Importance of daily data in long horizon inflation forecasting

- a MIDAS approach

Katrin Ekström 920605

Sofia Lundgren 920323

Bachelor Thesis in Economics (15 hp)

Department of Economics

Supervisor: Marcin Zamojski

Abstract

We examine the accuracy of forecast models for the monthly Euro area inflation, focusing on the MIDAS approach. We compare two mixed frequency models with four low frequency models, using fourteen mixed frequency variables sampled at daily or monthly frequency. Our data set covers the period of February 1999 until August 2017, and we use a 10-year rolling window to construct the forecasts. We use MIDAS models with one- respectively five-month lags, as these specifications provide the lowest average MSEs. Our findings show that the MIDAS model with five month lags perform better in-sample compared to the MIDAS model with one-month lag. The opposite applies for our out-of-sample forecasts. Furthermore, our findings suggest, in line with previous findings, that the MIDAS models perform well for short forecast horizons. On the contrary to previous research, we find that the MIDAS models provide worse forecasts than an AR(1) for longer forecast horizons.

Keywords: MIDAS, inflation, forecasting

Table of contents

1. Introduction

It is important that accurate and reliable inflation forecasts are provided. Policy makers, such as central banks and governments, who implement monetary and fiscal policies depend on reliable inflation forecasts. Similarly, investors trying to provide profitable investments and firms trying to set prices rely on accurate inflation forecasts (Ang, Bekaert & Wei, 2007). The importance of forecasting inflation more often as new information emerges has become increasingly important for market participants over the last years. From the perspective of central banks it is important to continuously consider updates about inflation when conducting monetary policies. Private market participants tend to update their market expectations more frequently as new information becomes available and as a result they adjust their investment strategies (Monteforte & Moretti, 2013; Andrade, Fourel, Ghysels & Idier, 2014).

Inflation data are reported at monthly frequency and as a result the usual way to forecast inflation have been based on models with monthly explanatory variables. Models that are based only on variables that are sampled with monthly frequency or averaged to the monthly frequency might conceal data that are released within months because of their initial construction. If only monthly data is considered when forecasting inflation this might provide less accurate forecasts compared to allowing for mixed frequency variables in the forecasting analysis (Monteforte & Moretti, 2013).

In this paper, we use monthly and daily time series data to examine the accuracy of different forecast models for the Euro area inflation, the Harmonised Index of Consumer Prices (HICP). The traditional procedure for time series regression analysis has been to use data sampled at the same frequency. To manage datasets of different frequencies it is common to use temporal aggregation to transform the higher frequency variable into a low frequency variable. In the transformation process, higher frequency time series lose much of the underlying information about the data (Rossana & Seater, 1995).

Macroeconomic variables are collected at low frequencies, either monthly or quarterly. In order to incorporate daily data series when conducting regression analysis for low frequency

macroeconomic variables, there is an increased interest in using models with mixed-frequency data (Andreou, Ghysels & Kourtellos, 2012). Regression models that permit variables to be sampled at different frequencies are proposed. Ghysels, Santa-Clara, and Valkanov (2004) develop the Mixed Data Sampling (MIDAS) model, where the dependent variable and the independent variables are allowed to be sampled at different frequencies. This enables a new set of analyses where more information can be extracted from higher frequency time series which can be used together with low frequency series.

During a month, new data is released continuously and this data may contain new information about the inflation. To improve and update inflation forecasts in real time, it might be favourable to incorporate financial indicators sampled at a higher frequency, such as movements in yield curves and interest rate spreads. This type of data is available on a daily basis and can capture current information about the inflation (Monteforte & Moretti, 2013; Mondugo, 2013). Recently, researchers apply the MIDAS approach in order to study inflation forecasting. The approach provides more accurate inflation forecasts compared to univariate models and low frequency Vector Autoregressive (VAR) models (Monteforte & Moretti, 2013).

In this paper, we test whether a mixed frequency model increase the forecast accuracy compared to low frequency models. Specifically we compare the monthly inflation forecast accuracy between two daily frequency MIDAS models and four models using data sampled at a monthly frequency. The dataset consist of 14 independent variables, 9 sampled monthly and 5 sampled daily.

Our results indicate that the models using daily data produce more volatile forecasts. For short forecasting horizons the mixed frequency models perform better than the others, as can be seen from the forecast's tracking performance of the realized inflation. For longer forecasting horizons however, a simple autoregressive model prove to produce the most accurate forecast.

We conclude that the use of a MIDAS model for monthly inflation forecasting might not be

optimal for a central bank, at least not using our variables and model specifications. This since they produce quite volatile forecasts and are more prone to overfitting.

The remainder of the paper is organized as follows: the next section provides an overview of the theories and literature within this field. The following section describes the dataset and the subsequent section explains the methods that we use. In the last two sections we present our results and conclusions.

2. Theory and concepts

2.1 Taylor rule and Phillips curve

Economic theory shows that there are different relations between inflation and other variables (Taylor, 1993; Phillips, 1958) and central banks are increasingly building their monetary policy on forms of the Taylor rule (Castro, 2011; Chen, Turnovsky & Zivot, 2014). Taylor (1993) proposes that monetary policies in the US can be explained by a relation between interest rates, inflation, and output gap. The Taylor rule implies that inflation can be controlled for through reactions in short term interest rates. For the purpose of forecasting inflation, the Phillips curve is more often used in the empirical literature compared to the Taylor rule. The Phillips curve is initially introduced by Phillips (1958) who shows that there is a relation between changes in wage rates and unemployment. Modern interpretations of the Phillips curve relate the actual unemployment rate to changes in the inflation rate, or relate an aggregated measure of economic activity to changes in the inflation rate (Atkeson & Ohanian, 2001; Stock & Watson 1999). Several researchers conduct studies based on different forms of the Phillips curve for forecasting inflation, see (Ang et al., 2007; Stock & Watson, 2007, 2008; Rumler & Valderrama, 2010; Dotsey, Fujita & Stark, 2015).

Stock and Watson (1999) conduct an extensive study where they use the conventional Phillips curve based on unemployment and an alternative generalized Phillips curve based on the real aggregated economic activity to study inflation forecasts. They find that the generalized Phillips curve provide more accurate inflation forecasts compared to the conventional Phillips curve. Other studies provide different results for forecasts based on the Phillips curve. Some findings show that the Phillips curve provides less accurate forecasts compared to different benchmark models (Atkeson & Ohanian, 2001; Dotsey et al., 2015; Rumler & Valderrama, 2010). Other findings show that the Phillips curve provides better forecast accuracy than benchmark models in some time periods and worse accuracy in other time periods (Stock & Watson, 2008).

2.2 Previous studies

Stock and Watson (2001) investigate the accuracy of VAR models when forecasting quarterly inflation. The study is conducted on quarterly data for the US using inflation, degree of unemployment and interest rate for the time period of 1969 to 2000. Similarly, Ang, Bekart and Wei (2007) use quarterly data for the variables inflation, GDP and unemployment rate. In their analysis, they also incorporate financial data for term structure for both short- and long-term together with data from three different inflation expectation surveys. Based on this data they compare several forecasting methods for quarterly inflation. In order to compare the outcome of different methods they use VAR and Autoregressive Moving Average (ARMA) models, forecast models containing information on term structures and forecast models based on the Phillips curve. Ang et al. (2007) either use the last observations or calculate the quarterly averages to include the monthly variables in the models.

Averaging of time series is also applied by Stock and Watson (2003, 2007). Stock and Watson (2007) forecast quarterly inflation in the US based on quarterly and monthly variables. In order to generate data series sampled at the same frequency they incorporate the quarterly average of the monthly variables. The macroeconomic variables included in their analysis are, among others, unemployment rate, logarithm of real GDP, capacity utilization rates and building permits. Similarly, Stock and Watson (2003) use macroeconomic variables and a large set of financial variables sampled at quarterly or monthly frequency in order to forecast quarterly inflation and GDP. The financial variables are variables such as nominal exchange rates, oil price, and stock price indices. They either incorporate the quarterly average of the monthly variables or use the last observation in each quarter to create data series with the same frequency.

2.3 The MIDAS approach

The availability of higher frequency data creates an increased interest in mixed-frequency models (Andreou et al., 2012). The MIDAS approach, which initially was introduced by Ghysels et al. (2004), is one approach that makes it possible to conduct regression analysis for mixed-frequency data. The original application of the MIDAS model was based on predicting

stock market volatility (Ghysels Santa-Clara & Valkanov, 2006; Ghysels, Sinko & Valkanov, 2007). Findings show that the MIDAS approach provide accurate out-of-sample forecasts for volatility (Ghysels et al., 2006). Asgharian, Hou and Javed (2013) provide similar results when studying short- and long-term volatility. They show that inclusion of low frequency macroeconomic variables within a GARCH-MIDAS regression improve the forecast ability of the long-term volatility compared to using a traditional volatility benchmark model such as a GARCH(1,1).

The use of the MIDAS approach increases as the MIDAS approach shows to be a proper method to use for both macroeconomic and finance related issues in order to handle the problem of mixed frequency time series (Ghysels et al., 2006; Andreou, Ghysels & Kourtellos 2013). Asimakopoulos, Paredes and Warmedinger (2013) compare fiscal time series forecasts provided by the MIDAS approach with forecast estimations from other models. Their findings suggest that the MIDAS-method, performs better than the other models and they conclude that the best way of conducting forecasts for fiscal time series is to incorporate data as it gets available over time.

Several more studies have adopted the MIDAS approach in order to forecast quarterly GDP based on a set of monthly explanatory variables, and found the method to provide forecasts of better quality. Marcellino and Schumacher (2011) conduct an analysis based on German quarterly GDP and perform forecasts using three different types of MIDAS models. Kuzin, Marcellino and Schumacher (2011) use 20 monthly variables to forecast quarterly GDP for the euro area. They examine if different MIDAS methods, a MIDAS and an AR-MIDAS, obtain more accurate forecasts in comparison with a mixed frequency VAR model. The findings show that the models work as complements to one another as the MIDAS performs better in short term and the mixed frequency VAR better on a longer forecasting horizon. Similarly, Clements and Galvão (2008) perform quarterly GDP forecasts but for the US. They use an Autoregressive (AR) model and an Autoregressive Distributed Lag (ADL) model to perform quarterly benchmark forecasts. These quarterly forecasts are compared with the forecasts from their MIDAS model, which includes monthly data. In their MIDAS-method they also incorporate an autoregressive (AR) term and estimates a AR-MIDAS model. The

empirical results show that when they incorporate monthly data and use the MIDAS approach the accuracy of the forecasts improves.

Andreou, Ghysels and Kourtellos (2013) provide a forecast analysis for quarterly GDP based on a large dataset of daily financial variables, monthly macroeconomic variables and quarterly variables. They investigate whether univariate models, models with macroeconomic data, models with financial data or if models that combine quarterly macroeconomic data with daily financial data obtain the best forecasts. Their findings suggest that quarterly forecasts of US GDP improve when daily data is included in a MIDAS model compared to the forecasts from models that only use low frequency variables. The root mean squared forecast errors are lower for MIDAS models compared to the outcome of a quarterly AR(1) model and a quarterly ADL model.

2.4 High frequency models for inflation

Mixed frequency models are also applied to test whether the accuracy of forecasting models for inflation increases. Modugno (2013) aim to forecast the monthly US and euro area inflation and include mixed frequency data that are sampled at monthly, weekly and daily frequencies. For the euro area Modugno (2013) uses oil bulletin price that is sampled at weekly frequency and data of world market prices for raw materials that is sampled at daily frequency. The findings suggest that models that only incorporate monthly data perform worse than models with mixed frequency data. Especially, the mixed frequency model provides the most accurate forecasts compared to the low frequency model and a random walk at a short horizon.

Some studies are conducted where the MIDAS approach is used to forecast inflation. Andrade et al. (2014) argue that frequent attention for macroeconomic conditions are important when conducting monetary policy decisions. They conduct a study where inflation risk in the euro area is predicted for two following quarters using daily data for financial variables such as oil prices, stock prices, exchange rates, money markets and bond markets. Their findings show that when only averages of variables with higher frequency are used for forecasting inflation

risk the forecast tend to be less accurate compared to when mixed frequencies are taken into account. They conclude that financial data contribute when inflation risks are forecasted.

Similarly, Monteforte and Moretti (2013) focus on inflation in a mixed frequency context but aim to study inflation forecasts for the euro area. They use the monthly variables lagged inflation, oil price and core inflation index together with the daily variables short- and long-term interest rates, spreads of interest rate, commodity prices and exchange rates. They construct three models that are forecasted using the MIDAS approach, and besides these three models univariate models such as a random walk and an AR(1) are used. Their findings show that when variables with daily frequencies are included, the MIDAS model obtains lower root mean squared forecast errors compared to univariate models.

3. Data

We use data that we collect from several sources (see Appendix 1). Our data covers the period of February 1999 until August 2017. Based on economic theory and previous literature we use 14 variables to forecast the Euro area inflation, HICP. Nine of the variables are sampled at monthly frequency and five variables are sampled at daily frequency.

3.1 Dependent variable

The European Central Bank (ECB) measures inflation using the Harmonised Index of Consumer Prices (HICP). This index ensures that the same method for calculating inflation is used for all countries within the European Union. HICP represents the average change in what households pay for a basket consisting of both goods and services (ECB, 2017a). HICP is used by researchers to forecast the monthly European area inflation (Monteforte & Moretti, 2013; Modugno, 2013) and therefore we use this measure as the dependent variable in our forecasting analysis.

3.2 Independent variables

Our set of 14 predictor variables are known to affect the inflation rate. By using this dataset, we aim to incorporate variables that contain information that affect the inflation and as a result contribute to our forecasts of the Euro area inflation.

We consider the monthly variable unemployment since unemployment is shown to affect the inflation rate (Stock & Watson, 2001; Ang et al., 2007). We use the unemployment rate, defined as the proportion of the population in the European area that is between 15 to 74 years and without work (Eurostat, 2017b). Another variable included in our model that can help to explain the inflation is the number of building permits (Stock & Watson, 2007; Sousa & Falagiarda, 2017), this since inflation usually rises when the economy flourishes and when the capacity of production gets strained (Andrade et al., 2014).

We also consider monetary aggregates such as M1, M2, and M3 since these aggregates are shown to be important for forecasting inflation (Falagiarda & Sousa, 2017) and other

macroeconomic variables that in turn affect inflation like GDP (Marsilli, 2014). M1 is defined by ECB (2017b) as a monetary aggregate that includes currency that is in circulation together with overnight deposits. M2 is the sum of M1 plus additional deposits with a maturity of up to two years, and deposits which can be redeemable up to three months. The last monetary aggregate, M3, is the broadest aggregate (that we consider) and includes M2 together with repurchase agreements, fund shares in the money market and debt securities that have a maturity of at most two years (ECB, 2017b).

We include a set of monthly variables for different types of interest rates, such as the short term 3-month EURIBOR rate and the long-term government bond interest rate. We also consider the interest spread, as the difference between long-term government bond interest rate in the Euro area and the 3-month short-term EURIBOR rate. We also include ECB's daily refinancing rate, the minimum interest rate on interbank loans (ECB, 2014).

We include a monthly data series for Production Price Index (PPI) since changes in PPI can help when forecasting consumer price indices (Clark, 1995). Furthermore, we include the daily USD/EUR exchange rate because of its effect on import and export prices (Monteforte and Moretti, 2013).

An additional factor that can be important for explaining the HICP index are changes in oil prices, such as changes in Brent oil price. Economic theory shows that shocks in oil prices affect the inflation rate (Gottfries, 2013) and researchers use the Brent oil prices for forecasting inflation (Andrade et al., 2014; Modugno, 2013; Monteforte & Moretti, 2013). Therefore, we include a daily data series for European Brent oil prices in our analysis. Other daily variables that we include are indices for the stock market, as stock market indices can provide valuable information about inflation expectations (Andrade et al., 2014; Stock & Watson, 1999). We consider both the Eurostoxx50 index and the S&P500 index, to represent the Euro area and US stock market respectively.

3.3 Descriptive statistics

Table 1 illustrates the descriptive statistics for our dataset during the period of February 1999 until August 2017. Our monthly data series consist of 223 observations and our daily data series consist of 6787 observations. Initially, the daily data series consisted of different numbers of observations, so we use interpolation to fill in the missing observations.

Table 1: Descriptive statistics for all variables in the analysis over the period of February 1999 until August 2017. The table include the mean, median, maximum and minimum values together with the standard deviations and the total number of observations for each variable.

Variables	Mean	Median	Max	Min	Std. dev.	Obs.
HICP	89.62	91.37	101.84	74.12	8.75	223
Unemployment	9.57	9.27	12.08	7.25	1.27	223
Building permits	172.87	160.5	342.30	73.40	81.84	223
M1	86.61	81.80	157.02	40.95	31.40	223
M2	85.85	91.74	128.99	47.97	24.35	223
M3	86.40	95.50	125.18	48.76	22.85	223
Euribor	2.00	2.12	5.11	-0.33	1.68	223
Long term interest rate	3.75	4.09	5.73	0.77	1.27	223
Interest spread	1.74	1.69	3.41	-0.55	0.97	223
PPI	169.84	174.10	208.30	122.30	27.44	223
Oil price	62.05	55.92	143.95	9.77	32.61	6787
Exchange rate	1.21	1.24	1.60	0.83	0.17	6787
Refinancing rate	1.94	2.0	4.75	0.05	1.45	6787
Eurostoxx	3258.80	3084.68	5464.43	1809.98	740.92	6787
S&P500	1409.74	1314.54	2482.76	679.28	393.61	6787

3.4 Data processing

We conduct an augmented Dickey Fuller (ADF) test to examine the stationarity of our time series. The null hypothesis for the test is that the data series is nonstationary (Gujarati & Porter, 2009). Based on the results of these tests and by observing plots of the data series, we include the data series for unemployment in second differences and the rest of the data series in first differences. The test results are presented in Appendix 2.

3.5 Data quality

The data that we include in our analysis have different characteristics. Some of the data series are not seasonally adjusted but due to lack of available data we use both. The series affected are oil prices, commodity prices and stock market prices. We leave it for future work to extend this study with data that is fully seasonally adjusted.

A weakness with the data is that we use revised data and not real-time data. The problem with using revised data is that it is not representative of the information available for real-time forecasters (Stark & Croushore, 2002; Bouwman & Jacobs, 2011). Furthermore, the accuracy of forecasting models for inflation is shown to be sensitive to the data vintage (Stark & Croushore, 2002).

One more weakness is that we have missing observations for some of the daily variables which is solved by interpolation. However the data show some strength with the choice of HICP as our dependent variable, since we ensure that inflation is measured in the same way across the Euro area.

4. Method

Traditionally regression models are applied on time series that is sampled at the same frequency, examples of such models are the AR- and VAR-models. Since many macroeconomic variables are released at different frequencies this has made it problematic to model e.g., the quarterly released GDP figures on monthly variables such as unemployment or inflation. To solve this problem different time aggregation schemes are used to convert variables to the same frequency, i.e., converting high frequency variables to a lower frequency, the most popular being that of a simple time average:

$$\bar{x}_t = \frac{1}{m} \sum_{j=1}^{m} L^{j/m} x_t^{(m)}$$

Where $x_t^{(m)}$ denotes a variable sampled m times more frequent than the basic (low) frequency time variable t. L^i is a lag operator with superscript in the equation's basic frequency. The lags of the higher frequency variables will then be denoted as fractions in the superscript, e.g., $L^{(t-1/m)} x_t = x_{t-1/m}$ will be the first high frequency lag at time t.

This time averaging has the drawback that the same weight is put on all the m past realizations of X, which means that we are not just missing out on the volatility of the variable but also the timing. This contradicts the common perception that more recent realizations of variables should be weighted differently.

4.1 MIDAS

Drawing on theory of distributed lag models, Ghysels et al. (2004) introduce an approach to incorporate data with different sampling frequencies. This approach use a more nuanced weighting scheme than the simple time average by imposing a distributional restriction on the weights. A simple version of their Mixed Data Sampling (MIDAS) regression model can be seen below, with one dependent variable y and one independent variable x. Here the β_i are the regression coefficients and $b(k; \theta)$ is the weighting function, where k is the lag number and θ is a vector of hyperparameters defining the shape of the weighting function. By

normalizing the weight, i.e. forcing to $b(k;\theta)$ to sum to unity, the coefficient β_1 can be identified.

$$y_{t+h} \;=\; \beta_0 \;+\; \beta_1 \sum_{k=1}^{m} b(k;\theta)L^{k/m}x_t^{(m)} \;+\; \varepsilon_{t+h}$$

Optimizing the weights based on the hyperparameters of the weighting function decrease the amount of parameters to estimate, which otherwise could be a problem when incorporating a large number of independent variables (Ghysels et al., 2007). As an example, modelling a monthly sampled variable on two months' worth of data generated by a daily sampled variable would amount to about 40 parameters to estimate. Further adding other independent variables sampled at daily frequency would rapidly increase the amount of parameters. The MIDAS approach would in such a scenario create a more parsimonious regression model given that the model is optimized over a few hyperparameters $\theta = [\theta_1, \theta_2 ..., \theta_j \,]$.

The distribution functions predominantly used in the MIDAS framework are the exponential Almon lag and Beta lag functions. The two procedures are popular in the MIDAS literature because of their ability to depict various distributional shapes requiring only a few parameters (Ghysels et al., 2007). Ghysels et al. (2007) argue that the exponential Almon lag is better suited for long horizon forecasting, hence that will be the distribution used in this paper.

The exponential Almon lag function is based on the Almon lag polynomial used in the distributed lag models literature (Almon, 1965) and defined by Ghysels et al. (2004) as:

$$b(k;\theta_1,\theta_2) \;=\; \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=1}^{m} exp(\theta_1 j + \theta_2 j^2)}$$

The specification of the function by Ghysels et al. (2004) use two parameters θ_1 and θ_2 but the number of parameters can be increased if the problem at hand benefits from a more complex weighting function. However, two parameters is the more popular choice and the flexibility of that specification is shown in Figure 1. There we can see that θ_1 determines the location of the polynomial while θ_2 specifies the slope.

Distribution of lag weights for different parameter values

Legend:
- $\theta_1 = 0.001, \theta_2 = -7e\text{-}04$
- $\theta_1 = 0.001, \theta_2 = -3e\text{-}04$
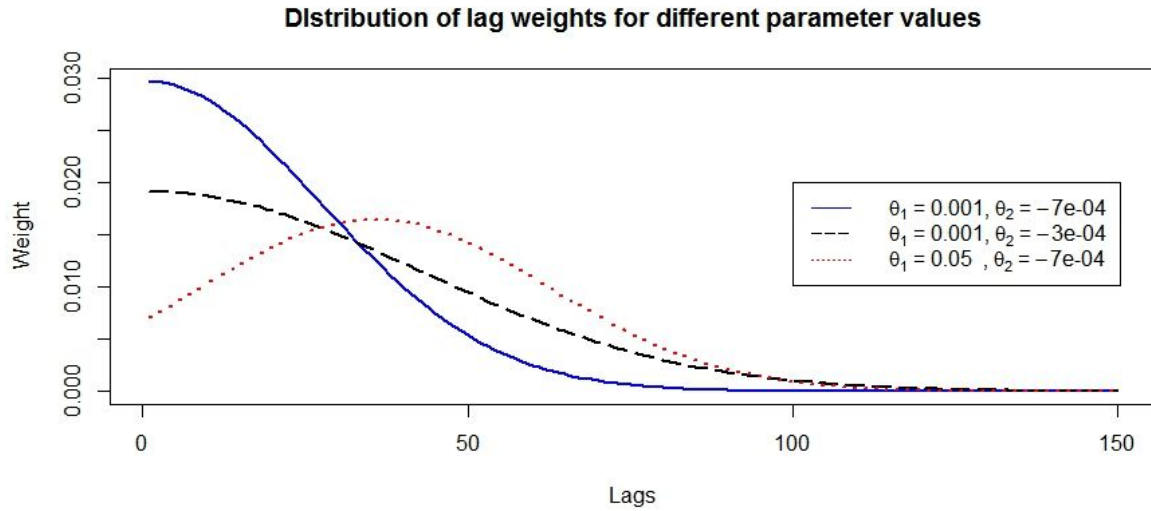- $\theta_1 = 0.05\ , \theta_2 = -7e\text{-}04$

Figure 1: Plot of the exponential Almon lag function for different parameter values.

4.2 Forecasting

In our analysis we divide our time series into two parts, one in-sample and one out-of-sample period. This creates a pseudo out-of-sample forecast, where we use the first part to train our models, i.e., estimating the parameters that best fit the data and the second part for testing the models' forecasting performance.

A rolling forecast window is also implemented. This means that the number of in-sample observations is fixed but the time frame is shifted when the forecasting horizon shifts. This is done by dropping the oldest observation in the in sample period, adding the first observation in the out-of-sample period instead and consequently dropping that observation from the out-of-sample period.

4.3 Measure of Accuracy

To evaluate the model performance both in- and out-of-sample we will use the Mean Square Error (MSE) and the Mean Absolute Percentage Error (MAPE)(as seen, e.g., in Chen, C., Twycross, J., & Garibaldi, J. M.(2017)). MSE of an estimation is the average squared difference between the realized and the estimated value, i.e., average squared error. The MSE used in our analysis can be seen below, with y_t representing the inflation at time t and $\widehat{y}_t$ being the estimated value of the same during the time period t = [1,T].

$$MSE(\widehat{y}_t) = \tfrac{1}{T} \sum_{t=1}^{T} (y_t - \widehat{y}_t)^2$$

MAPE also use a sum of the errors, but set in absolute terms and put in relation to the realized values by division, as can be seen below.

$$MAPE(\widehat{y}_t) = \tfrac{100}{T} \sum_{t=1}^{T} \frac{|y_t - \widehat{y}_t|}{y_t}$$

4.4 Diebold-Mariano test

To evaluate if the difference in performance between two forecasts is statistically significant we use the Diebold and Mariano (DM) test (1995). The forecasts performance is measured with a chosen loss function that is denoted by $L(e_{it})$ where e_{it} is the forecast error at time t associated with model i. When comparing two forecasts the test use the loss differential $d_{12t} = L(e_{1t}) - L(e_{2t})$ to compute the DM test statistic:

$$DM_{12} = \frac{\overline{d}_{12}}{\widehat{\sigma}_{d_{12}}}$$

Where $\overline{d}_{12}$ is the sample mean and $\widehat{\sigma}_{d_{12}}$ the sample variance of the loss differential. A quadratic loss function would thus be $L(e_{it}) = e_{it}^2$ and then the mean square error with T observations can be written $\overline{d}_{12} = \tfrac{1}{T} \sum_{t=1}^{T} L(e_{1t}) - L(e_{2t})$.

The test assumes the loss differential to be covariance stationary, which implies finite and constant mean and variance. As a consequence the test statistic can be compared to a standard normal distribution. (Diebold, 2015).

To test for equal forecast performance the null hypothesis is that the expected mean of the loss differential is zero, that is $E(d_{12t}) = 0$. When the test statistic is negative the first forecast performs better on average. But to see if the difference in performance is statistically significant for a chosen significance level the DM test statistic will be compared to the standard normal distribution.

5. Results

We use the first ten years of our data set, February 1999 to February 2009, as the in-sample period and the remaining observations until August 2017 as the out-of-sample period. This results in 120 monthly in-sample observations and 103 out-of-sample observations. The forecasts is further constructed with a 10 year rolling window.

5.1 MIDAS lag selection

We compare the in-sample and out-of-sample MSE for different lag lengths over a span of forecast horizons to find the optimal lag length in our MIDAS model. The resulting MSEs can be seen in Figure 2 and 3. The forecasts seem to perform similarly in-sample, the exception being a positive deviation at the nine month forecast horizon for the one month lag specification. The same deviation is present in the out-of-sample period, this time with a lower MSE compared to the others. A more stable forecast but not resulting from better pattern matching in-sample might be the result from accuracy measure characteristics. Larger deviations from the training data will be more heavily penalized with the MSE and might indicate that the 1 month lag forecast has some outliers but on average performs well. This interpretation is further strengthened by the out-of-sample MSE.

In addition, we test for heteroscedasticity of the MSE with a Bartlett's test. Since the forecasting horizon 12-24 months ahead is of special importance for central banks, this period is analyzed separately. For both time frames the null hypothesis, that the variances between the models differ significantly, is rejected at a 5% significance level with respective p-values of 0.9827 and 0.6003 (Table 2). Since the MSE variance is not significantly different between the models the lag length will be based on the MSE bias. The two model specifications with the lowest average MSE are, according to Table 3, the models that considered information from the previous month (MIDAS(1)) and the one with 5 month lag (MIDAS(5)). These are the mixed frequency models considered in the remainder of our analysis.

Table 2: Bartlett's test for heteroskedasticity, forecast MSE between models. The forecasts used was that for the first month, third month and then every third consecutive month up to the forecast 24 months ahead.

	All horizons	12-24 months ahead
Bartlett's K-squared	0.3977	2.7514
p-value	0.9827	0.6003

Table 3: Descriptive statistics forecast MSE using MIDAS models with different lag specifications. The forecasts used was that for the first month, third month and then every third consecutive month up to the forecast 24 months ahead.

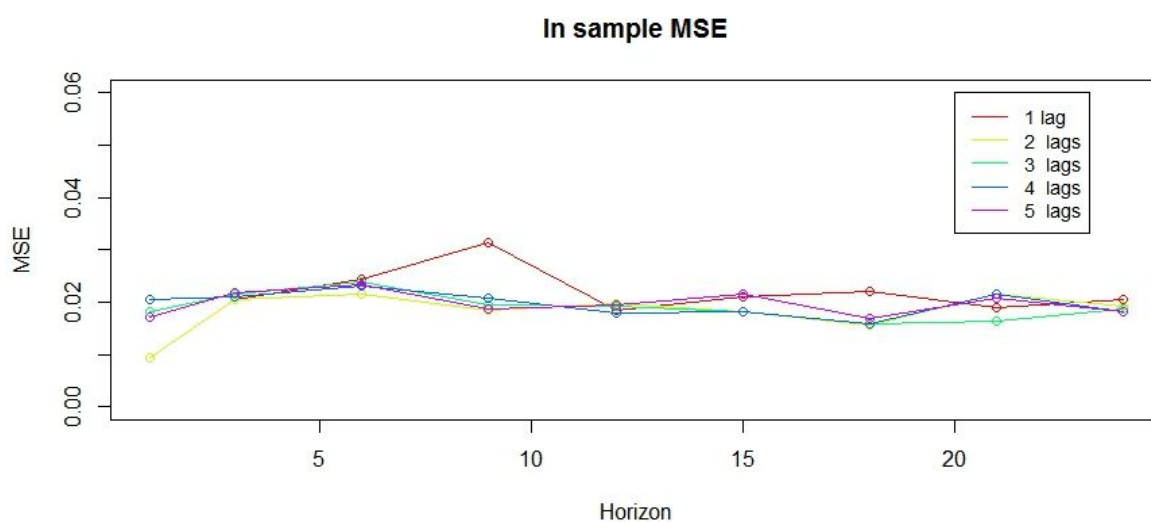| Lag length	Mean MSE, all horizons	Mean MSE, 12-24 months	Variance MSE, all horizons	Variance MSE, 12-24 months
1	0.0381	0.0455	0.0000810	0.00001907
2	0.0417	0.0486	0.0000901	0.00004266
3	0.0421	0.0512	0.0001202	0.00003912
4	0.0415	0.0496	0.0001097	0.00002650
5	0.0408	0.0470	0.0000911	0.00009386

Figure 2: MSE for different MIDAS model specifications using a ten year rolling window starting at February 1999.
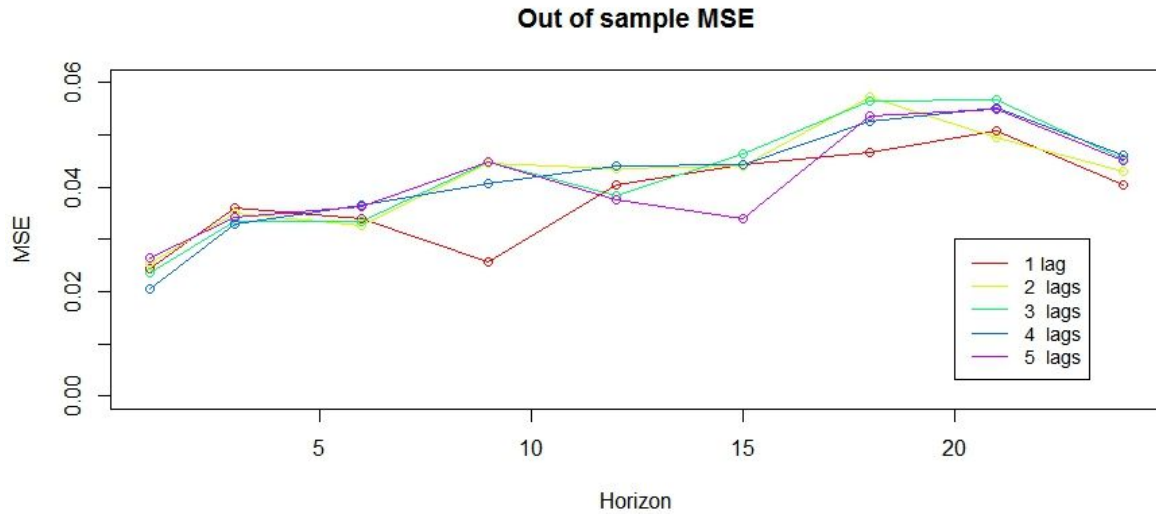
Out of sample MSE

Figure 3: MSE for forecasts using different MIDAS specifications. Forecasting from 2009-03 to 2017-08 using a 10 year rolling window.

5.2 Forecast Comparison

The models we compare our two MIDAS models with are an AR(1) and three MIDAS models using data sampled at the same (monthly) frequency. Two of the MIDAS models are implemented with a two month lag and an exponential Almon lag weight restriction. The time aggregates we use in these models is the monthly average, MIDAS(Mean), and the last observation in each month, MIDAS(Last Obs). The last MIDAS model use a two month lag but without a weighting function, this unrestricted model also use the monthly average of the daily data, U MIDAS(Mean).

5.2.1 MSE

The difference in goodness of fit in the in-sample period can be seen in Figure 4 and the forecasting performance in Figure 5, both measured by MSE. The AR(1) exhibit the worst goodness of fit considering the in-sample data but in contrast shows most promise in keeping the out-of-sample forecasting errors low and stable. Comparing the two high frequency MIDAS specifications the MIDAS(5) performs better in-sample than MIDAS(1) and in the out-of-sample it is the other way around. In the out-of-sample the worst forecast performance is produced by U MIDAS(Mean).
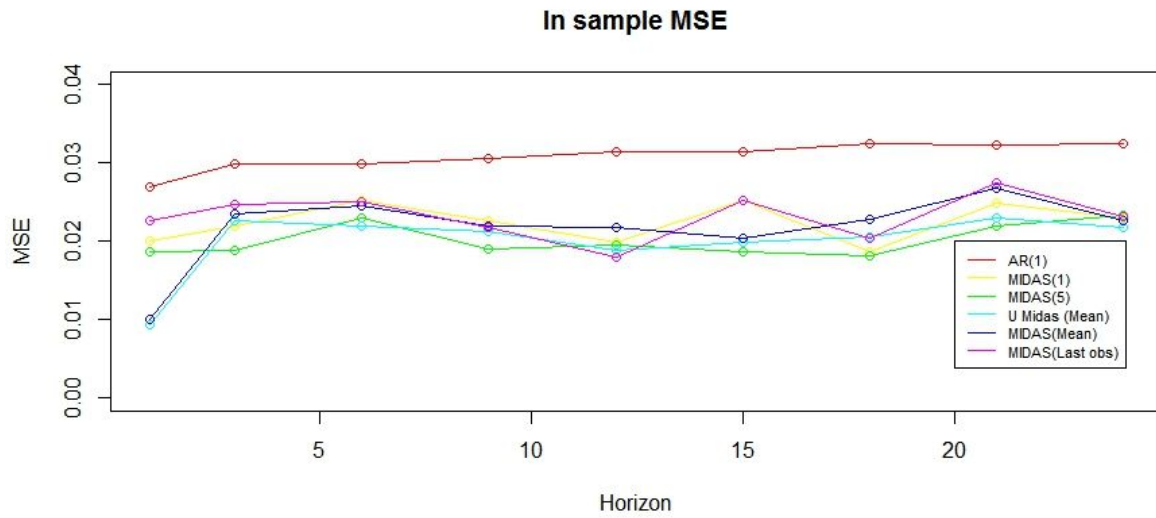
Figure 4: MSE for forecasts using five different models using a ten year rolling window starting at February 1999.
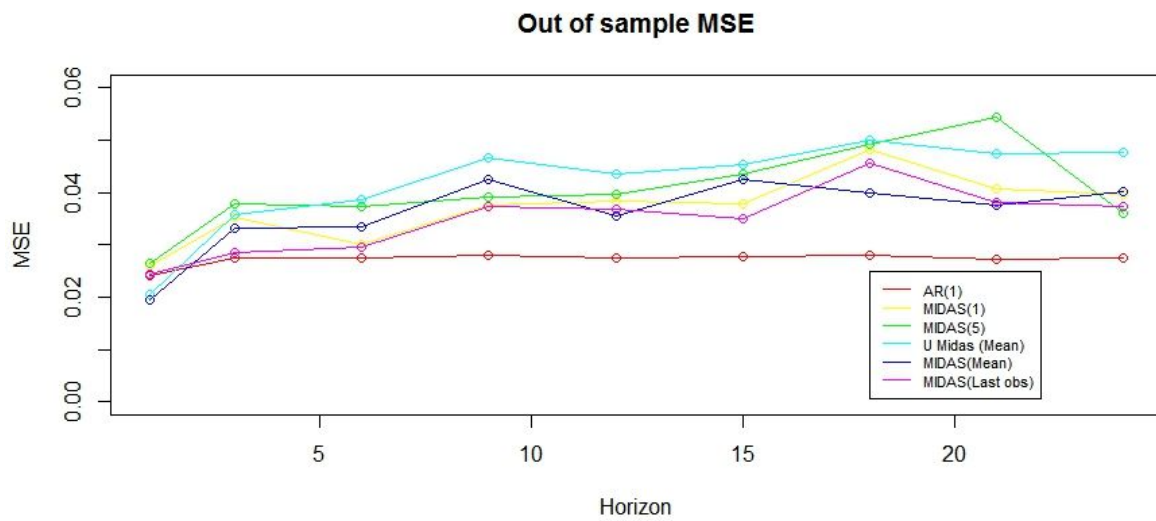

Figure 5: MSE for monthly forecasts using five different models. Forecasting from 2009-03 to 2017-08 using a 10 year rolling window.

5.2.2 Forecast plots

The resulting forecast from the six models can be seen in Figure 6 and Figure 7. The AR(1) displays consistently flat forecasts compared to the others, with the 1 month forecast horizon most variable, showing some lagged reaction to dips and highs.

Both the mixed frequency models show some good forecast performance for the 1 month horizon, but for other horizons MIDAS(5) display some significant deviation from the realized values and a high variance. In contrast MIDAS(1) show lower variance in the forecasts and some capability in predicting the ups and downs. Comparing the models for the 12 month horizon we observe that the MIDAS(1) provide a more flat forecast compared to MIDAS(5). Also comparing the two models for the 24 month forecast horizon, we see that the MIDAS(1) provide the best forecasts with MIDAS(5) showing quite large deviations from the realized change in HICP.

The MIDAS specifications using the monthly averages performs similarly to each other, differing in forecast variability with the U MIDAS(Mean) seemingly reacting more strongly to indications in the data for HICP movements. For the 1 month ahead forecast horizon the two models forecast the HICP quite well, on par with MIDAS(1) and better than MIDAS(5). MIDAS(1) seem to perform slightly better than the both mean-models however.

MIDAS(Last obs) often produce forecasts with less variance than the other same-frequency MIDAS models, showing more predictive power for some horizons and worse for others. MIDAS(mean) provide a somewhat better forecast on the 1 month horizon compared to the MIDAS(last obs) for instance.

The model displaying the least forecasts variation is thus the AR(1) and MIDAS(5) the one showing the most.

Figure 6: Forecast plots using AR(1), MIDAS(1), and MIDAS(5) for nine different forecasting horizons. The black line represents the realized change in HICP and the colored lines the model forecasts, the x-axis represents time and the y-axis the change.
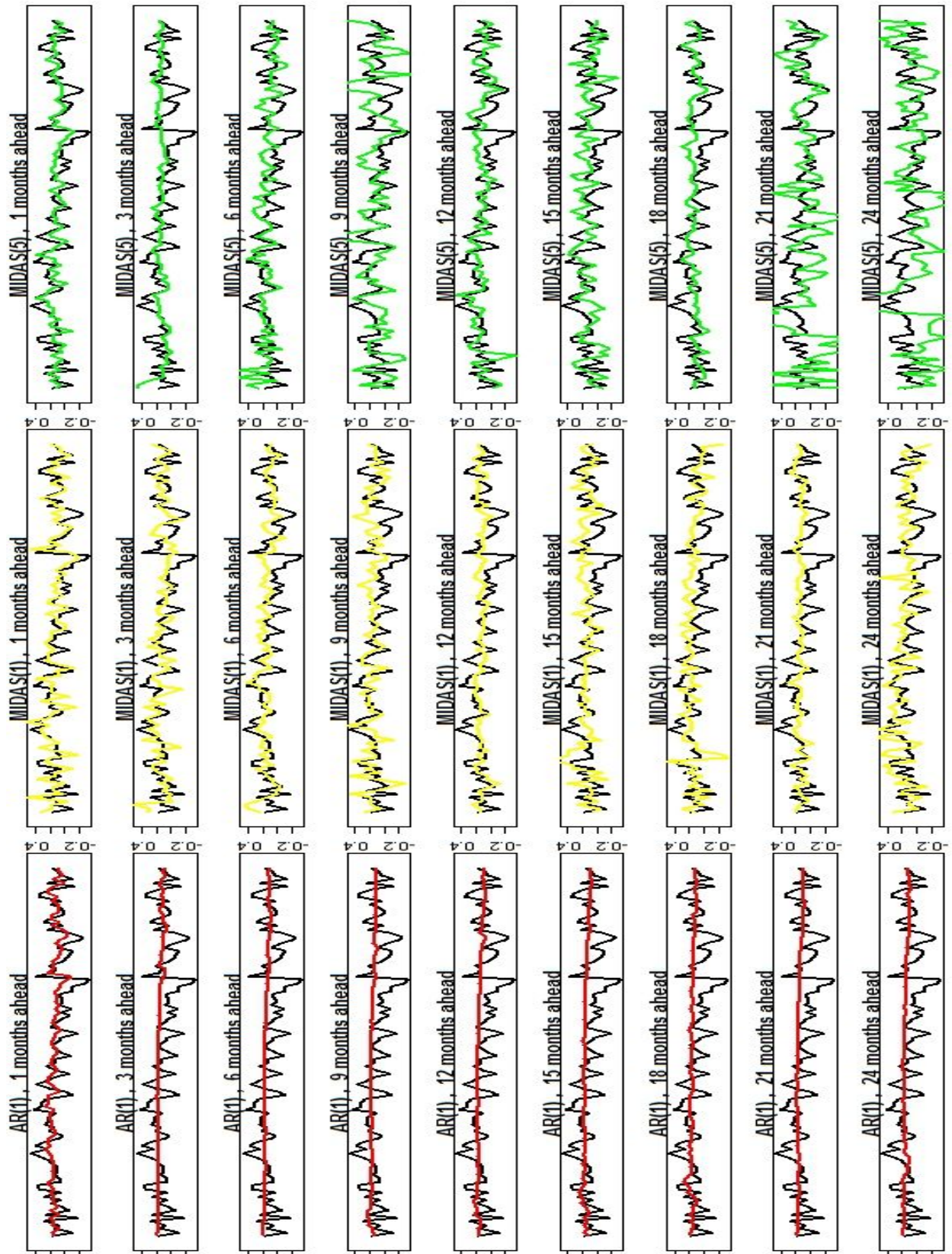
Figure 7: Forecast plots using U MIDAS(Mean), MIDAS(Mean), and MIDAS(Last obs) for nine different forecasting horizons. The black line represents the realized change in HICP and the colored lines the model forecasts, the x-axis represents time and the y-axis the change.

5.2.3 DM-tests

In order to evaluate if the difference in performance is statistically significant we conduct DM-tests. The accuracy measures we use are MSE and MAPE. Our forecasts predict the change in inflation and since the actual change in HICP is zero for two months in our out-of-sample we will omit these observations to be able to calculate the MAPE measure for this period.

Results for the two-sided DM tests comparing the forecasting performance of MIDAS(1) against all the other models can be seen in Table 4. A negative value indicate that the MIDAS(1) performs better than the model it is compared to and we can thus see that it is evaluated to be the better model almost twice as many times if we look at the MAPE as opposed to the MSE. The difference between MIDAS(1) and the models AR(1) and MIDAS(5) show more statistical significance than the rest. Where the AR(1) perform better overall but MIDAS(1) seem to have more predictive power in the short term. The comparison also seem to come out in favour of MIDAS(5) looking at the MSE and number of significant test statistics, but the number of negative statistics in the MAPE column favour MIDAS(1). The comparison between the two is thus a bit inconclusive.

We also compared all models against each other, Table 5,6, and 7 show the results for the time horizons 12, 24, and the averaged result for all the nine forecasting horizons used in the analysis. The results indicate that for the chosen time periods the AR(1) mostly produce significantly better forecasts than the other models. The forecasting performance of AR(1) also seem to be stable across all the forecast horizons.

MIDAS(1) show better relative results for 24 months ahead than 12 and the other way around for MIDAS(5). Looking at the forecast plots in Figure 6 this is to be expected for these horizons. This since they both display less volatility for forecast horizon 12, which in this case benefits MIDAS(5).

The models using monthly time aggregates and the MIDAS lag structure performs better than

MIDAS(1) and U MIDAS(Mean) for the chosen horizons. They also perform better than MIDAS(5) for the 24 months horizon. This trend is further seen on the aggregated level, with MIDAS(Last Obs) demonstrating the slightly better performance of the two. Combining this result with the forecast plots in Figure 7 we attribute this to the lower volatility for MIDAS(Mean) and MIDAS(Last Obs).

Table 4: Comparing the forecasts of the different models against the MIDAS(1) with a two-sided DM test. The table shows the test statistics for two measures of accuracy, MSE and MAPE. The statistical significance is indicated by "" - 10% "**" - 5% and "***" - 1%.*

	AR(1)		MIDAS(5)		U MIDAS (Mean)		MIDAS (Mean)		MIDAS (Last obs)	
Horizon	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
1	-1.05	-0.92	2.54 **	2.52 **	0.70	-0.71	0.76	0.53	-1.56	-0.16
3	0.70	-0.41	1.93 *	2.28 **	-1.21	-0.94	-0.26	0.26	0.66	1.58
6	1.96 *	-0.98	-2.63 **	-3.72 ***	-0.59	-2.34 **	-0.25	-1.13	1.70	-1.15
9	2.33 **	2.08 **	0.54	-1.26	1.07	-0.92	1.64	-0.02	1.82 *	2.20 **
12	2.24 **	1.93 *	1.33	3.49 ***	-0.68	0.11	1.73 *	1.77 *	0.92	1.30
15	2.14 **	2.59 **	2.17 **	3.25 ***	0.09	0.46	0.45	1.40	1.55	1.16
18	2.30 **	0.59	0.21	-1.80 *	-0.64	-1.98 *	1.23	-0.20	0.69	-0.44
21	1.94	0.81	-1.94	-4.72 ***	0.18	-1.34	1.48	-0.10	0.32	0.80
24	1.81 *	1.65	-0.68	-0.50	-1.15	-1.40	0.54	-0.74	0.71	0.32

Table 5: Comparing the forecasts for the 12 months ahead horizon using a two-sided DM-test. In compliance with the notation in the method chapter the rows represent model 1 and the columns model 2. MSE and MAPE are used as accuracy measures and the significance levels are represented by "" - 10% "**" - 5% and "***" - 1%*

	AR(1)		MIDAS(1)		MIDAS(5)		U MIDAS (Mean)		MIDAS (Mean)		MIDAS (Last Obs)	
	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
AR(1)	0	0	-2.24 **	-1.93 *	-1.24	1.60	-2.08 **	-1.87 *	-2.07 **	-0.88	-1.77 *	-1.50
MIDAS(1)	2.24 **	1.93 *	0	0	1.33	3.49 ***	-0.68	0.11	1.73 *	1.77 *	0.92	1.30
MIDAS(5)	1.24	-1.60	-1.33	-3.49 ***	0	0	-1.42	-3.69 ***	-0.58	-3.09 ***	-0.59	-2.45 **
U MIDAS (Mean)	2.08 **	1.87 *	0.68	-0.11	1.42	3.69 ***	0	0	1.89 *	2.59 **	1.16	1.12
MIDAS (Mean)	2.07 **	0.88	-1.73 *	-1.77 *	0.58	3.09 ***	-1.89 *	-2.59 **	0	0	-0.10	0.03
MIDAS (Last Obs)	1.77 *	1.50	-0.92	-1.30	0.59	2.45 **	-1.16	-1.12	0.10	-0.03	0	0

Table 6: Comparing the forecasts for the 24 months ahead horizon using a two-sided DM-test. In compliance with the notation in the method chapter the rows represent model 1 and the columns model 2. MSE and MAPE are used as accuracy measures and the significance levels are represented by "" - 10% "**" - 5% and "***" - 1%*

	AR(1)		MIDAS(1)		MIDAS(5)		U MIDAS (Mean)		MIDAS (Mean)		MIDAS (Last Obs)	
	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
AR(1)	0	0	-1.81 *	-1.65	-1.93 *	-1.97 *	-3.18 ***	-2.44 **	-2.93 ***	-1.96 *	-2.28 **	-1.28
MIDAS(1)	1.81 *	1.65	0	0	-0.68	-0.50	-1.15	1.40	0.54	-0.74	0.71	0.32
MIDAS(5)	1.93 *	1.97 *	0.68	0.50	0	0	0.08	-0.82	1.05	-0.23	1.24	0.74
U MIDAS (Mean)	3.18 ***	2.44 **	1.15	1.40	-0.08	0.82	0	0	2.31 **	2.40 **	1.94 *	1.58
MIDAS (Mean)	2.93 ***	1.96 *	-0.54	0.74	-1.05	0.23	-2.31 **	-2.40 **	0	0	0.46	0.95
MIDAS (Last Obs)	2.28 **	1.28	-0.71	-0.32	-1.24	-0.74	-1.94 *	-1.58	-0.46	-0.95	0	0

Table 7: Aggregated results from all horizons using the two-sided DM-test. In compliance with the notation in the method chapter the rows represent model 1 and the columns model 2. MSE and MAPE are used as accuracy measures and the significance levels are represented by "" - 10% "**" - 5% and "***" - 1%*

	AR(1)		MIDAS(1)		MIDAS(5)		U MIDAS (Mean)		MIDAS (Mean)		MIDAS (Last Obs)	
	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
AR(1)	0	0	-1.60	-0.82	-1.11	-0.86 **	-2.05 **	-1.69	-1.56 *	-0.66	-1.54	-0.46
MIDAS(1)	1.60	0.82	0	0	0.38	-0.05	-0.25	-1.01	0.81	0.20	0.76	0.62
MIDAS(5)	1.11	0.86 **	-0.38	0.05	0	0	-0.53	-0.53	0.01	0.25	0.10	0.52 *
U MIDAS (Mean)	2.05 **	1.69	0.25	1.01	0.53	0.53	0	0	1.64	2.27 **	1.08	1.41
MIDAS (Mean)	1.56 *	0.66	-0.81	-0.20	-0.01	-0.25	-1.64	-2.27 **	0	0	0.07	0.46
MIDAS (Last Obs)	1.54	0.46	-0.76	-0.62	-0.10	-0.52 *	-1.08	-1.41	-0.07	-0.46	0	0

6. Conclusion

In this paper we compare the forecast performance of six different models, using two high frequency MIDAS specifications, three low frequency MIDAS models and an AR(1) model. For the two high frequency models we use nine monthly and five daily variables, whereas in the other three MIDAS models all variables are included at monthly frequency, either by averaging or using the last observation. We find that the MIDAS(1) and the MIDAS(5) specifications provide the lowest average MSE, and we use these two models in our analysis to study the performance of MIDAS models that incorporate both daily and monthly data.

Our results for the goodness of fit in the in-sample period, shown in Figure 4, are in line with the findings by Andrade et al., (2014). Similar to their result the in-sample goodness of fit is improved when incorporating variables sampled at daily frequency. Which is to be expected, since more parameters can create a more intricate parameter structure making it possible to better emulate the data. On the contrary to the results by Andrade et al., (2014) we find that for our out-of-sample forecasts, shown in Figure 5, the AR(1) provides the lowest MSEs. The different results regarding the MSEs for the in-sample goodness of fit and the out-of-sample forecasts can occur as a consequence from the included variables in the analysis. This as the variables that are informative for the in-sample goodness of fit might not be as informative for the out-of-sample forecasts (Andrade et al. 2014). This might also result from overfitting, the model parameters more optimized for fitting known values than extracting information about general trends.

The findings from our analysis differs from findings in previous research. Previous research tend to show that the MIDAS approach outperforms same-frequency models and simpler univariate models when forecasting financial or macroeconomic variables (Asgharian et al., 2013; Asimakopoulos et al., 2013; Clements & Galvão, 2008; Andreou et al., 2013). Similarly, previous inflation forecasting research show that the MIDAS approach is better both for shorter and longer horizons (Monteforte & Moretti, 2014; Andrade et al., 2014). On the contrary to this, our analysis show that the MIDAS approach using daily data does not consistently outperform the other models, especially not for longer forecast horizons. The

mixed frequency MIDAS models, MIDAS(1) and MIDAS(5), does however provide more accurate forecasts in shorter horizons which is in line with previous research (Kuzin et al., 2011; Clements & Galvão, 2008).

According to our research the AR(1) model produce stable and more accurate monthly inflation forecasts than the other models. Consequently we argue that for a central bank this model would be the best to use out of the six in our analysis, when considering the specified time period and variables. But due to these opposing results to previous research we emphasize that it is important to conduct more research about inflation forecasting using the MIDAS approach.

Possible continuation and improvement to the research in this paper could look into improving accuracy and stability of forecasts by combining different ones from different methods. In our case it could prove interesting to, e.g., combine the stable AR(1) with the more volatile MIDAS(1). Further it could also be interesting to see if a combination of forecasts of the dependent variable using one independent variable at a time would provide a significant difference in forecasting performance, as suggested by Andreou et al. (2013)

Also, during our research we noticed that the algorithm estimating the model coefficients had some difficulties, sporadically resulting in substantial forecast errors. The instability might derive from the non-linear optimization algorithm. The models could consequently benefit from forecast trimming (Marcillino, 2007), you could for instance restrict the size of the allowed forecast difference between two consecutive periods. This will introduce more inertia for the more volatile forecasts. For a central bank this could prove valuable since they would rather have a stable forecasting model than a highly volatile one.

References

Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, *33*(1), 178-196.

Andrade, P., Fourel, V., Ghysels, E., & Idier, J. (2014). The financial content of inflation risks in the euro area. *International Journal of Forecasting*, *30*(3), 648-659.

Andreou, E., Ghysels, E., & Kourtellos, A. (2012). Forecasting with Mixed-Frequency Data. *In The Oxford Handbook of Economic Forecasting Oxford University Press*. DOI: 10.1093/oxfordhb/9780195398649.013.0009

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters look at daily financial data?. *Journal of Business & Economic Statistics*, *31*(2), 240-251.

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets or surveys forecast inflation better?. *Journal of Monetary Economics*, *54*(4), 1163-1212.

Asgharian, H., Hou, A. J., & Javed, F. (2013). The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach. *Journal of Forecasting*, *32*(7), 600-612.

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2013). *Forecasting fiscal time series using mixed frequency data.* (IDEAS Working Paper Series from RePEc No. 1550).

Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation?. *Quarterly Review - Federal Reserve Bank of Minneapolis*, *25*(1), 2-11.

Bouwman, K. E., & Jacobs, J. (2011). Forecasting with real-time macroeconomic data: The ragged-edge problem and revisions. *Journal of Macroeconomics*, *33*(4), 784-792.

Castro, V. (2011). Can central banks' monetary policy be described by a linear (augmented) Taylor rule or by a nonlinear rule?. *Journal of Financial Stability*, 7(4), 228-246.

Chen, C., Twycross, J., & Garibaldi, J. M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PLoS ONE, 12*(3).

Chen, Y. C., Turnovsky, S. J, & Zivot, E. (2014). Forecasting inflation using commodity price aggregates. *Journal of Econometrics*, *183*(1), 117-134.

Clark, T. E. (1995). Do producer prices lead consumer prices?. *Economic Review*, *80*(3), 25-40.

Clements, M. P., & Galvao, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *Journal of Business & Economic Statistics*, *26*(4), 546- 554.

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13*(3), 253-263.

Diebold, F. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. *Journal of Business & Economic Statistics, 33*(1), 1.

Dotsey, M., Fujita, S., & Stark, T. (2015). *Do Phillips curves conditionally help to forecast inflation?*. (IDEAS Working Paper Series from RePEc No.15-16)

ECB. (2014). The ECB`s negative interest rate.
https://www.ecb.europa.eu/explainers/tell-me-more/html/why-negative-interest-rate.en.html
[2017.12.14]

ECB. (2017a). Measuring inflation – the Harmonised Index of Consumer Prices (HICP).
https://www.ecb.europa.eu/stats/macroeconomic_and_sectoral/hicp/html/index.en.html
[2017.11.19]

ECB. (2017b). All glossary entries.
https://www.ecb.europa.eu/home/glossary/html/glossa.en.html
[2017.12.14]

ECB data warehouse. (2017). Statistical Data Warehouse European Central Bank. http://sdw.ecb.europa.eu/ [2017.12.04].

EUROSTAT (2017a). Eurostat Your key to European statistics. http://ec.europa.eu/eurostat/data/database [2017.12.04].

EUROSTAT. (2017b). Unemployment statistics.
http://ec.europa.eu/eurostat/statistics-explained/index.php/Unemployment_statistics
[2017.12.14]

Falagiarda, M., & Sousa, J. (2017). *Forecasting Euro Area Inflation Using Targeted Predictors: Is Money Coming Back?*. (ECB Working Paper Series No. 2015).

FRED (2017). FRED Economic Data. https://fred.stlouisfed.org/ [2017.12.04].

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models*. (IDEAS Working Paper Series from RePEc No. 2004s-20).

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1), 59-95.

Ghysels, E., Sinko, A. & Valkanov, R. (2007), MIDAS Regressions: Further Results and New Directions, *Econometric Review*, *26*(1), 53–90.

Gottfries, N. (2013). *Macroeconomics* (1. ed.). New York: Palgrave.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th. ed.). New York: Mc Graw Hill Education.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529-542.

Marcillino, M. (2007). A comparison of time series models for forecasting GDP growth and inflation. Working paper, IEP-Università Bocconi, IGIER and CEPR, 2007.

Marcellino, M., & Schumacher, C. (2011). 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.

Marsilli, C. (2014). *Variable Selection in Predictive MIDAS Models*. (IDEAS Working Paper Series from RePEc No. 520)

Modugno, M. (2013). *Now-casting inflation using high frequency data*. (IDEAS Working Paper Series from RePEc No. 1324).

Monteforte, L., & Moretti, G. (2013). Real‑Time Forecasts of Inflation: The Role of Financial Variables. *Journal of Forecasting*, *32*(1), 51-61.

Phillips, A W. (1958). The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861‑1957. *Economica*, *25*(100), 283‑ 299.

QUANDL. (2017). EURIBOR 3 Months (Daily). https://www.quandl.com/data/BOF/QS_D_IEUTIO3M-EURIBOR-3-Months-Daily [2017.12.04].

Rossana, R. J., & Seater, J. J. (1995). Temporal Aggregation and Economic Time Series. *Journal of Business & Economic Statistics*, *13*(4), 441-451.

Rumler, F., & Valderrama, M. T. (2010). Comparing the New Keynesian Phillips Curve with time series models to forecast inflation. *North American Journal of Economics and Finance*, *21*(2), 126-144.

Stark, T., & Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, *24*(4), 507-531.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics, 44*(2), 293-335.

Stock, J. H., & Watson, M. W. (2001). Vector Autoregressions. *The Journal of Economic Perspectives*, *15*(4), 101-115.

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. (2007). Why has U.S. inflation become harder to forecast?. *Journal of Money, Credit & Banking*, *39*(1), 3-33.

Stock, J. H., & Watson, M. W. (2008). *Phillips curve inflation forecasts.* (NBER Working Paper Series No. 14322).

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Series on Public Policy*, *39*, 195-214.

Yahoo Finance. (2017). Yahoo Finance. https://finance.yahoo.com/ [2017.12.04].

Appendix 1

Table A1: List of variables, frequency and source of collection.

Data series	Frequency	Collected from
HICP	Monthly	ECB data warehouse (2017)
Unemployment	Monthly	ECB data warehouse (2017)
Building permits	Monthly	ECB data warehouse (2017)
M1	Monthly	ECB data warehouse (2017)
M2	Monthly	ECB data warehouse (2017)
M3	Monthly	ECB data warehouse (2017)
Long term interest rate	Monthly	
Euribor 3 month	Monthly	Quandl (2017)
Interest rate spread	Monthly	ECB data warehouse (2017)
PPI	Monthly	FRED (2017)
Eurostoxx50	Daily	YAHOO Finance (2017)
Exchange rate USD/EUR	Daily	EUROSTAT (2017a)
Oil price	Daily	FRED (2017)
Refinancing rate EURO	Daily	ECB data warehouse (2017)
S&P500	Daily	YAHOO Finance (2017)

Appendix 2

In table A2, the p-values from the ADF test is presented. If the p-value is smaller than the level of significance equal to 0,05 we reject the null hypothesis. All data series except M1, M2, M3 and unemployment are stationary in first differences. For M1, M2 and M3 we study the plots of the data and conclude that there is a positive trend in the data series. The ADF-tests show that the data series M2 and M3 are stationary in first differences and that M1 is not stationary in either level or first differences. The plots of these three data series show that there is a positive trend in level but not in first difference. Therefore, we choose to include M1, M2 and M3 in first differences. For unemployment, we study the plot of the data in both level and first difference and find that there is a pattern over time. We conduct the ADF test on the second difference and find the data series to be stationary.

Table A2: The table shows the p-values from the ADF-test in level and first difference. For unemployment the result from taking the second difference is also presented.

Variable	ADF test with trend and intercept		
	Level	1st Difference	2nd Difference
HICP	0.99	0.00	
Unemployment	0.37	0.12	0.00
Building permits	0.87	0.00	
M1	0.99	0.33	
M2	0.01	0.83	
M3	0.03	0.45	
Euribor	0.22	0.00	
Long term interest rate	0.25	0.00	
Interest spread	0.34	0.00	
PPI	0.48	0.00	
Oil price	0.87	0.00	
Exchange rate	0.81	0.00	
Refinancing rate	0.80	0.00	
Eurostoxx	0.49	0.00	
S&P500	0.88	0.00	