing Results

<u>United States Core CPI – MIDAS Exponential Almon Regression Model</u>

Start = 2019(8), End = 2024(4)

A matrix: 14 × 3 of type dbl

	GVIF	Df	GVIF^(1/(2*Df))
mls(CPIAUCSL_ts, 0:1, 1)	9.62	2	1.76
mls(IQ_ts, 1, 1)	2.81	1	1.68
mls(DSPIC96_ts, 0:1, 1)	8.76	2	1.72
mls(BOPGSTB_ts, 1:2, 1)	4.78	2	1.48
mls(MTSDS133FMS_ts, 0:1, 1)	14.01	2	1.93
mls(AWHAEMAN_ts, 1, 1)	17.29	1	4.16
mls(JTSJOL_ts, 1:2, 1)	6.83	2	1.62
mls(TTLCONS_ts, 3, 1)	3.71	1	1.93
mls(DGORDER_ts, 0:1, 1)	8.51	2	1.71
mls(GACDFSA066MSFRBPHI_ts, 1, 1)	2.85	1	1.69
mls(UMCSENT_ts, 0, 1)	2.93	1	1.71
mls(UNRATE_ts, 1, 1)	23.11	1	4.81
mls(GASDESW_ts, 0:3, 5)	13.51	4	1.38
mls(ADPWNUSNERSA_ts, 0:4, 5)	9.46	5	1.25

Figure 17: United States Core CPI - Exponential Almon Model VIF Results

```
[1] "Mean:  -1.30854126335272e-17"
[1] "Standard Deviation:  0.131027048069022"
[1] "Skewness:  -0.269329686371663"
[1] "Kurtosis:  3.67671519523898"

        Jarque Bera Test

data:  a
X-squared = 1.7767, df = 2, p-value = 0.4113

NULL

        Breusch-Pagan test

data:  residuals(CORE_CPI_US_EQ2) ~ fitted(CORE_CPI_US_EQ2) + I((fitted(CORE_CPI_US_EQ2))^2)
BP = 2.3745, df = 2, p-value = 0.3051

        Breusch-Godfrey test for serial correlation of order up to 1

data:  residuals ~ fitted
LM test = 1.282, df = 1, p-value = 0.2575
```

Figure 18: United States Core CPI - Exponential Almon Model Residual Testing Results