A COMPARISON OF USING MIDAS AND LSTM MODELS FOR GDP NOWCASTING

Iva Glišić

Available at www.nbs.rs

Economic Research and Statistics Department

NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street

Telephone: (+381 11) 333 8000

www.nbs.rs

A comparison of using MIDAS and LSTM models for GDP nowcasting

Iva Glišić

Abstract: The paper elaborates on machine and deep learning methods, as well as mixed data sampling regression models, used for GDP nowcasting. The aim is to select an adequate model that shows better performance on the data used. The paper provides an answer to the question of whether the use of deep learning methods can improve GDP nowcasting compared to traditional econometric methods, as well as whether the use of specific high-frequency indicators improves the quality of the models used. The paper examines the selection of adequate indicators – both official and those from alternative sources, presents the framework of mixed data sampling regression models and deep learning models used for nowcasting, and gives an assessment of two such models on the example of Serbian GDP. Serbia's GDP was modelled for the period Q1 2016 – Q2 2023 and the end of the observed period (six quarters) was used for the forecast. Finally, two assessed models were compared – the mixed data sampling regression model and the LSTM neural network. A special focus is placed on ways to improve both models. The LSTM recurrent neural network model had a smaller forecast error, with the use of a combination of official and alternative (high-frequency) indicators, but the mixed data sampling regression model also proved to be a good tool for decision-makers, since its structure allows insight into the ongoing movements impacting GDP dynamics. The use of alternative indicators in nowcasting improved the projections through both presented models.

Keywords: GDP, nowcasting, MIDAS, neural networks, high-frequency indicators.
[JEL Code]: C32, C45, C53

Non-technical summary

The models used to obtain flash estimates of macroeconomic variables or the overall state of economic activity have gained in importance in recent years, especially in periods of major crises, when reliable information on the intensity of the effect of a certain phenomenon or measure is very important for economic activity. This type of modelling is related to the concept of nowcasting, used to obtain the flash estimate of an economic variable when official data are not yet available. GDP is a particularly useful variable for nowcasting. Although not without its flaws, GDP provides the most complete picture of the state of economic activity at a given time. One of the shortcomings is certainly the time lag in the publication of data for the current period, given that the official flash GDP estimate is available one month after the end of the reference quarter, while the official GDP estimate during that quarter does not exist. The nowcasting models aim to overcome these shortcomings.

One of the main questions in the assessment of such models concerns the relevant indicators that can adequately approximate the movement of GDP. Indicators suitable for use in these models must be highly correlated with the GDP indicator, but also more frequent and regular in publication. Given the need for the simultaneous use of mixed frequency data, mixed data sampling (MIDAS) regression models were developed. In addition to them, dynamic factor models (DFM), as well as their combinations, are often used in econometric analyses.

Given the need for a more precise estimate of economic activity, in addition to official indicators, alternative indicators are increasingly being used in literature and practice. Alternative indicators are usually big data and, as such, are differently structured and usually much more frequent and timely than the official ones. Nevertheless, working with such data introduces new complexities into the model, and one of the answers to the growing computational complexity of such models is the increasingly frequent use of machine learning, especially deep learning for the purpose of nowcasting.

The paper contains the assessment of two nowcasting models – MIDAS and LSTM, of which the former is a traditional econometric model, and the latter is a deep learning model. The models were assessed using a combination of official and alternative indicators as regressors, with the target projection variable being GDP growth. Both models adequately model GDP dynamics in the short run, with the LSTM neural network model proving to be more accurate, but also more difficult for evaluation and interpretation than the MIDAS model. In addition, since the LSTM model does not provide insight into the individual contributions of the variables used, the MIDAS model is much more useful in identifying triggers for changes in GDP. Also, by looking at the structure of the MIDAS model, it can be concluded that alternative indicators are statistically significant for GDP nowcasts. Due to the use of alternative indicators and the nature of macroeconomic data, both models have the problem of small samples and, consequently, overfitting to the data. As the database grows, it is expected that the performance of both assessed models, especially the deep learning model, will improve.

Contents:

1 Introduction

An economist is an expert who will explain tomorrow why the things he predicted yesterday did not happen today.

Laurence J. Peter

Economic forecasting is not an exact science, nor are its research results always the most precise, as the quote from the beginning of this paper suggests. Uncertainty and various risks affect the probability of realisation of the projected movement of a certain macroeconomic variable. The longer the prediction horizon, the more unreliable the projections are, and the same is true when uncertainty is high, as is the case in crisis periods, when, as a rule, macroeconomic projections are most closely observed and monitored. In crisis periods, the monitoring of official indicators of economic activity and their projections indicates the likelihood of deepening of or exiting from the current crisis.

The crisis caused by the covid pandemic, directly followed by the crisis triggered by the Ukraine war, showed that a large number of economic indicators are not timely, which further complicates the process of decision-making about adequate measures aimed at preserving a stable economic environment. This is especially true for the economic indicator such as GDP, as the first economic consequences of the covid pandemic were felt already in January (transportation issues), while the GDP indicator confirmed this only in late April or early May in most countries. Therefore, a focus in this period is placed on GDP nowcasting models. These models have been analysed in economic literature for a number of years, but they are especially useful in times of sudden changes in economic activity, as they provide a quick estimate of the current effect on GDP.

When assessing the GDP nowcasting models, the explanatory variables must be adequately selected. Already in that stage, two problems arise, the first relating to the fact that a large number of official indicators whose movement is correlated with the movement of GDP also have a publication lag. Another problem is that most of these indicators are expressed in a frequency (most often monthly) that differs from quarterly GDP data.

To model actual economic activity in real time, nowcasts often rely on high-frequency data, such as daily or weekly indicators. The first chapter of this paper deals with the selection of adequate indicators – both official and those from alternative sources, which almost as a rule have a shorter lag between the end of the reference period and the publication of the flash estimate of the results in the given period. The second chapter concerns traditional econometric methods that focus precisely on solving the problem of different frequencies of dependent and explanatory variables. In the second chapter, the framework of mixed data sampling regression models (MIDAS) will be presented, and one such model on the example of Serbian GDP will be assessed.

In a significant part of the literature cited in the paper, the authors refer to the growing number of available economic indicators that have not been published by the official statistical office of a country. The exponential growth in the availability of a huge amount of data is related to the term big data – those are large, complex and diverse datasets generated from different sources. These data are generated at high speed, are large (large data quantity) and diverse (different formats and types of data). Such data can be included in traditional econometric models and are primarily used in various machine learning models, especially deep learning. The third chapter pertains to the development and application of machine learning in econometrics, with a focus on deep learning models used in nowcasting. This chapter also assesses a deep learning model, specifically a type of a recurrent neural network. A neural network consists of connected nodes, i.e. neurons. These nodes make up layers, including an input layer, one or more hidden layers, and an output layer. Data flow through the layers, activating functions within each neuron whose outputs provide input on further use of the data. Recurrent networks allow communication that is not exclusively one-way between layers and nodes of the network. The paper uses the neural network LSTM (long short-term memory), which is particularly suitable for modelling time series for nowcasting Serbia's GDP.

In the last, fourth chapter, two assessed GDP nowcasting models are compared – MIDAS and LSTM, with a focus on methods for improving the used models.

The aim of the paper is to select an adequate set of indicators – both from official and alternative sources, for the purposes of finding Serbia's optimal GDP nowcasting model.

2 Selection of adequate variables for GDP nowcasting

The Serbian Statistical Office, as the main producer and disseminator of statistical data, publishes a large number of announcements, indicators, bulletins and similar reviews. Over a number of years, those were the only available data for economic analysis and macroeconomic projections. In time, various sentiment measures and analyses of the subjective expectations of economic entities (such as the inflation expectations indicator and ESI) were added.

In addition to indicators based on surveys, with the emergence of the internet, a large number of indicators became available through user interaction with webpages. An example are the search indicators on the Google engine, used in this paper. During the covid pandemic, it was very important to monitor the Google mobility index, which shows trends and changes in movements in certain regions. A topic of particular interest for economic analysis was the impact of restrictions aimed at preventing the covid fallout on the economy. Ilin et al. (2021) demonstrated that mobility indicators (such as Google's) can be used to assess the effectiveness of non-pharmaceutical intervention and to predict the spread of covid. They found that mobility data alone are sufficient to predict coronavirus infections at all geographic scales – from counties and cities, through states and provinces, to countries and eventually the entire world. In addition, models that exclude mobility data perform significantly worse, indicating the important role of mobility data as a high-frequency and alternative data source for forecasting.

In this paper, four official monthly indicators and five alternative, high-frequency indicators are proposed for the purpose of short-term GDP projections and nowcasting.

2.1 Statistical Office's official monthly indicators

The goal of this paper is to find the most adequate model for the purpose of short-term GDP projections and nowcasting. The most commonly used economic indicator is GDP as it represents the total value of all goods and services produced in a country in a certain period (usually a year).

In Serbia, according to the production approach, the services sector accounted for the largest portion of GDP (51% on average in 2016–2022), followed by the industry sector, with around 21%. In terms of expenditure, personal consumption accounted for the largest portion of GDP – around 70% on average in 2016–2022, while private investment made up around 16%, and government investment around 5% on average. Net exports are under a strong impact of external factors, primarily the prices of energy and cereals, as well as external demand of Serbia's largest trading partner – the EU countries, primarily Germany. In the last seven years, with the exception of 2016, net exports negatively contributed to total GDP: -1.2 pp on average.

GDP is the key indicator used by governments, policy makers, businesses and economists to monitor and analyse a country's economic performance and progress. As such, it has a great weight in deciding on the adequacy of current and direction of future economic policies, and the construction of economic sentiment. The Statistical Office publishes preliminary data on quarterly GDP outturn (the so-called flash estimate) one month after the end of the reference quarter (e.g. the flash estimate of GDP in the first quarter is published on the last day of April). This means that the first estimate of the state of the economy is only available at the end of the fourth month. Given the importance of this indicator, for decades already, great effort is invested in economic literature to assess models that will adequately predict the movement of GDP in the short run, but also estimate the current situation (nowcasting).

The Statistical Office collects data on industrial production, retail trade and goods exports and imports on a monthly basis. The first data on the movement of these indicators are available on the last day of the month following the reference month (at the end of February for the January performance). The fact that these data are more frequent and up-to-date than economic activity indicators makes them suitable for assessing the movement of total GDP, if it turns out that they have predictive power. A part of this paper deals with the selection of the optimal set of indicators for forecasting GDP dynamics.

The connection between industrial production and the overall trend of economic activity has long been the subject of research. This relationship was particularly strong in periods when the largest part of added value was that produced in the industrial sector, before the services sector took over primacy. Nevertheless, since a large number of domestic and external shocks affecting industrial production also influence economic activity as a whole, the dynamics of this indicator can still provide insight into the movement of entire economic activity.

Back in the 1980s, Stock & Watson (1988) assessed the dynamic factor model in order to obtain an index of the movement of total economic activity. For the purpose of constructing the index, data on industrial production, real personal income, manufacturing output, trade and employment were used.

Particularly interesting is the research conducted by the ECB in 2007, relating to the impact of the lag between the end of the reference period and the publication of the indicator flash estimate. Ignoring the differences in publication lag, the authors concluded that data on actual activity (especially industrial production) are the most important source of information. However, when their less timely publication is taken into account, data on actual activities become much less relevant, and survey data take their place. In very late GDP forecasts, made in the last month of the corresponding quarter and thereafter, the industrial production data still contain significant information (Bańbura & Rünstler, 2011).

The paper produced by the Fed of Dallas is also interesting – when assessing the model, they do not take into account the flash estimate or final data on economic indicators, but rather their combination, as well as the difference created in the revision of these data. This is illustrated by the model using monthly industrial production, employment and retail sales (Koenig et al., 2003).

In the papers by Stock & Watson (1988) and Koenig et al. (2003), in addition to the industrial production indicator, the retail trade indicator was also used for the purpose of assessing the overall economic activity.

The paper of the US Treasury is also interesting – one of the most important economic indicators is the dynamics of the retail sales index. Retail sales data are released mid-month, and the retail sales component goes directly into the GDP calculation for that quarter. If only one month of data is available (for forecasts at the beginning of the reference quarter), it is often assumed that the data for the remaining two months of the quarter will be unchanged or the trend observed will continue (Kitchen & Monaco, 2003).

Also, retail trade proved to be a statistically significant indicator for euro area GDP nowcasts using the Kalman filter in bridge models (Angelini et al., 2011), as well as for estimates of Canada's monthly GDP (Mourougan, 2006).

Many papers have explored the connection between foreign trade and economic activity as a whole. A widely known paper was published by Balassa (1985), who states that the purpose of including exports (together with capital and labour force) in the production function is to test the hypothesis that export orientation increases total factor productivity through beneficial effects on more efficient resource allocation, capacity utilisation, economies of scale, and technological change. The results obtained show that exports significantly contributed to the economic growth rate, but also much improved the explanatory power of the equations.

The Bank of France used the bridge model for the purpose of estimating monthly GDP, which includes both export and import variables, bearing in mind that the condition for including the variable in the model is at least the monthly frequency and a publication lag of less than two months after the end of the reference period. Apart from official data concerning imports and exports, European Commission survey data were also used. When it comes to the link between official import and export indicators and survey expectation indicators, the link with exports is quite direct, as this variable refers to the activity of France's economic partners and, as such, represents a substitute for external demand directed to France. The connection with imports is less obvious, but it is clear that higher imports from France will stimulate the activity of trading partners, with the economic cycles of European countries being quite aligned. When assessing the model, the authors concluded that foreign trade variables are extremely difficult to project due to their high volatility (Barhoumi et al., 2012).

For the purpose of Spain's GDP nowcasting and short-term forecast, a dynamic factor model was used, containing ten monthly indicators, eight of which relate to real activity, including goods imports and exports, as well as two survey-based data (Arencibia Pareja et al., 2020).

2.2 High-frequency indicators

The paper already touched upon the importance of timely data about the movement of economic activity for policy makers and economic agents, this being the most important input in making personal and business decisions. The purpose of a great body of research estimating GDP level before the official publication of the actual result is to find economic activity indicators which approximate this measure well, and with a smaller lag between the closing of the accounting period and data publication. These indicators gain particular importance in periods of departure from projected or long-term growth, during crises or sudden expansions. It was during the coronavirus-induced crisis that adequate business cycle indicators came to the fore as one of the main research issues. The potential candidates are the abovementioned Google trends and Google mobility index, as well as electricity consumption and production measures, air pollution measures, etc. Using some of the mentioned indicators, the models estimating real GDP growth at a given moment are assessed. In this paper, for the purpose of a short-term GDP projection, we used Google trends and electricity consumption indices, among other things.

2.2.1 Google trends indicator

One of the indicators increasingly used in the past years as an explanatory variable in the forecast and assessment of economic activity is the Google Trends indicator. Google Trends (*trends.google.com*) allow us to access the search dynamics of a specific term or a category in a specific territory at any given time, and the database itself stores data from 2004 onwards. The index calculation method is specific, as the database values do not represent the scale of the searches, but are normalised and range from 1 to 100, where 100 represents the point with the highest interest recorded in the reference period. One of the biggest advantages of using Google Trends is that the data are available at any time during the calculation period for the past performance, that is, in real time.

In his paper, Wołoszko (2020) used over 200 categories in developing weekly GDP estimates for OECD countries. Owing to the nature of neural networks, which are good at modelling various non-linear relationships, there was no need to make any assumption about the type of correlation between GDP growth rates and Google Trends. This paper focused particularly on the fact that the emergence of high-frequency indicators (though they may not have been created with macroeconomic modelling in mind) enabled the use of deep learning techniques that require a large database for optimal functioning. The model from the abovementioned paper successfully signalled the effect that the outbreak of the pandemic would have on economic activity of 45 out of 46 OECD countries and, overall, it had a lower quarterly GDP prediction error than the benchmark autoregression model. The results from the paper about the expected relationship between searches and GDP movements are in line with intuition (e.g. higher searches around *investment* category may signal higher GDP growth, while an increase in searches involving *crisis* and *recession* categories indicates lower growth).

Schmidt & Vosen (2009) were among the first to present in their paper a new indicator for private consumption as a component of GDP, based on Google Trends. Given that private consumption is the single most important component of GDP, timely and precise determination of its dynamic provides an insight into the current state of economic activity. Wu & Brynjolfsson (2015) used Google Trends as an explanatory variable in a model that predicts real estate prices, and claimed that economic predictions from this type of internet search based data can be applied to any market where internet search precedes a transaction, even when the transaction itself does not occur in virtual space. Choi & Varian (2012) state that the assertion made in their paper is not related to predicting future using Google Trends, but to estimating the present, thus entering the field of nowcasting. Google Trends have also been used in forecasting exchange rate trends (Markiewicz et al., 2018) and overcoming information gaps in developing countries (Narita & Yin, 2018).

A particularly interesting methodology of selection and processing of Google Trends data can be found in Wołoszko (2020), including the use of categories and not indices of searches created based on key terms, an approach applied in this paper as well. The paper also undertakes their transformation to address a few problems relating to the manner of data collection and presentation. The key problem with these data is that they show a relative significance of searches of a certain category relative to the total number of searches, which is then multiplied by a constant, so that at any time the index for the period when the search for that term was the most significant is 100, i.e.:

$$SVI_{ct} = \frac{SV_{ct}}{SVT_t} * C_c, \tag{1}$$

where *SVIct* is a relative share of searches by category *c* in period *t*. In this equation, only the denominator, i.e. the total number of searches, depends exclusively on the period *t*. The problem with this manner of index calculation is the fact that as the total number of searches increases, the relative share of relevant categories inevitably goes down, thus introducing bias in modelling. By transformation (1), we get:

$$svi_{ct} = log(SVI_{ct}) = sv_{ct} - svt_t + c_c. \tag{2}$$

Keeping in mind that *SVTt* remains unchanged for all categories in period *t*, we may isolate it by using the principal component analysis over the logarithmic series of *SVI*, which isolate the trend by means of the *HP* filter. The resulting first component is then recalculated to have the same mean and standard deviation as the average of the logarithmized *SVIs* and subtracted from the logarithmized *SVIs*. The transformation was carried out by using the y-o-y growth rates to overcome the seasonality problem, and indicators were taken from the source in the form of monthly indices.

In this paper, categories used as potential regressors are *business and industry, finance, real estate and travel.*

2.2.2 Electricity consumption

Apart from Google Trends, one of the high-frequency indicators proposed as a potential regressor is the variable pertaining to electricity consumption. To collect data on electricity consumption in the territory of Serbia we used the ENTSO-E Transparency Platform (European Network of Transmission System Operators for Electricity), a joint venture of 39 official European transmission system operators (TSO), including Serbia's Elektromreža Srbije. TSOs are agents working independently from other participants in the electricity market, responsible for electricity transmission through the main high voltage electrical grids. TSOs enable network access to electricity market participants and guarantee safe system operation and maintenance. Among other things, through membership in ENTSO-E, TSOs are obliged to submit certain data which are made publicly available on the website. Among these publicly available data one can also find the data about the actual total load in the area of TSO operation within one hour. These data must be submitted no later than one hour after the end of the accounting period, which makes them timely. The data may be downloaded directly from the platform or using *API*, as a more efficient option for download and processing of large databases such as this one. In this paper data were downloaded by software *R* and converted from hourly to monthly amounts.

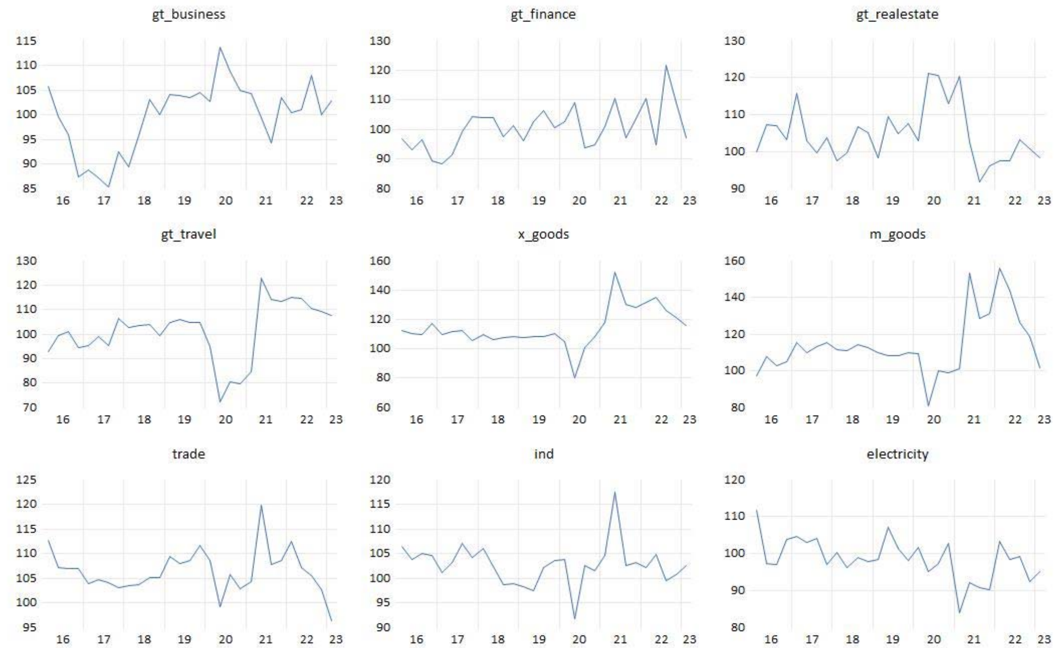
A question arises as to the relevance of including this piece of data in models for the projection of GDP growth rate, and/or justifiability of using electricity consumption as an indicator of economic activity. Not many papers have been published on this topic and one of them looked into the relationship between industrial output and electricity consumption, among other things. Using Johansen's cointegration methodology, Sun & Anwar (2015) concluded that there is a statistically significant long-term relationship when observing electricity consumption, industrial production and entrepreneurship, on the example of Singapore.

In empirical terms, the Serbian industry consumes over 30% of total available electricity (source: energy balances for 2021). It may for example be assumed that volatility in energy consumption can be attributed to shocks affecting industrial production, because the household sector is made up of a large number of small units whose consumption should be relatively stable. However, given that this paper deals with the projection of overall economic activity, there is no need to disaggregate total energy consumption, but only determine whether it represents a statistically significant explanatory variable in nowcasting and short-term GDP projections.

2.3 Transformation of used variables and checking of stationarity and multicollinearity

The data used in this paper were transformed in indices describing y-o-y dynamics, when it comes to quarterly and monthly data. As stated above, hourly data on electricity consumption were aggregated into monthly amounts and then transformed to y-o-y indices. Plots of series indicate potential stationarity around a constant.

Figure 1 **Graphical overview of independent variables**



A formal application of the ADF test without added lags, which tests the stationarity around a non-zero constant, with the significance level of 5%, indicates that the series are indeed stationary around a constant.

Further, it needs to be checked whether there is a problem of multicollinearity within the database. A certain degree of correlation between the indicators is certainly expected, considering that there is a great number of factors affecting the dynamics of all the mentioned indicators, mainly from the external environment (prices of products established in the global market, such as cereals and oil, the degree of external demand, imported inflation, etc.), but also from the domestic one (such as inflation rate). Based on the multicollinearity analysis, we can conclude there is a high degree of correlation between exports and imports of goods, which is somewhat expected, as well as between these variables and the Google Trend index concerning travel. Attention should be paid to these variables and their interaction when constructing the model.

3 Econometric model used in nowcasting

The approach of reducing a large number of variables or time series to just a few factors was first mentioned in 1946 (Burns & Mitchell, 1946), and concerns the business cycles analysis. Namely, even at that time it was defined that while there is a great number of indicators which can move procyclically or countercyclically relative to the phase the economy is in, there is still a certain general tendency toward recession or expansion, not necessarily noticeable in the same way as the number of the unemployed or the share of non-collectible loans. Further, these indicators of overall economic activity may be simultaneous (indicating the current state of the economy), leading (indicating the direction of future movement) or lagging (confirming tendencies that are underway or have occurred). As for the current projection of any variable, a successful forecasting requires us to determine a set of explanatory variables that best describe the movement of the forecasted variable. Ideal variables for these models would be those that are simultaneous and leading, and that predict current or future movements with great precision. Simultaneous indicators are particularly important if the lag period in publishing indicator value is shorter than in the variable that is being modelled.

The selection of potential regressors from the previous chapter is based on the principle of simultaneous indicators with the shorter lag period and leading indicators which indicate future changes in economic activity. Papers using the retail sale index have labelled it as a leading indicator, in the sense that it indicates emerging changes in the economy. We can add Google indices and electricity consumption, apart from the retail index, as leading indicators, while industrial production and export and import of goods are simultaneous indicators which have a relatively short lag period relative to GDP data.

3.1 Literature review

One of the most common models used for GDP nowcasting is a MIDAS model. This approach to data modelling was introduced for the first time in Ghysels et al. (2002), which presented the advantages of this approach for mixed data frequencies. According to authors, the typical regression models up to that time assumed the use of time series collected in regular and uniform time intervals. However, it is often the case that relevant information is contained in high-frequency indicators, while series that need to be modelled are of lower frequency, the typical example being macroeconomic data and the potential for modelling quarterly GDP data based on monthly indicators.

In their paper, Armesto et al. (2010) raised the question of the most efficient approach to mixed data modelling. They concluded that in models relying on a large number of explanatory variables, due to the problem of a large number of potential parameters, more parsimonious models may perform better, that is, aggregation models and MIDAS models, while the advantage of MIDAS models is that they also allow projections within the period.

Ghysels et al. also looked into the justifiability of incorporating financial data (which have a much higher frequency than macroeconomic variables) into macroeconomic projections. The premise is that a great number of daily financial time series contain information relevant for the further movement of economic activity that is lost through aggregation. The models used in the paper show that this hypothesis is accurate, i.e. that by using variations of MIDAS models with regressors from the financial sector, we arrive at a more reliable projection.

Frale & Monteforte (2010) combined two approaches in their work and proposed a mixed-frequency factor model, where high-frequency indicators are treated with a mixed data sampling regression model approach (FaMIDAS). They concluded that such models provide good assessments of monthly GDP and short-term quarterly projections.

3.2 Theoretical framework of MIDAS models

The theoretical framework was adopted from the papers by Ghysels et al. (2004), Ghysels et al. (2016) and Sinko (2008).

MIDAS models cannot be autoregression models due to the nature of AR models requiring equality of frequencies, but are most similar to distributed lag models.

A simple MIDAS model has the following structure:

$$Y_t = \beta_0 + B\left(L^{1/m}, \Theta\right) X_t^{(m)} + \varepsilon_t^{(m)} , \tag{3}$$

where m indicates how much the frequency of the explanatory variable is higher than the dependent variable, so in the case of modelling annual data with quarterly explanatory variables, m would be 4. Further,

$$B\left(L^{1/m}, \Theta\right) = \Sigma_{k=0}^{K} B(k, \Theta) L^{k/m} \tag{4}$$

$$L^{1/m} X_t^{(m)} = X_{t-1/m}^{(m)} \tag{5}$$

mathematically, there would be equality between $L^{k/m}$ coefficient multiplied with $X_t^{(m)}$ and the value of $X_t^{(m)}$ with *k/m* lags. In the example with annual and quarterly frequencies, this means that this year's value Y_t is projected to the database with quarterly data on the value $X_t^{(m)}$ up to the K^{th} quarter back. Theoretically, the number of parameters for assessment from the polynomial $B\left(L^{1/m}\right)$ is final for the purpose of simplification, keeping in mind that one b_k needs to be assessed for each period, and in case of high-frequency data (daily or hourly), that is an extremely large set of parameters to be assessed. Empirically, this problem was initially treated by aggregating high-frequency data so that their frequency would be of the same order as in a low-frequency, dependent variable, in the absence of models which allow for different frequencies. The aggregation approach would lead to neglecting some potentially important information through series transformation. In order to overcome this problem, vector Θ is introduced in MIDAS models in the function from which the model parameters are assessed. There are several potential specifications $B\left(L^{1/m}, \Theta\right)$ which may be used for the assessment of model parameters. Besides, the use of parametric function $B\left(L^{1/m}, \Theta\right)$ also enables a more efficient selection of lags that need to be included in the model. The selection of an adequate form of vector Θ actually makes it possible for the selection of the number of lags to be imposed by the structure of data themselves (Sinko, 2008).

In the matrix representation, an example of a regression model of data of different frequencies in which the dependent variable y_t is quarterly and whose dynamics is modelled using the lags of the dependent variable itself and the monthly variable x_τ and its lags, has the following form:

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \alpha_1 + \begin{bmatrix} x_6 & \cdots & x_1 \\ \vdots & \vdots & \vdots \\ x_{3n} & \cdots & x_{3n-5} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}. \tag{6}$$

Thus, in every quarter t, considering that $m = 3$, the dependent variable y_t is a linear combination of variables $x_{3t}, x_{3t-1}, x_{3t-2}$ from the current quarter m, and y_{t-1} and $x_{3(t-1)}, x_{3(t-1)-1}, x_{3(t-1)-2}$ from the previous quarter $t-1$.

In this way, a kind of frequency matching is performed, since the high-frequency variable x_τ is transformed into a lower-frequency vector $(x_{3t}, \ldots, x_{3t-5})^T$. It should be noted that for this type of transformation, the number of observations in variable x_τ should be exactly $3n$. In cases that are common in practice and which concern the so-called missing edges, i.e. lags and different times of publication of official data, it is necessary to either use balanced official data or estimate the missing values by using a model or applying the Kalman filter.

On the other hand, if we wanted to add to the model (6) another explanatory, high-frequency variable z, which is published on a weekly level, the model would need to be expanded. The use of weekly (as well as daily and business day data) reveals another limitation of multi-frequency regression models, i.e. the fact that these models do not allow variations in the number of observations within one calculation period m that refers to the dependent variable. This means that each month must strictly have four weeks, i.e. in this illustration each quarter must have 12 weeks. In the case of variable z_τ, for the purpose of frequency matching, the value of m equals 12. Apart from the abovementioned variables and lags, we are adding to the model observations $z_{12t}, z_{12t-1}, \ldots, z_{12t-11}$ from the current quarter m and $z_{12(t-1)}, z_{12(t-1)-1}, \ldots, z_{12(t-1)-11}$ from the previous quarter $t-1$. In matrix form:

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \alpha_1 + \begin{bmatrix} x_6 & \cdots & x_1 \\ \vdots & \vdots & \vdots \\ x_{3n} & \cdots & x_{3n-5} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix} + \begin{bmatrix} z_{24} & \cdots & z_1 \\ \vdots & \vdots & \vdots \\ z_{12n} & \cdots & z_{12n-23} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}. \quad (7)$$

The frequencies of some high-frequency variable x_τ are generally matched through transformation into a low-frequency vector $(x^{(i)}_{tm_i}, x^{(i)}_{tm_i-1}, \ldots, x^{(i)}_{tm_i-l})^T$, whereby we obtain a matrix representation of the general model in the following form:

$$\begin{bmatrix} y_l \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{l-1} & \cdots & y_{l-p} \\ \vdots & \vdots & \vdots \\ y_{n-1} & \cdots & y_{n-p} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \sum_{i=0}^{k} \boldsymbol{X}^{(i)} \begin{bmatrix} \beta^{(i)}_0 \\ \vdots \\ \beta^{(i)}_l \end{bmatrix} + \begin{bmatrix} \epsilon_l \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (8)$$

$$\boldsymbol{X}^{(i)} := \begin{bmatrix} x^{(i)}_{um_i} & x^{(i)}_{um_i-1} & \cdots & x^{(i)}_{um_i-l} \\ x^{(i)}_{(u+1)m_i} & x^{(i)}_{(u+1)m_i-1} & \cdots & x^{(i)}_{(u+1)m_i-l} \\ \vdots & \vdots & \cdots & \vdots \\ x^{(i)}_{tm_i} & x^{(i)}_{tm_i-1} & \cdots & x^{(i)}_{tm_i-l} \\ \vdots & \vdots & \cdots & \vdots \\ x^{(i)}_{(n-1)m_i} & x^{(i)}_{(n-1)m_i-1} & \cdots & x^{(i)}_{(n-1)m_i-l} \\ x^{(i)}_{nm_i} & x^{(i)}_{nm_i-1} & \cdots & x^{(i)}_{nm_i-l} \end{bmatrix}, \quad (9)$$

where n is the number of observations of the dependent variable, m is frequency, p is the number of lags of the dependent variable included in the model and u is the smallest integer to which $um_i - l > 0 \; and \; u > p$ applies.

When defining the MIDAS models, the notion of parametric functions was introduced. The selection of an adequate parametric constraint function leads to more desirable outcomes when it comes to model performances. The parametric constraint function and frequency matching are interdependent concepts which help to properly define the relationship between high-frequency and low-frequency variables. The goal is to approximate, as best as possible, the relationship between these variables despite their different frequencies.

The parametric constraint function plays the key role in frequency matching. It determines a functional form of the relationship between high-frequency and low-frequency variables. By imposing appropriate limitations, the function ensures that high-frequency information is appropriately mapped in the low-frequency domain.

The parametric constraint function provides a framework for defining how a low-frequency variable depends on the value or lags of the high-frequency variable. The function establishes a connection between two different frequency domains and allows for a meaningful interpretation and analysis of the relationship between variables observed at different frequencies.

There are several potential parametric constraint functions which may be used in modelling, including, inter alia, the exponential Almon polynomial, beta function (analogous to the probability function), Gompertz function, logarithmic Cauchy function. For the purposes of GDP forecast, considered in the remainder of this paper, Almon's exponential lag function is used (Almon, 1965).

This parametric function rests on the Weierstrass theorem. According to the Weierstrass theorem about extreme value (known as the extreme value theorem), used in mathematical analysis, if function *f(x)* is continuous on the closed and bounded interval [*a*, *b*], then *f(x)* must attain a maximum and a minimum on that interval, each at least once. Further, that function may be approximated by a polynomial *p(x)* of some order *P*. Take the following model for example:

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ \dots \ + \beta_n x_{t-n} + u_t, \quad t = 1, 2, \ \dots, T. \tag{10}$$

By applying the Weierstrass theorem, Shirley Almon concluded that the parameters $\beta_0, \beta_1, \ \dots, \beta_n$ can be approximated by the following polynomial of order *P*:

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_p i^P. \tag{11}$$

That is, by replacing (11) in (10),

$$y_t = a_0 x_t + (a_0 + a_1 + a_2 + \cdots + a_p) x_{t-1} + (a_0 + 2a_1 + 4a_2 + \cdots + 2^P a_p) x_{t-2} + \dots + (a_0 + na_1 + n^2 a_2 + \cdots + n^P a_p) x_{t-n} + u_t \tag{12}$$

where it applies:

$$
\begin{aligned}
y_t &= a_0 z_{0t} + a_1 z_{1t} + a_2 z_{2t} + \cdots + a_P z_{Pt} + u_t \\
z_{0t} &= \left(x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-n}\right) \\
&\vdots \\
z_{pt} &= \left(x_{t-1} + 2^P x_{t-2} + \cdots + n^P x_{t-n}\right) \quad .
\end{aligned} \tag{13}
$$

The use of Shirley Almon's polynomial when assessing the MIDAS model reduces the initial model with a large number of high-frequency parameters and their lags to a model with the smaller number of parameters for assessment.

3.3 Assessment of the MIDAS model

Standard statistical tests for establishing model validity and adequacy are used for assessing MIDAS models. It has already been determined that both the dependant variable (y-o-y GDP growth index) and all potential regressors (goods export and import, retail trade and industrial production, as well as indices of Google trends and electricity consumption) are stationary and that there is no significant problem of multicollinearity. Still, due to the relatively small sample (29 quarterly data and 85 monthly data for each potential regressor), variables adequate for the model needed to be selected. Also, the model was assessed based on the period Q1 2016 – Q4 2021, with the last six quarters being used for checking the validity of the forecast. Initially, the included variables were the Statistical Office's official indicators, as well as the electricity consumption index. The iterative process, monitoring the Akaike Criterion, showed that the best complement to official data is Google's index that pertains to business topics (gt_business), as well as the index for the topic of finance (gt_finance), in addition to the already present electricity consumption index.

A specificity of MIDAS models (as indeed of all models for GDP nowcasting) is that the primary indicator of the model's successfulness is the root mean forecast error (RMFE), even at the cost of including regressors that are not statistically significant by their p-value. Also, by monitoring the RMFE and the Akaike Criterion, we needed to decide whether a larger number of variables yields a greater contribution than a larger number of lags. As presented in the theoretical overview, the number of lags included in the model does not equal the number of model parameters as the impact of individual variables and their lags is assessed using a parameter function, in this case the Almon polynomial function. Given that we used monthly data, and that the quarterly GDP datum is being modelled, we chose the third order polynomial. In addition, the very procedure has the option of automatic selection of significant lags given that the model specification does not follow an already published model and is therefore not limited by the previously selected number of lags; rather, only the lags that turn out to be statistically significant in the given sample are included in the model. The initial model included all of the given potential regressors, and the principle of monitoring the RMFE together with the statistical significance of the given regressor iteratively excluded regressors, until the final model specification was selected. In regard to this, the model that performed the best on the base of proposed indicators is the following:

Figure 2 **The assessed MIDAS model**

Dependent Variable: GDP
Method: MIDAS
Date: 06/27/23 Time: 16:39
Sample (adjusted): 2016Q3 2021Q4
Included observations: 22 after adjustments
Method: PDL/Almon (polynomial degree: 3)
Automatic lag selection, max lags: 6
Chosen selection: 4 5 4

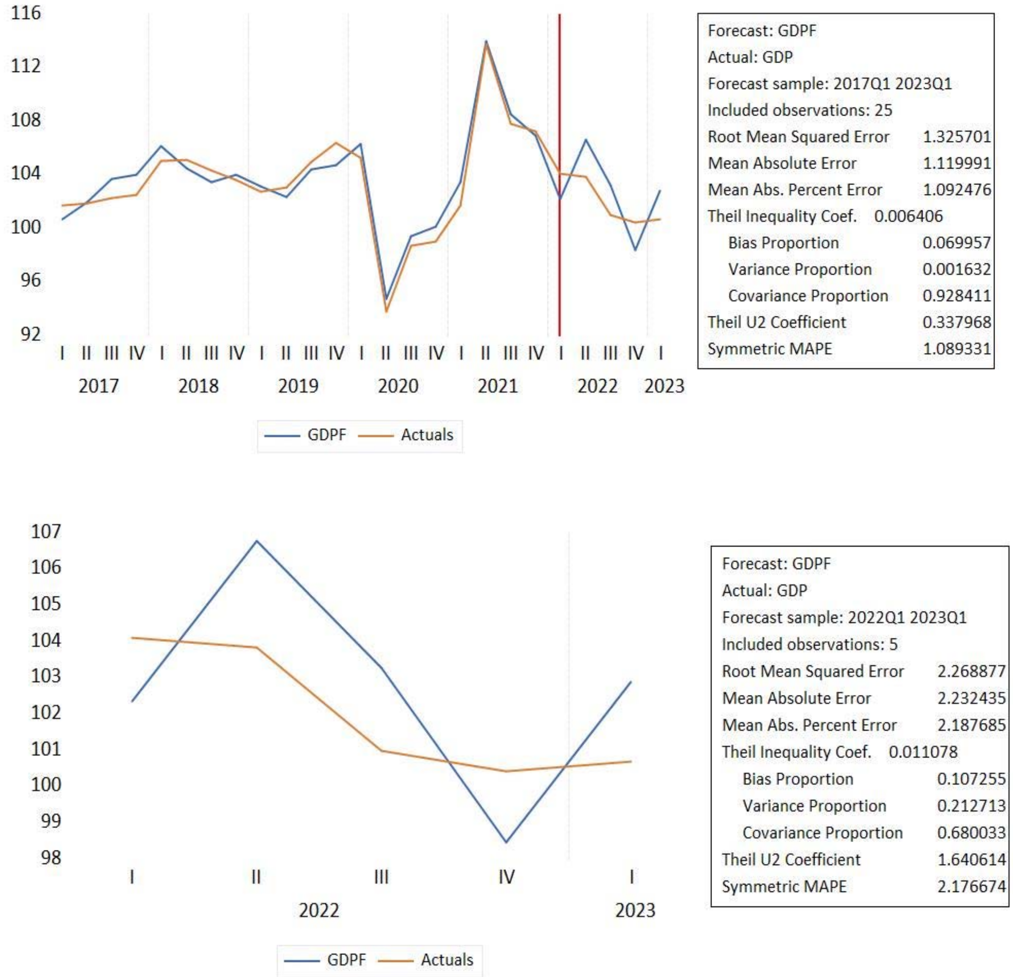
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP(-1)	0.701371	0.080774	8.683141	0.0000
Page: GT_BUSINESS Series: GT_BUSINESS Lags: 4				
PDL01	0.127017	0.224591	0.565549	0.5821
PDL02	0.075517	0.203254	0.371540	0.7167
PDL03	-0.041929	0.040506	-1.035135	0.3210
Page: IND Series: IND Lags: 5				
PDL01	-0.700033	0.173006	-4.046296	0.0016
PDL02	0.845103	0.144796	5.836520	0.0001
PDL03	-0.161418	0.025421	-6.349781	0.0000
Page: ELECTRICITY Series: ELECTRICITY Lags: 4				
PDL01	-0.242987	0.166338	-1.460801	0.1698
PDL02	0.301689	0.162097	1.861158	0.0874
PDL03	-0.068253	0.033511	-2.036719	0.0644

R-squared	0.939454	Mean dependent var	103.5311
Adjusted R-squared	0.894044	S.D. dependent var	3.829498
S.E. of regression	1.246536	Akaike info criterion	3.581570
Sum squared resid	18.64624	Schwarz criterion	4.077498
Log likelihood	-29.39727	Hannan-Quinn criter.	3.698396
Durbin-Watson stat	2.074969		

The final model includes one dependent variable lag (GDP(-1)), Google's business trend index (GT_BUSINESS) with four selected lags, the industrial output index with five selected lags (IND), as well as the electricity consumption index (ELECTRICITY) with four selected lags. All of the explanatory variables are monthly frequencies, while the regressand is quarterly. Thus, a total of 13 parameters were assessed using nine Almon coefficients.

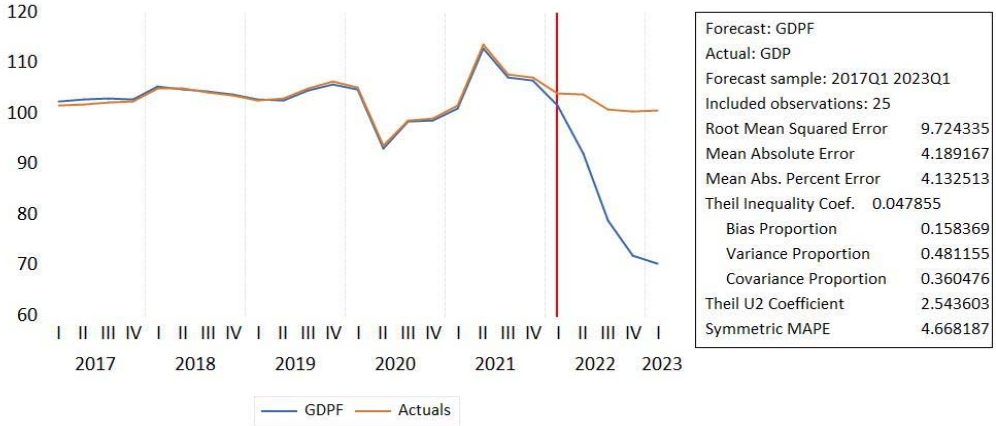
The selected model has the following performances in terms of both in-sample and out-of-sample error:

Figure 3 **Projections of the assessed MIDAS model and out-of-sample model projection**



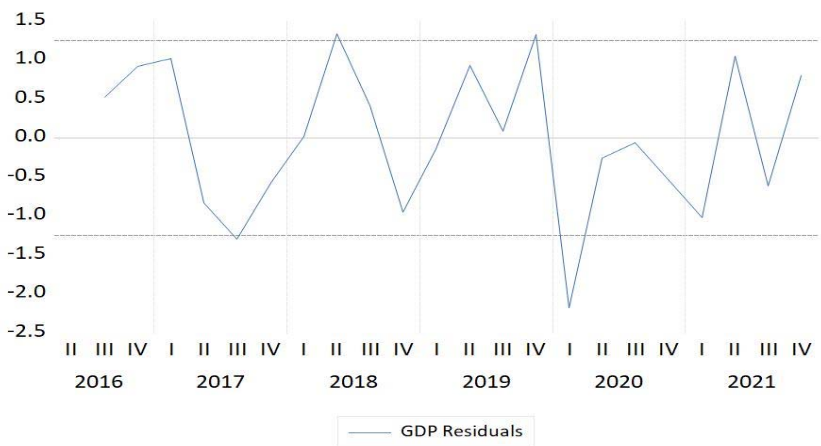
Individually, the Google trends indicator used in the model is not statistically significant, however, its exclusion has shown to be unfavourable to the performances of out-of-sample forecast. The potential reasons for the lack of statistical significance are reflected in the fact that this is a relatively small sample for a model of this size, in the way in which data about Google trends are being collected (though this was partly neutralised by the use of growth rates), as well as in the fact that the correlation rate between movements of this index and the industrial forecast index, though insufficiently high on its own to be problematic, is not negligible (-0.4). Moreover, MIDAS models assessed on such a small sample generally imply the issue of overfitting, i.e. the tendency to “remember” data instead of abstracting the key relations, which is reflected in a significantly smaller error within the sample, with a poorer forecast and bigger out-of-sample forecast errors. Exclusion of the business trends index leads to a significantly poorer out-of-sample forecast and the bigger presence of the problem of overfitting given that the model’s main goal is to achieve a forecast as accurate as possible, it is justified to include a variable that contributes to the reduction of the RMFE, though in itself it is not necessarily significant as an explanation of the dynamics of the dependent variable. Besides, if we compare the selected model containing two variables that are not official monthly indicators of economic activity with the model containing only the indicators published by the statistical office, we can see that the inclusion of Google's trends and electricity consumption in the model led to a better situation as regards this issue, but did not reduce the forecast error. Below, we can see the performance of the MIDAS model which contains regressors such as goods export and import, retail trade and industrial production, and where the issue of overfitting is evident.

Figure 4 **Projections of an alternative MIDAS model**



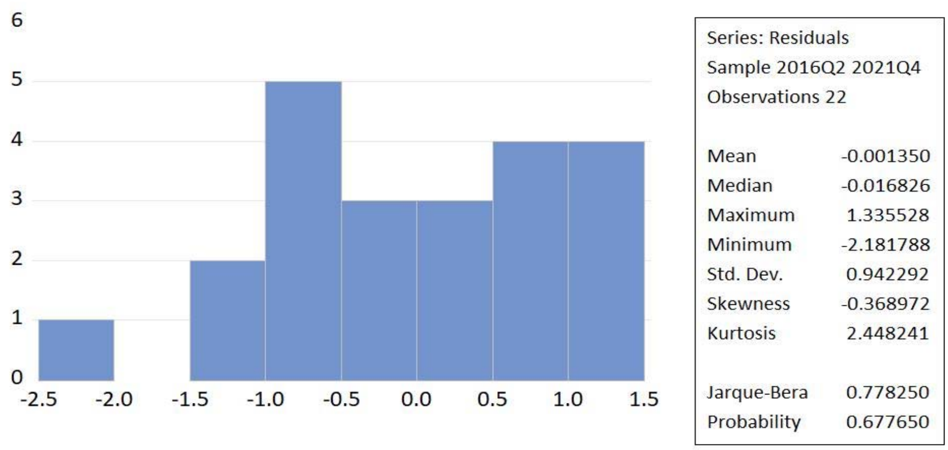
A further test of the model's adequacy is reflected in the testing of model residuals for the selected model, whose specification was previously shown in Figure 2 in this part.

Figure 5 **Residuals of the assessed MIDAS model**



Based on the chart, we can assume that the residuals are normally distributed, which can be verified using the Jarque-Bera normality test. The test results for the chosen MIDAS model are given below.

Figure 6 **Jarque-Bera normality test of residuals for the assessed MIDAS model**



The p-value for the Jarque-Bera normality test indicates the non-rejection of the zero hypothesis of the normality of residuals.

4 Machine learning models in econometrics

A specificity of the field of machine learning is partly reflected in the fact that there is no concrete overview of the history and development of this discipline (Plasek, 2016). The author stated an interesting assertion regarding the nature of the discipline, namely the fact that machine learning systems usually have the status of an experiment because they are used both as a confirmation and a counterweight to what we know. Still, the beginning of this discipline is usually associated with psychologist Frank Rosenblatt of the Cornell University, who invented a letter-recognition machine based on the ideas about the way the nervous system works. Called the Perceptron, the machine is considered the prototype of modern neural networks (Fradkov, 2020).

According to Murphy (2012), the goal of machine learning is the automatisation of the pattern-recognition process in data so that the patterns are relatively good at predicting future movements of the relevant variables. The need for this type of modelling arose with the growing amount of available data, i.e. it goes hand in hand with the big data concept, which is believed to be the first turning point in machine learning. The second turning point is associated with the costs of processing a large amount of data in increasingly complex systems, without adequate processors. Google and HeadsUp found a solution to this. The third turning point is the booming growth of deep learning.

Regardless of the lack of systematisation in this field, which is due to its explosive growth, Alzubi et al. (2018) underlined several key events that defined the development of this discipline, some of them being the formulation of the Turing test, presentation of the artificial intelligence concept and then of the algorithm for pattern presentation, triumph of the IBM's computer of Garry Kasparov in chess and the creation of neural networks able to recognise patterns and faces.

As the complexity of econometric models increases, which is a natural consequence of the growing complexity of interactions between various economic entities due to the emergence of the internet, cashless and online trade and similar features of the time we live in, including the increasing availability of databases, a need arose to improve existing models and make them more efficient. One way to achieve this is to introduce machine learning in econometrics, which we can define as the ability of the system to learn from data in a way that will ensure the automatisation of the model construction process and more efficiently resolve the problems that might occur (Janiesch et al., 2021).

Machine learning implies the application of algorithms that learn in iterations from problem-specific data, thus enabling an insight into the relations and patterns for which they are not explicitly programmed. There are several types of machine learning: supervised learning, unsupervised learning and reinforcement learning (Sarker, 2021). Supervised learning implies providing inputs and outputs so that the algorithm can identify the existing relations and create an adequate model describing those relations. Unsupervised learning requires the algorithm to find structure in data sets that are not marked in any way. Reinforcement learning involves a dynamic environment where the algorithm receives a feedback reaction from the system after performing an action, and the goal is to maximise the positive feedback reaction. Deep learning, which is in the focus of this paper, lies within machine learning.

Generally, we need three things to create a machine learning model:

- input, i.e. data to be analysed (e.g. if we are analysing a sentiment, we need textual data to assess the sentiment present);
- examples of expected output, i.e. the results we expect to get from data (textual data based on which we analyse the sentiment must have associated sentiment assessments);
- the manner of measuring the successfulness of the algorithm, i.e. the difference between the assessed and actual value based on which the manner of algorithm functioning will be adapted (this is actually the "learning" part of machine learning).

In fact, all these elements are also present in classic modelling, but this does not necessarily mean that a machine learning method can be applied to any econometrics problem. When an adequate model is applied, we certainly get satisfactory results when it comes to modelling in-sample data, as the model will "learn" to recognise the patterns present. However, it is questionable whether the model will be efficient in predicting values outside of the sample on which it was trained. Namely, machine learning models (and deep learning in particular) are highly prone to overfitting to data (Chollet et al., 2022), which we have already covered when assessing the MIDAS model.

When making a good machine learning model, a sufficiently large database is very helpful. In their paper, Zhang & Ling (2018) called machine learning a data-driven approach and emphasized that the key ingredient behind the success of these models is the recognition of patterns based on a sufficiently large number of learning inputs. Neural networks go beyond classical machine learning methods only when databases exceed a certain amount of data, and the model's performances begin to improve at a higher rate compared to classical algorithms.

Until a few years ago, the only data relevant for econometric analyses could be obtained from official statistical offices and relevant institutions. Today there is a multitude of high-frequency, relevant and available data that can be used to explain movements and predict economic variables, and machine learning allows us to assess the model that will most efficiently process such amount of input data.

Varian (2014) stated that the purpose of econometric analysis is in one of these four categories: prediction, summarising, assessment and hypothesis testing. Machine learning is primarily about predictions, therefore this paper assesses the quality of the machine learning model through its successfulness in predicting future movements of relevant variables.

From the very start, machine learning models were shown to be extremely useful in classification tasks. Some of the most famous machine learning algorithms are the following:

- Naïve Bayes, a classification algorithm based on the Bayes theorem, whose assumption is the mutual independence of characteristics of input data (hence the "naïve").

- Kernel methods, which include the support vector machines, a method that gained its current form in a paper from 1995 (Cortes & Vapnik, 1995), where it was defined as a new methodology for classification of two-class problems. The idea is that input vectors are mapped in a non-linear manner into a multidimensional space with characteristics, where the output is again a linear decision on input classification. The high level of generalisation with this approach is given special significance, and its essence is in finding optimal limits for decision-making as to the attribution to a specific class. In a multidimensional space, the question of setting optimal decision-making limits is not associated with determining the coordinates, but rather distances between the dots that represent inputs with the help of the kernel method. These models are often considered shallow representations of neural networks.

- Decision trees and tree-like methods, such as random forest and gradient boosting. The random forest algorithm involves many decision trees based on which the final output is created. Gradient boosting algorithms function in a similar manner and use the amplification of gradients to improve performance in dots where the previous model iterations turned out to be the weakest.

This paper deals specifically with the subsegment of machine learning that pertains to deep learning, more precisely neural networks, which has been increasingly present in time series modelling over the past decade. Deep learning is an approach to learning patterns from data which emphasizes the learning of sequential layers with growing efficiency. "Deep" in deep learning does not pertain to a better or deeper understanding that is reached by this approach, but rather it refers to the idea of the existence of sequential layers in the model. The depth of the model is defined by the number of layers contributing to the final outcome. Other appropriate names for this area could be *layered learning* and *hierarchical learning* (Chollet & Allaire, 2018). In theory, there is no upper limit for the number of layers through which learning can take place, but the main difference relative to other methods of machine learning is that classic machine learning implies one or maximum two layers of processed data, which is why such methods are sometimes called shallow learning.

4.1 Neural networks

Initially, the rise of neural networks unfolded as this method was improved for the purpose of better classification of images. The abrupt surge in usage happened when the networks turned out to be extremely useful for a series of problems, notably because in the process of creating an adequate model, they automated one of the key steps – the so-called feature engineering. Before neural networks, feature engineering implied processing data in a manner that would provide best performance outputs depending on the model used. Neural networks process input data alone, in the first layer, after which data are transformed in all subsequent layers until the last, output layer.

The explanation as to how neural networks function has been borrowed from Chollet et al. (2022). The basis of the neural network is the block unit called the perceptron, and several perceptrons placed in layers create a neural network. The perceptron is composed of four units:

- input layer,
- weights (values with which input data are weighted),
- weighted sum,
- activation function (which may or may not be activated upon the receipt of the weighted sum).

At the end of the process, an output layer is generated, whose dimensions are defined by the type of problem being resolved – if it is a regression or a binary classification, there will be only one output layer. In case of multifold classification, where the probability of belonging to a specific class is being modelled, the output layer will have as many nodes as there are classes. The type of the problem being modelled will determine the type of the activation function to be used. Some of the possible activation functions are listed below:

- Rectified Linear Unit (ReLu) – negative values are output as zero, and positive values as the maximum in the set;
- Leaky Rectified Linear Unit (Leaky ReLu) – accepts negative values, but with low weights which depend on the *a* coefficient;
- The Tanh activation function is non-linear, but has a narrow scope, hence it more often rejects important inputs and has activation values in the range [–1, 1];
- The sigmoid function is similar to the previous one, with the exception of activation values being in the [0, 1] range, therefore it is particularly suitable for modelling the probability problem;
- Softmax – activation values are added to 1, therefore it is suitable for multifold classification.

The minimum number of layers in a neural network is two, given that the input layer does not count, therefore the simplest, shallow neural network would be made up of an output layer and one hidden layer.

The network takes data from the input layer in batches, and the number of nodes is equal to the number of explanatory variables, which must be numerical. Several types of layers can be found in the hidden part. With dense layers, each node in the input layer is connected to each node in the output layer (fully connected layers). Convolutional layers are made up of filters which process a portion of input nodes, but always in such a manner than each input node is connected to at least one node in the hidden part of the network. This paper applied the model of the recurrent neural network (nodes can take their previous output values as input ones), more specifically the LSTM network (long short-term memory), which are the most suitable for working with time series, as we will present in the remainder of the chapter.

Each network must have a defined loss function, which is used to assess the quality of the model's prediction (successfulness). Depending on the number of output variables, we can have several loss functions in the model (one for each variable). The already known mean squared error (MSE) and mean absolute error (MAE) are applied in neural networks which are used for regression problems, while the binary cross entropy and hinge loss (a protocol similar to the support vector machines) are used for the problems of binary classification.

With the loss function, the optimisation protocol rounds up one neural network. With the help of the optimiser, weights are updated in each iteration so that the value of the loss function when generating a new output layer decreases. The most frequently used optimiser is the gradient descent. In fact, this is a simple principle of looking for the global minimum of the loss function by iteratively changing weights based on the previously realised output layers. As with all problems of this type, there is the issue of stopping at a local minimum, which is solved by the convex nature of the loss function and the learning rate. A high learning rate will move quickly down the convex loss function, and can thus "miss" the global minimum. A too low learning rate requires a large number of iterations to find the minimum, which means it is time-consuming. Finding the optimum learning rate has resulted in the improvement of the gradient descent optimiser and the creation of the ADAM optimiser (Adaptive Moment Estimation), which takes into account the previous values when calculating a new gradient.

Goulet Coulombe et al. (2022) compared different kinds of machine learning models based on data from the FRED base concerning industrial production, unemployment rates, inflation, the difference between the ten-year maturity rate of Treasury bills and the key policy rate, as well as real estate prices. They drew several conclusions, two of which are particularly relevant for this paper, namely the fact that the non-linearity of machine learning models is especially useful in situations of macroeconomic uncertainty, tight financial conditions and when the real estate bubble bursts, and the other conclusion concerns the amount of data in the model. The first conclusion is applicable in case of data-rich models, while smaller models have better performances if non-linear machine learning models are combined with the classic factor analysis.

Hopp (2022) said that of all machine learning models, the ones with neural networks recorded the largest growth. Still, of all areas in which neural network models are applied, the slowest progress has probably been recorded in current econometric projections, which is most likely the consequence of the same problems with which traditional econometric models are faced, such as multicollinearity, missing data, mixed frequencies and lags in data publication. In addition, a large amount of data and the addition of explanatory variables to models that are already experiencing some of these issues can further exacerbate the situation. Hopp concluded that the LSTM model has better performances that the DFM model because it does not rely on the invertibility of the matrix, and can thus process any data set or any combination of frequencies. One of the key advantages of using the LSTM model is the fact that it does not rely on the invertibility of the matrix, therefore its input can also be the singular matrices or non-squared matrices, which is often the case when it comes to real data. Also, as the given models use large databases, the invertibility of the matrix is often a very demanding condition technically and timewise, which is not necessarily possible to meet. In contrast, the shortcomings compared to the DFM model include the stochastic nature of the LSTM model, the lack of interpretability in coefficients and the fact that this is a "black box" of contributions of individual explanatory variables.

This shortcoming is referred to by Medeiros et al. (2005), placing neural networks in the group of models that leave a functional form unspecified. Though these models contain parameters (and a large number of those), they are not globally identified or assessed, and therefore cannot be interpreted.

Elsayed et al. (2021) asked whether we need deep learning models at all for projecting time series, and concluded that simpler models of classic machine learning are sometimes quite sufficient for modelling dynamics of univariant and multivariant time series.

4.1.1 LSTM neural networks

Siami-Namini et al. (2019) gave a brief overview of the LSTM model functioning. When the neural network is tasked with modelling interdependent data, i.e. data where the previous values affect the current ones (as is the case with time series data), it is desirable to use the models of recurrent neural networks (RNN). The RNNs differ from other neural networks by the presence of connections between layers that are not solely in one direction – from the input to the output layer. More specifically, the recurrent network "learns" by back propagation in the process of minimising loss function. Due to the issue of short memory of the previous state of inputs in classical RNN networks (the information is lost after only a few periods), the LSTM (long short-term memory) network model was developed. This model essentially functions the same as other RNN networks, with the addition of a special structure of the node itself, where there are several gates:

- input gate, that pertains to data entering an active cell;
- output gate, that processes data exiting an active cell;
- forget gate, which filters the relevant information received through the backward propagation method.

Though the RNNs have shown to be adequate for time series modelling, especially in the form of the LSTM model, a certain number of papers used other deep learning models for their current forecasts, and an overview of those can be found in Zheng et al. (2023). One of them is the model presented by Loermann and Maas (2019) in their paper, for the needs of assessing US GDP levels in the current quarter and several quarters ahead using the ANN model. The Artificial Neural Network (ANN) models are the simplest neural network models containing one hidden layer in their structure. A special advantage of using the ANN lies in the possibilities for choosing the relevant variables and lags using the grid structure by way of iterative exclusion of variables and lags that did not turn out to be statistically relevant for forecast improvement. Given that the ANN models have a pronounced issue with the local minimum in the model error minimisation process (which, in this case, is the mean squared error), 100 random ANN models were assessed and then included in one kernel model. This approach is called the ensemble operator approach. The model was then compared to a dynamic factorial model, and the ANN model had a much lower forecasting error.

The success of different kinds of neural networks in modelling time series has induced a number of authors to create new deep learning models suitable for this type of data. Thus, the Temporal Fusion Transformer (TFT) model was created. The transformer-type deep learning models are characterised by the adoption of the self-attention concept, i.e. the possibility to access inputs along the entire sequence in any moment, as well as the possibility of weighting them based on the learned material, which is what sets them apart from the RNNs (Zeng et al. 2022). This is the model that should solve the biggest issue with the LSTM – the relatively short network memory (several periods back). Still, longer-term memory is not always necessary, especially if the goal is modelling only a few periods ahead or estimating current levels (nowcasting).

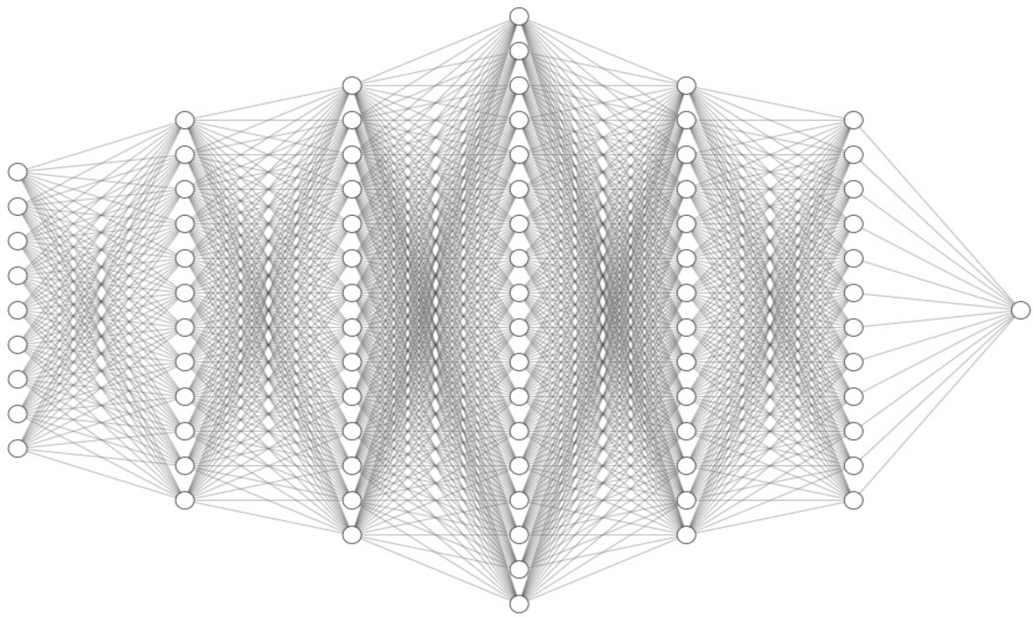
The above mentioned ensemble operator approach is a type of machine learning development relating to economic modelling by using a machine learning model to unify a larger number of already existing standard econometric models and/or machine learning models. The idea that a more reliable forecast can be obtained by averaging a larger number of models is not new in economics, but machine learning enables the weighting and selection of the most relevant models in a manner that minimises forecasting errors without *a priori* imposing the structure.

4.2 Assessing the LSTM model

The process of assessing LSTM models partly follows the methodology presented in Medeiros et al. (2005), though the model type is not the same. The approach boils down to the idea that the first assessed model should be as simple as possible, and then the performances of each ensuing, expanded model should be monitored. In this paper, the performances of each model iteration are compared using the mean squared error.

We assessed the LSTM RNN model with one input layer with nine regressors, and five hidden layers with 64, 128, 256, 128 and 64 nodes, and one node in the output layer. The structure of the assessed model's approximation is as follows.

Figure 8 **Approximation of the structure of the assessed LSTM model**



According to the already mentioned paper by Varian (2014), the problem of overfitting that often occurs with machine learning models is addressed by dividing the database into training data, validation data and assessment data. The model's efficiency is measured based on the reduction of the mean squared error, and the model was assessed in 1,000 epochs. The final value of the root mean squared error was 2.98, while the value for the validation set was 4.90.

The obtained error of out-of-sample forecast can be compared to the one that was obtained by assessing the MIDAS model. Namely, the exit from the LSTM model gives us information about three different mean squared errors – the first two are listed and pertain to the assessment part of the sample and the validation part of the sample, while an error obtained in the testing part of the sample in the LSTM model would correspond to out-of-sample error in the MIDAS model. The MIDAS model has an out-of-sample error of 2.27 (Figure 3 in part 3.3), while the comparable value with the LSTM model is 1.85.

The fact that the LSTM model has a larger validation base error compared to the training base confirms that the issue of overfitting to data is still present, and in case of this model it is almost impossible to avoid. The reason for this is the relatively small database which is a necessary precondition for successfully setting a machine learning model, in particular a deep learning one. It is additionally worrying that the root mean squared testing-based forecasting error has an extremely low value compared to the training and validation bases, although it shows greater precision of the LSTM model compared to the MIDAS. The reasons for such inconsistency between models are several:

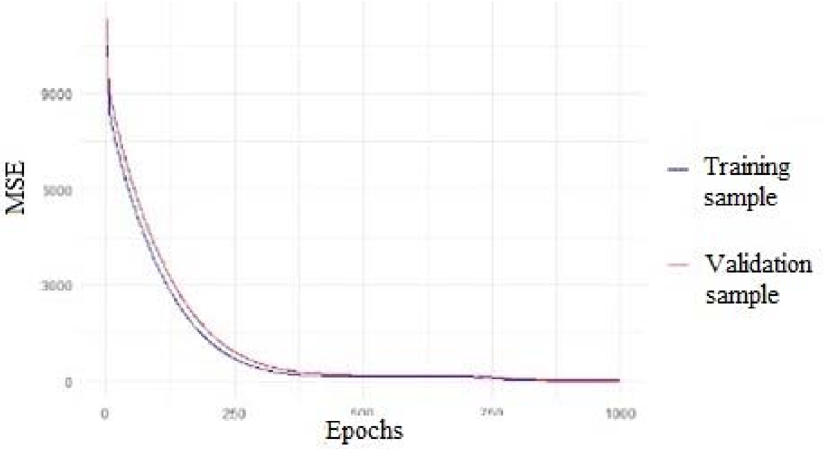
- Besides the already mentioned problem of base size, an additional issue is the fact that the validation set is much smaller than the training set, while the testing set contains only ten monthly regressor data for each series;

● Variance in the testing base is much smaller than the variance in the testing and validation bases;

● This type of problem occurs sometimes when the neural network is trained using the Keras/TensorFlow. When using the batch gradient descent, the value of error in each epoch is the average of all batches processed until that point. Hence, the high error value at the beginning shall affect the entire error through the average, regardless of how much it decreased during the model training. On the other hand, this problem is not present when testing model successfulness as in that case only the final error value is taken into account.

Below are model performances on the training and validation sets.

Figure 9 **Performances of the neural network model on the training and validation parts of the sample**



5 Conclusion

In their papers, Varian (2014) and Murphy (2012) stated that the basic measure of adequacy of a machine learning model is forecasting success. This is certainly not the only criterion when it comes to traditional econometric models, but it is the goal of the models used for nowcasting. The goal of assessing the MIDAS model and the neural network model is the assessment of GDP current movements, therefore it is justifiable to compare these two models using the root of the mean squared forecasting error.

Based on the out-of-sample forecasting error, we can conclude that the LSTM neural network model is more successful at assessing current GDP movements and the near-term GDP dynamics (in this case, up to five quarters ahead). Still, we stated the reasons why the results of the LSTM model should be taken with reservations because of the small sample, low variance, and the applied software. The root mean squared error on the database for LSTM validation is significantly higher than the forecasting error of the MIDAS model.

The contribution of the MIDAS model is significant when it comes to the differentiation between regressors that are statistically significant and the ones that are not, as the LSTM model has unspecified parameters, i.e. insight into its structure is not possible. A specificity of the nowcasting model is also reflected in the manner of assessing the statistical significance of the explanatory variable, given that the justifiability of including regressors whose test of statistical significance has a high p-value is explained by the improvement of the forecast. The final form of the MIDAS model contains three variables, one of which is an official indicator of the statistical office, industrial production index, while two variables are alternative indicators of economic activity, Google's business trends index related to corporate-related searches, as well as the measure of electricity consumption. Interestingly, the industrial production index can be associated with the industrial sector, while the business trends index can intuitively be associated with the services sector which together make up almost three-quarters of total GDP, on the production side. Electricity consumption as an indicator can be associated with all branches of GDP on the production side, and can thus include the fluctuations not included in industrial production or Google's business trends index in the industry and service sectors, though it can contain shocks that affect the construction and agriculture sectors which, together with the above mentioned branches and net taxes, comprise the entire GDP.

This paper made several references to sources that underscore sufficiently large databases as a necessary condition for a successful machine learning model. The phrase "a sufficiently large base" means nothing on its own if it is not accompanied by a specific number. Sadly, there is not an accurate amount of data that would guarantee that a machine learning model will function in the best possible manner and give the best results. It is a fact that deep learning models were partly made precisely in order to process very large amounts of inputs generated each second, primarily through internet use. Therefore the bases used for the training of deep learning models have up to several million inputs, which is impossible to achieve with macroeconomic data. Nevertheless, deep learning models will record an improvement in their performances by adding new and good quality inputs (there has been talk about counterproductivity if bases are expanded with data characterised with issues such as missing segments, collinearity, autoregression, etc). This assertion is true not only for machine or deep learning models, but also for traditional econometric models, with which it is easier to come by meaningful representations of reality if there is a "sufficient" amount of data. There is no specific number, but it is almost certain that the database used in this paper does not meet the size criterion that would justify the transition from traditional econometric to deep learning models. In contrast, the complexity and non-linearity present in this base, reflected in a large number of potential regressors, their varying frequency and complex dynamics and problematics of GDP forecasting, partly justify the attempt to model alternative approaches relative to classic econometric models.

One of the reasons for a small amount of data is the inclusion of alternative indicators since their series are significantly shorter than the official macroeconomic indicators of economic activity. Still, these indicators have been shown to be statistically significant (electricity consumption) and better results were achieved when they were included in the model (Google trends) than when a comparable model was used based solely on official statistics. In addition, the complexity in the manner of collecting, processing, publishing and the frequency of these indicators also speaks in favour of using a machine learning model. Hence, if the number of available indicators and the length of the available series continue to grow, we can also expect an increase in performances of both traditional and machine learning models.

In short, the assessed MIDAS model and the neural network model adequately model GDP dynamics in the short term, with the LSTM model turning out more efficient if the root mean squared forecasting error is used for comparison. Both models displayed a tendency towards overfitting to available data, which is mostly a consequence of the relatively small database for assessing the models. Still, though the LSTM model was shown to be better, if we bear in mind the computational demands and the inability to analyse the significance of individual parameters, the MIDAS model gives us more than an adequate short-term projection of Serbia's GDP dynamics. As one of the hypotheses of this paper is to verify the statistical significance of alternative parameters in combination with official macroeconomic indicators, after an insight into the MIDAS model we can conclude that Google's business trend indicator and the electricity consumption indicator from the ENTSO-E platform contribute to the improvement of the short-term projection. Besides, the alternative indicators used in combination with the industrial production indicator, in a simplified manner, model the individual branches within the production approach to the assessment of total GDP. Essentially, even though a deep learning model has turned out to be better at projecting Serbia's short-term GDP dynamics, because of its structure and transparency, the MIDAS model is a valuable input for economic policy makers, corporates and individuals who benefit from the timely and adequate economic activity assessment as well. As the database grows, the performances of both assessed models are expected to be increasingly better, especially the deep learning model.

References

Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33 (1), Article 1. https://doi.org/10.2307/1911894.

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, 1142, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012.

Anesti, N., Galvão, A. B. & Miranda-Agrippino, S. (2018). Uncertain Kingdom: Nowcasting GDP and its Revisions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3278618.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. & Rünstler, G. (2011). Short-term forecasts of euro area GDP growth. *The Econometrics Journal*, 14 (1), C25–C44. https://doi.org/10.1111/j.1368-423X.2010.00328.x.

Arencibia Pareja, A., Gomez-Loscos, A., Luis López, M. D. & Perez-Quiros, G. (2020). A Short Term Forecasting Model for the Spanish GDP and its Demand Components. *Economía*, 43 (85), 1–30. https://doi.org/10.18800/economia.202001.001.

Armesto, M. T., Engemann, K. M. & Owyang, M. T. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St. Louis Review.*

Balassa, B. (1985). Exports, policy choices, and economic growth in developing countries after the 1973 oil shock. *Journal of Development Economics*, 18 (1), 23–35. https://doi.org/10.1016/0304-3878(85)90004-5.

Bańbura, M. & Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27 (2), 333–346. https://doi.org/10.1016/j.ijforecast.2010.01.011.

Barbaglia, L., Consoli, S. & Manzan, S. (2021). Forecasting GDP in Europe with Textual Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3898680.

Barhoumi, K., Darné, O., Ferrara, L. & Pluyaud, B. (2012). Monthly GDP Forecasting Using Bridge Models: Application for the French Economy. *Bulletin of Economic Research*, 64, s53–s70. https://doi.org/10.1111/j.1467-8586.2010.00359.x.

Burns, A. & Mitchell, W. (1946). *Measuring Business Cycles*. NBER.

Chernis, T. & Sekkel, R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics*, 53 (1), 217–234. https://doi.org/10.1007/s00181-017-1254-1.

Choi, H. & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88, 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x.

Chollet, F. & Allaire, J. J. (2018). *Deep learning with R*. Manning.

Chollet, F., Kalinowski, T. & Allaire, J. J. (2022). *Deep learning with R* (Second edition). Manning.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20 (3), 273–297. https://doi.org/10.1007/BF00994018.

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S. & Schmidt-Thieme, L. (2021). *Do We Really Need Deep Learning Models for Time Series Forecasting?* (arXiv:2101.02118). arXiv. http://arxiv.org/abs/2101.02118.

Fradkov, A. (2020). Early History of Machine Learning. *IFAC-PapersOnLine*, 53 (2), 1385–1390. https://doi.org/10.1016/j.ifacol.2020.12.1888.

Frale, C. & Monteforte, L. (2010). *FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure*.

Ghysels, E., Kvedaras, V. & Zemlys, V. (2016). Mixed Frequency Data Sampling Regression Models: The *R* Package *MIDASR*. *Journal of Statistical Software*, 72 (4), Article 4. https://doi.org/10.18637/jss.v072.i04.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2002). *The MIDAS Touch: Mixed Data Sampling Regression Models*.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004). *Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies* (No. w10914; Issue w10914, p. w10914). National Bureau of Economic Research. https://doi.org/10.3386/w10914.

Giannone, D., Agrippino, S. M. & Modugno, M. (2013). *Nowcasting China Real GDP*.

Giannone, D., Reichlin, L. & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55 (4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010.

Goulet Coulombe, P., Leroux, M., Stevanovic, D. & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37 (5), Article 5.

Hopp, D. (2022). Economic nowcasting with long short-term memory artificial neural networks. *Journal of Official Statistics*, 38 (3), 847–873.

Ilin, C., Annan-Phan, S., Tai, X. H., Mehra, S., Hsiang, S. & Blumenstock, J. E. (2021). Public mobility data enables COVID-19 forecasting and management at local and global scales. *Scientific Reports*, *11*(1), 13531. https://doi.org/10.1038/s41598-021-92892-8.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31 (3), 685–695. https://doi.org/10.1007/s12525-021-00475-2.

Kitchen, J. & Monaco, R. (2003). Real-time Forecasting in Practice: The U.S. Treasury Staff's Real-time GDP Forecast System. *U.S. Department of the Treasury*.

Koenig, E. F., Dolmas, S. & Piger, J. (2003). The Use and Abuse of Real-Time Data in Economic Forecasting. *Review of Economics and Statistics*, 85 (3), 618–628. https://doi.org/10.1162/003465303322369768.

Loermann, J. & Maas, B. (2019). *Nowcasting US GDP with artificial neural networks*.

Markiewicz, A., Verhoeks, R. C., Verschoor, W. F. C. & Zwinkels, R. C. J. (2018). *The Winner Takes it All: Measuring Attention and Predicting Exchange Rates with Google Trends*.

Medeiros, M. C., Terasvirta, T. & Rech, G. (2005). Building Neural Network Models for Time Series: A Statistical Approach. *Journal of Forecasting*, 25 (1), Article 1.

Moody, J., Levin, U. & Rehfuss, S. (1993). *Predicting the U.S. Index of Industrial Production (Extended Abstract)*.

Mourougane, A. (2006). *Forecasting Monthly GDP for Canada*. OECD. https://doi.org/10.1787/421416670553.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Narita, F. & Yin, R. (2018). *In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries* [IMF Working Papers].

Plasek, A. (2016). On the Cruelty of Really Writing a History of Machine Learning. *IEEE Annals of the History of Computing*, *38*(4), 6–8. https://doi.org/10.1109/MAHC.2016.43.

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2 (3), 160. https://doi.org/10.1007/s42979-021-00592-x.
Schmidt, T., & Vosen, S. (2009). Forecasting Private Consumption: Survey-based Indicators vs. Google Trends. *Ruhr Economic Papers*.
Siami-Namini, S., Tavakoli, N. & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *2019 IEEE International Conference on Big Data (Big Data)*, 3285–3292. https://doi.org/10.1109/BigData47090.2019.9005997.
Sinko, A. (2008). *Some Applications of Mixed Data Sampling Regression Models*.
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. https://doi.org/10.1016/bs.hesmac.2016.04.002.
Stock, J., & Watson, M. (1988). *A Probability Model of The Coincident Economic Indicators* (No. w2772; p. w2772). National Bureau of Economic Research. https://doi.org/10.3386/w2772.
Sun, S. & Anwar, S. (2015). Electricity consumption, industrial production, and entrepreneurship in Singapore. *Energy Policy*, 77, 70–78. https://doi.org/10.1016/j.enpol.2014.11.036.
Turnovsky, S. J. (2000). *Methods of macroeconomic dynamics* (2. ed). MIT Press.
Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28 (2), 3–28. https://doi.org/10.1257/jep.28.2.3.
Woloszko, N. (2020). *Tracking activity in real time with Google Trends* (OECD Economics Department Working Papers No. 1634; OECD Economics Department Working Papers, Vol. 1634). https://doi.org/10.1787/6b9c7518-en.
Wu, L. & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2022293.
Zeng, A., Chen, M., Zhang, L. & Xu, Q. (2022). *Are Transformers Effective for Time Series Forecasting?* (arXiv:2205.13504). arXiv. http://arxiv.org/abs/2205.13504.
Zhang, Y. & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. *Npj Computational Materials*, *4*(1), 25. https://doi.org/10.1038/s41524-018-0081-z.
Zheng, Y., Xu, Z. & Xiao, A. (2023). Deep learning in economics: A systematic and critical review. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-022-10272-8.