MASTER'S THESIS

Interuniversity Master in Statistics and Operations Research UPC-UB

Title: From MIDAS to Deep Learning: A comprehensive benchmark of big data economic forecasting models

Author: Miguel Benítez Humanes

Advisor: Dr. Argimiro Arratia (UPC), Dr. Alex Cebrián (NTT Data)

Speaker: Dr. Ariel Duarte

Department: Computer Science

University: Universitat Politècnica de Catalunya (School of Mathematics and Statistics)

Date: December 2023

Academic year: 2023-2024

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

UNIVERSITAT DE BARCELONA

From MIDAS to Deep Learning: A comprehensive benchmark of big data economic forecasting models*

Miguel Benítez Humanes

December 2023

Abstract

This thesis proposes a new benchmark including some of the most popular economic forecasting models. Among other features, it tests the forecasting ability of a text-based indicator, of sophisticated mixed data sampling (MIDAS) models and of several neural network architectures. The benchmark shows that, of all the models considered, a simple one-factor dynamic factor model (DFM) using only economic variables is the best model to predict gross domestic product growth. For inflation, the best model in the benchmark is the ARIMA model. However, this study finds that in periods with high uncertainty like 2020-2022, long term forecasts improve when using more complex DFMs for gross domestic product, and appropriate neural network architectures for inflation.

*The code and most of the data used to obtain the results presented in this thesis are publicly available in the following GitHub repository: `https://github.com/miguelBhumanes/FinalMasterThesis.git`

I am grateful for the invaluable comments and guidance of my research supervisors: Dr. A. Arratia and Dr. A. Duarte.

I also thank the useful comments of Dr. M Linares.

Contents

1 Context of this Thesis

This final master thesis is the result of a research assistantship with Dr Argimiro Arratia, Associate Professor at the Department of Computer Science of Universitat Politècnica de Catalunya (UPC), and Dr. Ariel Duarte, Associate Lecturer at UPC. In the context of this assistantship all the work here presented has been undertaken only by myself, with the guidance of Dr. Arratia and Dr. Ariel Duarte, who benefit from my work due to its complementarity with their research projects.

The assistantship began in 2023 with the goal of exploring the potential of text based financial indicators to improve economic forecasts. More specifically, the initial idea was to explore the utility of these indicators using Mixed Data Sampling (MIDAS) regression models for forecasting. After a throughout literature review of economic forecasting methods, the focus of the research assistantship shifted to creating a comprehensive benchmark of economic forecasting models.

This thesis represents a first version of that comprehensive benchmark, which summarizes all the results obtained until December 2023. This version already includes numerous alternative Dynamic Factor Models (DFMs) and many neural network architectures. In addition it includes basic ARIMA models and a LSTM autoregressive model to assess the relative value of dealing with big data sets for forecasting.

The main objective of this thesis is to present an initial source that economists can refer to when in need of developing new forecasting models. More specifically, a comprehensive benchmark where several forecasting models are consistently compared using the same data and training strategies. Logically, the benchmark is most useful when more models are included. Therefore all the code and most of the data[1] used to generate the results of the thesis are publicly available in a GitHub repository (miguelB-humanes/FinalMasterThesis). This will aid other economists in expanding the benchmark in the future, by trying different architectures with the same data, or more variations of the presented models.

[1]The text-based indicator used in this study is proprietary data of Dr. Argimiro Arratia and is therefore not available

2 Introduction

The expectations of economic agents are crucial in their decision making, especially, those regarding income and prices. Therefore, one main task expected from economists is forecasting the future value of gross domestic product (GDP) growth and inflation. Traditionally, economists have used economic aggregates and key variables like the reference interest rate as regressors of linear models that forecast future values of economic variables. However, more recent approaches try to improve the forecasts in at least three ways. First, the use of databases with a large number of predictors, which capture many different aspects of the economy. Second, the inclusion of alternative indicators in these databases, such as text derived indicators. Third, the use of model architectures able to adjust for the different frequencies of the data (like MIDAS models) and the potentially non linear relationships between the data (like neural networks).

While the number of research papers on economic forecasting is large, most of them focus only on one of the three mentioned approaches. For instance, most papers make use of only one or a few architectures to forecast GDP. Even when some economists have undertaken very useful efforts to prepare a comprehensive benchmark like in Hopp (2022), the level of completeness is still below ideal. Any fruitful attempt at improving economic forecasting should begin with a throughout review of the results of existing methods. Therefore, there is a need for an updated and more comprehensive benchmark of economic forecasting models.

This thesis proposes that a sufficiently comprehensive benchmark should satisfy at least the following characteristics: first, it should include both linear and non-linear methods. And the non-linear part should be exhaustive, as there exist many competing architectures. Second, the the methods should make use of large databases like the ones available to economists nowadays. Additionally, these databases should include, if possible, alternative indicators such as text based indicators. Thirdly, since the different indicators are typically in very different frequencies (quarterly, monthly, daily,...) a good quality benchmark would study the improvement of the models under sophisticated methods of frequency alignment like mixed data sampling (MIDAS) and more naïve approches. Finally, most papers in the literature focus exclusively on forecasting GDP. However, there exists a similar social interest in forecasting inflation. Therefore, a complete benchmark should compare the ability to forecast at least both GDP growth and inflation.

This thesis aims to present a benchmark that satisfies the stated requirements. The initial ambition of this study was to assess the potential of a text-based attention indicator to improve forecasts of economic variables. This will be achieved in this thesis. However, the forecast improvement of any indicator is specific to each model. Therefore, the thesis had to inevitably deal with the question of what is the best forecasting model. Leveraging the work done to answer that question, my ambition has been to repurpose my results into a report close to the ideal benchmark that has been described. This should result in a bigger contribution to the literature: a useful initial point of consultation for any economist with the job of employing new methods or data to improve economic forecasts.

The benchmark presented in this thesis first considers two very simple models: ARIMA models and an autoregressive LSTM model. Then it tests a large number of Dynamic Factor Models (DFMs). Lastly, it includes four of the most popular neural network architectures (Feed Forward, LSTM, GRU, Self Attention) trying different frequency alignment techniques. The results point out to one of the simplest DFM (the one presented in Bok, Caratelli, Giannone, Sbordone, and Tambalotti (2018)) being the most practical forecaster of GDP in the benchmark (with simple autoregressive models being also a very good alternative). On the other hand, a simple ARIMA model appears to be the best forecaster of inflation out of all the models compared. However, when doing long term forecasts (1 year ahead) on the 2020-2022 period, results differ. It appears that more complex models are able to achieve better results. These results are appealing: it seems that some of the simplest models are the best forecasting tool. Only when in need of a long term forecast during uncertain economic periods should economists care for developing more complex models.

This thesis gives strong backing to this claim. All the models presented were tested in a consistent way, using the same data and strategies for training. Therefore the reader can compare results in a much more reliable way than by reading different papers in the literature, which feature rather narrow explorations of the forecasting model space and use different datasets, training strategies. By giving researchers the ability to reliably compare a reasonably wide range of the best performing models in the previous literature, this thesis aims to make a useful contribution to the body of research focusing on economic variable forecasting. In addition, the thesis gives clear guidance on the best complex models in case of doing long term forecasts during uncertain periods, and makes publicly available the code to promptly implement them.

The rest of the sections of this thesis are organized as follows. Section 3 presents a thorough literature review of economic forecasting to motivate the use of the models and data used in this benchmark. In Section 4 the database used for the models is explained, as well as the way in which the data has been processed. Section 5 presents the ARIMA forecasting models and a non-linear autoregressive model. Section 6 introduces a very comprehensive benchmark of linear DFM models and their results. Section 7 similarly presents a less extense but still comprehensive benchmark of non-linear models and their results. Finally, in Section 8 the main conclusions are presented, limitations of this thesis and the ways in which the benchmark could be improved in the future.

3 Literature Review

A good starting point for a literature review of economic forecasting is the famous Lucas critique (Lucas Jr, 1976). According to it, any attempt at estimating a variable directly as a function of other economic data (especially aggregates) is assuming a given economic structure which may change. Considering the critique, Dynamic Stochastic General Equilibrium (DSGE) models should be an appropriate forecasting method. They introduce microeconomic foundations that create a more stable economic structure. GDP growth and inflation oscillations depend only on exogenous random (or policy) shocks and a set of parameters that are calibrated to fit real data.

However, DSGE models do not manage to produce consistent reliable forecasts of either GDP or inflation (Edge, Gürkaynak, Reis, & Sims, 2010). Moreover, to best match the real data they often rely on unrealistic assumptions that do not necessarily have strong microeconomic evidence (Romer, 2019a). On the other hand, relatively simpler models based on regressions, like Vector Autoregressions (VARs) can achieve similar results (Christoffel, Coenen, & Warne, 2010). Therefore, at least for forecasting purposes regression methods are justified.

VARs are certainly a very popular approach in economic forecasting. Nonetheless, regular VARs are unsuitable for data sets with a large number of variables. In spite of Bayesian VARs being able to handle a bigger number of parameters, when dealing with potentially hundreds of variables other approaches tend to be preferred in the literature. Namely, Dynamic Factor Models (DFMs) like the ones explained in Stock and Watson (2016). Notably, as De Mol, Giannone, and Reichlin 2008 show, a BVAR with a Gaussian prior tends to generate forecasts that are very correlated to a DFM.

DFMs assume that all the economic variables in the data set are explained by a reduced set of latent variables plus some random disturbances. If these disturbances follow a distribution with a covariance matrix proportional to the identity matrix, then such a model can be estimated by doing a regression where the factors are the (normalized) Principal Components Analysis (PCA) of the variables included in the database. (Stock & Watson, 2016). Ardia, Bluteau, and Boudt (2019) follow this approach when estimating a DFM to predict GDP with remarkable results. For forecasting horizons under 2 quarters, their errors are significantly lower than those corresponding to the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve

Bank of Philadelphia, according to the data reported by Bok et al. (2018). Their longer horizon forecasts are nonetheless worse than the SPF. Although, in general, forecasting ability beyond that horizon is very limited. The root mean squared error (RMSE) for GDP growth is close to 2.5% (Bok et al., 2018).

There are at least two potential reasons why the model by Ardia et al. (2019) achieves good results. First, they make use of an extended data set with financial and text-derived indicators, besides economic data. This constitutes a motivation to use financial data and text-derived indicators. On the contrary, Bok et al. (2018) argue against the inclusion of financial data. They cite Knotek and Zaman (2017), who do confirm that financial variables tend to worsen one quarter ahead forecasts and increase volatility of predictions. Nonetheless, Knotek and Zaman (2017) also claim financial variables improve forecasts further in the future thanks to their forward looking nature. They also claim the one quarter ahead forecast error decreases as the selection of financial is sufficiently complete and of good quality. In particular, they mention Stock and Watson (2003) as a comprehensive survey of financial indicators, which will serve as the base for the financial data set selected in this study.

Another reason for the good results of Ardia et al. (2019) might be their use of Mixed Data Sampling (MIDAS) frequency alignment. Financial and text-derived data can have very high frequency (even several observations per minute). Specifically, in Ardia et al. (2019) and in this study the frequency of both financial data and text indicators will be daily. While it is possible to directly regress on lags of variables that have different frequency (called Unrestricted MIDAS or U-MIDAS approach (Foroni, Marcellino, & Schumacher, 2011)) when the frequency imbalance is as large as the one between monthly and daily data this does not yield the best forecasting accuracy (Foroni et al., 2011). A naïve approach is to simply take the average for each month. This means taking weighted averages where all weights are restricted to be equal. A more refined alternative is to calibrate the weighting scheme to fit the data as well as possible. This is the essence of MIDAS models as presented in Ghysels, Santa-Clara, and Valkanov (2004). In the case of DFMs, Marcellino and Schumacher (2010) demonstrate that using MIDAS rather than a equally-weighted approach is a more sensible frequency alignment option.

One caveat of DFMs is their implicit assumption of linearity. When there is a large number of features, the assumption of all relationships being linear might be too limiting. Indeed, Hopp (2022) finds that Long-Short

Term Memory (LSTM) neural networks outperform a BVAR, and a DFM in nowcasting GDP growth. A neural network can be seen as a non-linear extension of a standard multivariate linear regression. In fact, the very common logarithmic transformation of the dependent variable is equivalent to an exponential activation function of a linear combination of the regressors' values. Doing this several times both in parallel (several nodes) and sequentially (several layers) creates a standard feed-forward multi-layer perceptron artificial neural network. So using neural network forecasting models is justified. They have the potential to be equivalent non-linear extensions of DFMs that capture additional patterns in the data, improving forecasts.

Theoretically, a neural network is a better choice to predict economic variables as functions of other regressors. These architectures were proven to be universal approximators with as little as one hidden (intermediate) layer by Hornik, Stinchcombe, and White (1989). This means that, provided they have sufficient nodes, they are able to approximate any finite dimension continuous function that defines the relation between the regressors and the dependent variable within a closed set of values. More recently, Kidger and Lyons (2020) have proven this holds true for most choices of activation functions, including rectified linear units (RelU). RelU is the most popular activation function due to its computational efficiency and good results, and the one used for all models in this paper. Therefore in theory, if all the relevant regressors are included and the target variable is a relatively stable function of those regressors, neural networks are the best type of model to predict economic variables. This provides one good reason to focus on neural networks in this benchmark.

Meanwhile, a potential improvement over standard feed-forward neural network architectures is the introduction of layers that consider the time structure of the time series inputs. Recurrent neural networks (RNNs) introduce this aspect by inputting two values to each node: the value of the corresponding time observation and an output value (called "hidden-state") coming from the node corresponding to the previous time period. LSTM neural networks are a type of RNN that combines two hidden states (one for long term memory and other for short term memory) with the inputs to obtain the recurrent predictions (Hochreiter & Schmidhuber, 1997). It is able to solve some technical problems of RNNs (namely, vanishing/exploding gradients) while also achieving good results, as in Hopp (2022). More recently, Gated Recurrent Units (GRUs) were introduced by Cho et al. (2014) as another effective RNN that uses less parameters than LSTM while still solving the same problems.

Moving aside from RNNs, in his famous paper Vaswani et al. (2017) suggest the use of non-recurrent neural networks to predict patterns with a temporal dependency. The multi-head attention layers they describe capture the dependency of data by computing dot products of embeddings (weighted transformations) of the data input vectors, rather than computing hidden states. This alternative layer type is much faster to train and has enabled the development of popular large language models like Open AI's GPT Model. As far as known, multi-head attention layers have not been used to forecast GDP or inflation. Therefore, some of the models presented will make use of this architecture to test it.

The existence of non-linear relationships is not necessarily exclusive to data published at the same frequency. Neural network models should be able to use a combination of monthly economic data with alternative information like daily text-derived indicators. To this extent, some papers like Xu, Zhuo, Jiang, and Liu (2019), Choi, Cho, and Kim (2020) and Xu, Liu, Jiang, and Zhuo (2021) successfully extend feed-forward and LSTM neural networks with MIDAS to handle the differences in frequency. Their results are in line with those of Marcellino and Schumacher (2010) and Ardia et al. (2019) for DFMs, suggesting that the MIDAS frequency alignment yields superior forecasting ability.

Finally, if a neural network forecasting model can be seen as the non-linear extension of a linear regression, the non-linear improvement of a DFM should have some sort of dimensionality reduction mechanism. One option is to directly apply the machine learning algorithm on PCA factors. This is equivalent to having one initial "PCA layer" that performs this operation on the raw data and then outputs the factors to subsequent layers. This could be improved with other dimensionality reduction options. Hauzenberger, Huber, and Klieber (2023) test several of those options and concludes the Autoencoder (AE) is the best dimensionality reduction technique for inflation forecasting. However, due to time and data constraints, this thesis did not test these methods and relied on using the PCA factors as the only method for dimensionality reduction.

Navigating the zoo of GDP and inflation forecasting models is an intensive task that spans beyond economics into other fields like Computer Science and Mathematics. Therefore, a consistent benchmark of the best performing models would be very useful for researchers. According to this literature review, it seems that DFMs and several types of neural network architectures should be included as they appear to yield the best results.

Finally, ARIMA models and a non-linear autoregressive model are included as “simplest model” elements in the benchmark. The rest of the models will be assessed not only on the relative performance between them but also on their ability to beat these very simple models.

4 Data

In this section, the data of this study is presented. The following subsections will explain the reasons to select all the individual variables included in the data set, the methods used to process them and the reasons for choosing those methods[2]. The last subsection includes a detailed explanation of how the factors used both for the DFMs and neural network models are created through PCA.

4.1 Dependent Variables

The models of this study will try to forecast the GDP growth and the inflation over one year. For that, first the quarterly real GDP and the monthly CPI Index data are downloaded from the Federal Reserve Bank of St Louis Economic Database (or FRED)[3]. The final two forecasting targets are obtained by computing the rate of growth over one year for each of the series. This transformation is the most common in the literature. It would be convenient to be able to assume that the series is already stationary after this transformation, as it would make this study more comparable.

Fortunately, following the Box Cox methodology (Box, Jenkins, Reinsel, & Ljung, 2015) there does not seem to be any strong reason not to make this assumption:

- First, the variance can be assumed homoskedastic. While there are episodes of high variance in both GDP growth and inflation, it is easy to observe that these correspond to brief periods coinciding with the Great Financial Crisis and the Covid pandemic (Figures 1 and 2). It is sensible to assume that these episodes are outlier negative economic shocks followed by a predictable economic recovery of a similar magnitude, rather than a changing variance of the random error term.
- Second, the two series downloaded are the seasonally adjusted versions of the data, available at the FRED. So there is no need of additional transformations to deal with the seasonality of the original series.

[2]Appendix A contains a list with the selected economic and financial indicators, as well as their source and their transformation.

[3]GDP available on https://fred.stlouisfed.org/series/GDPC1, and consumer price index available on https://fred.stlouisfed.org/series/CPIAUCSL

- Third, the plots show how both series are roughly mean stationary. Formal tests are done to reinforce this claim (Table 1). The ADF test p-values are very high suggesting the series are not mean stationary. However the p-values decrease when excluding all data after 2019 (Covid pandemic years) and is even below 5% for inflation. In addition, the KPSS tests for both series cannot reject that they are stationary. Therefore it is decided not to take any additional difference.

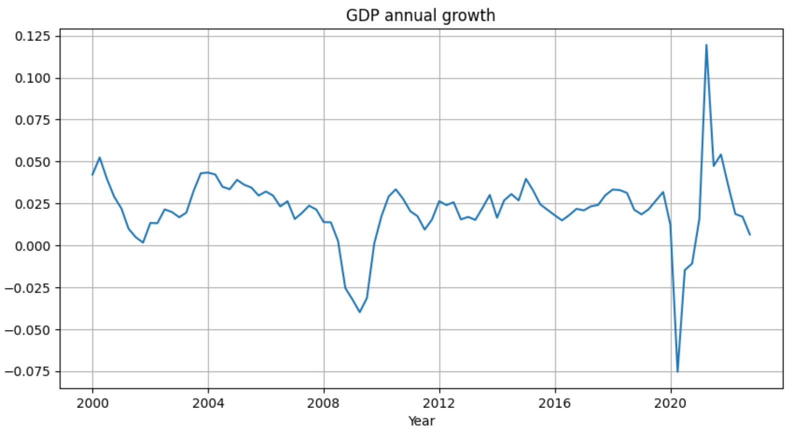

Figure 1: GDP Annual Growth

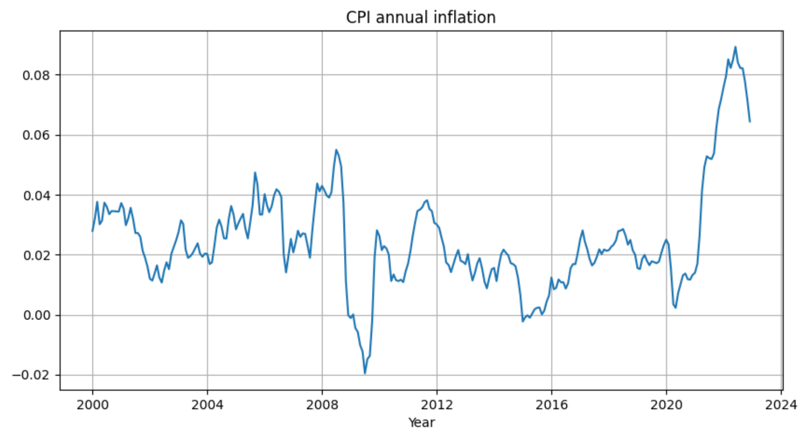

Figure 2: Annual CPI Inflation

	ADF Test	KPSS Test
GDP Growth	0.1769	0.1000*
Inflation	0.1789	0.1000*
GDP Growth (ex. 2020-2022)	0.1349	
Inflation (ex. 2020-2022)	0.0418	

Table 1: ADF and KPSS p-values for GDP Growth and Inflation

*Note: KPSS p-values are higher than 0.1 but the Python package used (statsmodels) does not display the exact result above that threshold.

4.2 Data Set of Regressors

The data set of regressors used to explain those two variables consists of observations of 281 variables from January 4, 2000 to December 31, 2022. The 281 variables result from the sum of 44 monthly and quarterly economic indicators, 37 daily financial indicators and 200 daily text indicators. It may be argued that the number of observations is relatively small for the number of features in the data set. Especially, when considering that complex neural networks with many parameters tend to require large amounts of data to reach their best performance (Sun, Shrivastava, Singh, & Gupta, 2017). In the case of this study, the limiting factor is the number of text-derived indicator observations in the data set provided. They represent more than half of the features, yet their data is only available since year 2000.

This study considers the most interesting option is to proceed with this "wide rectangular" data set. First, many economists using alternative indicators will usually encounter the same problem. As their contributions may depend entirely on using text-derived or other alternative indicators, it is more interesting that this benchmark explores how the different models compare in finding non linear relationships between this and more traditional data. The comparison will proceed in a "scarce" data setting similar to the one they may face. Fortunately, the use of PCA factors as a way of reducing the dimensionality of the data shows some ability of dealing with this problem.

4.3 Economic Regressors

It is clear that economic indicators should be included as features to forecast economic variables such as GDP and inflation. In Bok et al. (2018) a selection of indicators is provided, which is also the one used by Hopp (2022). Another starting point could be the database created by McCracken and Ng (2016) which also includes several disaggregated data indicators. This is for instance part of the data set used by Ardia et al. (2019). Considering that the data set will be limited to observations beginning in 2000, it is better not to use all the disaggregated indicators to avoid an excessive number of parameters. Bok et al. (2018) claim that the use of disaggregated data does not improve forecasting results. Unfortunately, avoiding parameter proliferation in the models will come at the expense of not testing this claim for non-linear models.

However, this study introduces some modifications to the set of economic indicators selected by Bok et al. (2018). First, the indicators for which the Federal Reserve does no longer publish information are dropped. These are 4 indicators corresponding to those published by the Institute for Supply Management (ISM) and the ADP National Employment Report. These indicators are replaced with a manufacturing confidence index and survey indexes for the non-manufacturing sector for the US, made publicly available by the OECD for the US. Considering the results obtained, these alternative indicators (which roughly measure the same target) appear to be good substitutions.

Second, additional indicators are added to the initial data set suggested by Bok et al. (2018). These are the University of Michigan survey-derived indexes for inflation expectations and consumer confidence, several measures of the amount of money in circulation, the federal funds rate and the S&P Case-Shiller index of house prices. There are reasons to include those indicators. The monthly survey results by the University of Michigan, and the Case-Shiller index of house prices are included due to their popularity among economists and traders. Meanwhile, the reason to add monetary variables and the federal funds rate is motivated by strong theoretical and empirical evidence of monetary variable changes having real effects (Romer, 2019b). Additionally, the reference interest rate is a staple variable in most theoretical economic models, and most empirical ones. Therefore in this study it is considered to be worth to add these variables to the initial selection done by Bok et al. (2018).

It is customary to introduce the variables in stationary form, even for Machine Learning models, as in Hopp (2022). This is also the approach taken in this study. It is possible that there are some relationships between GDP growth, the inflation rate and the levels of some of the regressors. However, testing which subset of variables need to left in non-stationary form is too computationally consuming to be part of the discussion of this benchmark. Once downloaded, the variables are passed through an algorithm that determines the optimal transformation to convert the series to stationary form. The available transformations are four: (1) no transformation, (2) level change, (3) percentage change, (4) level change of the percentage change.

The algorithm that converts data to stationary form first fits several ARIMA models. Then, the number of differences done to the original series is the one of the model that minimizes the Akaike's Information Criterion. When the differentiation is to be done on normal data, the growth rate is computed. This is more common than doing the level difference, and is also the approach taken by other papers like Bok et al. (2018) or De Mol et al. (2008). However when the algorithm recommends differencing a data point that is already a growth rate or a percentage, the level change of that percentage quantity is computed instead. The algorithm also tests for potential seasonality of the series. Since all the series that usually exhibit seasonal patterns were already downloaded in a seasonally adjusted form from the Federal Reserve of St. Louis database, it is not surprising that no additional transformation was required.

A final transformation is the imputation of missing values to the data set. A common heuristic consists of imputing the mean of the series (since it is stationary). A more sensible option, and the method used in this study, is to use the EM-PCA algorithm described in Marcellino and Schumacher (2010). Another source of missing values are those corresponding to the months with no publication for quarterly data. In this study the more simple approach of linearly interpolating the stationary quarterly series is chosen. Another possible option could have been to estimate a state-space model to impute the missing monthly observations of the quarterly data.

4.4 Financial Regressors

The inclusion of financial data has the potential to improve predictions of economic variables in short to medium term horizons thanks to the forward looking nature of the financial markets (Knotek & Zaman, 2017). Also according to Knotek and Zaman (2017) the inclusion of financial variables could increase the volatility of the shortest term forecasts if the selection of financial variables is relatively noisy or uninformative. Following their advice, the selection of financial variables of Stock and Watson (2003) is used as a starting point.

The definitive list of financial variables differs only slightly from that of Stock and Watson (2003). The main difference being the inclusion of additional variables due to them containing considerable predictive power. The included variables are: (1) the spreads between several points of the US yield curve and the yield of their correspoding inflation-protected securities and (2) the credit spread on corporate bonds for different ratings. The spread between government securities and their corresponding inflation-protected securities is a very good measure of the expectations of agents about future inflation. Similarly, the general level of credit spread carries information about agent's expectations of general default probabilities. Since these are related to general economic conditions, they are expected to contain signals of future GDP growth, inflation. Indeed, Mueller (2009) finds that credit spreads have predictive power.

The financial information can potentially be very noisy. Therefore the data is processed in a way that preserves as much signal information as possible. First, all the data is downloaded either from the Federal Reserve of St. Louis Database or from Bloomberg. These sources are widely trusted to provide accurate historical financial information. Second, the data is cleaned. For this, all missing values corresponding to weekends are eliminated. Then, the logarithmic first difference is computed, which is a standard transformation in finance to make the series stationary. However, for the series which are already in percentage terms like spreads or yields, the first difference in levels is computed instead. There are still some missing values, which are imputed using the same EM-PCA algorithm used to input missing economic data. The resulting data set has close to six thousand observations of daily financial data from January 2000 to December 2022.

4.5 Text-Based Data: Attention Indicator

The text-derived indicator this study uses is one that measures the attention given by different news outlets to a set of topics. The attention metric was created using a Latent Dirichlet Allocation (LDA) on a corpus including news from the Wall Street Journal, Barron's and Smart Money. LDA is an unsupervised algorithm that estimates the topics present in a piece of news (Blei, Ng, & Jordan, 2003). The data set accessed included a daily attention metric for two hundred different topics. The metrics were normalized to have values between 0 and 1. In addition, the log return is taken to obtain stationary series.

4.6 PCA Factors

As Stock and Watson (2016) explain, in a DFM all the variables of the economy are assumed to be driven by a set of latent variables plus some random disturbances. And if these disturbances follow a distribution with a covariance matrix proportional to the identity matrix, then such a model can be estimated by doing a regression where the factors are the (normalized) PCA factors of the variables included in the database. In addition, the PCA factors are a great way of reducing the dimensionality of the data for other models like the neural network models in Section 7.

While this should be done for all the variables of the database, the approach in this thesis is to do it separately for each kind of data (Economic, Financial and Text). The main reason is the need of a frequency alignment of Financial and Text daily data to the monthly frequency to merge it with the Economic data. The choice of frequency alignment (simple averages or a MIDAS model) could influence the way in which the factors are build. It would also affect the way in which the EM-PCA algorithm imputes the missing values. In addition, separate factors are created for both Financial and Text variables, despite being on the same frequency. When PCA was performed on the joint dataset, the obtained factors loaded almost exclusively on variables of either of the two datasets. Therefore, the approach selected is to separate their PCA factors, which simplifies testing the relative ability of financial and text regressors in the model. This ability to test alternative models with and without each one of the types of variables is also another argument for creating separate PCA factors.

The main drawback of doing the factors separately is the potential correlation between factors when created in this way. Fortunately, their correlation is mostly very low (Figure 3). By construction, all PCA factors (economic, financial, text) have 0 Pearson correlation between each other.

Surprisingly, the correlation between Text and Financial indicators is almost 0, too. Financial factors and Economic factors, logically, exhibit some correlation. However, the value is rather low. In addition to Pearson correlation, Kendall correlation is used (Figure 4). Kendall correlation is non parametric and should better capture non-linear relations between the factors. It seems that the correlation is still low, with the exception of some correlation between the Text factors. In conclusion, the correlation introduced by doing the PCA factors separately will not likely be an issue for the models.

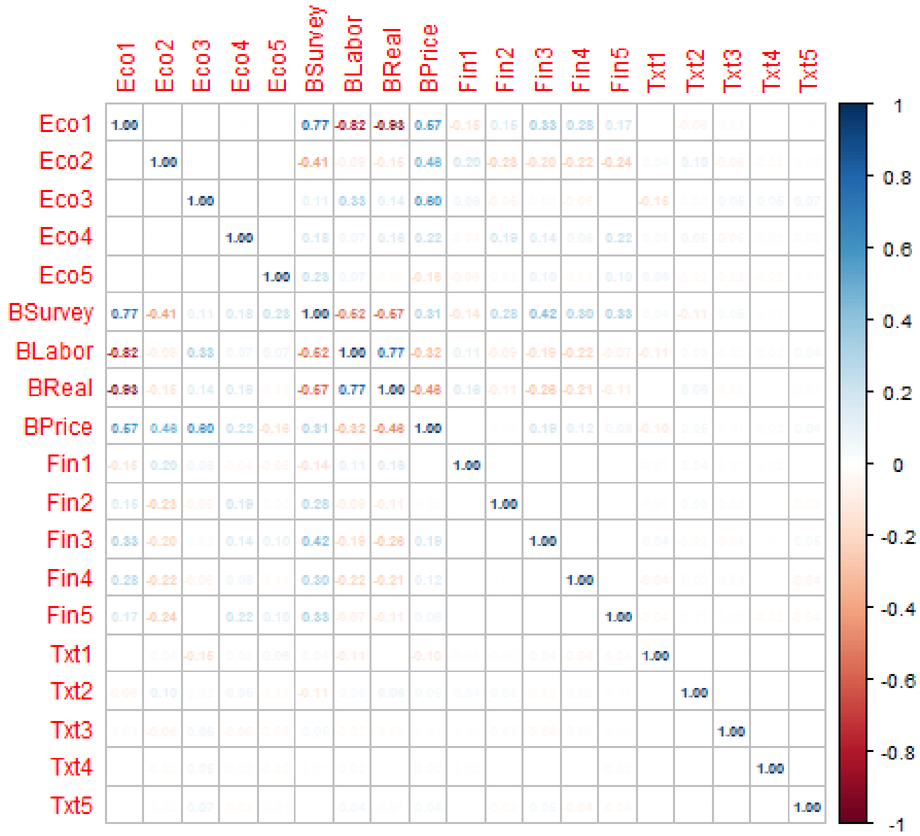

Figure 3: Pearson Correlation Plot of Factors

Note: "Eco", "Fin" and "Txt" stand for the economic, financial and text-based PCA factors, which go from the first one (1) to the fifth (5) for each type. The variables starting by "B" correspond to the four blocks included.

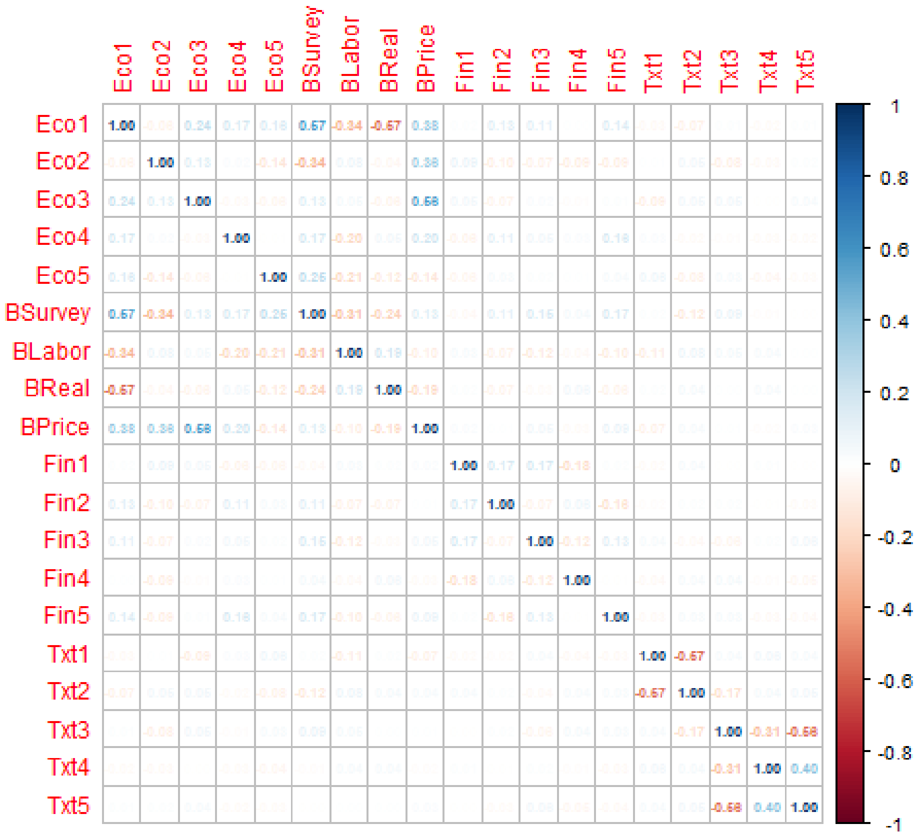

Figure 4: Kendall Correlation Plot of Factors

The correlation is higher however for the “blocks”. These “blocks” correspond to the first PCA factor of datasets that include only economic variable of a certain type. For instance, the “Labor” block is the first PCA factor resulting from the dataset including variables like Non-Farming Payroll Employment and the unemployment rate. These factors are recommended by Bok et al. (2018) to “control for idiosyncrasies in particular subgroups of series”. Specifically, they add three blocks: one for real economic variables (e.g. production, consumption, exports...), one for survey variables (e.g. indexes of confidence) and one for labor market-related variables (e.g. unemployment). This study chooses to include them and test their potential to improve predictions despite their correlation with the economic factors. In addition, a “price” block with information of all the variables related to prices in the economy is added, which could be useful for inflation forecasting. Since the blocks are using the same variables as the economic blocks, it is natural to find high correlations. Additionally, since the way in which they are built does not ensure independence between the blocks, there are also high correlations between some of the blocks.

Another relevant decision is how many dimensions to choose. In other words, how many latent variables are assumed to drive the economy. In Bok et al. (2018) it is recommended to use only the first PCA factor for the economic data set (plus the blocks). In Ardia et al. (2019) the approach is the same, but including also financial information in the dataset. Their approach makes a strong assumption that economic variables are solely determined by a single latent variable (which seems to proxy the general level of economic activity). It is decided to select the first five PCA dimensions for each dataset:

- **Economic variables:** The first economic PCA factor is indeed a proxy for general economic growth, being positively correlated with consumption, production or employment. However, this factor captures only 23% of the total variance in the dataset. The first 5 dimensions manage to capture a more comprehensive 47%. The rest of PCA factors also appear to capture relevant movements in the economy. For instance, factor 3 and 4 are very positively correlated with inflation and inventories and could proxy them[4].

- **Financial variables:** The first 5 PCA factors capture 60% of the total variance, with the first PCA factor capturing 26%. Factor 1 is correlated with interest rate increases, decreases in spreads and increases in the market index. Therefore it seems to also be a proxy for economic/market growth. Factor 2 is correlated with interest rate increases, increases in spreads and decreases in bond values. This could be a proxy of interest rate hikes by the Central Bank. The remaining three factors capture less clear financial market patterns.

- **Text variables:** The first factor captures 9% of the total variance and the first 5 PCA only 30%, out of the 200 topics available. The topics are not labelled therefore it is impossible to interpret the PCA factors.

In summary, the data is processed into 15 factors (5 for each category) plus 4 blocks. The increase in the amount of data captured by the regressors relative to most of the papers using DFMs could be adding relevant information or just noise. Fortunately, this thesis will be able to test it by trying

[4]Appendix A includes a table that briefly describes these factors

the models with only the first Economic PCA Factor and sequentially adding other factors as regressors. This should be yet another valuable contribution of this thesis.

5 Autoregressive Models

There is a significant effort in gathering and processing all the data required for the big data methods included in this benchmark. A very relevant question is whether this is required at all. In other words: can GDP growth and inflation data be forecasted without using any external regressor? Economic variables are obviously the result of many factors. But the interactions could be so difficult to capture that in practice a simple autoregressive model of the variable of interest and its residuals could be a sufficiently good forecasting model. Therefore,the benchmark begins by including the RMSE for an ARIMA model. This will serve as reference to assess whether the rest of the methods, which include extra information, truly result in an increase in forecasting ability.

Since GDP growth and inflation have already been confirmed as stationary, the next step is selecting the appropriate ARIMA model. For this, a stepwise search of the best model is implemented, with the objective to minimize Akaike's Information Criterion. The selected models will be estimated for the whole sample and also for the sample excluding the years between 2020 and 2022 to remove the effect of Covid outliers. In both cases, the samples are separated such that the first 80% of data is selected as training data, and the remaining 20% as test data. The RMSEs for one quarter ahead (a good indicator of short term forecasting ability) and one year ahead (a good indicator of long term forecasting ability) are shown on the tables 2 and 3.

In both this section and in the remaining sections of the paper RMSE is used as the indicator for forecasting quality. This metric is aligned with the preferences of economic agents in need of the forecasts, who are likely to dislike bigger deviations relatively more than they dislike small forecasting errors. In addition, RMSE is the most popular measure of forecasting accuracy in the literature, with most papers cited only reporting it and not other measures. So it is the best metric to facilitate the comparison with other papers.

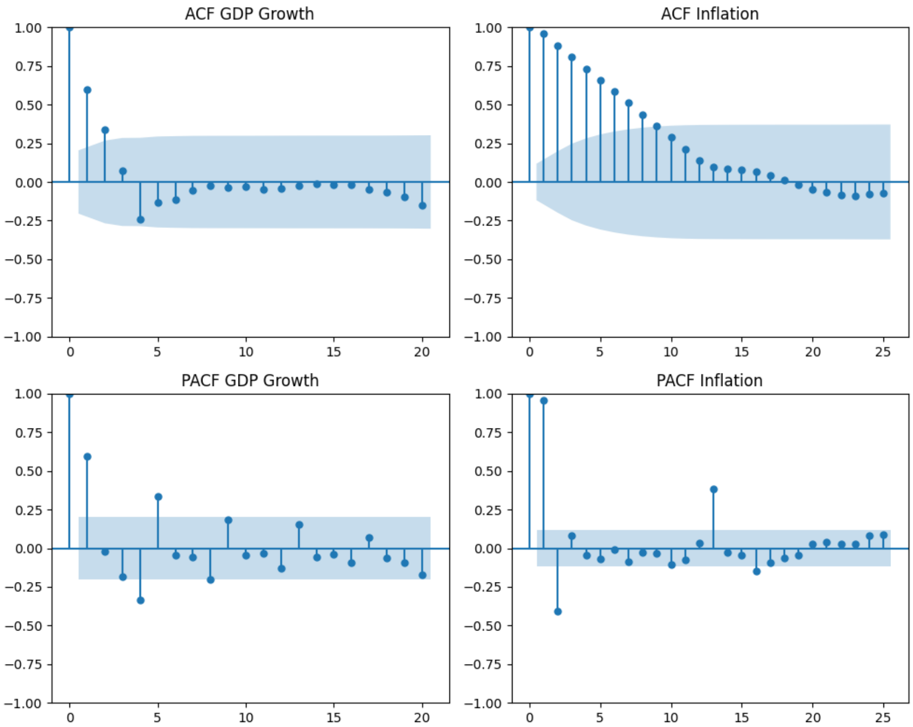

Figure 5: ACF and PACF

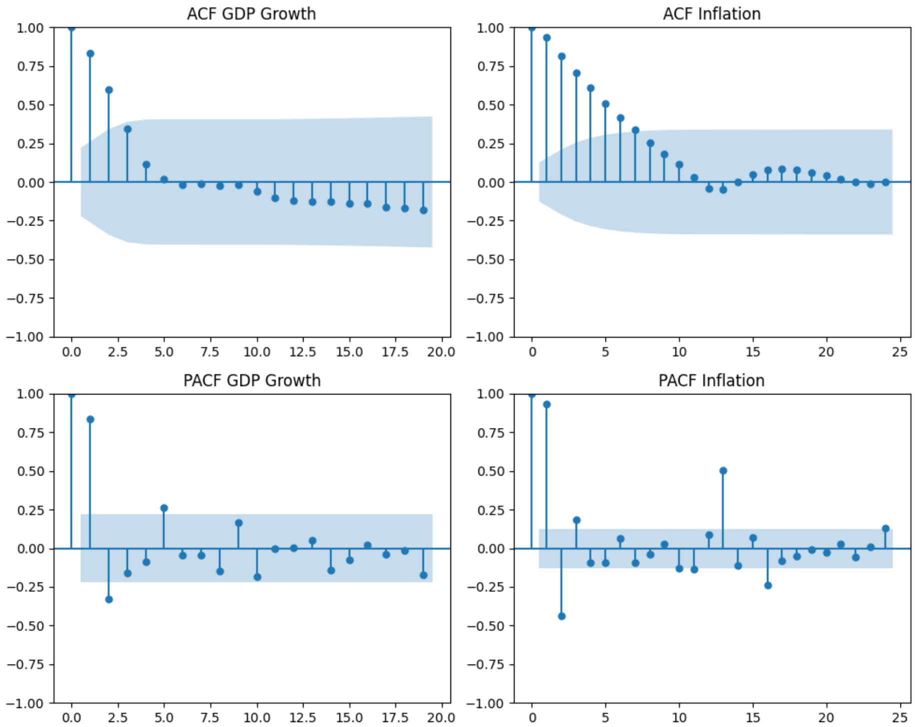

Figure 6: ACF and PACF for samples excluding 2020-2022

In addition to the stepwise procedure, the Autocorrelation (ACF) and Partial Autocorelation (PACF) functions of GDP growth and inflation (Figures 5 and 6) are observed to suggest alternative ARIMA models. The figures show slowly decreasing autocorrelation functions and partial autocorrelations that quickly reach 0. This suggests that the series are autoregressive. Therefore, in addition to the ARIMA models suggested by the stepwise procedure, an AR(5) and AR(13) models are tested for GDP growth and inflation respectively. These lags are quite significant in the PACFs of both series, and correspond to the value of exactly one year before. It makes economic sense that GDP growth and inflation could be affected by previous values up to one year before.

	AR	Auto ARIMA	LSTM
GDP Growth	4.00 AR(5)	3.55 AR(1)	3.53
GDP Growth (ex. Covid)	0.36 AR(5)	0.36 AR(5)	0.46
Inflation	1.14 AR(13)	1.21 ARMA(1,3)	2.64
Inflation (ex. Covid)	0.50 AR(13)	0.41 AR(3)	0.47

Table 2: One Quarter Ahead RMSE (%)

	AR	Auto ARIMA	LSTM
GDP Growth	5.24 AR(5)	5.09 AR(1)	3.73
GDP Growth (ex. Covid)	0.85 AR(5)	0.85 AR(5)	0.64
Inflation	3.81 AR(13)	3.29 ARMA(1,3)	3.30
Inflation (ex. Covid)	0.73 AR(13)	0.42 AR(3)	0.45

Table 3: One Year Ahead RMSE (%)

Finally, considering the possibility that GDP and inflation could be autoregressive in a non-linear way, a LSTM neural network is added. This neural network takes the previous year of information, in a way equivalent to the AR(5) and AR(13) of the table. It is trained in the same way as models in Section 7.

The results point out some key findings. First, it is clear that eliminating the period 2020-2022 results in better predictions. However, these years are not dropped because it is interesting to test which models perform better during periods of high uncertainty. Researchers should expect to get worse forecasting results when testing the models during these periods. Second, simple ARIMA models appear to be better than the LSTM autoregressive model for inflation forecasting, although their RMSEs are close. On the other hand, the LSTM autoregressive model appears to be better in the case of GDP forecasts, with the exception of 1Q ahead forecasts when excluding the 2020-2022 period. In the next sections, it will be made clear that given their simplicity and their relative results, both the ARIMA and the autoregressive LSTM model are among the best and most practical economic forecasting models.

6 Dynamic Factor Models

From the literature review it may be concluded that DFMs are likely the most popular linear forecasting model in economics. It provides a straight forward way to handle data sets with a large number of features.

In this section the alternative specifications of a DFM model to predict GDP growth and inflation are tested. It will be implemented by simply doing a linear regression on the PCA factors created with the selected variables.

6.1 Introduction to DFMs

Equation 1: Dynamic factor model

$$Y_t = f(F_t, F_{t-1}, F_{t-2}, ...) + V_t \quad , \quad V_t \sim N(0, R)$$

$$F_t = g(F_{t-1}, F_{t-2}, ...) + W_t \quad , \quad W_t \sim N(0, Q)$$

DFMs consider the different indicators used as the observed time series of a state space model (Y_t). In this state space model, the state variables that determine all the series are a series of factors that represent latent economic variables (F_t). Often, DFMs will assume functions f and g linear, and consider only one factor and one lag representing the current general economic activity, determining the evolution of all observed series ($Y_t = \beta(F_t^1) + V_t$).

If the variables in the W vector are independent, this model can be estimated by doing a linear regression where the factors are the (normalized) PCA factors of the variables included in the database. (Stock & Watson, 2016). This is the approach taken in this study: first, the PCA factors are created as explained in Section 4. Then, a linear regression of the factors is implemented using basic R functions or the “midasr” R package (Kvedaras & Zemlys-Balevicius, 2022) in the case of using a MIDAS frequency alignment.

6.2 Introducing Mixed Data Sampling

The dataset of regressors includes several variables in a daily frequency, while most economic variables are available in a monthly frequency. Aligning the different frequencies of the data is likely one of the most crucial issues for economic forecasters.

The simplest approach is taking the simple average of all the subperiods to align to the lowest frequency. For instance, forecasters usually take the average of three months to move the data to a quarterly frequency. This is the approach taken by Bok et al. (2018), and the most popular choice.

However, this approach implicitly assumes that all subperiod observations affect the dependent variable with the same intensity. There are economic reasons for departing from this assumption. Consider as an example the monthly inflation data for predicting GDP next quarter. Clearly, inflation expectations affect the prices of negotiated contracts between economic agents and should ultimately have some effect on economic growth. But the most recent unexpected changes in economic indicators tend to have a relatively larger impact on those expectations and their final effect on GDP (Fatas, 2018). Therefore, at least in some cases it is expected that the most recent data is weighted relatively more.

The first alternative is to directly regress the lower frequency dependent variable on the higher frequency regressor. This does not assume any weighting and instead lets the data decide the optimal relative impact of each subperiod. This method, usually called "U-MIDAS" (Foroni et al., 2011) leads to parameter proliferation. Unfortunately, when the frequencies differ considerably (like in the case of quarterly and daily data, for instance) the parameter proliferation leads to a big variance of estimations (Foroni et al., 2011). The optimal choice in this case consists in restricting the weights while preserving some degree of data-driven weighting.

The Mixed Data Sampling (MIDAS) model proposed by Ghysels et al. (2004) can be easily understood as a refinement of the simplest approach consisting in just doing a simple average of the subperiods. In the equations below, y_t represents a dependent variable in a low frequency and the $x_{s(t)}$ represents one regressor in a higher frequency. The chosen direction of alignment is to express the higher frequency variables (like the daily text-based indicator) in the lowest frequency reported (quarterly, for instance).

Equation 2: MIDAS model

$$y_t = g(\sum_{i=1}^{k} w_i x_{s(t)-i}; \beta) + \epsilon_t = \beta^{'}(\sum_{i=1}^{k} w_i x_{s(t)-i}) + \epsilon_t$$

$$= \beta_0 + \beta_1(\sum_{i=1}^{k} w_i x_{s(t)-i}) + \beta_2(\sum_{i=1}^{k} w_i x_{s(t-1)-i}) + ... + \beta_N(\sum_{i=1}^{k} w_i x_{s(t-N+1)-i}) + \epsilon_t$$

$$\sum_{i=1}^{k} w_i = 1 \quad , \quad w_i = h(i, \theta) \quad , \quad s(t) = \sum_{j=1}^{t} m_j$$

The t indicates low frequency lags of y. Meanwhile, x is expressed in a higher frequency: $s(t)$ is the sum of m_j at t, where m_j is the number of high frequency periods each low frequency period j has. Therefore, the model above assumes a linear dependence (a linear g function) of the frequency aligned regressor. Each frequency aligned point of the regressor is multiplied by a corresponding beta parameter. On the other hand, k is the number of high frequency lags considered to aggregate the variable. k may be equal, smaller or bigger than m_j, although in the models tested it is set to be equal.

The alignment is done using a vector of weights w_i of length k. These are the weights by which the k lags considered of the low frequency variable are multiplied to do the alignment. When all the weights are restricted to be equal, then the results are are the same as the ones obtained by doing a simple average of the subperiods. When the weights are unrestricted, the model turns into a U-MIDAS.

Ghysels et al. (2004) propose using a weighting function h that determines each of the w_i according to their lag (i), but reducing the parameter space to only the length of the θ vector. This preserves the flexibility of U-MIDAS but limits the problem of parameter proliferation. While there are many alternatives for h, this thesis uses only the most popular function (the normalized exponential Almon function with two parameters). There is a potential of improving forecasts using other alignment functions. But due to time constraints this is reserved for future research.

Equation 3: Normalized Exponential Almon

$$h(i, \theta) = \frac{exp(\theta_1 \cdot i + \theta_2 \cdot i^2)}{\sum_{i=1}^{k} \exp(\theta_1 \cdot i + \theta_2 \cdot i^2)}$$

Extending a DFM to include a MIDAS frequency alignment requires creating separate PCA factors in the different frequencies as done in Section 4, and then using them as regressors in a MIDAS model. This is one of the methods suggested in Marcellino and Schumacher (2010) who claim to improve forecasts over a standard DFM.

6.3 Competing Models

Considering the comprehensiveness of the benchmark that this thesis is trying to elaborate, and the insights from the literature review, there is a big space of potential DFM models to test:

- The DFM model will be used to forecast GDP and inflation.
- Additionally both short term (one quarter ahead) and long term (one year ahead) forecasting ability should be assessed.
- Given the clear impact of the uncertainty in the period 2020-2022, this study will test DFMs for both the sample that includes and the sample that excludes this period. This will allow to test which forecasting models performs the best under economic stability, uncertainty.
- Bok et al. (2018) use the blocks described in Section 4. This may or may not work in the more complex DFMs that this thesis tests. Therefore both the option of including and excluding the blocks is tested.
- The benchmark also includes the variations with the average and the MIDAS frequency alignment for the DFM.
- DFMs typically use only 1 PCA factor, but as explained in Section 4, more PCA factors could result in better forecasts. Ideally, regularization methods like elastic net or lasso would be used to avoid excessive variance when including many regressors. Since this is not possible when using the "midasr" R package, a more rudimentary approach is chosen. The models are tested using either only one PCA factor all

the first five instead. The version with only one PCA factor for each type of variable (economic, financial and text) will serve as a sort of regularized version of the model.

- Finally, to test the forecasting power of each type of variables (especially the text-based indicator) four kinds of model are tested: one with only economic factors, one with economic and financial factors, one with economic and text factors, and a final one with all the types of factors. Comparing the performance of each type, the thesis will make some conclusions about the value of including other type of regressors in economic forecasting.

All the combinations result in 256 models. A significantly comprehensive benchmark.

6.4 Model Training

In a similar way to how other models in this thesis are trained, the data is separated into a train and a test sample. The first 80% of observations is the train sample and the remaining 20% will be used as a test sample to compute the RMSE.

The models are estimated using basic R functions for linear regression and also the "midas_r" and "average_forecast" functions of the "midasr" R package to do the MIDAS regressions (Kvedaras & Zemlys-Balevicius, 2022). The RMSEs are computed using the a rolling window approach. That means that the train sample is used to estimate only the first observation of the test sample. After this is done, the train sample drops its first observation at the beginning and appends the first observation of the test sample at its end. Then, the second observation of the test sample is predicted with a model re-estimated on the updated train sample. Forecasting one period ahead is straight forward using this method.

When the forecast is for a further horizon, there is the possibility of assuming the predictions as true data and iterating the predictions forward until obtaining the prediction for the horizon of interest. Nonetheless, since the data set of regressors contains variables that have "forward looking nature" (like for instance stock prices, which capture information about expectations of the future), this study decides instead to directly set the future period to be predicted as the dependent variable. In other words: $Y_{t+k} = f(F_t)$. Therefore, when predicting one quarter ahead, the model directly uses the y corresponding to the next quarter, or the data 3 months ahead as dependent variable. When predicting one year ahead, it does the same with the y corresponding to 4 quarters ahead.

The MIDAS models in the benchmark use exclusively the normalized exponential Almon (nealmon) function. The function weights are optimized using the "optim" algorithm in R. The two weights of the nealmon function are initialized to the zero vector, which corresponds to a simple average frequency alignment. In a few models, the optimization fails to converge during the rolling window forecast. In some of the models, this problem is easily solved by initializing the weights to the values obtained in the initial model estimation.

6.5 DFM Results

The DFM models are assessed according to their RMSE for the test sample. In Appendix B, the complete set of results can be observed. In this section, the main observations and the results of the best DFM models are presented.

Looking at the table of RMSEs (Tables 4-8), this thesis finds that, in general, DFM models are not able to beat ARIMA forecasts for inflation, while they were able to do so for some models of GDP. This observation remarks the power of simple ARIMA models to predict inflation, already noted in Section 5.

A notable exception comes when predicting inflation one year ahead in the sample including the years of the pandemic. As observed in Table 8, all the best performing DFMs achieved values well below ARIMA and the autoregressive LSTM model. Notoriously, all these models use additional regressors besides economic variables, and do also use the MIDAS frequency alignment.

Note: In the tables 4-8, "E" stands for Economic Factors. "F" for Financial Factors and "T" for Text-based Factors. The suffix 1 means that only the first PCA factor was used, while 5 means that the first five were used instead. "B" indicates that the DFM used the Blocks described in Section 4, while

"NB" indicates that they were not used. Finally, "EW" means that the frequency alignment was done using the average (equal weighting), while "MIDAS" indicates that MIDAS was used. E1 B EW corresponds to the model used in (Bok et al., 2018). Below, the alternative without blocks is also compared.

	RMSE (%)
E1 NB EW	3.00
E1T1 NB EW	3.00
E5F5 NB MIDAS	3.02
E1 B EW	3.00
E1 NB EW	3.26

Table 4: Best Models to forecast One Quarter Ahead GDP Growth (testing on 2020-2022). In addition, Bok model with and without blocks.

	RMSE (%)
E1F1T1 NB EW	0.49
E1F1 NB EW	0.50
E1F1 B EW	0.50
E1F1T1 B EW	0.50
E5F5T5 B EW	0.50
E1 B EW	0.53
E1 NB EW	0.53

Table 5: Best Models to forecast One Quarter Ahead GDP Growth (dropping 2020-2022). In addition, Bok model with and without blocks.

	RMSE (%)
E5F5 NB EW	2.90
E5 NB EW	2.92
E5F5T5 NB EW	2.92
E1 B EW	3.58
E1 NB EW	3.85

Table 6: Best Models to forecast One Year Ahead GDP Growth (testing on 2020-2022). In addition, Bok model with and without blocks.

	RMSE (%)
E1 B EW	0.62
E1 NB EW	0.64
E1 NB MIDAS	0.64
E5F5T5 NB EW	0.64
E1 B EW	0.62
E1 NB EW	0.64

Table 7: Best Models to forecast One Year Ahead GDP Growth (dropping 2020-2022). In addition, Bok model with and without blocks.

	RMSE (%)
E1F1 B MIDAS	2.91
E1T1 B MIDAS	2.94
E1F1T1 B MIDAS	2.96
E5F5T5 B MIDAS	2.97
E1 B EW	3.12
E1 NB EW	3.20

Table 8: Best Models to forecast One Year Ahead CPI Inflation (testing on 2020-2022). In addition, Bok model with and without blocks.

Regarding GDP forecasts, except for long term predictions in uncertain times, the model by Bok is a very good model. It is always either among the best 3 forecasters, or close to them. Additionally, the simpler version of the Bok model that excludes blocks does not increase the RMSE excessively. This is good news for forecasters: a relatively simple model is consistently among the best performers.

However, it seems that the model with 5 PCA factors achieves a much better result in the 2020-2022 sample for one year ahead predictions. This confirms the hypothesis that including additional PCA factors can lead to better predictions. The absolute best result is achieved by the model also including 5 Financial PCA factors. However the improvement is minimal. In practice, it does not seem necessary to add non-economic variables.

The table of results in Appendix B is very extense and it does not show clear patterns. Apparently, whether some refinements (like adding more PCA factors, doing MIDAS frequency alignment, adding blocks or using non economic regressors) improve forecasts or not are highly dependent on the forecasting horizon and the uncertainty of the period. This highlights the value of doing a benchmark like the one elaborated in this section. It also adds more merit to the model by Bok, which is able to perform well across most settings.

In general, it can observed that using a MIDAS frequency alignment over an equally weighted (EW) alignment tends to worsen the results for GDP forecasting, while it improves them for inflation forecasting. GDP MIDAS DFMs are better than their EW counterparts only around 30% of the time, while the figure is closer to 75% for inflation DFMs. Also, adding text factors only improves RMSEs in around 40% of the models both for GDP and inflation forecasting. However, it is important to emphasize that this is highly dependent on the specific prediction target. It is possible that using MIDAS alignment and text-based indicators in a DFM could improve the results in some settings not appearing on this benchmark.

7 Neural Network Models

Neural network models for economic forecasting have gained popularity over the last few years. While neural networks are more complex and convergence to the global optimum value of the parameters is not ensured, they are able to capture non-linearities in the data.

It is clear that there are non-linear relations between economic variables. For instance, the effect of net public spending on GDP (or fiscal multiplier), one of the most studied relationships in economics, appears to depend on many other variables like whether the changes come from taxes or expenditure savings, the phase of the business cycle and the level of interest rate as Alesina, Azzalini, Favero, Giavazzi, and Miano (2018) find. While they use a linear model, a more scalable way of considering these non-linearities for all variables is using neural network models.

The explanation of the different neural network architectures would cause this thesis to drift away from its core research idea. Instead, this section will focus on the refinements added to the popular architectures used (feed forward, LSTM, GRU and Multi-Head Attention). In case the reader might need more context, the handbook by Goodfellow, Bengio, and Courville (2016) is a well explained source for the first three architectures. Additionally, the original paper by Vaswani et al. (2017) is the perfect source to understand the way in which self attention layers work.

7.1 Frequency Alignment with MIDAS Layers

Frequency alignment is also a necessary step in non-linear models. In this study, I align the data including a processing layer at the beginning of all the neural networks.

The input to the models is a tuple of two matrices corresponding to a year of observations of the regressors: one 12x5 matrix containing 12 lags of the five monthly economic factors and another 259x10 matrix containing 259 lags of five financial and five text-based factors. The frequency alignment layer seeks to convert those two matrices into a single input matrix used to predict the target variable. The alignment is done to the quartertly frequency. Therefore the processed output will be a 4x15 layer (4 quarters for 15 factors in total).

The first approach to align frequencies is to take quarterly averages for each of the data sets. The operation transforms the 259x10 matrix into a 4x10 matrix, and the 12x5 matrix of economic variables into a 4x5 matrix. These two are concatenated by columns to obtain the desired 4x15 input matrix. This approach is labelled "Equally Weighted Frequency Alignment (EW)" in this benchmark. It is implemented in with TensorFlow by means of a convolution operation that restricts all weights to be equal, like in a simple average. The code creates a custom convolution layer very similar to the standard Conv2D TensorFlow layer, which implements this operation and the specified frequency alignment.

The alternative approach suggested by the literature was using MIDAS to align the frequencies. To achieve a training process equivalent to the one mathematically detailed by Xu et al. (2019), another custom convolutional layer is created. Using a TensorFlow convolution layer, the weights are determined to be equal to those of a normalized exponential Almon function, with two trainable parameters. These parameters will be jointly trained with the rest of the model. As far as known while writing this thesis, this is the only publicly available implementation of a neural network with a MIDAS frequency alignment that is jointly trained with the rest of the neural network. Being implemented on TensorFlow in Python, the most popular neural network programming framework, this adds to the value of this benchmark for forecasters.

In total, two different layers are applied to each of the MIDAS neural networks, one for each type of input. The data set of daily variables is passed through a "daily to quarterly" MIDAS layer that transforms it into a 4x10 matrix. Similarly, the monthly data set is passed through another equivalent layer that transforms it into a 4x5 frequency aligned input. Again, the inputs are concatenated to obtain the desired 4x15 shape. Figure 7 below graphically describes the process.

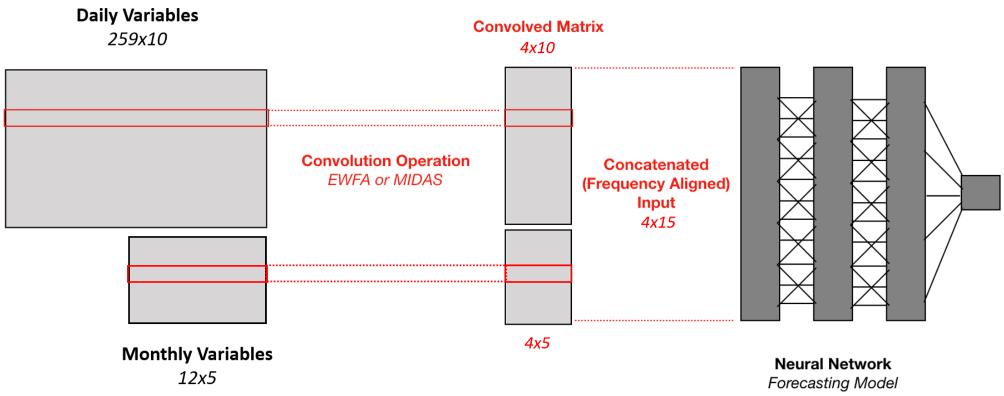

Figure 7: The matrix of daily variables is convolved to obtain a matrix in the monthly frequency. The convolution can use equal weighting (EWFA) or MIDAS weights. The convolved output is then concatenated with the matrix of monthly variables to obtain the frequency aligned input matrix.

7.2 Dimensionality Reduction

The immediate approach to create a neural network model that uses a big data set for economic forecasting could have been to directly include the individual variables as regressors. Theoretically, a neural network, if well optimized, would select weights that disregard the effect of variables that only add noise. However, in practice, the ability of neural networks to do feature selection is imperfect. For the case of this data set, the direct use of individual variables resulted in a lot of noise and very bad predictions.

Since the number of observation is very limited, and the architectures are rather complex relative to the number of observations that are available, the decision in this thesis is not to do any trainable dimensionality reduction (like, for instance, an encoder model). Instead, the same PCA factors that were used for the DFMs are used. Fortunately, this approach to reducing the dimension of the dataset does achieve reasonable forecasting results.

7.3 Competing Models

Ideally, the benchmark presented would include a set of models as big as the one presented for DFMs. However, it takes much longer to compute the RMSEs of neural networks: the optimization takes longer, and each model requires a grid search of hyperparameters (which implies re-fitting the same network over and over again with different hyperaparameters).

Due to time constraints, this thesis has chosen to consider only models in which 15 PCA factors (five for each type of variable) are included as regressors. Additionally, the models are only tested for samples including the 2020-2022 period of the Covid pandemic. According to Hopp (2022), neural network models appear to be better predictors for periods of high uncertainty like the Covid pandemic. Knowing this, and the relatively good results simpler models achieved in prior sections for the periods excluding Covid, this thesis opted to study the period in which neural networks can contribute the most to improve forecasting results. Testing other sample periods and different combinations of regressors as inputs is left for future research.

The different models consider the following features:

- First, all models are tested for GDP growth and inflation forecasting, following one of the most important requisites specified to consider this benchmark as comprehensive.
- Second, both short term (one quarter ahead) and long term (one year ahead) forecasting ability are tested.
- Third, four different neural network architectures are tried: feed-forward, LSTM, GRU and Multi-Head (or self) Attention.
- Finally, each neural network considers an equal weighting (EW) and a MIDAS frequency alignment layer.

In total, 64 models are tested.

7.4 Training

To train the models, the data is first processed so that every observation of the target variable is matched with a year's worth of lags of the predictors. For example, the one quarter ahead predictors of the inflation in March 2006 (Q1) are all the values within the first and last quarter of 2005 (both included). In addition, all variables are normalized by subtracting the minimum and dividing by their range, in order for all values to lie between zero and one.

Then, the observations are divided into three subsets: a training set, a validation set and a test set. The different sets represent 68%, 12% and 20% of the observations, respectively. The split is performed selecting the first 80% of observations as input for the model. During the training, the model fitting algorithm separates 15% of that input as validation set, which will be used to evaluate the prediction error of several hyper-parameters of the models and tune them. The last 20% of observations is used as test set. The prediction error of the tuned and trained model on that test set is the value reported in the result tables.

The hyper-parameters of the models that are tuned are three: the number of nodes (output dimension) of the hidden layers, the number of hidden layers and the percentage of dropout. They are all tried in a grid search procedure repeated for each different model. The number of nodes is chosen from one of the following quantities: (10,50,100). It is not appropriate to consider greater numbers as the ratio of trainable weights to training data observations would be too high. Next, the number of input layers is selected from (1,2,3). Again, the number of layers is kept very small to avoid the proliferation of parameters. Thirdly, the dropout in the models can be any of the following: (0%, 10%). Dropout is the most common technique in Machine Learning to increase the robustness of the predictions, and is used as a way to regularize the model.

The activation function used in each layer could be considered another "hyperparameter" that can be optimally chosen. However, for simplicity, this is not tuned in the grid search. Instead, the rectified linear unit (RelU) activation function is used in all hidden layers. This is the most common choice for neural networks due to its computational efficiency (their gradient is the one of a linear function, or zero). Only in the final output layer a different activation function is used: a sigmoid in this case. Since the dependent variable has been normalized to range between 0 and 1, sigmoid is an appropriate choice, which is also popularly used for the final output layer of neural networks.

The models are implemented using the TensorFlow (Version 2.15) package in Python, which also integrates the popular Keras framework. The combination of Python and TensorFlow is one of the most popular (if not the first) approaches to implement Machine Learning models. Therefore it is the one used to elaborate this benchmark, to maximize the usability for future researchers.

The model weights are optimized using the Adaptive Momentum Estimation Algorithm (Adam) of Kingma and Ba (2014). It is a numeric optimization method like stochastic gradient descent. But it updates the weights considering a moving average and variance of the current and past gradients. This gives the optimization process "momentum" which contributes to Adam being able to achieve better results relative to other popular optimization algorithms. (Kingma & Ba, 2014). The chosen learning rate is a tenth of the standard value: 10^{-4}. This relatively smaller learning rate required of more iterations to converge to minima. Therefore the limit number of epochs (complete passes through the entire data set) is increased to 500. The optimization of the models appeared to work the best when increasing the number of epochs and reducing the learning rate, relative to standard values. Lastly, the initialization for the algorithm is random, except for the MIDAS parameters, which are initialized to the zero vector (equally weighting scheme).

In training the models, it is crucial to avoid overfitting the data. Especially when having such a high limit for the number of epochs in the training of the model. To do so, a callback is implemented in the code, which stops training the models when the validation value of the validation error has not decreased after several epochs. While the training set error may continue decreasing, the code prevents the weights being updated after the error in the validation set stops improving. In addition, the callback also makes sure that the weights which lead to the smallest validation error are restored.

The small number of available observations and the small learning rate used inevitably lead to the need of passing through the data several times. This thesis considers the callback implemented is a reasonable enough safeguard to prevent overfitting.

7.5 Neural Network Results

The results for all the neural networks in Table 9 point out to a rather limited capacity of these non linear models to beat the previously presented models. Neural network models appear to only be generally the best model when doing long term forecasts of inflation in the 2020-2022 sample. For all other prediction targets, either the autoregressive models or DFMs achieved better results.

There is also considerable heterogeneity in the results, however there are two patterns worth mentioning. First, MIDAS versions of the neural networks tend to obtain lower RMSEs than their equally weighted counterparts (60% of the time). Second, the GRU MIDAS architecture is arguably the best one. It achieves the best forecast for long term inflation in the covid sample, and a very competitive forecast for one quarter ahead GDP growth in the covid sample. In also tends to be the best or among the best architectures of all the tested ones.

	GDP 1Q	**GDP 1Y**	**Inf. 1Q**	**Inf. 1Y**
ANN (EW)	3.38	3.49	2.14	3.27
ANN (MIDAS)	3.31	3.60	2.14	2.97
LSTM (EW)	3.45	3.48	2.52	3.12
LSTM (MIDAS)	3.43	3.60	2.35	2.71
GRU (EW)	3.26	**3.41**	2.47	2.77
GRU (MIDAS)	**3.07**	**3.47**	**2.39**	**2.54**
SA (EW)	3.49	**3.39**	2.55	2.69
SA (MIDAS)	3.33	3.48	2.60	2.96

Table 9: RMSEs (%) for all Neural Network architectures of the benchmark. "1Q" indicates one quarter ahead predictions. "1Y", stands for one year ahead predictions.

8 Conclusion

To present the conlcusions of the study, this section begins with a table for each of the forecasting targets treated in the benchmark (Tables 10-17). Each table includes a selection of the most relevant models: the best ARIMA, non linear autoregressive, DFM and neural network model. Then, the main insights and contributions of this study are listed. Finally, the main limitations of this study are explained, with recommendations for future expansion, improvement of this benchmark.

8.1 Tables of selected models

Note: In the tables, "E" stands for Economic Factors in the DFM models. "F" for Financial Factors and "T" for Text-based Factors. The suffix 1 means that only the first PCA factor was used, while 5 means that the first five were used instead. "B" indicates that the DFM used the Blocks described in Section 4, while "NB" indicates that they were not used. Finally, "EW" means that the frequency alignment was done using the average (equal weighting), while "MIDAS" indicates that MIDAS was used. "NN" stands for "Neural Network".

	RMSE (%)
ARIMA	3.55
Autoreg. LSTM	3.53
Bok Model	**3.00**
Bok Model (NB)	3.26
E1T1 NB EW	**3.00**
NN GRU MIDAS	3.07

Table 10: Most relevant models: 1Q ahead GDP Growth (2020-2022)

	RMSE (%)
ARIMA	**0.36**
Autoreg. LSTM	0.46
Bok Model	0.53
Bok Model (NB)	0.53
E1F1T1 NB EW	0.49

Table 11: Most relevant models: 1Q ahead GDP Growth (ex. 2020-2022)

	RMSE (%)
ARIMA	5.09
Autoreg. LSTM	3.73
Bok Model	3.58
Bok Model (NB)	3.85
E5F5 NB EW	**2.90**
E5 NB EW	**2.92**
NN SA EW	3.39

Table 12: Most relevant models: 1Y ahead GDP Growth (2020-2022)

	RMSE (%)
ARIMA	0.85
Autoreg. LSTM	**0.64**
Bok Model	**0.62**
Bok Model (NB)	**0.64**

Table 13: Most relevant models: 1Y ahead GDP Growth (ex. 2020-2022)

	RMSE (%)
ARIMA	**1.14**
Autoreg. LSTM	2.64
Bok Model	3.21
Bok Model (NB)	3.14
E5F5 NB MIDAS	2.82
NN ANN MIDAS	2.17

Table 14: Most relevant models: 1Q ahead inflation (2020-2022)

	RMSE (%)
ARIMA	**0.41**
Autoreg. LSTM	0.47
Bok Model	0.62
Bok Model (NB)	0.64
E1T1 B MIDAS	0.58

Table 15: Most relevant models: 1Q ahead Inflation Growth (ex. 2020-2022)

	RMSE (%)
ARIMA	3.29
Autoreg. LSTM	3.30
Bok Model	3.12
Bok Model (NB)	3.20
E1F1 B MIDAS	2.91
NN GRU MIDAS	**2.54**

Table 16: Most relevant models: 1Y ahead Inflation (2020-2022)

	RMSE (%)
ARIMA	**0.42**
Autoreg. LSTM	0.45
Bok Model	0.74
Bok Model (NB)	0.64
E1F1 NB MIDAS	0.58
E1T1 NB MIDAS	0.58

Table 17: Most relevant models: 1Y ahead Inflation (ex. 2020-2022)

8.2 Main insights

In this study a comprehensive benchmark has been created, including an extense selection of the best performing economic forecasting models. The consistent comparison of these models allows to derive some insights.

First, the DFM model proposed by (Bok et al., 2018) is not only a relatively easy to implement model (only economic variables, without a sophisticated frequency alignment and only one PCA factor) but also the best performing model overall. As it can be observed in the tables, it is a competitive model, if not the best, for predicting GDP in all settings except for long term forecasts of GDP in uncertain periods like 2020-2022. On the other hand, an even simpler version of this model, one that drops the so called "blocks", is still a competitive model choice. Otherwise, when the sample does not include uncertain periods like 2020-2022, an even simpler ARIMA model is also very well performing. It even obtains the lowest RMSE for one quarter ahead forecasting.

In addition, inflation also appears to be best predicted in practice with a simple ARIMA model. The only exception is again the case of long term forecasts when testing on the 2020-2022 period. Considering the effort associated to processing and dealing with data sets of several regressors, the accuracy shown by these simple models is good news for forecasters.

However, this thesis has also shown that more complex models may be needed when predicting economic variables one year ahead, during periods of high volatility. When forecasters find themselves in a context like the Covid pandemic, the benchmark suggests more complex models give more reliable forecasts. For GDP growth forecasting, considering more economic factors in the DFM appears to be the best choice. In the case of inflation, most of the tested neural networks architectures achieve better results, with the GRU MIDAS architecture being the best performer.

Another important aspect of this thesis is the use of an alternative text based indicator. Appendix B shows it is not possible to claim that the use of the text-based factors results in better forecasts in general. However, the RMSEs different model specifications obtain appear to depend on the specific target of predictions. Therefore, this thesis does not refute the potential of text indicators for improving forecasts. Indeed, some of the best performing models in the benchmark included text factors. Although it seems that in most cases, models improve little by adding them, and many times they obtain similar results to models that add only financial variables instead.

Finally, MIDAS extensions to the models, as a more sophisticated approach to frequency alignment, have been tested. Inflation DFMs and neural networks do seem to benefit from this refinement. However, that is not the case when predicting GDP growth with DFMs.

In conclusion, this study obtains clear recommendations for economic forecasters. They should start their forecasting exercises by using simple autoregressive models for both GDP and inflation, or may also use the simple DFM by Bok et al. (2018) for GDP forecasting. During uncertain economic periods, they may benefit from using a GRU MIDAS neural network for inflation forecasting, or extend their DFMs with more than one PCA economic factor for GDP growth forecasting.

In addition to giving support for these convenient recommendations, this benchmark is also valuable due to the guidance it provides for economists trying to improve these results: they can easily see which models have worked better than others when tested in the same conditions. Moreover, the code and most of the data used are readily available on a public repository for future researchers to use and improve upon.

Despite some limitations of this benchmark (explained in the next section), this thesis considers the first version here presented satisfactorily achieves the proposed objective of elaborating a comprehensive and useful initial reference for economic forecasters.

8.3 Limitations of this study

To complete this thesis, and maximize the usefulness for future researchers, the reader should also be aware of the limitations of this study and the resulting benchmark. Outlying these limitations will help clarify the ways in which the benchmark can be completed and improved by future research.

An initial and clear limitation is the number of models that have not been included in the benchmark, despite being popular in forecasting and potentially being able to achieve lower RMSEs. For instance, BVARs have been described in the literature, and are quite popular among economists. Their slightly higher complexity compared to DFMs, and the approximately equivalent results of the two types of models (De Mol et al., 2008) justified their exclusion from the benchmark. Yet, they may lead to better results under some settings.

Similarly, the set of non linear models is relatively limited. While the study does test all of the popular architectures as of 2023, it does not include any non linear method that is not a neural network. Tree based methods are very popular, and many times they perform better when data is scarce, like in this case. Despite the poor results found by (Hopp, 2022) for some tree based methods, future researchers are encouraged to consider alternative non-linear methods. There exists a possibility that simpler non linear methods can yield sufficiently good predictions.

Another way of expanding the benchmark is through testing more variations of the existing architectures. By trying new variations, researchers may find a model that better extracts the signals of the data to succesfully predict some of the targets. It is recommended to start by testing more neural network architectures variations. Starting with testing different sets of regressors, future researchers may find that dropping some of the factors the models improve RMSEs, as happened in the case of the DFMs presented in this study. Due to time constraints this thesis could not test all the possible combinations of inputs (Economic, Financial and Text variables) and their resulting RMSEs. Neural Networks did only improve RMSEs for a specific target of inflation. However, it could be that when excluding all variables except the economic factors (for instance), the RMSEs improve, as happened with most DFMs. In addition, exploring a bigger hyperparameter space and also other configurations of the included architectures might result in improvements in RMSEs.

Besides trying different combinations of factors, future researchers may also consider including different regressors and ways to reduce the dimensionality of the dataset. The selection of regressors was carefully done attending to economic theory and the existing literature, as explained in Section 4. Nevertheless, it is difficult to know when the selected set of regressors will result in the lowest noise-to-signal ratio. Future researchers are encouraged to add, drop variables from the selected data set.

Especially, it is recommended to start by dropping the text based factors, or at least substituting them by other text indicators which have more observations. In this study, the chosen windows for training the neural networks and the number of layers, nodes included was limited by a small number of observations that could only use data after the year 2000. This was done due to prioritizing a prediction setting similar to that some researchers with limited data for alternative indicators may find.

But researchers with different priorities could for example select only those regressors available since the 1970s (many economic and financial indicators are) and would duplicate the observations while reducing the number of features. This allows to enrich the network architectures, improve their learning and ultimately achieve better results. One important architecture refinement is the possibility of using encoding layers, jointly trained with the rest of the model, to reduce the dimensions in a more data-driven way. As shown by Hauzenberger et al. (2023), this dimensionality reduction technique is the one that apparently leads to the best forecasting results.

The last significant limitation of this benchmark comes from the rather simplistic approach taken for regularization. The "midasr" package used to estimate the MIDAS DFMs did not provide a way to use elastic net regularization techniques. The best alternative found was to use either only a very limited set of regressors (only the first PCA factors) or all of them. To be consistent, this approach is extended to the rest of DFMs. Considering the most relevant DFMs were those for GDP forecasting, which do not tend to benefit from MIDAS, it appears a promising next step could be to use equal weighting and do the linear regressions with an elastic net regularization.

Similarly, dropout was the only regularization technique considered for neural networks. And due to the time-intensiveness of the grid search only two values were included in the grid search (0% and 10%). Future researchers could try other regularizations like different dropout rates, but especially elastic net regularization of the parameters estimated for the neural networks.

This is interesting for two reasons: first, it is easiliy implemented with TensorFlow in Python. Second, it has the potential of doing feature selection (by setting some weights to zero) without the need of trying several combinations (the approach taken in this benchmark).

To summarize, there are three main ways in which this benchmark may be improved. First, including different models. Second, testing more variations (either of the model, or of its inputs) of the included models. Third, using more sophisticated regularization techniques. And more specifically, using L1 or elastic net regularization in non-MIDAS models. Given the straight forward nature of the third option, this thesis suggests starting with this one as the first line of improvement for any interested researcher, while the other two are more suitable for later research.

References

Alesina, A., Azzalini, G., Favero, C., Giavazzi, F., & Miano, A. (2018). Is it the "how" or the "when" that matters in fiscal adjustments? *IMF Economic Review*, *66*, 144–188.

Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, *35*(4), 1370–1386.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, *10*, 615–643.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control.* John Wiley & Sons.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Choi, E., Cho, S., & Kim, D. K. (2020). Power demand forecasting using long short-term memory (lstm) deep-learning model for monitoring energy sustainability. *Sustainability*, *12*(3), 1109.

Christoffel, K. P., Coenen, G., & Warne, A. (2010). Forecasting with dsge models. *ECB Working Paper*.

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*(2), 318–328.

Edge, R. M., Gürkaynak, R. S., Reis, R., & Sims, C. A. (2010). How useful are estimated dsge model forecasts for central bankers?[with comments and discussion]. *Brookings Papers on Economic Activity*, 209–259.

Fatas, A. (2018). Self-fulfilling pessimism: The fiscal policy doom loop. *VoxEU. org*, *28*.

Foroni, C., Marcellino, M. G., & Schumacher, C. (2011). U-midas: Midas regressions with unrestricted lag polynomials. *Bundesbank Discussion Paper Series 1*.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. *CIRANO Working Papers*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. (`http://www.deeplearningbook.org`)

Hauzenberger, N., Huber, F., & Klieber, K. (2023). Real-time inflation forecasting using non-linear dimension reduction techniques. *International Journal of Forecasting*, *39*(2), 901–921.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hopp, D. (2022). Benchmarking econometric and machine learning methodologies in nowcasting. *arXiv preprint arXiv:2205.03318*.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Kidger, P., & Lyons, T. (2020). Universal approximation with deep narrow networks. In *Conference on learning theory* (pp. 2306–2327).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Knotek, E. S., & Zaman, S. (2017). Nowcasting us headline and core inflation. *Journal of Money, Credit and Banking*, *49*(5), 931–968.

Kvedaras, V., & Zemlys-Balevicius, V. (2022). midasr [Computer software manual]. Retrieved from `https://cran.r-project.org/web/packages/midasr/midasr.pdf` (Version 0.8)

Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-rochester conference series on public policy* (Vol. 1, pp. 19–46).

Marcellino, M., & Schumacher, C. (2010). Factor midas for nowcasting and forecasting with ragged-edge data: A model comparison for german gdp. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550.

McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Mueller, P. (2009). Credit spreads and real activity. In *Efa 2008 athens meetings paper.*

Romer, D. (2019a). *Advanced macroeconomics* (5th ed.). McGraw-Hill. (Chapter 7, p. 364)

Romer, D. (2019b). *Advanced macroeconomics* (5th ed.). McGraw-Hill. (Chapter 6)

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of economic literature*, *41*(3), 788–829.

Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415–525). Elsevier.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the ieee international conference on computer vision* (pp. 843–852).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Xu, Q., Liu, S., Jiang, C., & Zhuo, X. (2021). Qrnn-midas: A novel quantile regression neural network for mixed sampling frequency data. *Neurocomputing*, *457*, 84–105.

Xu, Q., Zhuo, X., Jiang, C., & Liu, Y. (2019). An artificial neural network for mixed frequency data. *Expert Systems with Applications*, *118*, 127–139.

A Table of Regressors

This appendix includes a list of all the economic and financial regressors used in this study.

A.1 Economic Indicators

The extense list of economic variables is separated according to the "block" in which they were included. Each variable includes a small description, its source and its transformation between parenthesis.

A.1.1 Real Variables

- Real Gross Domestic Product: sum of value of all goods produced in the economy. Available on FRED as GDPC1. (% increase)
- Manufacturers' New Orders: Durable Goods. Available on FRED as DGORDER. (% increase)
- Advance Retail Sales: Retail Trade and Food Services. Available on FRED as RSAFS. (% increase)
- New One Family Houses Sold: United States. Available on FRED as HSN1F. (% increase).
- New Privately-Owned Housing Units Started: Total Units. Available on FRED as HOUST. (% increase)
- Industrial Production: Total Index. Available on FRED as INDPRO. (% increase)
- Total Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Inventories: a measure of inventories in the economy. Available on FRED as I42IMSM144SCEN. (level change on % increase)
- Total Construction Spending: Total Construction in the United States. Available on FRED as TTLCONS. (level change on % increase)

- New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units. Available on FRED as PERMIT. (% increase)
- Capacity Utilization: Total Index: estimate of total production capacity used, for all industries. Available on FRED as TCU. (no transformation).
- Total Business Inventories. Available on FRED as BUSINV. (level change on % increase)
- Real Personal Consumption Expenditures: measure of aggregate household consumption, adjusted for inflation. Available on FRED as PCEC96 (% increase)
- Manufacturers' Value of Shipments: Durable Goods. Available on FRED as AMDMVS. (% increase)
- Manufacturers' Unfilled Orders: Total Manufacturing. Available on FRED as AMTMUO. (% increase)
- Manufacturers' Total Inventories: Durable Goods. Available on FRED as AMDMTI. (level change on % increase)
- Real gross domestic income: sum of all incomes in the economy. Available on FRED as A261RX1Q020SBEA. (% increase)
- Real Disposable Personal Income: aggregate income available to households for consumption (after taxes, for instance). Available on FRED as DSPIC96. (% increase)
- Exports of Goods and Services, Balance of Payments Basis. Available on FRED as BOPTEXP. (% increase)
- Imports of Goods and Services, Balance of Payments Basis. Available on FRED as BOPTIMP. (% increase)

A.1.2 Survey Variables

- Current General Business Conditions; Diffusion Index for New York: Answers about the General Business Conditions for the State of New York in the Empire State Manufacturing Survey. Available on FRED as GACDISA066MSFRBNY. (no transformation)

- Current General Activity; Diffusion Index for Federal Reserve District 3: Philadelphia. Percentage of answers saying "increase" minus percentage of answers saying "decrease" in questions about General Business Activity in Manufacturing Business Outlook Survey. Available on FRED as GACDFSA066MSFRBPHI. (no transformation)
- University of Michigan: Inflation Expectation. Available on FRED as MICH. (level difference)
- University of Michigan: Consumer Sentiment. Available on FRED as UMCSENT. (level difference)
- Manufacturing Composite Indicator of Confidence. Available on OECD.stat. (level difference)
- Non manufacturing Bussiness Situation Current Activity. Available on OECD.stat. (no tranformation)
- Non manufacturing Raw Materials Stocks Tendency. Available on OECD.stat. (level difference)
- Non manufacturing Employment Tendency. Available on OECD.stat (level difference)
- Non manufacturing Orders Intentions Tendency. Available on OECD.stat. (no transformation)

A.1.3 Labor Variables

- All Employers Non Farm: number of workers in the economy. Available on FRED as PAYEMS. (% increase).
- Unemployment Rate. Percentage of the labor force (excluding military personnel) looking for employment. Available on FRED as UNRATE. (level difference)
- Nonfarm Business Sector: Unit Labor Costs for All Workers: index of unit labor costs (labor costs divided by productivity of workers). Available on FRED as ULCNFB. (% increase).
- Job Openings: Total Nonfarm. Available on FRED as JTSJOL. (% increase)

A.1.4 Price Variables

- Consumer Price Index for All Urban Consumers: All Items in U.S. City Average: index of cost of a basket of goods that represents average consumption of urban consumers. Available on FRED as CPIAUCSL (level difference of % increase)
- Producer Price Index by Commodity: Final Demand: measures de price level that producers pay for inputs. Available on FRED as PPIFIS. (level difference of % increase)
- Import Price Index (End Use): All Commodities: price level of non-military goods traded between the US and other countries. Available on FRED as IR. (% increase)
- Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index). Available on FRED as PCEPILFE. (level difference of % increase)
- Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average: also known as "core inflation". Available on FRED as CPILFESL. (level difference of % increase)
- Nonfarm Business Sector: Unit Labor Costs for All Workers. Available on FRED as ULCNFB. (% increase)
- Personal Consumption Expenditures: Chain-type Price Index. Available on FRED as PCEPI. (level difference of % increase)
- Export Price Index (End Use): All Commodities. Available on FRED as IQ. (% increase)
- University of Michigan: Inflation Expectation. Available on FRED as MICH. (level difference)
- Monetary Base; Total. Total currency in circulation, plus Federal Reserve balances. Available on FRED as BOGMBASE. (% increase)
- M1: Monetary Base, plus money available in several types of bank current deposits. Available on FRED as M1SL. (% increase)
- M2: M1, plus money available in several types of savings accounts. Available on FRED as M2SL. (% increase)

- S&P/Case-Shiller U.S. National Home Price Index: popular index of house prices. Available on FRED as CSUSHPINSA. (% level difference of % increase)

- Federal Funds Effective Rate: interest rate on overnight loans. Main interest rate targeted by the monetary policy of the Federal Reserve. Available on FRED as FEDFUNDS. (level difference).

A.1.5 Brief Description of PCA Factors

	% Variance Explained	**Most correlated variables**
PCA Eco. 1	23%	In general, positively correlated with all variables of production, consumption and expectations of growth
PCA Eco. 2	10%	Price indexes (+) Retail Sales (+) M1 (+) Production (-) Future orders and producers' sentiment (-)
PCA Eco. 3	7%	Several price indexes (+) Inflation expectations (+)
PCA Eco. 4	5%	Several types of inventories (+)
PCA Eco. 5	4%	Did not show a high correlation with any variable

Table 18: Economic Variables PCA Factors

*Note: "+" stands for positive and "-" negative correlation.

A.2 Financial Indicators

The financial variables include (1) several points of the yield curve (several interest rates), (2) all available credit spreads at FRED, (3) the spreads between inflation protected securities and their standard counterparts, (4) exchange rates of the currencies of the main international trade partners of the US and (5) a stock market index for main asset classes (stocks, investment grade bonds, high yield bonds, commodities. For commodities I also add the price of gold and crude oil, due to their popularity and economic relevance).

A.2.1 Variables

- Market Yield on U.S. Treasury Securities at 1-Month Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS1MO. (level difference)
- Market Yield on U.S. Treasury Securities at 3-Month Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS3MO. (level difference)
- Market Yield on U.S. Treasury Securities at 6-Month Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS6MO. (level difference)
- Market Yield on U.S. Treasury Securities at 1-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS1. (level difference)
- Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS2. (level difference)
- Market Yield on U.S. Treasury Securities at 5-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS5. (level difference)
- Market Yield on U.S. Treasury Securities at 7-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS7. (level difference)

- Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS10. (level difference)
- Market Yield on U.S. Treasury Securities at 20-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS20. (level difference)
- Market Yield on U.S. Treasury Securities at 30-Year Constant Maturity, Quoted on an Investment Basis. Available on FRED as DGS30. (level difference)
- Secured Overnight Financing Rate. Available on FRED as SOFR. (level difference)
- Spread between inflation indexed and non inflation indexed bond - 5 years. Both indicators available at FRED as DFII5 and DGS5. (level difference)
- Spread between inflation indexed and non inflation indexed bond - 7 years. Both indicators available at FRED as DFII7 and DGS7. (level difference)
- Spread between inflation indexed and non inflation indexed bond - 10 years. Both indicators available at FRED as DFII10 and DGS10. (level difference)
- Spread between inflation indexed and non inflation indexed bond - 20 years. Both indicators available at FRED as DFII20 and DGS20. (level difference)
- Spread between inflation indexed and non inflation indexed bond - 30 years. Both indicators available at FRED as DFII30 and DGS30. (level difference)
- ICE BofA AAA US Corporate Index Option-Adjusted Spread. Available on FRED as BAMLC0A1CAAA. (level difference)
- ICE BofA AA US Corporate Index Option-Adjusted Spread. Available on FRED as BAMLC0A2CAA. (level difference)
- ICE BofA A US Corporate Index Option-Adjusted Spread. Available on FRED as BAMLC0A3CA. (level difference)

- ICE BofA BBB US Corporate Index Option-Adjusted Spread. Available on FRED as BAMLC0A4CBBB. (level difference)
- ICE BofA BB US High Yield Index Effective Yield. Available on FRED as BAMLH0A1HYBB. (level difference)
- ICE BofA B US High Yield Index Effective Yield. Available on FRED as BAMLH0A2HYB. (level difference)
- ICE BofA CCC & Lower US High Yield Index Effective Yield. Available on FRED as BAMLH0A3HYCEY. (level difference)
- EUR/USD: Euro. Available on Bloomberg as EURUSD:CUR. (log difference)
- JPY/USD: Japanese Yen. Available on Bloomberg as JPYUSD:CUR. (log difference)
- GBP/USD: Pound Sterling. Available on Bloomberg as GBPUSD:CUR. (log difference)
- CNY/USD: Renminbi. Available on Bloomberg as USDCNY:CUR. (log difference)
- CAD/USD: Canadian Dollar. Available on Bloomberg as USDCAD:CUR.
- MXN/USD: Mexican Peso. Available on Bloombgerg as USDMXN:CUR. (log difference)
- KRW/USD: Korean Won. Available on Bloomberg as USDKRW:CUR. (log difference)
- S&P 500. Both the price and the volume of the index are included as indicators. Available on Bloomberg as SPX:IND. (log difference)
- iShares Core U.S. Aggregate Bond ETF. Available on Bloomberg as AGG:US
- iShares Broad USD High Yield Corporate Bond ETF. Available on Bloomberg as USHY:US. (log difference)
- S&P-GSCI Commodity Index Future. Available on Bloomberg as SPGSCI:IND
- Crude oil futures price. Available on Bloomberg as CL1:COM. (log difference)

- Gold futures price. Available on Bloomberg as GC1:COM. (log difference)

A.2.2 Brief Description of PCA Factors

	% Variance Explained	Most correlated variables
PCA Fin. 1	26%	US Government Bond Yields 2-30 years (+) S&P 500 (+) Spreads BBB-B (-) Investment Grade Bonds (-)
PCA Fin. 2	13%	US Government Bond Yields (+) USD value vs other currencies (+) Spreads (+) Investment Grade and High Yield Bonds (-)
PCA Fin. 3	8%	Short term interest rates (-) Spreads (+) Inflation Expectations (+)
PCA Fin. 4	7%	Short term interest rates (+) Inflation Spreads (+)
PCA Fin. 5	6%	Short term interest rates (+) USD value vs other currencies (-) Gold (+)

Table 19: Financial Variables PCA Factors

*Note: “+” stands for positive and “-” negative correlation.

B Tables of DFMs RSMEs

Note: In the tables, "E" stands for Economic Factors in the DFM models. "F" for Financial Factors and "T" for Text-based Factors. The suffix 1 means that only the first PCA factor was used, while 5 means that the first five were used instead. "B" indicates that the DFM used the Blocks described in Section 4, while "NB" indicates that they were not used. Finally, "EW" means that the frequency alignment was done using the average (equal weighting), while "MIDAS" indicates that MIDAS was used. "NC" stands for "Non-Convergence" of the model.

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	**3.00**	3.43	3.26	3.80
E1F1	3.28	3.09	3.31	3.53
E1T1	**3.00**	52.11	3.26	4.12
E1F1T1	3.28	4.87	3.30	3.57
E5	3.75	3.94	4.28	3.17
E5F5	3.76	3.02	4.13	3.80
E5T5	3.79	3.75	4.44	3.80
E5F5T5	3.76	3.12	4.24	3.63

Table 20: RMSEs (%) for 1Q ahead GDP growth forecasts (2020-2022)

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	0.53	0.56	0.53	0.71
E1F1	0.50	0.67	0.50	0.66
E1T1	0.53	0.63	0.54	0.76
E1F1T1	**0.49**	0.69	0.50	0.67
E5	0.59	0.77	0.58	0.76
E5F5	0.54	0.85	0.55	0.70
E5T5	0.55	0.87	0.55	1.11
E5F5T5	0.56	2.85	0.50	NC

Table 21: RMSEs (%) for 1Q ahead GDP growth forecasts (ex. 2020-2022)

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	3.85	3.95	3.58	3.37
E1F1	3.55	3.59	3.89	3.34
E1T1	3.83	3.71	3.63	3.24
E1F1T1	3.51	4.03	4.06	3.57
E5	2.92	3.67	2.98	3.72
E5F5	**2.90**	4.90	3.03	NC
E5T5	2.98	3.54	3.00	NC
E5F5T5	2.92	4.40	3.00	3.19

Table 22: RMSEs (%) for 1Y ahead GDP growth forecasts (2020-2022)

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	0.64	0.64	0.62	0.80
E1F1	0.72	0.73	0.77	0.86
E1T1	0.66	**0.59**	0.68	0.77
E1F1T1	0.72	0.67	0.80	0.86
E5	0.85	0.80	1.03	1.03
E5F5	0.67	1.05	0.82	0.87
E5T5	0.83	1.03	1.03	0.79
E5F5T5	0.64	NC	0.84	1.13

Table 23: RMSEs (%) for 1Y ahead GDP growth forecasts (ex. 2020-2022)

Note: The models which have only economic variables do not need any frequency alignment (all variables are already in the monthly frequency). I simply report the same result for EW and MIDAS in this case.

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	3.14	3.14	3.21	3.21
E1F1	3.17	3.17	3.28	3.09
E1T1	3.14	3.19	3.23	3.14
E1F1T1	3.21	3.18	3.27	3.09
E5	3.06	3.06	3.40	3.40
E5F5	3.10	**2.82**	3.30	3.09
E5T5	3.05	2.89	3.42	3.35
E5F5T5	3.09	2.87	3.29	3.07

Table 24: RMSEs (%) for 1Q ahead Inflation growth forecasts (2020-2022)

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	0.64	0.64	0.62	0.62
E1F1	0.66	0.64	0.64	0.63
E1T1	0.64	0.60	0.62	**0.58**
E1F1T1	0.66	0.65	0.65	0.65
E5	0.66	0.66	**0.58**	**0.58**
E5F5	0.63	0.61	**0.58**	0.65
E5T5	0.67	0.63	0.60	0.61
E5F5T5	0.66	0.71	**0.58**	0.66

Table 25: RMSEs (%) for 1Q ahead Inflation growth forecasts (ex. 2020-2022)

	NB EW	**NB MIDAS**	**B EW**	**B MIDAS**
E1	3.20	3.20	3.12	3.12
E1F1	3.17	3.35	3.10	**2.91**
E1T1	3.20	3.25	3.12	2.97
E1F1T1	3.16	3.39	3.10	2.94
E5	3.24	3.24	3.19	3.19
E5F5	3.25	3.20	3.20	3.04
E5T5	3.25	3.29	3.20	3.00
E5F5T5	3.25	3.20	3.20	3.04

Table 26: RMSEs (%) for 1Y ahead Inflation growth forecasts (2020-2022)

	NB EW	NB MIDAS	B EW	B MIDAS
E1	0.64	0.64	0.74	0.74
E1F1	0.64	**0.58**	0.75	0.65
E1T1	0.64	**0.58**	0.75	0.65
E1F1T1	0.65	**0.58**	0.75	0.65
E5	0.67	0.67	0.76	0.76
E5F5	0.63	0.59	0.71	0.67
E5T5	0.67	0.64	0.75	0.64
E5F5T5	0.65	0.69	0.72	0.65

Table 27: RMSEs (%) for 1Y ahead Inflation forecasts (ex. 2020-2022)